Dynamic Regulatory Graph Learning for Histology-to-Spatial Transcriptomics

Yikai Luo[1] **Peng Zhang**[1]* **Heyang Zhao**[1] **Hongming Shan**[2] **Wenjian Wang**[1]
[1]School of Computer and Information Technology, Shanxi University
[2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University

Abstract

Virtual spatial transcriptomics aims to predict spatial gene expression from histopathology images. However, gene expression is not governed by morphology alone; regulatory relationships are structured, context-dependent, and shaped by the local tissue microenvironment. Existing visual-to-expression models and fixed-prior graph methods struggle to align molecular dependencies with spatially varying histological contexts. We propose DragH2ST, a dual-stage retrieval-augmented dynamic regulatory graph learning framework that constructs task-specific regulatory manifolds and modulates the use of the prior and gene representations at the spot level according to local histology. Before training, DragH2ST retrieves and integrates heterogeneous biomedical knowledge to build a task-specific regulatory prior, guiding gene representations toward biologically plausible and context-adaptive regulatory structures. To capture microenvironment-dependent regulation, context-gated graph rewiring modulates the shared GAT-encoded gene basis using histological context, enabling adaptive regulatory decoding with gene-wise gates. At inference, DragH2ST retrieves morphologically similar historical spot prototypes as case-level references for interpreting predictions and assessing biological plausibility. Experiments on multiple benchmarks show that DragH2ST improves prediction accuracy, preserves gene-gene regulatory topology, and captures spatially heterogeneous expression patterns.

1 Introduction

Pathology is undergoing a paradigm shift from experience-driven qualitative morphological assessment to data-driven quantitative molecular inference [2, 11, 32]. Hematoxylin and eosin (H&E)-stained whole slide images (WSIs), the clinical gold standard for cancer diagnosis, contain rich morphological cues reflecting cellular atypia, tissue architecture, and tumor-stroma interactions. However, traditional morphology-based analysis paradigms struggle to directly quantify the complex molecular regulatory programs underlying these visual patterns, particularly in characterizing gene expression with significant spatial heterogeneity.

Spatial transcriptomics has addressed this gap by measuring gene expression while preserving tissue spatial organization. However, its widespread application in clinical practice remains significantly hindered by high costs, complex workflows, and strict sample requirements [7, 14, 27, 31]. To bridge the gap between morphological phenotypes and molecular profiles, virtual spatial transcriptomics has emerged, aiming to infer spatially resolved gene expression directly from H&E-stained images.

Although pioneering studies such as ST-Net [9] and Hist2ST [38] show that CNN/ViT-based models can predict transcriptomic signals from morphology, most existing methods still formulate the task

*Corresponding author: pengzhang@sxu.edu.cn.
Code: https://github.com/YikaiLuo05/DragH2ST.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

as direct visual-to-expression regression. Such a formulation mainly learns statistical associations between histological appearance and gene abundance, but provides limited mechanisms for modeling why genes co-vary or how their dependencies change across tissue contexts. In practice, gene expression is not determined by morphology alone; it is regulated through structured and context-dependent molecular interactions shaped by the local tissue microenvironment, including cellular composition, stromal organization, and immune infiltration. Therefore, treating genes as independent outputs or relying on fixed dependencies can lead to context-misaligned predictions, especially when the same regulatory program is rewired across tumor, stromal, and immune-enriched regions.

Accurate histology-to-spatial transcriptomics thus requires moving beyond direct mapping and explicitly reasoning over biologically grounded regulatory graphs. Importantly, such regulatory structures should not be freely inferred from visual correlations alone, as morphology–expression associations can be noisy, dataset-specific, and biologically underdetermined. Instead, they should be constructed from external biomedical knowledge to provide credible regulatory scaffolds, and then adapted to local histological contexts. Static biological priors offer useful constraints, but they are often task-agnostic and insensitive to spot-level microenvironmental variation. Task-conditioned retrieval provides a practical way to derive task-specific regulatory priors from heterogeneous biomedical knowledge, while spot-level graph rewiring further adapts these priors according to local histology.

Motivated by this, we propose DragH2ST (Dynamic Regulatory Graph Learning for Histology-to-Spatial Transcriptomics), which formulates histology-to-spatial transcriptomics prediction as dynamic regulatory graph learning. DragH2ST retrieves biomedical knowledge to construct task-specific regulatory priors, modulates the shared GAT-encoded gene representations for spot-specific decoding using local histological context, and retrieves morphologically similar historical prototypes at inference for case-based interpretation.

Contributions. Our contributions are: (**i**) We formulate histology-to-spatial transcriptomics prediction as dynamic regulatory graph learning, integrating task-specific regulatory prior construction with spot-level histology-conditioned graph rewiring; (**ii**) We introduce context-gated dynamic rewiring (CGDR), which modulates a shared GAT-encoded gene basis according to local histological context for efficient spot-specific decoding; and (**iii**) We develop a dual-graph framework with noise-robust ranking supervision that jointly models the gene regulatory and spatial neighborhood graphs, improving robustness to sequencing depth variation and batch effects.

2 Related works

Purely vision-driven mapping from local regression to generative modeling. Early virtual spatial transcriptomics methods mainly learned direct mappings from histology to gene expression. Representative examples include CNN or MIL regression models such as ST-Net [9] and HE2RNA [29], structure-aware architectures such as HisToGene and Hist2ST [26, 38], and later extensions that model richer tissue organization, including NH2ST, HyperST, and iStar [28, 39, 40]. More recent work also strengthens visual encoding or generation through pathology foundation models and generative modeling, such as DeepSpot, DeepSpot2Cell, and HaDM-ST [18, 24, 25]. However, these methods still largely treat prediction as a one-way morphology-to-expression mapping and do not explicitly adapt biological structure to the local microenvironment. In contrast, our work couples retrieved biological priors with context-gated graph rewiring, allowing prior-encoded gene representations to contribute adaptively to decoding according to local visual context.

Static priors and multimodal alignment to bridge the semantic gap. Another line of work reduces the semantic gap by introducing multimodal alignment and biological priors. Alignment-based methods such as BLEEP, mclSTExp, and CarHE map images, expression profiles, and auxiliary cellular signals into shared latent spaces [22, 37, 42]. Prior-guided approaches instead inject Gene Ontology, protein interaction, pathway, or regulatory information, as in M2TGLGO, PRALINE, P-NET, PEaRL, and related spatial regulation models [5, 6, 8, 16, 20, 23, 30]. Related spatial-omics models include HEIST for hierarchical graph representation learning [19] and stDiff for spatial expression imputation using single-cell transcriptomics [15]. These methods improve semantic grounding, but most still depend on globally shared static priors or alignment objectives that are not explicitly adaptive to task-specific knowledge needs and spot-level microenvironmental variation. In contrast, our framework retrieves task-relevant biomedical knowledge during training, modulates

GAT-encoded gene representations for spot-specific decoding, and uses prototype retrieval at inference for case-based interpretation.

3 Methodology

3.1 Problem formulation

We formulate histology-to-spatial transcriptomics prediction as dynamic regulatory graph learning over two coupled structures: a spatial tissue graph and a gene regulatory manifold. Let a tissue section be represented as $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s, A_s)$, where each node $v_i \in \mathcal{V}_s$ denotes a spot with histopathology patch $X_i \in \mathbb{R}^{H_{\text{img}} \times W_{\text{img}} \times C}$, spatial coordinates, and normalized expression vector $Y_i \in \mathbb{R}^{N_{\text{gene}}}$ over a task-specific gene set $q = \mathcal{S}_g$. Conventional methods learn $Y_i = f_\theta(X_i)$, ignoring spatial tissue context and structured gene-gene dependencies.

DragH2ST formulates expression prediction as inference jointly conditioned on tissue organization and task-specific biological structure. DP-RAG constructs a task-specific prior graph $A_{\text{prior}} = \mathcal{R}(q, \mathcal{K})$ from a multi-source biomedical knowledge pool $\mathcal{K}$, while CGDR generates spot-specific modulation that induces a context-aware gene basis $Z^{(i)}$. Let $h_i = \mathcal{V}_\theta(X_i)$ and $h_i^{\text{final}} = [\mathcal{C}_\theta(\{h_j\}_{v_j \in \mathcal{V}_s}, A_s)]_i$. The resulting predictor integrates the direct morphology–expression mapping with decoding over the context-aware gene manifold:

$$\hat{Y}_i = (1 - \lambda_{mix})\hat{Y}_i^{lin} + \lambda_{mix}\hat{Y}_i^{reg}. \tag{1}$$

Here, $\hat{Y}_i^{lin}$ captures the direct relationship between spatial tissue representations and expression abundance, while $\hat{Y}_i^{reg}$ incorporates gene associations encoded by the biological prior and local context modulation. Both predictions use h_i^{final}, and $\lambda_{mix} \in [0, 1]$ controls their relative contributions. The prior construction, context modulation, and prediction branches are detailed below.

3.2 Dual-stage retrieval: task-specific prior construction and prototype-based evidence

Dynamic prior-guided retrieval-augmented generation (DP-RAG). Before training, we construct a task-level prior to address the limitations of globally shared single-source graphs. As illustrated in Fig. 1(b), DP-RAG uses the target gene set defined by the current task as its query to retrieve and fuse $N_{KG} = 4$ complementary external knowledge sources, represented by similarity matrices $\{S_k\}_{k=1}^{N_{KG}}$, where each $S_k \in \mathbb{R}^{N_{gene} \times N_{gene}}$. Gene2Vec captures data-driven co-expression patterns, GO quantifies functional relevance, STRING PPI provides protein-interaction information, and MSigDB captures shared pathway membership (detailed definitions in Appendix C). We then perform weighted fusion over these four sources to obtain S_{fused} and construct the consensus gene adjacency matrix A_{prior}:

$$S_{\text{fused}} = \frac{\sum_{k=1}^{N_{KG}} w_k S_k}{\sum_{k=1}^{N_{KG}} w_k}, \qquad A_{prior} = \mathcal{P}_{t,\text{top-k}}(S_{\text{fused}}) \tag{2}$$

where w_k is the source fusion weight for the k-th knowledge source, selected via Optuna-based validation search. The operator $\mathcal{P}_{t,\text{top-k}}$ clips the fused scores to $[0, 1]$ and sets diagonal entries to one. It then applies a similarity threshold t and row-wise top-k selection, binarizes the retained connections, and adds self-loops. Since the input query gene sets differ across prediction tasks (or datasets), this query-driven mechanism ensures that the resulting A_{prior} is no longer a globally static graph, but rather a dynamically assembled, task-specific prior topology. This provides a robust foundation for the subsequent spot-level manifold rewiring.

Inference-stage prototype retrieval. To provide case-based references for molecular predictions, we use the non-parametric retrieval mechanism shown in Fig. 1(d). An external memory bank $\mathcal{M}$ stores tissue morphology together with measured gene expression profiles. Offline K-Means clustering of training-set spot features $\{h_j\}$ yields N_{proto} clusters; representative tissue patches nearest to the cluster centers form the prototype library $\mathcal{M} = \{(h_m^{proto}, y_m^{proto})\}_{m=1}^{N_{proto}}$, where h_m^{proto} and y_m^{proto} denote the visual feature and measured gene expression of each prototype.

After predicting expression for a query spot x_q (with feature h_q), we retrieve its $K_{retrieval}$ nearest prototypes using cosine similarity:

$$\mathcal{N}_{retrieval}(x_q) = \{(h_k^{proto}, y_k^{proto}) \mid k \in \text{Top-K}_{retrieval}(\text{sim}(h_q, h_m^{proto}))\} \tag{3}$$

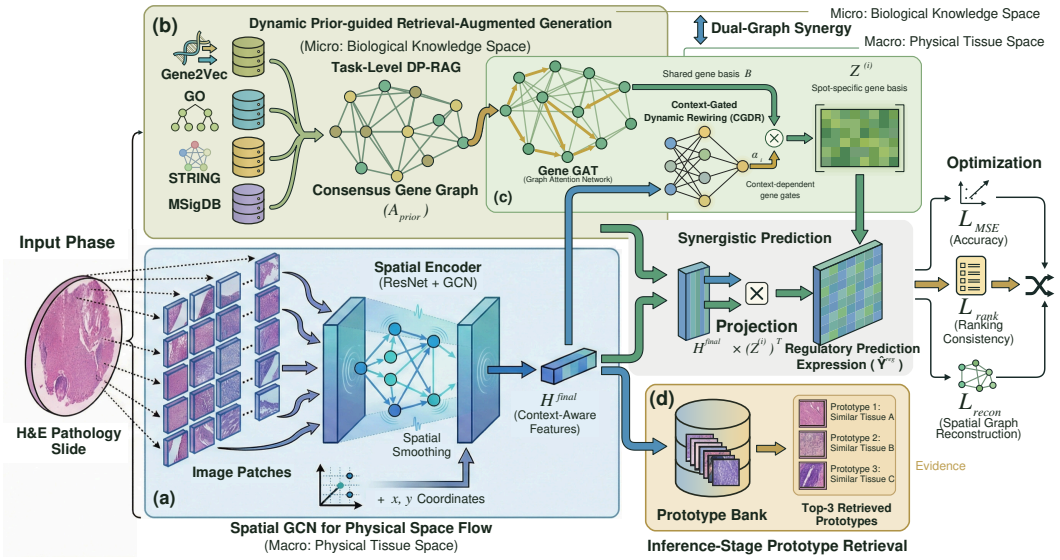


Figure 1: Overview of DragH2ST. (a) The Spatial Encoder models tissue structure through a ResNet50 [10] visual backbone and Spatial GCN. (b) Before training, DP-RAG integrates multi-source knowledge to construct a task-level gene prior. (c) Gene GAT encodes the shared prior into a gene basis, which CGDR modulates according to local visual context for spot-specific decoding. (d) At inference, the model retrieves historical prototypes to interpret predictions. The final prediction combines linear regression from spatial visual features with prior-guided regulatory decoding through a weighted sum.

The retrieved prototypes retain both the morphological features and measured expression profiles of historical tissue samples. Comparing the query with these cases helps interpret its prediction in terms of tissue morphology and molecular expression (see Appendix D for a visualization).

3.3 Context-gated dynamic rewiring (CGDR)

To adapt the task-specific prior graph A_{prior} to local tissue heterogeneity, we introduce Context-Gated Dynamic Rewiring (CGDR), as shown in Fig. 1(c) and Fig. 2. The key idea is to encode the shared biological scaffold with Gene GAT and use local histological context to modulate the resulting gene basis for each spot. In this way, CGDR realizes spot-level dynamic regulatory adaptation through context-conditioned gene decoding.

Given the spatially enhanced visual feature $h_i^{final} \in \mathbb{R}^{D_v}$ of the i-th spot, a lightweight MLP produces a raw gene-wise gate:

$$\begin{aligned} \alpha_i^{raw} &= \sigma\Big(\mathcal{F}_{gate}(h_i^{final})\Big), \\ \alpha_i &= 0.5 + \alpha_i^{raw}, \end{aligned} \tag{4}$$

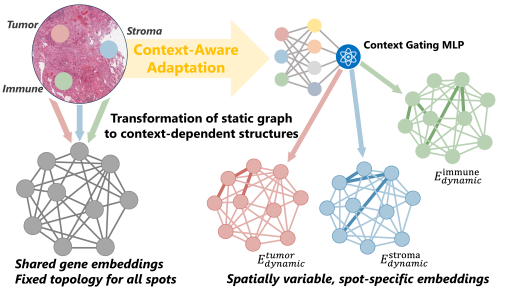


Figure 2: CGDR predicts spot-specific gene gates from spatial visual context to modulate the shared GAT-encoded gene basis under different microenvironments.

where $\mathcal{F}_{gate}(h) = W_{g2}\text{ReLU}(W_{g1}h + b_{g1}) + b_{g2}$, with $W_{g1} \in \mathbb{R}^{D_{hidden} \times D_v}$, $W_{g2} \in \mathbb{R}^{N_{gene} \times D_{hidden}}$, and learnable biases b_{g1}, b_{g2}. The raw sigmoid output $\alpha_i^{raw} \in (0, 1)$ is shifted to obtain the actual Soft Gating vector $\alpha_i \in (0.5, 1.5)$. Thus, $\alpha_{i,u} < 1$ attenuates the embedding of gene u in the current microenvironment, whereas $\alpha_{i,u} > 1$ amplifies it. This perturbation-style design avoids exact zero-valued gates and retains a nonzero multiplicative coefficient for each gene channel.

Let $B = \mathrm{GAT}(E_{\mathrm{global}}, A_{\mathrm{prior}}) \in \mathbb{R}^{N_{\mathrm{gene}} \times D_v}$ be the shared gene basis encoded from the global embeddings and biological prior. CGDR applies the gene-wise gates to this basis:

$$Z^{(i)} = B \odot (\alpha_i \mathbf{1}^T) = \mathrm{diag}(\alpha_i) B, \tag{5}$$

where $\odot$ denotes element-wise multiplication and $\mathbf{1}^T$ broadcasts α_i across the basis dimension. The spot-specific basis $Z^{(i)}$ is used for regulatory decoding. The shared GAT captures prior-guided gene associations, and the local gates determine their contribution to expression prediction in each tissue context.

This mechanism realizes soft adaptation of prior-guided gene decoding: A_{prior} specifies the biological relations encoded by GAT, and α_i controls the contribution of each encoded gene representation to the current spot. As shown in Appendix A.2, basis modulation is equivalent to scaling the regulatory predictions gene-wise, giving a linear additional modulation cost. Coupled with DP-RAG, it forms a dual-level dynamic mechanism: DP-RAG provides the task-level biological scaffold, and CGDR adapts its effective utilization to local tissue microenvironments.

3.4 Dual-graph synergistic architecture

As shown in Fig. 1, DragH2ST bridges physical tissue structure and molecular regulatory networks through a dual-graph synergistic architecture. Specifically, we construct a spatial adjacency graph to model the physical proximity between spots, and a gene prior graph to capture biological interactions within the feature space. By integrating these two distinct topologies, the model utilizes orthogonal and complementary information from both domains.

Spatial GCN for physical space flow to smooth microenvironmental noise. To smooth noise and leverage the spatial continuity of tissues, we construct a spatial adjacency graph $\mathcal{G}_{spatial}$ based on the physical coordinates of spots, with its adjacency matrix denoted as $A_{spatial}$. Specifically, we use a graph convolutional network (GCN) to perform spatial smoothing on the visual features:

$$H^{(l+1)}_{spatial} = \mathrm{Norm}_2\Big(\mathrm{ReLU}\Big(D_\epsilon^{-1} A_{spatial} H^{(l)}_{spatial} \Theta^{(l)}_{spatial}\Big)\Big) \tag{6}$$

where $H^{(l)}_{spatial} \in \mathbb{R}^{N_{spot} \times D_v}$ is the node feature matrix of the l-th layer, with $H^{(0)}_{spatial}$ containing the input spot representations. The diagonal matrix D_ϵ has entries $(D_\epsilon)_{ii} = \max(\sum_j (A_{spatial})_{ij}, 10^{-8})$, so $D_\epsilon^{-1} A_{spatial}$ performs mean aggregation over spatial neighbors. Here, $\Theta^{(l)}_{spatial}$ is the learnable weight matrix, and Norm_2 denotes row-wise ℓ_2 normalization after the ReLU activation. Aggregating the layer-wise representations yields the spatially enhanced visual context $H^{final} \in \mathbb{R}^{N_{spot} \times D_v}$, allowing each spot to perceive the surrounding macroscopic tissue structure.

Gene GAT for feature space flow to refine molecular semantics. Unlike the GCN which aggregates information in physical space, the gene flow focuses on consistency within the gene feature space. We apply a shared GAT to the global gene embeddings under the biological prior:

$$B = \mathrm{GAT}(E_{\mathrm{global}}, A_{prior}) \tag{7}$$

Here, A_{prior} serves as an attention mask over the retrieved biological relations, and $B \in \mathbb{R}^{N_{gene} \times D_v}$ is the shared gene basis in the spatial feature dimension. CGDR then produces the context-aware basis $Z^{(i)}$ through Eq. (5).

Manifold projection and prediction fusion. The spatially enhanced feature h_i^{final} (the i-th row of H^{final}) feeds both prediction branches. The linear branch applies layer normalization followed by a linear mapping, while the regulatory branch projects the feature onto the context-aware gene basis:

$$\begin{aligned} \hat{Y}_i^{lin} &= \mathrm{LN}(h_i^{final}) W_{lin} + b_{lin}, \\ \hat{Y}_i^{reg} &= h_i^{final} (Z^{(i)})^T. \end{aligned} \tag{8}$$

Here, LN denotes layer normalization, $W_{lin} \in \mathbb{R}^{D_v \times N_{gene}}$ and $b_{lin} \in \mathbb{R}^{N_{gene}}$ are the linear head parameters, and the GAT output shares the feature dimension of h_i^{final}. Equation (1) fuses direct spatial regression with prior-guided regulatory prediction. The fusion coefficient λ_{mix} is fixed for each training run and specified in the experiment configuration.

3.5 Optimization objectives

We propose a composite objective function that balances expression prediction, distributional fit, and spatial structural consistency. The total loss $\mathcal{L}_{total}$ is defined as a weighted sum:

$$\mathcal{L}_{total} = \lambda_{mse}\mathcal{L}_{MSE} + \lambda_{zinb}\mathcal{L}_{zinb} + \lambda_{rank}\mathcal{L}_{rank} + \lambda_{pcc}\mathcal{L}_{pcc} + \lambda_{recon}\mathcal{L}_{recon} \tag{9}$$

where $\lambda_{\cdot}$ controls each loss weight. We use a coarse-to-fine search strategy to balance multi-objective training and keep major loss components at comparable scales in the early stage.

Distributional and regression losses. To accommodate the zero-inflated nature of single-cell data, we employ the Zero-Inflated Negative Binomial (ZINB) loss. The likelihood of observing expression y_{ig} is given by:

$$P(y_{ig} \mid \mu_{ig}, \theta_{ig}, \pi_{ig}) = \pi_{ig}\delta_0(y_{ig}) + (1-\pi_{ig})\mathrm{NB}(y_{ig}|\mu_{ig}, \theta_{ig}) \tag{10}$$

where $\mu_{ig}, \theta_{ig}, \pi_{ig}$ denote the predicted mean, dispersion, and zero-inflation probability, respectively, and $\delta_0(\cdot)$ is the Dirac delta function concentrated at zero. We use the MSE loss to supervise expression prediction:

$$\mathcal{L}_{MSE} = \frac{1}{N_{spot}N_{gene}} \sum_{i=1}^{N_{spot}} \sum_{g=1}^{N_{gene}} (\hat{Y}_{ig} - Y_{ig})^2,$$

where $\hat{Y}_{ig}$ and Y_{ig} are the final predicted and target normalized expression values, respectively.

Structural and consistency constraints. To mitigate batch effects and preserve relative expression rankings, we use the STRank loss of Nishimura et al. [23], which compares predicted (P_{ig}) and target (Q_{ig}) expression distributions for each gene across spots by cross-entropy:

$$\mathcal{L}_{rank} = \sum_{i=1}^{N_{spot}} \sum_{g=1}^{N_{gene}} -Q_{ig}\log(P_{ig}) \tag{11}$$

To preserve spatial neighborhood structure, we reconstruct the coordinate-based spot adjacency $A_{spatial}$ from the spatial representations H^{final}. With row-wise ℓ_2 normalization denoted by $\bar{H}$, the reconstruction loss is

$$\mathcal{L}_{recon} = \frac{1}{N_{spot}^2} \left\| \sigma(\bar{H}\bar{H}^T) - A_{spatial} \right\|_F^2, \tag{12}$$

where σ is the element-wise sigmoid and $\|\cdot\|_F$ is the Frobenius norm. This loss fits pairwise spot similarities to spatial connectivity; Appendix A.1 analyzes its regularization of the spatial representations. Finally, we directly maximize the linear correlation between predictions and ground truth via a Pearson Correlation Coefficient (PCC) loss:

$$\mathcal{L}_{pcc} = 1 - \frac{1}{N_{gene}} \sum_{g=1}^{N_{gene}} \mathrm{Corr}(\hat{Y}_{:,g}, Y_{:,g}) \tag{13}$$

4 Theoretical analysis

4.1 Dual-graph structural regularization

The two graphs provide complementary structural constraints. DP-RAG constructs a task-specific A_{prior} that restricts gene attention to retrieved biological relations. In the spatial branch, $\mathcal{L}_{recon}$ fits the coordinate-based adjacency $A_{spatial}$ using the normalized spot representations. We analyze this spatial regularization through the following decomposition.

Proposition 4.1. *For unit-norm spot representations and a fixed nonnegative spatial adjacency, the sigmoid-based reconstruction loss decomposes into a nonlinear spatial smoothness term, a squared similarity term, and a constant. The spatial term increases with squared feature distance along spatial edges.*

As derived in Appendix A.1, the spatial term encourages nearby spots to have similar representations, while the squared similarity term discourages uniformly large pairwise similarities. This spatial objective complements the biological structure supplied by the gene attention mask.

Table 1: Average Pearson Correlation Coefficient (PCC) of the top 50 Highly Expressed Genes (HEG) and top 50 Highly Variable Genes (HVG) compared to ground truth expression. **Bold** indicates the best performance. Full detailed results including more metrics are provided in Appendix G.1.

	HER2ST		DLPFC		cSCC	
Method	HVG	HEG	HVG	HEG	HVG	HEG
ST-Net [9]	0.096±0.0439	0.101±0.0489	0.096±0.0423	0.101±0.0353	0.186±0.0975	0.180±0.0864
HisToGene [26]	0.123±0.0942	0.051±0.0732	0.042±0.0611	0.023±0.0418	0.127±0.0666	0.071±0.0383
Hist2ST [38]	0.204±0.0985	0.232±0.0671	0.060±0.0255	0.060±0.0283	0.211±0.0668	0.213±0.0629
BLEEP [37]	0.255±0.0372	0.245±0.0387	0.271±0.0417	0.266±0.1171	0.254±0.1718	0.245±0.1531
TANGLE [12]	0.135±0.0315	0.143±0.0762	0.245±0.0551	0.196±0.0391	0.093±0.0502	0.091±0.0532
OmiCLIP [3]	0.103±0.0437	0.189±0.0420	0.035±0.0183	0.242±0.1125	0.266±0.1397	0.249±0.1324
M2TGLGO [30]	0.220±0.0695	0.200±0.0969	0.204±0.0416	0.243±0.0425	0.197±0.0688	0.193±0.0575
GR2ST [41]	0.156±0.0443	0.100±0.0615	0.246±0.1220	0.193±0.0634	0.200±0.1701	0.130±0.0973
STimage [35]	0.265±0.2007	0.268±0.2248	0.175±0.0893	0.168±0.0884	0.312±0.1321	0.279±0.1148
DragH2ST(Ours)	**0.292**±0.0713	**0.315**±0.0446	**0.281**±0.0530	**0.281**±0.0490	**0.329**±0.1016	**0.281**±0.1053

Table 2: Ablation Study (Summary). We report the PCC on the all genes set (All) and the top-50 Highly Variable/Expressed Genes (HVG50/HEG50). Full detailed results including more metrics are provided in Appendix G.2.

	HER2ST			DLPFC			cSCC		
Variant	All	HVG50	HEG50	All	HVG50	HEG50	All	HVG50	HEG50
DragH2ST	**0.2142**±0.0980	**0.292**±0.0713	**0.315**±0.0446	**0.0794**±0.0137	**0.281**±0.0530	**0.281**±0.0490	**0.2514**±0.0628	**0.329**±0.1016	**0.281**±0.1053
w/o DP-RAG	0.2029±0.0976	0.285±0.0775	0.306±0.0486	0.0737±0.0145	0.265±0.0535	0.263±0.0503	0.1757±0.0702	0.283±0.0938	0.248±0.0711
w/o CGDR	0.2112±0.0972	0.288±0.0634	0.309±0.0388	0.0754±0.0137	0.275±0.0510	0.274±0.0470	0.2254±0.0723	0.308±0.0846	0.271±0.0834
w/o Spatial GCN	0.2097±0.0954	0.267±0.0723	0.302±0.0463	0.0749±0.0139	0.269±0.0531	0.268±0.0497	0.2045±0.0499	0.291±0.0837	0.269±0.0630

4.2 CGDR: dynamic soft rewiring and computational efficiency

Building on the soft-gating mechanism, we analyze how CGDR conditions information flow on the local tissue context and the cost of its additional modulation. Explicitly constructing a dense adjacency matrix for each spot requires $O(N_{gene}^2)$ entries.

Proposition 4.2. *For a shared GAT-encoded gene basis B, CGDR induces a context-aware basis $Z^{(i)} = \mathrm{diag}(\alpha_i)B$. Decoding with this basis is exactly equivalent to gene-wise scaling of the shared-basis regulatory prediction. With fixed visual and hidden dimensions, gate generation and output modulation add $O(N_{gene})$ work per spot.*

As derived in Appendix A.2, the positive gates yield a diagonal modulation of the gene basis and its effective contributions to local decoding. GAT encoding is shared across spots, and the linear bound describes the additional gate generation and output modulation. This supports microenvironment-adaptive use of the biological prior with an efficient context-conditioned decoder.

5 Experiments and results

5.1 Experimental setup

Datasets. We evaluate DragH2ST on three public spatial transcriptomics datasets: HER2ST [1], DLPFC [21], and cSCC [13]. HER2ST features highly heterogeneous breast cancer regions, testing

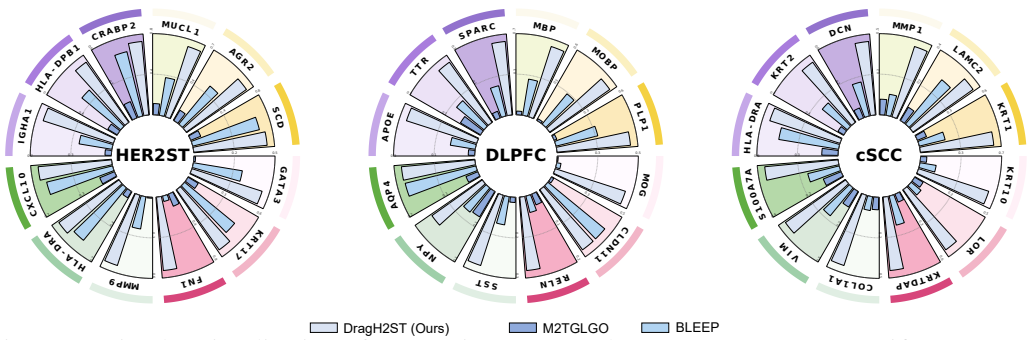


Figure 3: Circular visualization of Gene-wise PCC. Each sector represents a specific gene (e.g., *MUCL1*, *FN1*), and the bars within each sector illustrate the PCC performance of different models.

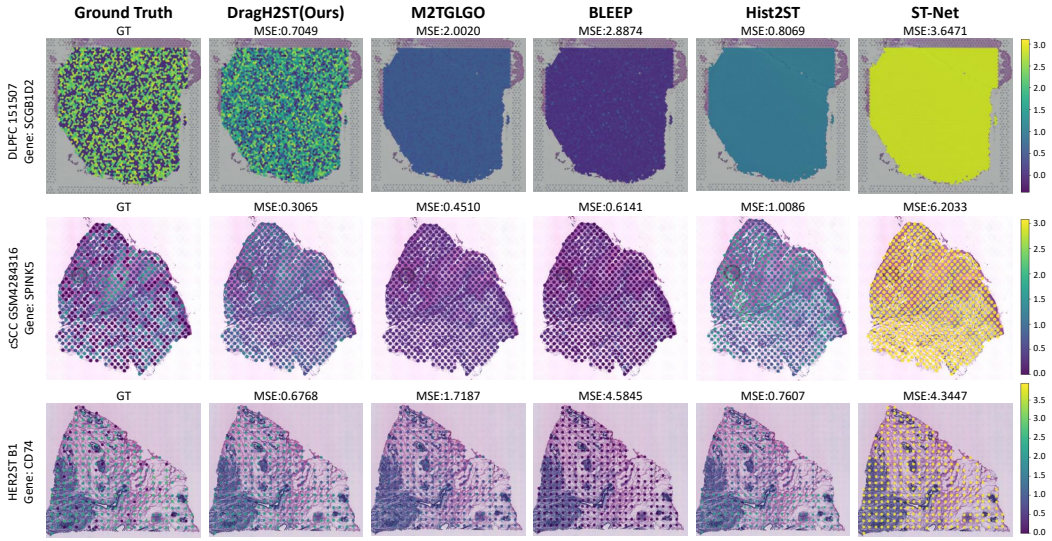


Figure 4: Visualization of spatially resolved gene expression for marker genes.

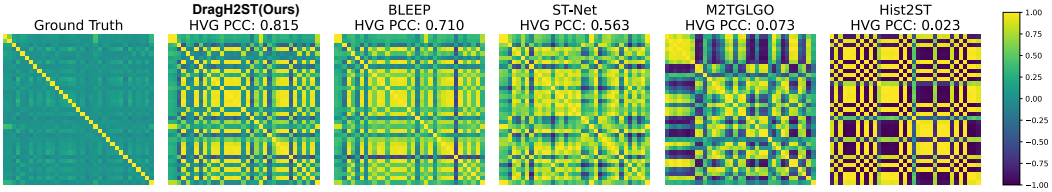


Figure 5: Visualization of gene–gene correlation (GGC) matrices. All panels use the same $[-1, 1]$ color scale, which encodes correlation sign and magnitude. Positive and negative regions, stripes, and module boundaries show correlation structure; the reported PCC measures agreement between predicted and ground-truth GGC matrices.

the model's dynamic adaptability to complex microenvironments. DLPFC exhibits strict laminar cortical structures, serving as a benchmark for maintaining structural integrity under sequencing noise. cSCC involves complex epithelial tumor architectures, testing the model's capability to delineate fine-grained tumor–stroma boundaries. Dataset statistics and preprocessing procedures are provided in Appendix B.1.

Implementation details. We adopted a Leave-One-Slide-Out Cross-Validation strategy for each dataset. For gene expression prediction, we selected HVGs as the prediction targets. We use a shared hyperparameter configuration across the three datasets. Detailed training configurations and hyperparameters are provided in Appendix B.2.

Evaluation metrics. Following the protocols in [30, 38], we use Pearson correlation coefficient (PCC), PCC on Highly Variable Genes (HVGs), and PCC on Highly Expressed Genes (HEGs) as the primary metrics, and Mean Squared Error (MSE) to measure reconstruction error. We further report PCC on Lowly Variable Genes (LVGs) and Lowly Expressed Genes (LEGs) to assess robustness. Higher PCC and lower MSE indicate better performance.

5.2 Quantitative comparison with SOTA methods

We compared DragH2ST with methods based on pure visual features (such as pure CNNs), methods for modeling long-range or local spatial dependencies (such as Transformers/Graphs), and methods based on multimodal or static prior knowledge to validate the effectiveness of DragH2ST.

Gene-wise performance analysis. As shown in Table 1, our method achieves higher PCC on the reported gene subsets across all three datasets. The full model combines a task-specific prior, spatial information, and spot-conditioned gene representations to model context-dependent expression patterns. Competitive methods such as M2TGLGO [30] and BLEEP [37] also perform well on HEG and HVG subsets; results on additional gene subsets and metrics are provided in Appendix G.

Furthermore, we evaluate DragH2ST across multiple metrics and gene subsets. It consistently outperforms all baselines on most indicators, with especially notable gains in heterogeneous tissues such as HER2ST and cSCC. Detailed results are provided in Appendix G.

Specific marker gene analysis. Furthermore, we analyzed performance on representative marker genes with distinct biological roles and spatial patterns. For each dataset, we curated genes that mark key anatomical or pathological structures, covering functional categories such as structural proteins, cell-type markers, and signaling molecules (detailed criteria and gene lists are provided in Appendix H). As shown in Fig. 3, DragH2ST achieves higher gene-wise PCC than the compared baselines for representative markers, including *MUCL1* and *FN1* in HER2ST, *MBP* in DLPFC, and *KRT1* in cSCC.

5.3 Visualization results

Spatially resolved gene expression visualization. Fig. 4 compares predicted spatial expression for representative marker genes. Methods such as Hist2ST sometimes produce smoother expression maps that miss sharp transitions and heterogeneous hotspots. The full DragH2ST framework better recovers complex spatial patterns while preserving fine-grained tissue structures. These visualizations illustrate the spatial predictions obtained by combining spatial graph modeling with context-conditioned gene representations.

Gene-gene correlation (GGC) analysis. Fig. 5 visualizes the GGC matrices for the top highly variable genes. The ground truth matrix exhibits distinct modular patterns, representing co-regulated gene programs. DragH2ST accurately reconstructs these regulatory modules, achieving a high correlation coefficient (PCC = 0.815) with the ground truth structure. In contrast, baseline methods such as Hist2ST struggle to capture these complex inter-gene dependencies, resulting in noisy or unstructured correlation matrices (e.g., Hist2ST, PCC = 0.023). This demonstrates that our dual-stage retrieval framework not only improves point-wise prediction accuracy but also effectively preserves the biological topology of the transcriptome.

Biological validity and spatial heterogeneity of gene associations. For supplementary analysis, we evaluate gene association patterns derived from DP-RAG priors and CGDR gates from functional and spatial perspectives. The results show biologically supported gene associations and spot-level spatial heterogeneity consistent with local tissue context. More detailed module enrichment analyses, subnetwork visualizations, and evaluations of spatial heterogeneity are provided in the Appendix F.

5.4 Ablation study

Table 2 summarizes the component ablations. Removing DP-RAG lowers PCC, indicating that task-relevant biological priors help expression prediction. With DP-RAG retained, removing CGDR also reduces predictive performance, suggesting that local histology-conditioned modulation further complements the task-level prior. Removing Spatial GCN likewise reduces performance, showing the contribution of neighborhood information. Together, these results support complementary roles for the biological prior, local histological context, and spatial structure.

6 Conclusion

In this paper, we introduced DragH2ST, a framework that formulates virtual spatial transcriptomics as dynamic regulatory inference beyond static priors. Through dual-stage retrieval, DragH2ST integrates task-level biological knowledge before training and provides prototype cases for interpretation at inference. DP-RAG constructs a task-specific biological prior, while CGDR modulates the shared GAT-encoded gene basis according to local histology for context-conditioned expression decoding. Together with a dual-graph architecture and multi-objective optimization, the full model improves robustness to batch effects and sequencing noise. Experiments show strong predictive performance, gene relationships supported by biological knowledge, and spatially heterogeneous expression patterns. These findings support its use for molecular phenotyping research in computational pathology.

Acknowledgments and Disclosure of Funding

Funding. This work was supported in part by the National Natural Science Foundation of China (62506220, 62476157, 62471148), Fundamental Research Program of Shanxi Province (202403021222024), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2024L002), Special Project for Patent Transformation in Shanxi Province (20250016), and 2025 High-Quality Industrial Development Special Program for Scientific Intelligence "Baituan Baixiang" (Second Batch) (2025-GZL-RGZN-BTBX-02015).

Competing interests. The authors declare no competing interests.

References

[1] Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z. Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nature Communications*, 12(1):6012, 2021. doi: 10.1038/s41467-021-26271-2.

[2] Maria Addolorata Bonifacio and Maria Addolorata Mariggiò. New trends in pathology: From cell morphology to molecular medicine. *International Journal of Molecular Sciences*, 24(14): 11743, 2023. doi: 10.3390/ijms241411743.

[3] Weiqing Chen, Pengzhi Zhang, Tu N. Tran, Yiwei Xiao, Shengyu Li, Vrutant V. Shah, Hao Cheng, Kristopher W. Brannan, Keith Youker, Li Lai, et al. A visual-omics foundation model to bridge histopathology with spatial transcriptomics. *Nature Methods*, 22(7):1568–1582, 2025. doi: 10.1038/s41592-025-02707-1.

[4] Jingcheng Du, Peilin Jia, Yulin Dai, Cui Tao, Zhongming Zhao, and Degui Zhi. Gene2vec: distributed representation of genes based on co-expression. *BMC Genomics*, 20(Suppl 1):82, 2019. doi: 10.1186/s12864-018-5370-x.

[5] Yiwen Du, Kun Xu, Siwen Zhang, Lanming Chen, Zhenhao Liu, and Lu Xie. Construction of gene regulatory networks based on spatial multi-omics data and application in tumor-boundary analysis. *Genes*, 16(7):821, 2025. doi: 10.3390/genes16070821.

[6] Haitham A. Elmarakeby, Justin Hwang, Rand Arafeh, Jett Crowdis, Sydney Gang, David Liu, Saud H. AlDubayan, Keyan Salari, Steven Kregel, Camden Richter, et al. Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598(7880):348–352, 2021. doi: 10.1038/s41586-021-03922-4.

[7] Daniela Grases and Eduard Porta-Pardo. A practical guide to spatial transcriptomics: lessons from over 1000 samples. *Trends in Biotechnology*, 44(5):1230–1242, 2026. doi: 10.1016/j.tibtech.2025.08.020.

[8] Max Hallemeesch, Marija Pizurica, Paloma Rabaey, Olivier Gevaert, Thomas Demeester, and Kathleen Marchal. Prior knowledge injection into deep learning models predicting gene expression from whole slide images. arXiv preprint arXiv:2501.14056, 2025.

[9] Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4(8):827–834, 2020. doi: 10.1038/s41551-020-0578-x.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/cvpr.2016.90.

[11] Arfa Jabin, Jyoti Singh Kirar, and Shandar Ahmad. AI-based methods for modelling whole-slide imaging data in cancer diagnosis and transcriptome profile prediction. *BMC Artificial Intelligence*, 1(1):16, 2025. doi: 10.1186/s44398-025-00018-1.

[12] Guillaume Jaume, Lukas Oldenburg, Anurag Vaidya, Richard J Chen, Drew FK Williamson, Thomas Peeters, Andrew H Song, and Faisal Mahmood. Transcriptomics-guided slide representation learning in computational pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9632–9644, 2024. doi: 10.1109/cvpr52733.2024.00920.

[13] Andrew L Ji, Adam J Rubin, Kim Thrane, Sizun Jiang, David L Reynolds, Robin M Meyers, Margaret G Guo, Benson M George, Annelie Mollbrink, Joseph Bergenstråhle, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, 182 (2):497–514.e22, 2020. doi: 10.1016/j.cell.2020.05.039.

[14] Mikko J. Lammi and Chengjuan Qu. Spatial transcriptomics, proteomics, and epigenomics as tools in tissue engineering and regenerative medicine. *Bioengineering*, 11(12):1235, 2024. doi: 10.3390/bioengineering11121235.

[15] Kongming Li, Jiahao Li, Yuhao Tao, and Fei Wang. stDiff: a diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Briefings in Bioinformatics*, 25(3): bbae171, 2024. doi: 10.1093/bib/bbae171.

[16] Yurui Li, Jin Chen, and Haohan Wang. Spatially varying cell-specific gene regulation network inference. *bioRxiv*, 2025. doi: 10.1101/2025.10.17.683188.

[17] Arthur Liberzon, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P. Mesirov. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27(12): 1739–1740, 2011. doi: 10.1093/bioinformatics/btr260.

[18] Xuepeng Liu, Zheng Jiang, Pinan Zhu, Hanyu Liu, and Chao Li. HaDM-ST: Histology-assisted differential modeling for spatial transcriptomics generation. arXiv preprint arXiv:2508.07225, 2025.

[19] Hiren Madhu, João Felipe Rocha, Tinglin Huang, Siddharth Viswanath, Smita Krishnaswamy, and Rex Ying. HEIST: A graph foundation model for spatial transcriptomics and proteomics data. arXiv preprint arXiv:2506.11152, 2025.

[20] Sejuti Majumder, Saarthak Kapse, Moinak Bhattacharya, Xuan Xu, Alisa Yurovsky, and Prateek Prasanna. PEaRL: Pathway-enhanced representation learning for gene and pathway expression prediction from histology. arXiv preprint arXiv:2510.03455, 2025.

[21] Kristen R. Maynard, Leonardo Collado-Torres, Lukas M. Weber, Cedric Uytingco, Brianna K. Barry, Stephen R. Williams, Joseph L. Catallini, Matthew N. Tran, Zachary Besich, Madhavi Tippani, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*, 24(3):425–436, 2021. doi: 10.1038/s41593-020-00787-0.

[22] Wenwen Min, Zhiceng Shi, Jun Zhang, Jun Wan, and Changmiao Wang. Multimodal contrastive learning for spatial gene expression prediction using histology images. *Briefings in Bioinformatics*, 25(6):bbae551, 2024. doi: 10.1093/bib/bbae551.

[23] Kazuya Nishimura, Haruka Hirose, Ryoma Bise, Kaito Shiku, and Yasuhiro Kojima. Learning relative gene expression trends from pathology images in spatial transcriptomics. arXiv preprint arXiv:2512.06612, 2025.

[24] Kalin Nonchev, Sebastian Dawo, Karina Silina, Holger Moch, Sonali Andani, Tumor Profiler Consortium, Viktor H Koelzer, and Gunnar Rätsch. DeepSpot: Leveraging spatial context for enhanced spatial transcriptomics prediction from H&E images. *medRxiv*, 2025. doi: 10.1101/2025.02.09.25321567.

[25] Kalin Nonchev, Glib Manaiev, Viktor H Koelzer, and Gunnar Rätsch. DeepSpot2Cell: Predicting virtual single-cell spatial transcriptomics from H&E images using spot-level supervision. *bioRxiv*, 2025. doi: 10.1101/2025.09.23.678121.

[26] Minxing Pang, Kenong Su, and Mingyao Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *bioRxiv*, 2021. doi: 10.1101/2021.11.28.470212.

[27] Tancredi Massimo Pentimalli, Nikos Karaiskos, and Nikolaus Rajewsky. Challenges and opportunities in the clinical translation of high-resolution spatial transcriptomics. *Annual Review of Pathology: Mechanisms of Disease*, 20(1):405–432, 2025. doi: 10.1146/annurev-pathmechdis-111523-023417.

[28] Mingcheng Qu, Yuncong Wu, Donglin Di, Yue Gao, Tonghua Su, Yang Song, and Lei Fan. Spatially gene expression prediction using dual-scale contrastive learning. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15974 of *Lecture Notes in Computer Science*, pages 574–584. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-032-05182-0_56.

[29] Benoît Schmauch, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro, Aurélie Kamoun, Meriem Sefta, Sylvain Toldo, Mikhail Zaslavskiy, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nature Communications*, 11(1):3877, 2020. doi: 10.1038/s41467-020-17678-4.

[30] Hang Shi, Changxi Chi, Peng Wan, Daoqiang Zhang, and Wei Shao. Multi-modal topology-embedded graph learning for spatially resolved genes prediction from pathology images with prior gene similarity information. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20810–20819, 2025. doi: 10.1109/cvpr52734.2025.01938.

[31] Kelly D Smith, David K Prince, James W MacDonald, Theo K Bammler, and Shreeram Akilesh. Challenges and opportunities for the clinical translation of spatial transcriptomics technologies. *Glomerular Diseases*, 4(1):49–63, 2024. doi: 10.1159/000538344.

[32] Albrecht Stenzinger, Maximilian Alber, Michael Allgäuer, Philipp Jurmeister, Michael Bockmayr, Jan Budczies, Jochen Lennerz, Johannes Eschrich, Daniel Kazdal, Peter Schirmacher, et al. Artificial intelligence and pathology: From principles to practice and future applications in histomorphology and molecular profiling. *Seminars in Cancer Biology*, 84:129–143, 2022. doi: 10.1016/j.semcancer.2021.02.011.

[33] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005. doi: 10.1073/pnas.0506580102.

[34] Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L Gable, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, 2023. doi: 10.1093/nar/gkac1000.

[35] Xiao Tan, Onkar Mulay, Jacky Xie, Samual MacDonald, Taehyun Kim, Chenhao Zhou, Zherui Xiong, Samuel X. Tan, Nan Ye, Amy McCart Reed, et al. Robust and interpretable prediction of gene markers and cell types from spatial transcriptomics data. *Nature Communications*, 17 (1):1781, 2026. doi: 10.1038/s41467-026-68487-0.

[36] The Gene Ontology Consortium, Suzi A Aleksander, James Balhoff, Seth Carbon, J Michael Cherry, Harold J Drabkin, Dustin Ebert, Marc Feuermann, Pascale Gaudet, Nomi L Harris, et al. The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1):iyad031, 2023. doi: 10.1093/genetics/iyad031.

[37] Ronald Xie, Kuan Pang, Sai Chung, Catia Perciani, Sonya MacParland, Bo Wang, and Gary Bader. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 70626–70637, 2023. doi: 10.52202/075280-3095.

[38] Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5):bbac297, 2022. doi: 10.1093/bib/bbac297.

[39] Chen Zhang, Yilu An, Ying Chen, Hao Li, Xitong Ling, Lihao Liu, Junjun He, Yuxiang Lin, Zihui Wang, and Rongshan Yu. HyperST: Hierarchical hyperbolic learning for spatial transcriptomics prediction. arXiv preprint arXiv:2511.22107, 2025.

[40] Daiwei Zhang, Amelia Schroeder, Hanying Yan, Haochen Yang, Jian Hu, Michelle Y. Y. Lee, Kyung S. Cho, Katalin Susztak, George X. Xu, Michael D. Feldman, et al. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nature Biotechnology*, 42(9):1372–1377, 2024. doi: 10.1038/s41587-023-02019-9.

[41] Jingli Zhou, Siyuan Li, Rui Han, Xuan Wang, Yadong Wang, and Junyi Li. GR2ST: Spatial transcriptomics prediction based on graph-enhanced multimodal contrastive learning. *Bioinformatics*, 42(5):btag209, 2026. doi: 10.1093/bioinformatics/btag209.

[42] Jiawei Zou, Kai Xiao, Zexi Chen, Jiazheng Pei, Jing Xu, Tao Chen, Likun Hou, Chunyan Wu, Yunlang She, Zhiyuan Yuan, et al. Predicting spatial transcriptomics from H&E image by pretrained contrastive alignment learning. *bioRxiv*, 2025. doi: 10.1101/2025.06.15.659438.

A Proofs of theoretical analysis

A.1 Proof of spatial structural regularization

Recap: Proposition 4.1 characterizes the spatial reconstruction loss in Eq. (12) as a nonlinear neighborhood regularizer together with a squared similarity term.

Proof. Let $A = A_{spatial}$ be fixed and nonnegative, $N = N_{spot}$, and $\bar{h}_i$ denote the unit-norm representation of spot i. Define $S = \sigma(\bar{H}\bar{H}^T)$. Since

$$\bar{h}_i^T \bar{h}_j = 1 - \frac{1}{2}\|\bar{h}_i - \bar{h}_j\|^2, \tag{14}$$

we can define a distance penalty $\psi(t) = \sigma(1) - \sigma(1 - t/2)$ for $t \in [0, 4]$. It satisfies $\psi(0) = 0$ and

$$\psi'(t) = \frac{1}{2}\sigma(1 - t/2)\big[1 - \sigma(1 - t/2)\big] > 0. \tag{15}$$

Thus the nonnegative neighborhood energy

$$\begin{aligned} \mathcal{R}_{spatial} &= \sum_{i,j} A_{ij}\psi\big(\|\bar{h}_i - \bar{h}_j\|^2\big) \\ &= \sigma(1)\sum_{i,j} A_{ij} - \sum_{i,j} A_{ij}S_{ij} \end{aligned} \tag{16}$$

increases with feature distance on spatial edges. Expanding the implemented loss gives

$$\begin{aligned} \mathcal{L}_{recon} &= \frac{1}{N^2}\left(\|S\|_F^2 - 2\sum_{i,j} A_{ij}S_{ij} + \|A\|_F^2\right) \\ &= \frac{1}{N^2}\left(2\mathcal{R}_{spatial} + \|S\|_F^2 + C_A\right), \end{aligned} \tag{17}$$

where $C_A = \|A\|_F^2 - 2\sigma(1)\sum_{i,j} A_{ij}$ is independent of the spot representations. The first term encourages neighborhood smoothness, while $\|S\|_F^2$ penalizes large pairwise similarity scores. This establishes the decomposition for the spatial reconstruction objective, including its sigmoid transformation. □

A.2 Analysis of dynamic soft rewiring

Recap: CGDR applies histology-derived gene gates to a shared GAT-encoded basis, yielding context-conditioned regulatory decoding.

Proof. Let $B = \mathrm{GAT}(E_{\mathrm{global}}, A_{prior}) \in \mathbb{R}^{N_{gene} \times D_v}$ and $D_i = \mathrm{diag}(\alpha_i)$ for spot i. The context-aware basis is $Z^{(i)} = D_i B$. For the row-vector spot representation h_i^{final}, regulatory decoding satisfies

$$\hat{Y}_i^{reg} = h_i^{final}(D_i B)^\top = \left(h_i^{final} B^\top\right) D_i = \left(h_i^{final} B^\top\right) \odot \alpha_i^\top. \tag{18}$$

Thus, scaling each row of the gene basis is exactly equivalent to scaling the corresponding regulatory output. The implementation uses the latter form and evaluates the shared GAT once per tissue forward pass.

The effective contribution of the biological prior can also be expressed for the single-layer ReLU GAT. Let β_{uv} be its shared attention coefficient from gene v to gene u, and e_v the input gene embedding. Since $\alpha_{i,u} > 0$, positive homogeneity of ReLU gives

$$\begin{aligned} z_u^{(i)} &= \alpha_{i,u}\,\mathrm{ReLU}\left(\sum_{v \in \mathcal{N}(u)} \beta_{uv} W e_v\right) \\ &= \mathrm{ReLU}\left(\sum_{v \in \mathcal{N}(u)} \alpha_{i,u}\beta_{uv} W e_v\right). \end{aligned} \tag{19}$$

The resulting effective contribution matrix is

$$C^{(i)} = D_i\beta, \qquad C^{(i)}_{uv} = \alpha_{i,u}\beta_{uv}, \qquad \sum_v C^{(i)}_{uv} = \alpha_{i,u}. \tag{20}$$

Here, β describes shared GAT attention, and $C^{(i)}$ describes its effective contribution to the spot-specific decoder. Positive diagonal scaling preserves the prior-supported sparsity and modulates each gene's aggregate contribution according to local tissue context.

With fixed visual and hidden dimensions, generating N_{gene} gates and scaling the regulatory outputs takes $O(N_{gene})$ additional work per spot. The equivalent output-scaling implementation reuses the shared basis B throughout the tissue forward pass. □

The gate interval $(0.5, 1.5)$ enables soft attenuation and amplification of the prior-guided gene representations around unit scaling.

B Detailed experimental setup

B.1 Datasets and preprocessing

Datasets. We utilized three datasets across different tissue types primarily based on 10x Visium and ST platforms:

- **HER2ST [1]**[1]: Focused on HER2-positive breast cancer, comprising 36 slides. This dataset features highly heterogeneous tumor regions, requiring the model to distinguish fine-grained spatial gene expression changes driven by local tumor pathology rather than relying solely on global statistical regularities. We followed previous protocols [38] to select 12 slides for evaluation.
- **DLPFC [21]**[2]: Consists of 12 Visium slides from the dorsolateral prefrontal cortex. The cortex exhibits a strict and unique laminar structure. This dataset serves as a benchmark for structural consistency, testing the model's ability to recover anatomical integrity against variations in sequencing depth.
- **cSCC [13]**[3]: Contains 12 slides of human cutaneous squamous cell carcinoma. It features complex epithelial tumor structures and dense stromal infiltration. We used it to evaluate the model's capability in delineating fine-grained tumor-stroma boundaries and handling histological structures significantly different from other datasets.

Preprocessing. For all datasets, we followed standard preprocessing pipelines. Image patches were cropped to 112×112 pixels and resized to 224×224 before feature extraction with a ResNet50[10] visual encoder. For gene expression, we first applied library-size normalization, followed by the base-10 logarithmic transformation $\log_{10}(x + 1)$. For HER2ST, we selected 785 HVG following the protocol of [38]; for DLPFC and cSCC, we selected 1000 HVG as prediction targets. We adopted a Leave-One-Slide-Out Cross-Validation strategy for robust evaluation.

B.2 Implementation details

The model was implemented using the PyTorch Lightning framework and trained on NVIDIA A800 GPUs. We used the Adam optimizer with an initial learning rate of 7.5×10^{-6} and a dropout rate of 0.2 for the spatial encoder. Training ran for at most 500 epochs with a batch size of 1, random seed 12000, and early stopping based on validation PCC with patience 50.

DP-RAG priors were constructed before training with a similarity threshold of 0.290 and row-wise top-k selection with $k = 65$. The source fusion weights were 0.242 for Gene2Vec, 0.096 for GO, 0.098 for STRING, and 0.181 for MSigDB. The prediction fusion coefficient was $\lambda_{mix} = 0.749$. The loss weights were $\lambda_{mse} = 26.427$, $\lambda_{zinb} = 0.071$, $\lambda_{rank} = 0.525$, $\lambda_{pcc} = 1.967$, and $\lambda_{recon} = 0.165$. Prototype retrieval used 32 clusters per training slide and selected the nearest prototypes by cosine similarity.

[1] https://github.com/almaan/her2st.git
[2] https://research.libd.org/spatialLIBD/
[3] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144239

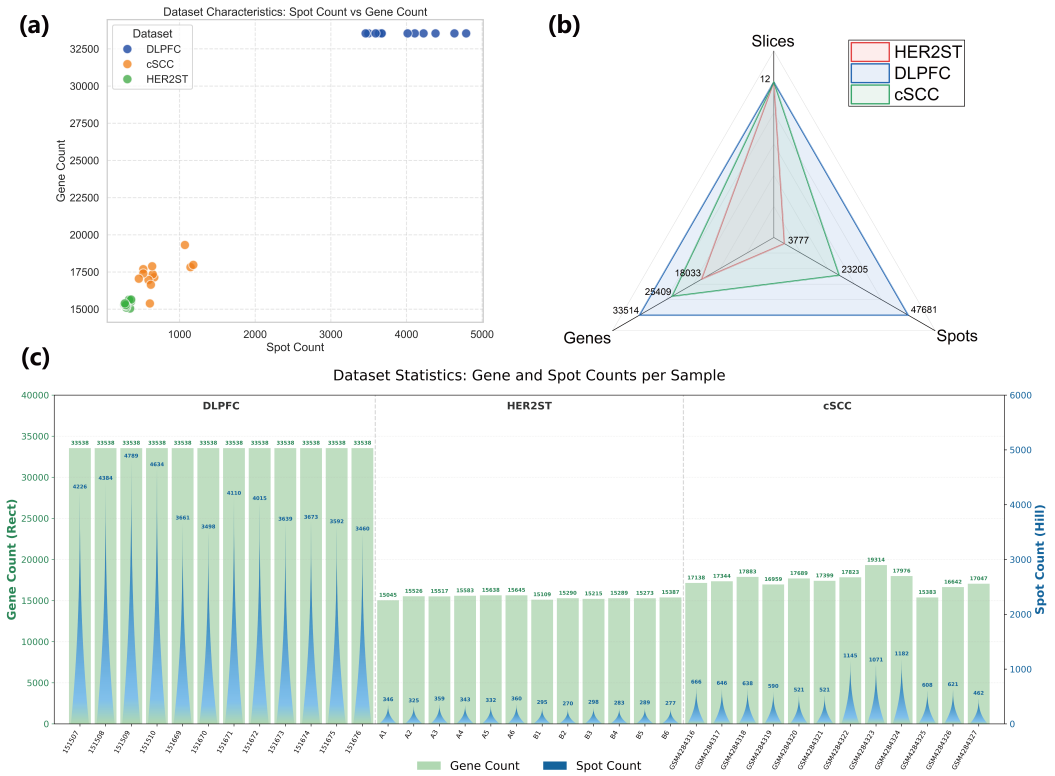


Figure 6: Detailed statistics of the three benchmarks: DLPFC, HER2ST, and cSCC. The dual-axis plot visualizes the Gene Count (green bars, left axis) and Spot Count (blue hills, right axis) for each sample. This visualization highlights the variation in data sparsity and sequencing depth across different tissue types.

C DP-RAG implementation and task-level dynamicity

Unlike traditional static-prior methods that rely on a single-source, fixed, and globally shared prior graph, DP-RAG does not depend on a predefined gene graph that remains unchanged across tasks when constructing the consensus prior gene graph A_{prior}. Instead, it uses the target gene set of the current prediction task as the query condition to dynamically retrieve, fuse, and filter task-relevant biological relations from a multi-source knowledge pool, thereby forming a task-specific prior topology. After obtaining the multi-source relation matrices relevant to the current task, we perform weighted fusion to obtain the fused similarity matrix S_{fused}:

$$S_{\text{fused}} = \frac{\sum_{k=1}^{4} w_k S_k}{\sum_{k=1}^{4} w_k} \tag{21}$$

where S_k denotes the similarity matrix constructed from source k over the target gene set of the current task, and w_k denotes the corresponding fusion weight for that knowledge source. Since different tasks involve different target gene sets and associated biological relations, the resulting similarity matrices and prior topologies vary accordingly.

The dynamicity of DP-RAG lies in **task-level prior construction** driven by the target gene set. In other words, within the same leave-one-slide-out fold, all spots share the same base prior graph; however, when the prediction task changes, the retrieved candidate relations, fusion results, and the final prior topology also change accordingly because the target gene set is different. Therefore, DP-RAG is not a simple reuse of a globally fixed static graph, but rather a task-driven process of multi-source knowledge reorganization.

The four knowledge sources are constructed as follows:

1. **Gene2Vec** [4] (S_{g2v}): We utilize pre-trained gene embeddings obtained from large-scale transcriptomic corpora (based on GEO meta-analysis). The similarity is computed as the cosine similarity between gene embedding vectors, capturing data-driven co-expression patterns.
2. **Gene Ontology** [36] (S_{GO}): We parse the Gene Ontology (GO) hierarchy (`go-basic.obo`) and human annotations (`goa_human.gaf`). For GO terms t and u in the same namespace, we compute an ancestor-overlap similarity:

$$s(t,u) = \frac{|\operatorname{Anc}(t) \cap \operatorname{Anc}(u)|}{\max(|\operatorname{Anc}(t)|+1, |\operatorname{Anc}(u)|+1)},$$

   where $\operatorname{Anc}(t)$ is the set of ancestors of t in the GO hierarchy. Term pairs from different namespaces receive zero similarity. The gene-level score averages the positive term-pair similarities across the two genes' annotation sets, with zero assigned when no positive score is available. Diagonal entries of S_{GO} are set to one.
3. **STRING PPI** [34] (S_{PPI}): We utilize protein-protein interaction (PPI) data from the STRING database (v12.0). After mapping genes to STRING proteins, we combine direct interaction strength with shared-neighbor similarity as $(S_{PPI})_{ij} = 0.7d_{ij} + 0.3J_{ij}$ for $i \neq j$. Here, d_{ij} is the retained interaction's `combined_score` divided by 1000 (zero for an absent edge), and $J_{ij} = |\mathcal{N}_i \cap \mathcal{N}_j|/|\mathcal{N}_i \cup \mathcal{N}_j|$ is the Jaccard similarity of the retained protein-neighbor sets (zero for an empty union). Pairs with missing protein mappings receive zero similarity, and diagonal entries of S_{PPI} are set to one.
4. **MSigDB** [17, 33] (S_{MSigDB}): We employ the C2 (Canonical Pathways) subset of the Molecular Signatures Database (MSigDB). The similarity between genes u and v is calculated using the Jaccard Index of the sets of pathways they participate in:

$$S_{uv} = \frac{|\mathcal{P}_u \cap \mathcal{P}_v|}{|\mathcal{P}_u \cup \mathcal{P}_v|}$$

Finally, we clip the fused scores to $[0,1]$ and set diagonal entries to one. We then apply a similarity threshold t and row-wise top-k selection, binarize retained connections, and add self-loops to obtain A_{prior} for GAT propagation. This means that candidate relations from multi-source knowledge are not inherited indiscriminately; instead, only the local prior topology most relevant and reliable for the current task is preserved under the constraint of the task objective.

On top of this task-level prior construction, the subsequent CGDR further performs spot-level local context-adaptive modulation, and together they form the dual-level dynamic mechanism emphasized in this work.

D Inference-stage prototype retrieval for visual evidence

Prototype retrieval offers case-based references for interpreting molecular predictions. After prediction, the model retrieves morphologically similar tissue patches with measured expression profiles from the training-set reference library.

For a query spot in a dense tumor region, retrieved prototypes show similar tissue morphology, including a high nuclear-to-cytoplasmic ratio and dense cellularity, together with measured expression patterns such as high *SCGB2A2* and *MBP*. Comparing these cases with the query links its prediction to recognizable morphological patterns and observed molecular profiles.

E CGDR implementation and spot-level dynamicity

E.1 Implementation and spot-level dynamicity

Unlike DP-RAG, whose dynamicity is expressed at the task level through the construction of a shared prior graph for the current prediction target, CGDR is responsible for the **spot-level dynamicity** of the framework. Within the same leave-one-slide-out fold, all spots share the base prior graph A_{prior} constructed by DP-RAG. CGDR does not perform an additional external knowledge retrieval for each spot; instead, it adaptively modulates how this shared task-level prior is utilized according to the

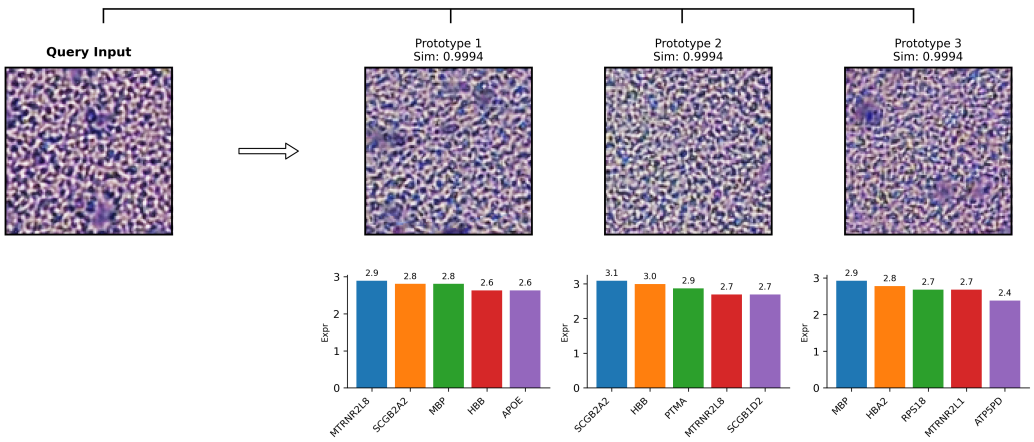


Figure 7: Visualization of prototype retrieval (RAG). Given a query tissue patch (left), the figure shows the three most similar historical prototypes retrieved from the training set based on visual features. Their similarity scores (Sim) and measured expression profiles provide case-based molecular references for interpreting the prediction.

local visual context of each individual spot. In other words, the dynamicity introduced by CGDR does not arise from per-spot prior retrieval, but from **spot-specific prior utilization and rewiring**.

Specifically, for the i-th spot, let $h_i^{final} \in \mathbb{R}^{D_v}$ denote the spatially enhanced visual representation. We introduce a lightweight gating network $\mathcal{F}_{gate}$ to generate a spot-specific gene modulation vector:

$$\alpha_i^{raw} = \sigma(W_{g2} \cdot \mathrm{ReLU}(W_{g1} h_i^{final} + b_{g1}) + b_{g2}), \tag{22}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function. To avoid the training instability caused by hard binary gating, the raw sigmoid output is not used directly as the final modulation coefficient. Instead, it is further mapped to a Soft Gating interval through affine rescaling. Concretely, if the raw sigmoid output is denoted by α_i^{raw}, then the actual modulation coefficient used for graph rewiring is defined as

$$\alpha_i^{soft} = 0.5 + \alpha_i^{raw}, \tag{23}$$

which gives $\alpha_i^{soft} \in [0.5, 1.5]$. This design enables soft suppression when $\alpha < 1$ and soft enhancement when $\alpha > 1$, while avoiding exact zero-valued gates that could disrupt gradient propagation.

The global embeddings $E_{global} \in \mathbb{R}^{N_{gene} \times D_e}$ are first encoded by the shared Gene GAT. The resulting basis is then modulated by the spot-specific gates:

$$B = \mathrm{GAT}(E_{global}, A_{prior}), \qquad Z^{(i)} = \mathrm{diag}(\alpha_i^{soft}) B. \tag{24}$$

CGDR turns the shared prior-guided gene basis into a context-conditioned decoder. In the implementation, the same operation is evaluated as gene-wise scaling of the regulatory prediction $h_i^{final} B^\top$, as derived in Appendix A.2. Together with DP-RAG, this forms a dual-level dynamic mechanism spanning task-level prior construction and local adaptation.

E.2 Empirical efficiency analysis of CGDR

For the supplementary efficiency analysis, we evaluate CGDR, which applies context gates to the shared gene basis after GAT encoding. Table 3 compares its training time, peak memory, and predictive performance with explicit per-spot dense graph construction.

Gate generation and feature scaling have $O(N_{gene})$ additional cost for fixed feature dimensions; the table reports the measured training costs of the compared implementations.

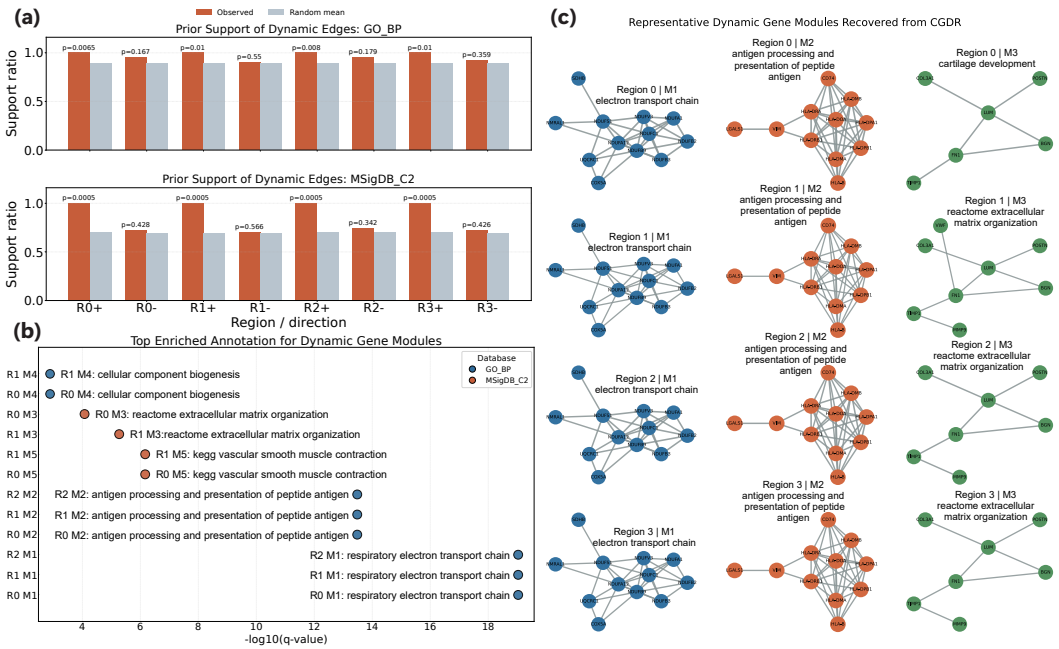


Figure 8: **Biological validity of the gene association graph derived from DP-RAG priors and CGDR gates.** **(a)** Prior support of dynamically reinforced edges versus random edge sets. **(b)** Top enriched annotations of dynamic gene modules. **(c)** Representative dynamic gene module subnetworks.

Table 3: Efficiency comparison between CGDR and explicit per-spot graph construction.

Dataset	fold	Nodes	Method	Edges	Train Time (s)	Peak Mem (MB)	PCC-All
HER2ST	7	785	CGDR	2053	437.39	5841.1	0.3482
HER2ST	7	785	Per-Spot Graph	616225	2141.11	11201.44	0.3517
cSCC	2	1000	CGDR	13909	1158.02	19960.14	0.2428
cSCC	2	1000	Per-Spot Graph	1000000	9914.48	49880.42	0.2242

As shown in Table 3, compared with explicit per-spot dense dynamic graph construction, CGDR offers substantial advantages in both training time and peak memory consumption, while maintaining overall comparable predictive performance. For example, on HER2ST fold 7, CGDR reduces the training time from 2141.11 s to 437.39 s and the peak memory usage from 11201.44 MB to 5841.1 MB. On cSCC fold 2, the training time is reduced from 9914.48 s to 1158.02 s, while peak memory usage decreases from 49880.42 MB to 19960.14 MB. Meanwhile, the PCC-All values of CGDR remain comparable to those of explicit dynamic graph construction on both datasets, indicating that the proposed design can significantly improve efficiency while largely preserving the predictive effectiveness of dynamic rewiring.

Taken together, the **spot-level dynamicity** embodied by CGDR does not arise from reconstructing an entirely new prior graph for each spot, but from **spot-specific gating/modulation over a shared task-level prior**. This design preserves adaptability to local microenvironmental variation while avoiding the prohibitive computational cost of explicit dense dynamic graph construction, thereby providing empirical support for efficient spot-level dynamic rewiring.

F Biological validity and spatial organization of the dual-level dynamic synergy between DP-RAG and CGDR

For supplementary biological analysis, we use CGDR, with context gates applied to the shared gene basis after GAT encoding. We examine the learned gene-association structure through the functions

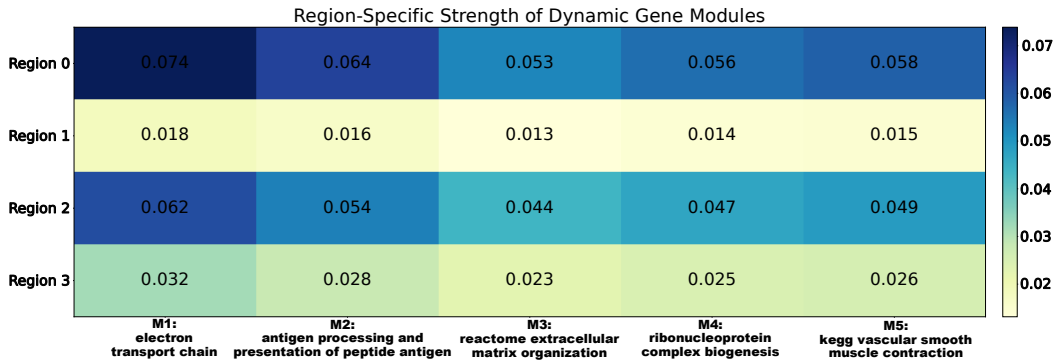


Figure 9: **Regional variation in module strength under CGDR gating.** Heatmap of mean within-module edge-weight increase across pseudo-regions.

of dynamic gene modules, external prior support and regional variation, and the spatial organization of spot-level graph deviations. DP-RAG constructs a task-relevant biological prior, and CGDR modulates the shared gene basis according to local histological context. These analyses characterize the functional coherence and spatial organization of the resulting gene associations.

F.1 Functional characterization and biological interpretability of dynamic gene modules

We first examined representative modules in this gene association graph through functional enrichment analysis. As shown in Fig. 8(b), these modules are consistently enriched for biologically meaningful functional themes across regions, including *respiratory electron transport chain*, *antigen processing and presentation*, and *extracellular matrix organization*, as well as *cellular component biogenesis* and *vascular smooth muscle-related pathways* in selected regions. These enrichments show that the gene association modules align with representative biological processes across tissue regions.

To further illustrate the interpretability of these dynamic modules, Fig. 8(c) visualizes representative module-level subnetworks. Distinct functional themes correspond to gene sets with clear biological semantics. For example, the module associated with respiratory electron transport is primarily composed of mitochondrial respiratory-chain genes such as *NDUFA*, *NDUFB*, *NDUFS*, *UQCRC1*, *COX5A*, and *SDHB*; the antigen processing and presentation module is enriched with immune-related genes including *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*, *HLA-DRB1*, and *CD74*; and the extracellular matrix organization module contains stromal remodeling genes such as *COL3A1*, *LUM*, *BGN*, *FN1*, *POSTN*, *TIMP1*, and *THBS1*. Together, these observations suggest that the locally strengthened connections in the learned dynamic graph are not formed arbitrarily, but are organized around functionally coherent gene groups, thereby supporting the biological interpretability of the model.

F.2 External prior support and regional variation of dynamic graph changes

To examine the biological annotations of strengthened edges, we compared dynamically reinforced edges with random edge sets using `GO_BP` and `MSigDB_C2`. As shown in Fig. 8(a), reinforced edges have higher annotation support than random edges across multiple regions, with statistically significant differences in several cases. The strengthened gene associations are therefore consistent with the functional knowledge used in the analysis.

We next examined regional differences in the average increase in within-module edge weights, as shown in Fig. 9. Even within the same functional module, some pseudo-regions exhibit stronger edge enhancement than others. This regional variation reflects differential modulation of gene relationships by local histological context. Together with the annotation-support comparison in Fig. 8(a), these findings support complementary roles for DP-RAG and CGDR. DP-RAG supplies a task-relevant biological prior, while CGDR modulates the shared gene basis according to local tissue context.

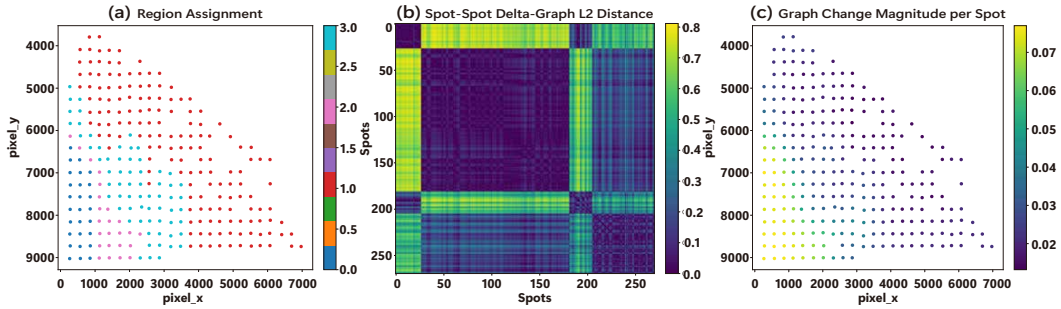


Figure 10: **Spatial organization of spot-level changes in the gene association graph. (a)** Pseudo-region assignment of spatial spots. **(b)** Pairwise spot–spot delta-graph distance heatmap. **(c)** Spatial distribution of graph change magnitude per spot.

F.3 Spatial organization of spot-level graph deviations

Beyond regional variation, we further investigated how individual spots deviate from the shared task-level prior graph. We first partitioned spots into pseudo-regions based on their spatial coordinates, as shown in Fig. 10(a). Based on this partition, Fig. 10(b) presents the pairwise delta-graph L2 distance between spots. The resulting distance matrix exhibits a clear block structure that aligns well with the pseudo-region assignments: spots within the same region tend to display more similar graph deviation patterns, whereas spots from different regions show substantially larger structural differences. Gene association patterns derived from CGDR gates vary across spots and exhibit spatial organization consistent with local tissue context.

In addition to pairwise graph distances, we visualized each spot's graph-change magnitude relative to the shared prior. Fig. 10(c) shows a non-uniform but spatially structured distribution of changes. Graph-change magnitude is positively associated with local tissue heterogeneity (Spearman $\rho = 0.589$), consistent with spatially dependent modulation. Together, Figs. 10(a)–10(c) show how spot-level use of the shared, task-relevant prior varies across tissue locations.

G Additional experimental results

In this section, we provide a more detailed evaluation of our model's performance on varying gene subsets, including All Genes, Highly Variable Genes (HVG), and Highly Expressed Genes (HEG) at different filtering thresholds (Top-50, 100, 300).

G.1 Detailed comparison with SOTA methods

Table 4 presents a comprehensive quantitative comparison of DragH2ST against state-of-the-art baselines across all three datasets (HER2ST, DLPFC, cSCC). We report the Pearson Correlation Coefficient (PCC) for varying gene subsets, including all target genes, filtered subsets of Highly Variable Genes (HVG-50, 100, 300) and Highly Expressed Genes (HEG-50, 100, 300), as well as the challenging Lowly Variable Genes (LVG) and Lowly Expressed Genes (LEG). As observed, **DragH2ST consistently outperforms baselines across most metrics, especially on global (All Genes) and hard-to-predict genes (LVG/LEG).** This validates that our dual-stage retrieval framework, by incorporating external biological knowledge and dynamic structural rewiring, effectively regularizes predictions in data-sparse regimes and captures complex spatial heterogeneity better than purely data-driven methods.

G.2 Detailed ablation study results

In this subsection, we provide the full breakdown of our ablation study results across all evaluated metrics and datasets. We evaluate model variants by removing key components, including DP-RAG, CGDR, and Spatial GCN. For the main ablation analysis, we focus on the modules directly associated with the proposed dual-level dynamic prior construction and spatial modeling. The table below shows

Table 4: Comparison of PCC on different gene sets across three datasets. All denotes the full gene list; HVG/HEG-K denotes Top-K Highly Variable/Expressed Genes. LVG/LEG refers to the 300 Lowly Variable/Lowly Expressed Genes.

HER2ST

Model	All	HVG300	HVG100	HVG50	HEG300	HEG100	HEG50	LVG300	LEG300
ST-Net [9]	0.0451±0.0161	0.079±0.0288	0.092±0.0369	0.096±0.0439	0.081±0.0305	0.096±0.0391	0.101±0.0489	0.0157±0.0079	0.0188±0.0176
HisToGene [26]	0.1529±0.1123	0.132±0.0823	0.112±0.0752	0.123±0.0942	0.122±0.0879	0.075±0.0716	0.051±0.0732	0.1612±0.1451	0.1443±0.1371
Hist2ST [38]	0.1312±0.0833	0.196±0.1150	0.208±0.1117	0.204±0.0985	0.205±0.1168	0.220±0.0936	0.232±0.0671	0.0533±0.0374	0.0477±0.0305
BLEEP [37]	0.0899±0.0217	0.163±0.0309	0.216±0.0316	0.255±0.0372	0.165±0.0384	0.213±0.0388	0.245±0.0387	0.0403±0.0161	0.0349±0.0075
TANGLE [12]	0.1400±0.0867	0.150±0.0731	0.149±0.0536	0.135±0.0315	0.150±0.0873	0.150±0.0721	0.143±0.0762	0.1304±0.0936	0.1342±0.0798
OmiCLIP [3]	0.0984±0.0332	0.100±0.0344	0.099±0.0393	0.103±0.0437	0.131±0.0360	0.162±0.0371	0.189±0.0420	0.0986±0.0356	0.0716±0.0291
M2TGLGO [30]	0.0757±0.0289	0.142±0.0598	0.191±0.0682	0.220±0.0695	0.147±0.0676	0.190±0.0833	0.200±0.0969	0.0221±0.0088	0.0185±0.0063
GR2ST [41]	0.0645±0.0315	0.082±0.0292	0.124±0.0335	0.156±0.0443	0.075±0.0356	0.097±0.0480	0.100±0.0615	0.0646±0.0441	0.0600±0.0393
STimage [35]	0.1397±0.1472	0.225±0.2244	0.256±0.2271	0.265±0.2007	0.225±0.2254	0.261±0.2307	0.268±0.2248	0.0640±0.0706	0.0642±0.0704
DragH2ST(Ours)	**0.2142**±0.0980	**0.265**±0.0937	**0.286**±0.0790	**0.292**±0.0713	**0.277**±0.0897	**0.305**±0.0612	**0.315**±0.0446	**0.1657**±0.0955	**0.1538**±0.0940

DLPFC

Model	All	HVG300	HVG100	HVG50	HEG300	HEG100	HEG50	LVG300	LEG300
ST-Net [9]	0.0235±0.0124	0.069±0.0377	0.093±0.0420	0.096±0.0423	0.069±0.0377	0.095±0.0414	0.101±0.0353	0.0002±0.0001	0.0002±0.0001
HisToGene [26]	0.0133±0.0207	0.022±0.0349	0.035±0.0527	0.042±0.0611	0.019±0.0322	0.020±0.0364	0.023±0.0418	0.0075±0.0114	0.0086±0.0114
Hist2ST [38]	0.0099±0.0047	0.030±0.0149	0.045±0.0202	0.060±0.0255	0.033±0.0158	0.049±0.0237	0.060±0.0283	0.0001±0.0001	0.0001±0.0001
BLEEP [37]	0.0656±0.0304	0.158±0.0808	0.234±0.1106	0.271±0.0417	0.156±0.0790	0.226±0.1056	0.266±0.1171	0.0034±0.0023	0.0036±0.0025
TANGLE [12]	**0.1446**±0.0286	0.169±0.0343	0.204±0.0419	0.245±0.0551	0.174±0.0329	0.183±0.0317	0.196±0.0391	**0.1293**±0.0251	**0.1257**±0.0261
OmiCLIP [3]	0.0498±0.0199	0.048±0.0204	0.041±0.0189	0.035±0.0183	0.135±0.0574	0.215±0.1039	0.242±0.1125	0.0481±0.0187	0.0028±0.0016
M2TGLGO [30]	0.0967±0.0157	0.113±0.0192	0.202±0.0324	0.204±0.0416	0.113±0.0192	0.220±0.038	0.243±0.0425	0.0174±0.0091	0.0188±0.0103
GR2ST [41]	0.1104±0.0277	0.151±0.0514	0.204±0.0905	0.246±0.1220	0.146±0.0376	0.163±0.0503	0.193±0.0634	0.0814±0.0148	0.0817±0.0172
STimage [35]	0.0485±0.0351	0.088±0.0649	0.127±0.0742	0.175±0.0893	0.087±0.0645	0.124±0.0732	0.168±0.0884	0.0007±0.0012	0.0004±0.0007
DragH2ST(Ours)	0.0794±0.0137	**0.182**±0.0373	**0.258**±0.0472	**0.281**±0.0530	**0.182**±0.0375	**0.260**±0.0485	**0.281**±0.0490	0.0137±0.0030	0.0135±0.0030

cSCC

Model	All	HVG300	HVG100	HVG50	HEG300	HEG100	HEG50	LVG300	LEG300
ST-Net [9]	0.0872±0.0523	0.158±0.0909	0.179±0.1005	0.186±0.0975	0.158±0.0902	0.180±0.0974	0.180±0.0864	0.0265±0.0184	0.0251±0.0182
HisToGene [26]	0.1271±0.0554	0.143±0.0549	0.138±0.0550	0.127±0.0666	0.130±0.0511	0.098±0.0477	0.071±0.0383	0.1044±0.0552	0.1098±0.0613
Hist2ST [38]	0.0987±0.0413	0.173±0.0663	0.201±0.0708	0.211±0.0668	0.175±0.0666	0.203±0.0694	0.213±0.0629	0.0287±0.0144	0.0265±0.0142
BLEEP [37]	0.0831±0.0522	0.168±0.1012	0.243±0.1506	0.254±0.1718	0.166±0.0956	0.240±0.1354	0.245±0.1531	0.0257±0.0174	0.0238±0.0176
TANGLE [12]	0.0958±0.0475	0.103±0.0505	0.096±0.0519	0.093±0.0502	0.104±0.0514	0.098±0.0504	0.091±0.0532	0.0897±0.0448	0.0849±0.0456
OmiCLIP [3]	0.1954±0.0874	0.235±0.1013	0.246±0.1116	0.266±0.1397	0.229±0.0953	0.238±0.1069	0.249±0.1324	0.1630±0.0771	0.1671±0.0918
M2TGLGO [30]	0.0321±0.0117	0.070±0.0248	0.139±0.0443	0.197±0.0688	0.071±0.0253	0.141±0.0423	0.193±0.0575	0.0094±0.0052	0.0068±0.0048
GR2ST [41]	0.0971±0.0670	0.132±0.0955	0.173±0.1422	0.200±0.1701	0.121±0.0842	0.126±0.0930	0.130±0.0973	0.0743±0.0533	0.0761±0.0593
STimage [35]	0.2007±0.1206	0.269±0.1436	0.305±0.1473	0.312±0.1321	0.251±0.1307	0.256±0.1040	0.279±0.1148	0.1235±0.0857	0.1225±0.0858
DragH2ST(Ours)	**0.2514**±0.0628	**0.292**±0.0672	**0.318**±0.0932	**0.329**±0.1016	**0.287**±0.0642	**0.287**±0.0802	**0.281**±0.1053	**0.2184**±0.0744	**0.2121**±0.0725

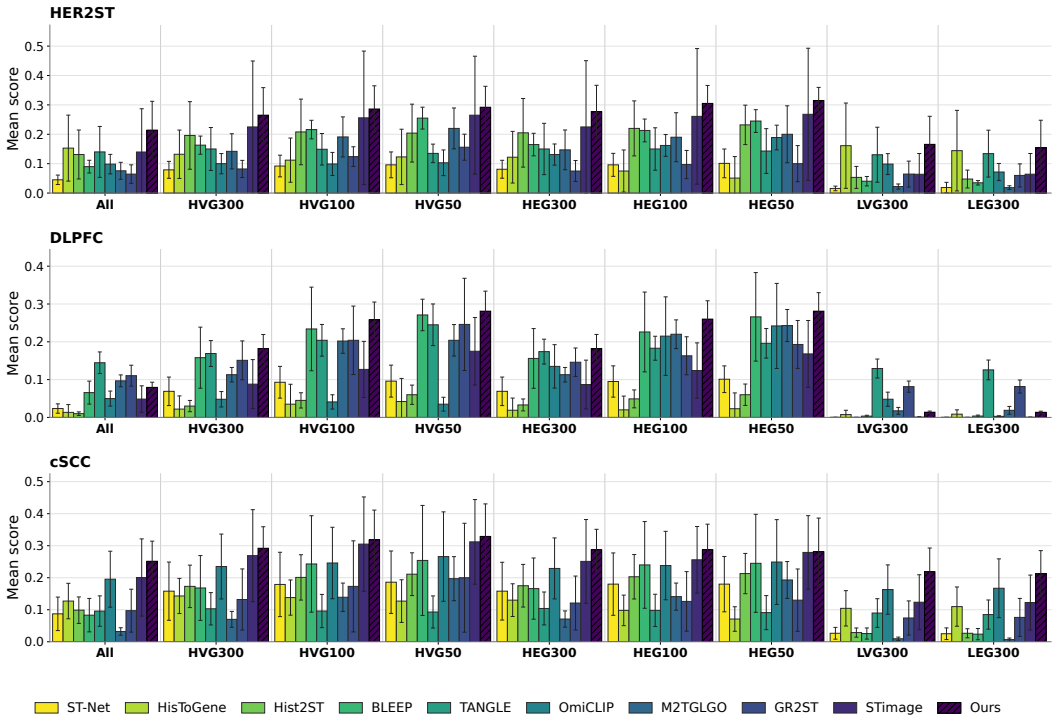


Figure 11: Visual comparison of different models on the benchmark datasets.

the performance of each model variant on the full gene set (All) as well as varying subsets of HVG and HEG.

Table 5: Detailed Ablation Study of PCC and MSE on HER2ST, DLPFC, and cSCC Datasets. Metrics report PCC for varying gene subsets: All genes (PCC-All), top-K Highly Variable/Expressed Genes (HVG/HEG-K), and bottom 300 Lowly Variable/Expressed Genes (LVG300/LEG300). The last column reports the MSE on all genes.

HER2ST										
Variant	PCC-All	HVG300	HVG100	HVG50	HEG300	HEG100	HEG50	LVG300	LEG300	MSE
DragH2ST	0.2142±0.0980	0.265±0.0937	0.286±0.0790	0.292±0.0713	0.277±0.0897	0.305±0.0612	0.315±0.0446	0.1657±0.0955	0.1538±0.0940	0.296±0.0227
W/O DP-RAG	0.2029±0.0976	0.262±0.0975	0.270±0.0865	0.285±0.0775	0.237±0.0941	0.287±0.0679	0.306±0.0486	0.1609±0.0919	0.1493±0.0898	0.300±0.0273
W/O CGDR	0.2112±0.0972	0.264±0.0902	0.282±0.0735	0.288±0.0634	0.257±0.0866	0.298±0.0568	0.309±0.0388	0.1487±0.0979	0.1470±0.0960	0.298±0.0236
W/O Spatial GCN	0.2097±0.0954	0.247±0.0922	0.245±0.0795	0.267±0.0723	0.243±0.0884	0.288±0.0614	0.302±0.0463	0.1426±0.0933	0.1404±0.0915	0.298±0.0254
DLPFC										
Variant	PCC-All	HVG300	HVG100	HVG50	HEG300	HEG100	HEG50	LVG300	LEG300	MSE
DragH2ST	0.0794±0.0137	0.182±0.0373	0.258±0.0472	0.281±0.0530	0.182±0.0375	0.260±0.0485	0.281±0.0490	0.0137±0.0030	0.0135±0.0030	0.159±0.0150
W/O DP-RAG	0.0737±0.0145	0.175±0.0386	0.244±0.0488	0.265±0.0535	0.175±0.0387	0.246±0.0496	0.263±0.0503	0.0123±0.0031	0.0121±0.0031	0.169±0.0071
W/O CGDR	0.0754±0.0137	0.179±0.0365	0.252±0.0453	0.275±0.0510	0.179±0.0367	0.254±0.0466	0.274±0.0470	0.0132±0.0030	0.0132±0.0028	0.166±0.0066
W/O Spatial GCN	0.0749±0.0139	0.177±0.0374	0.247±0.0478	0.269±0.0531	0.177±0.0376	0.249±0.0491	0.268±0.0497	0.0127±0.0031	0.0126±0.0030	0.167±0.0062
cSCC										
Variant	PCC-All	HVG300	HVG100	HVG50	HEG300	HEG100	HEG50	LVG300	LEG300	MSE
DragH2ST	0.2514±0.0628	0.292±0.0672	0.318±0.0932	0.329±0.1016	0.287±0.0642	0.287±0.0802	0.281±0.1053	0.2184±0.0744	0.2121±0.0725	0.239±0.0327
W/O DP-RAG	0.1757±0.0702	0.224±0.0623	0.267±0.0682	0.283±0.0938	0.231±0.0667	0.246±0.0649	0.248±0.0711	0.1502±0.0489	0.1383±0.0623	0.251±0.0642
W/O CGDR	0.2254±0.0723	0.2748±0.0485	0.299±0.0735	0.308±0.0846	0.268±0.0376	0.268±0.0747	0.271±0.0834	0.205±0.0864	0.1991±0.0712	0.258±0.0479
W/O Spatial GCN	0.2045±0.0499	0.249±0.0692	0.278±0.0673	0.291±0.0837	0.233±0.0732	0.264±0.0677	0.269±0.0630	0.1628±0.0767	0.1573±0.0814	0.281±0.0552

Table 5 reports the component ablations for all target genes and the evaluated gene subsets. Removing DP-RAG lowers performance across gene subsets, showing the contribution of task-relevant biological priors. With DP-RAG retained, removing CGDR also reduces performance, supporting local histology-conditioned modulation of the task-level prior. Removing Spatial GCN increases prediction error and lowers correlation, indicating a complementary role for neighborhood information. The full model maintains strong overall performance across datasets and gene subsets, supporting the complementary contributions of the biological prior, local histological context, and spatial structure.

H Marker gene selection criteria

To provide a comprehensive and biologically interpretable evaluation of our model, we curated a set of representative marker genes for each dataset. The selection criteria prioritize genes that: (1) exhibit distinct spatial patterns corresponding to key anatomical or pathological structures; (2) represent diverse functional categories (e.g., structural proteins, signaling molecules, immune markers); and (3) are well-established in the literature for the respective tissue type. Table 6 summarizes the selected genes and their biological relevance for each dataset.

I Limitations, future work and broader impacts

Despite the consistent advantages of DragH2ST across multiple benchmarks, its generality is still partly bounded by the coverage, granularity, and reliability of available biomedical knowledge resources. While our multi-source retrieval-augmented prior construction alleviates the limited coverage and potential bias of conventional static priors, external knowledge bases remain incomplete, heterogeneous, and continuously evolving. Future work could incorporate richer, more fine-grained, and better-curated knowledge resources to further improve model applicability across broader biological scenarios.

DragH2ST may have positive societal impact by making spatial molecular profiling more accessible from routine H&E histology and by supporting interpretable molecular phenotyping in computational pathology. However, because the predicted gene expression profiles are inferred rather than directly measured, they should not be used as standalone clinical evidence without expert review, bias assessment, and prospective validation.

Table 6: Selected Marker Genes for Each Dataset and Their Biological Relevance.

Dataset	Gene	Category	Biological Relevance
HER2ST	MUCL1	Tumor/Epithelial	Breast-specific mucin; high in breast cancer
	AGR2	Tumor/Epithelial	ER-positive tumor marker; survival-associated
	SCD	Tumor/Epithelial	Lipid metabolism; active in proliferating cancer cells
	GATA3	Tumor/Epithelial	Luminal subtype transcription factor; duct marker
	KRT17	Tumor/Epithelial	Keratin 17; basal-like cells and ductal structures
	FN1	Stroma	Fibronectin; strongest stromal marker; defines tumor boundary
	MMP9	Stroma/Invasion	Matrix metalloproteinase; invasion at tumor-stroma interface
	HLA-DRA	Immune	MHC-II; antigen-presenting cells (macrophages)
	CXCL10	Immune	Chemokine; T-cell recruitment; inflammatory microenvironment
	IGHA1	Immune	Immunoglobulin; plasma cells (B-cell lineage)
	HLA-DPB1	Immune	MHC-II; high spatial variability in immune-active regions
	CRABP2	Tumor Subtype	Retinoic acid signaling; marks specific tumor subpopulations
DLPFC	MBP	White Matter	Myelin basic protein; strongest white matter marker
	MOBP	White Matter	Oligodendrocyte marker; distinguishes white/gray matter
	PLP1	White Matter	Myelin structural protein; neural signal conduction
	MOG	White Matter	Outer myelin layer; oligodendrocyte differentiation
	CLDN11	White Matter	Tight junctions in CNS myelin; neural insulation
	RELN	Layer Marker	Layer I marker (Cajal-Retzius cells); cortical layering
	SST	Neuronal	Somatostatin; inhibitory interneurons in deep layers
	NPY	Neuronal	Neuropeptide Y; interneuron marker with scattered pattern
	AQP4	Glial	Aquaporin 4; astrocyte marker at blood-brain barrier
	APOE	Glial	Apolipoprotein E; astrocyte/microglia; Alzheimer's risk gene
	TTR	Choroid Plexus	Transthyretin; marks choroid plexus near ventricles
	SPARC	Vascular	Enriched in vascular endothelium and microglia
cSCC	MMP1	Invasion	Collagenase; degrades ECM for tumor invasion
	LAMC2	Invasion	Laminin $\gamma 2$; basement membrane; invasion front marker
	KRT1	Differentiation	Spinous layer marker; normal differentiation state
	KRT10	Differentiation	Pairs with KRT1; defines skin structural layers
	LOR	Differentiation	Loricrin; terminal differentiation (granular/cornified layer)
	KRTDAP	Differentiation	Keratinocyte differentiation-associated protein
	KRT2	Differentiation	Top-ranked gene; late differentiation/keratinization
	COL1A1	Stroma	Type I collagen; major dermal component
	DCN	Stroma	Decorin; collagen organization; defines stroma regions
	VIM	Stroma/EMT	Vimentin; fibroblasts; also marks EMT in invasive cancer
	S100A7A	Inflammation	Psoriasin family; upregulated in inflammatory skin/cSCC
	HLA-DRA	Immune	MHC-II; antigen-presenting cells in tumor periphery

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes].

   Justification: The main claims in the abstract and introduction accurately reflect the contributions and scope of the paper. The paper clearly presents DragH2ST as a framework for histology-to-spatial transcriptomics prediction, including task-level biomedical knowledge retrieval, spot-level context-gated graph rewiring, dual-graph synergistic learning, and inference-stage prototype retrieval, and these claims are supported by the experimental results and analyses.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes].

   Justification: Appendix I discusses limitations arising from the coverage, granularity, and reliability of external biomedical knowledge resources. It also discusses the need for validation before clinical use of predicted gene expression profiles.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes].

   Justification: The paper analyzes dual-graph structural regularization and CGDR in Section 4, with derivations in Appendix A. For unit-norm spot representations and a fixed nonnegative spatial adjacency, the sigmoid-based reconstruction loss decomposes into a nonlinear neighborhood term, a squared similarity term, and a constant; the CGDR analysis establishes the equivalence between gene-basis modulation and regulatory-output scaling and specifies the additional gating cost.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes].

   Justification: The main text and Appendix B describe the datasets, preprocessing, model architecture, training objectives, and evaluation protocol. The open-source code repository linked on the first page provides the implementation and detailed experiment configurations needed to reproduce the main results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes].

   Justification: Links to the public datasets are provided in Appendix B.1. Our code is publicly available through the repository linked on the first page, together with experiment configurations and instructions for reproducing the main results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes].

   Justification: The paper and Appendix B summarize the main training and test settings, including preprocessing, leave-one-slide-out cross-validation, optimization, and the main hyperparameters. Detailed data-split, hyperparameter, and model-selection configurations are provided in the open-source code repository.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes].

Justification: The main quantitative comparisons and ablation studies report results in the form of mean $\pm$ standard deviation. These error bars summarize variability across the evaluation folds under the leave-one-slide-out cross-validation setting and are reported for the main metrics supporting the paper's claims.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes].

Justification: The paper provides sufficient information on the computational resources needed to reproduce the experiments. Appendix B.2 specifies that the model was trained on NVIDIA A800 GPUs, with a maximum of 500 epochs and batch size 1. Table 3 additionally reports training time and peak memory consumption for CGDR and the explicit per-spot graph construction baseline. These measurements describe the computational resources used in the efficiency experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes].

Justification: The research conforms to the NeurIPS Code of Ethics. The experiments are conducted on publicly available spatial transcriptomics and histopathology datasets, and the work does not involve new human subject recruitment, clinical intervention, privacy-sensitive data release, or unauthorized data collection.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes].

Justification: The paper discusses both the potential positive impact of improving accessibility to spatial molecular profiling and the potential risk of over-reliance on inferred molecular predictions without proper clinical validation in Appendix I.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A].

Justification: The paper does not release high-risk generative models, large-scale scraped datasets, pre-trained language models, or image generators that pose a high risk of misuse. The work focuses on predictive modeling for spatial transcriptomics using public biomedical datasets, so this item is not applicable.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes].

    Justification: The paper properly credits the creators and original owners of all existing assets used in this study, including public datasets, pretrained models, and external biological knowledge resources. These assets are cited with source links and are used in accordance with their respective licenses and terms of use.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes].

    Justification: The released code is accompanied by documentation describing environment setup, data preparation, experiment configuration, and training and evaluation procedures. The repository is linked on the first page.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing experiments or new research with human subjects. All experiments are conducted using publicly available datasets, so this item is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A].

Justification: The paper does not conduct new human subject research, collect new biological samples, perform clinical interventions, or release privacy-sensitive subject data. The experiments are based on publicly available datasets, so additional IRB approval is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A].

Justification: The core method development, experiments, and scientific conclusions of this research do not involve LLMs as important, original, or non-standard components. Any use of LLMs solely for writing, editing, or formatting does not affect the methodology, scientific rigor, or originality of the work, and therefore does not require declaration.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.