MANGO: FLOW MATCHING ON THE MULTIPLICATIVE GEOMETRY OF GENE EXPRESSION FOR SPATIAL TRANSCRIPTOMICS PREDICTION FROM HISTOLOGY

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting spatial molecular profiles from routine H&E images can extend spatial omics to large pathology cohorts, but existing methods commonly rely on Euclidean generative paths, fixed tissue graphs, and uniformly weighted gene objectives. These choices overlook the multiplicative geometry of nonnegative gene expression, prevent tissue neighborhoods from adapting to the evolving molecular state, and bias learning toward abundant genes. We propose MANGO, an expression-conditioned Riemannian graph flow for histology-based spatial-omics prediction. MANGO performs conditional flow matching on the multiplicative geometry of gene expression and predicts the clean expression endpoint, yielding closed-form geodesic corruption and geometry-preserving denoising. A dynamic multi-view graph Transformer jointly models spatial proximity, morphological similarity, and the current multiplicative expression state, and reconstructs the graph layer by layer. An expression-aware Riemannian objective further balances supervision across genes. Across 17 transcriptomics datasets spanning spot-level, Visium HD, and cell-level resolutions, MANGO ranks first on 10 and second on the remaining seven, while achieving the highest average PCC in all four evaluation groups. It also leads in four representative all-gene settings and reaches PCC values of 0.409 and 0.292 on spatial metabolomics and proteomics, surpassing the strongest baselines by 0.176 and 0.088, respectively. These results demonstrate that coupling expression-aware geometry with adaptive tissue structure enables accurate and resolution-robust spatial-omics inference from histology.

1 INTRODUCTION

Spatial transcriptomics (ST) jointly profiles molecular abundance and tissue location, enabling the study of spatially organized cell states, tissue niches, and tumor microenvironments (Ståhl et al., 2016; Xiao & Yu, 2021). Yet generating paired histology–transcriptomics data remains substantially more expensive and less routine than acquiring hematoxylin and eosin (H&E)-stained slides. Predicting spatial gene expression from H&E images therefore offers a scalable way to extend molecular analysis to large pathology cohorts (He et al., 2020; Wang et al., 2025). The task is nevertheless fundamentally underdetermined: visually similar regions may exhibit different molecular states, while the expression at one location depends not only on its own morphology but also on the surrounding tissue organization.

Early approaches treated locations independently or aggregated only a small spatial neighborhood (Pang et al., 2021). More recent slide-level methods capture broader context through graph neural networks, hierarchical aggregation, or attention (Ganguly et al., 2025; Qu et al., 2025; Fu et al., 2025). In parallel, diffusion and flow-matching models learn a conditional distribution over expression rather than a single deterministic regression target, reducing the tendency to average over plausible molecular profiles (Chen & Lipman, 2024; Zhu et al., 2025; Huang et al., 2025; Si et al., 2026). These developments establish spatial context and generative modeling as complementary ingredients for histology-to-transcriptomics prediction. However, they are commonly combined without explicitly accounting for the geometry of expression or for how spatial relationships should evolve along the generative trajectory.

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

We identify three coupled limitations. First, gene expression is nonnegative, sparse, and strongly scale-dependent, but conventional Euclidean corruption and interpolation do not explicitly respect its multiplicative structure. A unit displacement near zero has a different biological meaning from the same displacement at high abundance, and an unconstrained trajectory may traverse regions that are poorly matched to the data distribution. Second, graph neighborhoods are typically fixed from physical coordinates or morphology before generation. Such a graph cannot adapt when the current expression estimate reveals that adjacent locations belong to different molecular niches, or that morphologically related locations share a latent state. Third, standard objectives are dominated by highly expressed genes. This imbalance is particularly damaging when evaluation is averaged across genes, because weak but spatially informative signals receive comparatively little supervision.

To address these limitations, we propose an expression-conditioned Riemannian graph flow framework. We endow expression with a multiplicative geometry under which geodesics are linear in log coordinates and remain in the valid shifted domain. Rather than directly regressing a vector field, the model predicts the clean expression endpoint at each time; the corresponding Riemannian conditional vector field follows analytically from the logarithmic map. This endpoint parameterization provides a direct denoising target while retaining the geometry-aware flow-matching objective.

The endpoint predictor couples morphology, space, and the current expression state through a dynamic multi-view graph. Each location is initialized by the sum of a time embedding, a pathology-feature embedding, and an embedding of its noisy expression in log coordinates. At every graph layer, we fuse row-normalized spatial distance, morphological cosine distance, and multiplicative expression distance to select neighbors. Messages retain the views separately through spatial directions and distances, signed pathology-feature differences, and signed log-expression differences, which modulate both attention scores and values. Each layer produces an updated clean-expression estimate that is used to reconstruct the graph for the next layer. Finally, an inverse-expression gene weighting balances supervision across genes without changing the overall loss scale.

We evaluate the framework across spot-level, Visium, Visium HD, and cell-level benchmarks spanning multiple tissues and sequencing platforms. The method obtains the highest average Pearson correlation in all four evaluation regimes and remains effective as spatial resolution and graph size increase. These results support the central premise that expression-aware geometry and adaptive multi-view tissue context are mutually reinforcing rather than independent design choices.

Our key contributions are summarized as follows:

- **Geometry-aware conditional generation.** We formulate histology-conditioned expression generation through the intrinsic geometry of expression space and adopt clean-endpoint prediction to parameterize a Riemannian conditional flow, enabling closed-form geodesic corruption and denoising.
- **Dynamic multi-view tissue modeling.** We introduce an expression-conditioned graph Transformer that jointly uses spatial, morphological, and multiplicative expression relations for neighbor selection and edge-aware attention, and progressively rebuilds the graph from layer-wise expression estimates.
- **Balanced and resolution-robust prediction.** We derive an expression-weighted Riemannian objective that strengthens supervision for low-expression genes, and demonstrate consistent gains across diverse spatial resolutions (e.g., Visium HD, Xenium) and modalities (spatial metabolomics and proteomics).

2 Related Work

Discriminative Spatial Prediction. Early methods formulate histology-to-expression inference as independent patch-wise regression. ST-Net (He et al., 2020) and HisToGene (Pang et al., 2021) learn local morphology-to-expression mappings, but do not explicitly coordinate predictions across an entire tissue section; recent benchmarks consequently highlight the advantage of slide-level context modeling (Wang et al., 2025). Contemporary approaches incorporate this context through different architectural priors. MERGE (Ganguly et al., 2025) uses spatial and feature-space clustering to construct a hierarchical graph for long-range communication. NH^{2}ST (Qu et al., 2025) combines dual-scale contrastive learning with cross-attention to align pathology and expression representations. GHIST (Fu et al., 2025) instead trains its visual encoder end-to-end for single-cell expression

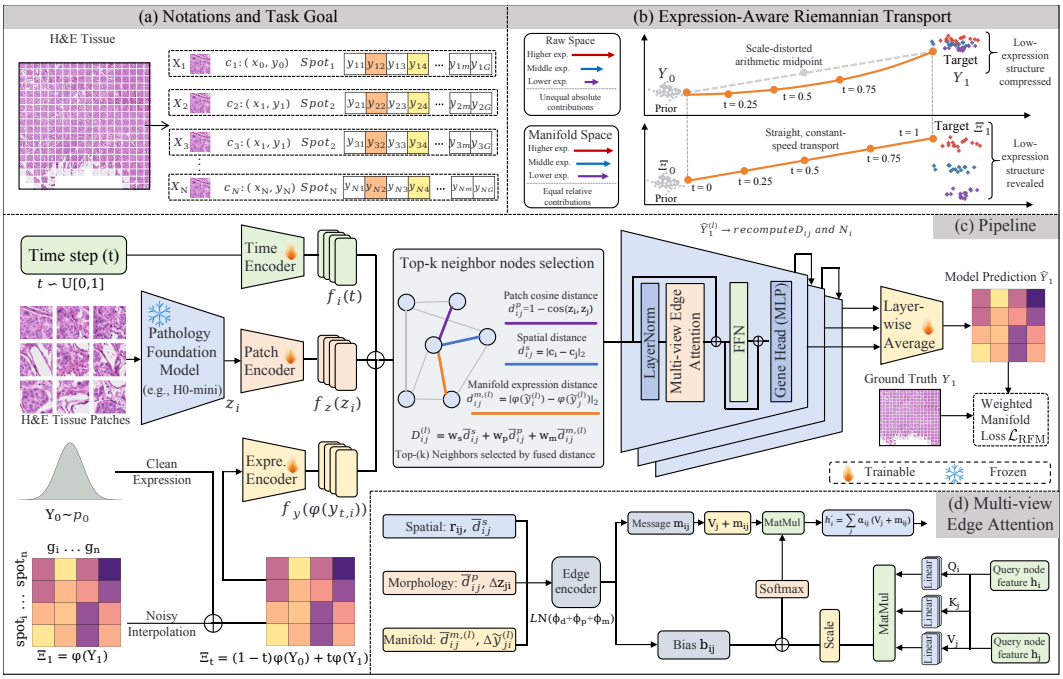


Figure 1: Overview of the proposed expression-conditioned Riemannian graph flow. (a) Notation and the histology-to-expression prediction task. (b) Multiplicative-geometry transport replaces an unconstrained Euclidean path. (c) Time, morphology, and noisy expression initialize each node, followed by layer-wise dynamic multi-view graph refinement. (d) Spatial, morphological, and multiplicative edge cues jointly modulate graph attention.

prediction. These methods substantially improve contextual reasoning, yet still learn a deterministic map from histology to a single expression estimate. Moreover, their tissue relations are constructed from fixed coordinates, image features, or predefined hierarchies; the molecular state being inferred cannot in turn revise the communication topology. This separates tissue-context modeling from expression refinement and limits the representation of state-dependent molecular dependencies.

Generative Spatial Prediction. Generative approaches instead model the conditional distribution of spatial expression, allowing multiple molecular outcomes to be associated with similar histology. Stem (Zhu et al., 2025) uses conditional diffusion to capture expression variability. STFlow (Huang et al., 2025) extends generation to the whole slide with flow matching and a zero-inflated negative-binomial prior. FLAG (Si et al., 2026) separates spatial graph encoding from gene-level diffusion and aligns the resulting representations with a pretrained foundation model. HINGE (Fang et al., 2026) adapts a single-cell foundation model into a conditional generator through masked diffusion. Although these methods move beyond point estimation, two structural limitations remain. First, the transport or corruption process is generally defined in an unconstrained Euclidean space, which does not reflect the nonnegative and scale-dependent geometry of expression. Second, tissue context is typically fixed before or outside the generative trajectory, preventing the evolving molecular state from refining spatial and morphological interactions. The resulting mismatch between expression geometry and tissue topology motivates a conditional generator that couples intrinsic expression transport with state-dependent graph refinement.

3 Preliminaries

3.1 Histology-Conditioned Spatial Expression Inference

Given an H&E-stained tissue section, we partition the image into N patches centered at the measured spatial locations. A slide is represented as $\mathcal{D} = \{\mathbf{X}, \mathbf{C}, \mathbf{Y}\}$, where $\mathbf{X} \in \mathbb{R}^{N\times H\times W\times 3}$ contains the

image patches, $\mathbf{C} \in \mathbb{R}^{N\times 2}$ their spatial coordinates, and $\mathbf{Y} \in \mathbb{R}^{N\times G}_{\geq 0}$ the normalized expression of G target genes. Their ith entries are $\mathbf{X}_i$, $\mathbf{c}_i$, and $\mathbf{y}_i$. We learn the slide-level distribution $p(\mathbf{Y} \mid \mathbf{X}, \mathbf{C})$ rather than independent spot-wise regressions, capturing tissue-level dependence and the one-to-many mapping from morphology to expression.

To condition this distribution, a frozen pathology foundation model extracts $\mathbf{z}_i = f_{\mathrm{PFM}}(\mathbf{X}_i) \in \mathbb{R}^D$ for every location. Stacking the location-wise embeddings gives $\mathbf{Z} = [\mathbf{z}_1^\top, \ldots, \mathbf{z}_N^\top]^\top \in \mathbb{R}^{N\times D}$; these features are cached across folds and sampling runs.

3.2 Conditional Flow Matching

Let $\mathbf{H} = (\mathbf{Z}, \mathbf{C})$ collect the slide-level histological and spatial context, let p_0 be a tractable source distribution, and let $p_1(\cdot \mid \mathbf{H})$ denote the expression distribution. Flow matching (Chen & Lipman, 2024) seeks a time-dependent vector field whose flow map Ψ_t pushes p_0 toward $p_1(\cdot \mid \mathbf{H})$:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\Psi_t(\mathbf{Y}_0;\mathbf{H}) &= \mathbf{v}_t(\Psi_t(\mathbf{Y}_0;\mathbf{H}),\mathbf{H}), \quad & \Psi_0(\mathbf{Y}_0;\mathbf{H}) &= \mathbf{Y}_0, \\ p_t(\cdot \mid \mathbf{H}) &= (\Psi_t)_{\#}p_0, & p_{t=1}(\cdot \mid \mathbf{H}) &= p_1(\cdot \mid \mathbf{H}), \end{aligned} \tag{1}$$

where $(\Psi_t)_{\#}$ denotes the pushforward measure. Because the marginal field $\mathbf{v}_t$ is unavailable, choose a coupling $\pi_{\mathbf{H}} \in \Pi(p_0, p_1(\cdot \mid \mathbf{H}))$, where Π denotes the set of couplings between its arguments, and an endpoint-conditioned path $p_t(\cdot \mid \mathbf{Y}_0, \mathbf{Y}_1)$ generated by a known field $\mathbf{u}_t(\cdot \mid \mathbf{Y}_0, \mathbf{Y}_1)$ with boundaries $\mathbf{Y}_0$ and $\mathbf{Y}_1$. Marginalization gives

$$\begin{aligned} p_t(\mathbf{Y} \mid \mathbf{H}) &= \int p_t(\mathbf{Y} \mid \mathbf{Y}_0, \mathbf{Y}_1)\,\mathrm{d}\pi_{\mathbf{H}}(\mathbf{Y}_0, \mathbf{Y}_1), \\ \mathbf{v}_t(\mathbf{Y}, \mathbf{H}) &= \frac{1}{p_t(\mathbf{Y} \mid \mathbf{H})}\int \mathbf{u}_t(\mathbf{Y} \mid \mathbf{Y}_0, \mathbf{Y}_1)p_t(\mathbf{Y} \mid \mathbf{Y}_0, \mathbf{Y}_1)\,\mathrm{d}\pi_{\mathbf{H}}(\mathbf{Y}_0, \mathbf{Y}_1). \end{aligned} \tag{2}$$

The marginal velocity is therefore the posterior mean of the conditional velocities, yielding the simulation-free objective

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{\substack{t\sim\mathcal{U}[0,1],\,(\mathbf{Y}_0,\mathbf{Y}_1)\sim\pi_{\mathbf{H}} \\ \mathbf{Y}_t\sim p_t(\cdot\mid\mathbf{Y}_0,\mathbf{Y}_1)}}\left[\|\mathbf{v}_\theta(\mathbf{Y}_t,\mathbf{H},t) - \mathbf{u}_t(\mathbf{Y}_t \mid \mathbf{Y}_0,\mathbf{Y}_1)\|_F^2\right]. \tag{3}$$

where $\|\cdot\|_F$ is the Frobenius norm. The conventional Euclidean interpolant $\mathbf{Y}_t = (1-t)\mathbf{Y}_0 + t\mathbf{Y}_1$, whose conditional velocity is $\mathbf{Y}_1 - \mathbf{Y}_0$, assumes additive straight-line expression dynamics. We instead learn a positive-domain path with an evolving tissue graph.

4 Method

Overview. Figure 1 summarizes MANGO: endpoint flow matching in a multiplicative geometry, layer-wise dynamic multi-view graph denoising, and an expression-aware Riemannian objective with geometry-preserving sampling.

4.1 Riemannian Flow Matching for Gene Expression

Because additive displacement is scale dependent for nonnegative expression, we perform transport with a fixed shift $s > 0$ on the domain $\mathcal{M}_s = \{\mathbf{y} : y_g + s > 0,\ \forall g\}$ using the global chart $\varphi(\mathbf{y}) = \log(\mathbf{y} + s\mathbf{1})$ and inverse $\varphi^{-1}(\mathbf{x}) = \exp(\mathbf{x}) - s\mathbf{1}$. Pulling the Euclidean inner product back through this chart gives, with $\mathbf{1}$ denoting the all-ones vector,

$$g_{\mathbf{y}}(\mathbf{u},\mathbf{v}) = \left\langle \frac{\mathbf{u}}{\mathbf{y}+s\mathbf{1}}, \frac{\mathbf{v}}{\mathbf{y}+s\mathbf{1}} \right\rangle = \sum_{g=1}^{G}\frac{u_g v_g}{(y_g+s)^2}. \tag{4}$$

The metric measures relative changes, and the isometry φ maps geodesics on $\mathcal{M}_s$ to straight lines in log coordinates. Unlike raw Euclidean distance, this normalization prevents highly abundant genes from setting the transport scale and exposes relative differences among low-expression genes.

Following Riemannian flow matching (Chen & Lipman, 2024), a prior sample $\mathbf{Y}_0$ and endpoint $\mathbf{Y}_1$ define

$$\begin{aligned} \boldsymbol{\Xi}_t &= (1-t)\varphi(\mathbf{Y}_0) + t\varphi(\mathbf{Y}_1), \qquad \mathbf{Y}_t = \varphi^{-1}(\boldsymbol{\Xi}_t), \\ \mathbf{u}_t(\mathbf{Y}_t \mid \mathbf{Y}_0, \mathbf{Y}_1) &= (\mathbf{Y}_t + s\mathbf{1}) \circ \frac{\varphi(\mathbf{Y}_1) - \varphi(\mathbf{Y}_t)}{1-t}. \end{aligned} \tag{5}$$

The first line is the log-chart geodesic and the second its tangent velocity on $\mathcal{M}_s$, where $\circ$ denotes elementwise multiplication. The global chart yields a closed-form interpolation whose inverse remains in the positive domain, avoiding a numerical geodesic solver. We use a wrapped Gaussian prior for all spatial omics benchmarks.

Rather than regress the tangent field directly, the denoiser predicts the clean endpoint $\widehat{\mathbf{Y}}_1 = f_\theta(\mathbf{Y}_t, \mathbf{Z}, \mathbf{C}, t)$ and induces

$$\mathbf{v}_\theta(\mathbf{Y}_t, \mathbf{H}, t) = (\mathbf{Y}_t + s\mathbf{1}) \circ \frac{\varphi(\widehat{\mathbf{Y}}_1) - \varphi(\mathbf{Y}_t)}{1-t}. \tag{6}$$

This equivalent parameterization retains an interpretable expression target and provides the endpoint estimate used to refine the tissue graph.

4.2 Expression-Conditioned Dynamic Multi-View Graph Denoiser

For location i, we initialize $\mathbf{h}_i^{(0)} = f_t(t) + f_z(\mathbf{z}_i) + f_y(\varphi(\mathbf{y}_{t,i}))$, where f_t is a sinusoidal timestep embedder followed by an MLP, f_z is a linear projection, and f_y is a LayerNorm–MLP encoder. Including the current expression exposes the state being denoised to every graph layer instead of making topology morphology-driven.

Let $\widetilde{\mathbf{y}}_i^{(0)} = \mathbf{y}_{t,i}$ and $\widetilde{\mathbf{y}}_i^{(\ell)} = \widehat{\mathbf{y}}_{1,i}^{(\ell)}$ for $\ell \geq 1$, so that $\widetilde{\mathbf{y}}_i^{(\ell)}$ is the latest expression state available to layer ℓ. To construct the graph for the next layer, each candidate edge is evaluated from spatial, pathology, and manifold views: $d_{ij}^s = \|\mathbf{c}_i - \mathbf{c}_j\|_2$, $d_{ij}^p = 1 - \cos(\mathbf{z}_i, \mathbf{z}_j)$, and $d_{ij}^{m,(\ell)} = \|\varphi(\widetilde{\mathbf{y}}_i^{(\ell)}) - \varphi(\widetilde{\mathbf{y}}_j^{(\ell)})\|_2$, respectively. We divide each distance by its per-node mean over the candidate set and rank edges using $D_{ij}^{(\ell)} = w_s \bar{d}_{ij}^s + w_p \bar{d}_{ij}^p + w_m \bar{d}_{ij}^{m,(\ell)}$, where $w_s, w_p, w_m \geq 0$ are the three view weights. The k candidates with the smallest scores form $\mathcal{N}_i^{(\ell)}$. Sparse candidate retrieval avoids an $N \times N$ matrix, while rebuilding from the latest expression estimate couples topology and expression refinement (Wang et al., 2019). Thus morphology can connect phenotypically related regions, while the manifold view can separate adjacent locations whose inferred molecular states diverge.

The graph topology uses a fused score, but message passing retains the views separately. For each selected edge (i, j), we encode

$$\begin{aligned} \mathbf{e}_{ij}^{(\ell)} = \mathrm{LN}\Big(&\phi_d\big[\mathbf{r}_{ij}, \bar{d}_{ij}^s, \bar{d}_{ij}^p, \bar{d}_{ij}^{m,(\ell)}\big] + \phi_p(\mathbf{z}_j - \mathbf{z}_i) \\ &+ \phi_m\big(\varphi(\widetilde{\mathbf{y}}_j^{(\ell)}) - \varphi(\widetilde{\mathbf{y}}_i^{(\ell)})\big)\Big), \end{aligned} \tag{7}$$

where $\mathbf{r}_{ij} = (\mathbf{c}_j - \mathbf{c}_i)/(\|\mathbf{c}_j - \mathbf{c}_i\|_2 + \epsilon)$ denotes the spatial direction, with $\epsilon > 0$ preventing division by zero. Here, $[\cdot]$ denotes concatenation, LN is LayerNorm, and ϕ_d, ϕ_p, ϕ_m are the distance, pathology-difference, and manifold-difference MLPs. The signed differences preserve directional information absent from scalar distances, allowing attention to distinguish opposing feature changes of similar magnitude.

In attention head a of width d_a, node states are projected as $\mathbf{q}_i^a = \mathbf{W}_Q^a \mathbf{h}_i^{(\ell)}$, $\mathbf{k}_j^a = \mathbf{W}_K^a \mathbf{h}_j^{(\ell)}$, and $\mathbf{v}_j^a = \mathbf{W}_V^a \mathbf{h}_j^{(\ell)}$. The edge representation modulates both attention scoring and value transport:

$$\begin{aligned} \alpha_{ij}^{a,(\ell)} &= \mathrm{softmax}_{j \in \mathcal{N}_i^{(\ell)}}\left(\frac{\langle \mathbf{q}_i^a, \mathbf{k}_j^a \rangle}{\sqrt{d_a}} + [\mathbf{W}_b \mathbf{e}_{ij}^{(\ell)}]_a\right), \\ \mathbf{o}_i^{a,(\ell)} &= \sum_{j \in \mathcal{N}_i^{(\ell)}} \alpha_{ij}^{a,(\ell)} \left(\mathbf{v}_j^a + [\mathbf{W}_v \mathbf{e}_{ij}^{(\ell)}]_a\right). \end{aligned} \tag{8}$$

The two projected edge terms are the head-specific scalar bias and value message. Concatenated heads pass through output, residual, and feed-forward blocks; each gene head predicts $\widehat{\mathbf{y}}_{1,i}^{(\ell)}$ for the

next graph update. For L layers, the final endpoint is $\widehat{\mathbf{Y}}_1 = L^{-1}\sum_{\ell=1}^{L}\widehat{\mathbf{Y}}_1^{(\ell)}$. This recurrent update makes message passing responsive to the inferred molecular trajectory rather than fixed throughout denoising.

4.3 Expression-Aware Objective and Riemannian Sampling

To prevent abundant genes from dominating a heterogeneous panel, we compute a detached batch statistic in log coordinates,

$$\mu_g = \frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}}\varphi(\mathbf{Y}_1)_{ig}, \qquad \widetilde{w}_g = (\mu_g + \epsilon_w)^{-\gamma}, \tag{9}$$

where $\mathcal{V}$ contains valid, non-padded locations, $\epsilon_w > 0$ stabilizes the weights, and $\gamma \geq 0$ controls reweighting. Normalizing to unit mean, clipping at $w_{\max}$, and renormalizing gives w_g. Substituting Eq. 6 and its target counterpart into the metric in Eq. 4 cancels the state-dependent factor $(\mathbf{Y}_t + s\mathbf{1})$. The weighted Riemannian field-matching loss therefore reduces to

$$\mathcal{L}_{\mathrm{RFM}} = \frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}}\sum_{g=1}^{G} w_g \frac{\left[\varphi(\widehat{\mathbf{Y}}_1)_{ig} - \varphi(\mathbf{Y}_1)_{ig}\right]^2}{(1-t)^2}. \tag{10}$$

The factor $(1-t)^{-2}$ converts endpoints to velocities, while w_g balances genes without changing the average loss scale.

At inference, initialize $\mathbf{Y}_{t_0} \sim p_0$ and update directly in log coordinates:

$$\mathbf{\Xi}_{t+\Delta t} = \mathbf{\Xi}_t + \frac{\Delta t}{1-t}\left[\varphi(\widehat{\mathbf{Y}}_1) - \mathbf{\Xi}_t\right], \quad \mathbf{Y}_{t+\Delta t} = \varphi^{-1}(\mathbf{\Xi}_{t+\Delta t}), \tag{11}$$

where $\mathbf{\Xi}_t = \varphi(\mathbf{Y}_t)$ and Δt is the time-grid increment. The inverse chart keeps all states on $\mathcal{M}_s$; repeated updates yield the prediction, and different prior draws yield samples.

5 Datasets and Evaluation

Datasets. We evaluate on spatial transcriptomics across three resolutions: spot-level HEST-1k (Jaume et al., 2024), high-resolution Visium HD (18 HEST-1k slides across COAD, LUAD, SOC, and IDC, one slide per patient), and cell-level Xenium using the SpatialEx splits (Liu et al., 2026). Spot-level and Visium HD tasks use patient-stratified leave-one-patient-out cross-validation. We also test transfer to spatial metabolomics (SMA MALDI-MSI FMP-10, 3 slides, leave-one-mouse-out (Vicari et al., 2024)) and proteomics (HTAPP-MBC CODEX, 13 patients, leave-one-patient-out (Klughammer et al., 2024)).

Baselines and Evaluation. We organize the competing methods into discriminative and generative baselines. The discriminative group includes UNI (Chen et al., 2024b), H0-mini (Filiot et al., 2025), MERGE (Ganguly et al., 2025), NH^{2}ST (Qu et al., 2025), and SpatialEx (Liu et al., 2026); the generative group includes STFlow (Huang et al., 2025), Stem (Zhu et al., 2025), and HINGE (Fang et al., 2026). UNI and H0-mini are evaluated using PCA-reduced ridge regression following the HEST protocol (Jaume et al., 2024). For the spot-level and Visium-HD experiments, feature-based methods (Stem, SpatialEx, STFlow, MANGO and NH^{2}ST) use the same H0-mini features, whereas MERGE and HINGE retain their official model configurations. We report PCC between predicted and ground-truth profiles for the top 50 highly variable features after $\log 1\mathrm{p}$ normalization (Jaume et al., 2024), averaged over three random seeds.

Implementation Details. MANGO is trained with Adam using batch size 2 and a learning rate of 5×10^{-4} for at most 100 epochs. We evaluate after every epoch, apply early stopping with patience 20, and clip the gradient norm to 1.0. For image input, histology patches are augmented with random horizontal and vertical flips, 90° rotations, and mild brightness, contrast, and saturation jitter. For flow matching, training times are sampled uniformly from $[0, 1]$. We use a wrapped Gaussian prior with standard deviation 0.5. Inference uses a uniform five-point grid from 0.01 to 1.0 with Riemannian Euler updates, and we retain the checkpoint with the highest training PCC.

6 RESULTS

6.1 GENE EXPRESSION PREDICTION

Table 1 compares MANGO with discriminative and generative baselines across 17 datasets. MANGO ranks first on 10 datasets and second on the remaining seven, while achieving the best average PCC in all four resolution evaluation groups. The leading baseline varies substantially across datasets, whereas MANGO remains consistently among the top two. These results show that the method remains competitive as the spatial resolution and platform change. At the group level, MANGO exceeds the strongest competing average by 0.022, 0.034, and 0.029 on spot-level Xenium, Visium, and Visium HD, respectively, and remains first at cell level (0.380 versus 0.376). The advantage therefore persists across resolutions rather than being driven by a single platform.

Table 1: The results of gene expression prediction. Datasets are systematically categorized by spatial resolution and sequencing platforms. **Bold** and underlined values indicate the best and second-best results, respectively. Additional MSE and MAE metrics are provided in Appendix.

			Discriminative Models					Generative Models			
			UNI	H0-mini	MERGE	NH^2ST	SpatialEx	STFlow	Stem	HINGE	**OURS**
Spot-level (Patch & Bin)	Xenium	IDC	$0.572_{.096}$	$0.586_{.083}$	$0.553_{.067}$	$0.579_{.089}$	$0.467_{.093}$	$\underline{0.601}_{.075}$	$0.493_{.074}$	$0.567_{.056}$	$\mathbf{0.602}_{\mathbf{.067}}$
		PAAD	$0.479_{.093}$	$0.495_{.057}$	$0.451_{.068}$	$0.385_{.095}$	$0.387_{.015}$	$\underline{0.506}_{.029}$	$0.382_{.086}$	$0.444_{.040}$	$\mathbf{0.537}_{\mathbf{.068}}$
		SKCM	$0.624_{.049}$	$0.596_{.082}$	$0.618_{.069}$	$0.580_{.098}$	$0.627_{.034}$	$\underline{0.636}_{.071}$	$0.629_{.065}$	$0.523_{.040}$	$\mathbf{0.701}_{\mathbf{.059}}$
		LUNG	$0.552_{.023}$	$0.549_{.051}$	$0.473_{.051}$	$0.511_{.051}$	$0.501_{.014}$	$\mathbf{0.628}_{\mathbf{.024}}$	$0.495_{.026}$	$0.566_{.044}$	$\underline{0.618}_{.008}$
	Xenium Avg.		0.557	0.557	0.524	0.514	0.496	$\underline{0.593}$	0.500	0.525	**0.615**
	Visium	PRAD	$0.234_{.124}$	$0.306_{.048}$	$0.275_{.086}$	$0.294_{.070}$	$0.307_{.021}$	$\underline{0.376}_{.013}$	$0.268_{.013}$	$0.351_{.063}$	$\mathbf{0.431}_{\mathbf{.004}}$
		COAD	$0.275_{.026}$	$0.257_{.061}$	$0.233_{.088}$	$0.232_{.048}$	$0.300_{.012}$	$0.326_{.009}$	$0.168_{.093}$	$\mathbf{0.351}_{\mathbf{.063}}$	$\underline{0.346}_{.009}$
		READ	$0.200_{.036}$	$0.226_{.046}$	$0.221_{.043}$	$0.156_{.033}$	$0.188_{.035}$	$\mathbf{0.311}_{\mathbf{.012}}$	$0.138_{.044}$	$0.165_{.055}$	$\underline{0.282}_{.010}$
		CCRCC	$0.186_{.097}$	$0.186_{.093}$	$0.215_{.094}$	$0.184_{.088}$	$0.252_{.077}$	$0.251_{.089}$	$0.169_{.071}$	$\underline{0.267}_{.087}$	$\mathbf{0.372}_{\mathbf{.085}}$
		LYMPH	$0.258_{.051}$	$0.263_{.042}$	$0.226_{.069}$	$0.255_{.048}$	$\mathbf{0.325}_{\mathbf{.051}}$	$0.310_{.055}$	$0.170_{.048}$	$0.254_{.055}$	$\underline{0.314}_{.054}$
	Visium Avg.		0.231	0.248	0.234	0.224	0.274	$\underline{0.315}$	0.183	0.278	**0.349**
	Visium HD	Breast	$0.122_{.077}$	$0.116_{.071}$	$0.089_{.068}$	$0.099_{.063}$	$0.076_{.045}$	$\underline{0.178}_{.111}$	$0.058_{.069}$	$0.149_{.082}$	$\mathbf{0.233}_{\mathbf{.147}}$
		Lung	$0.431_{.099}$	$0.442_{.095}$	$0.409_{.086}$	$0.425_{.085}$	$0.343_{.099}$	$\underline{0.477}_{.064}$	$0.343_{.099}$	$0.420_{.050}$	$\mathbf{0.481}_{\mathbf{.056}}$
		Colon	$0.275_{.048}$	$0.256_{.074}$	$0.366_{.088}$	$0.366_{.092}$	$0.348_{.077}$	$\underline{0.381}_{.130}$	$0.380_{.285}$	$0.327_{.145}$	$\mathbf{0.406}_{\mathbf{.113}}$
		Ovary	$0.370_{.135}$	$0.358_{.130}$	$0.422_{.040}$	$\mathbf{0.511}_{\mathbf{.049}}$	$0.437_{.029}$	$0.436_{.130}$	$0.361_{.008}$	$0.284_{.074}$	$\underline{0.467}_{.047}$
	Visium HD Avg.		0.300	0.293	0.322	0.350	0.301	$\underline{0.368}$	0.286	0.295	**0.397**
Cell-level	Xenium	Breast	$0.419_{.012}$	$0.412_{.010}$	$0.425_{.005}$	$0.424_{.007}$	$0.421_{.012}$	$\underline{0.428}_{.006}$	$0.374_{.001}$	$0.404_{.019}$	$\mathbf{0.456}_{\mathbf{.004}}$
		Skin	$0.065_{.047}$	$0.027_{.040}$	$\mathbf{0.270}_{\mathbf{.007}}$	$0.188_{.003}$	$0.220_{.005}$	$0.212_{.035}$	$0.037_{.084}$	$0.181_{.036}$	$\underline{0.233}_{.008}$
		Colon (M)	$0.494_{.009}$	$0.505_{.014}$	$\mathbf{0.554}_{\mathbf{.006}}$	$0.495_{.001}$	$0.518_{.001}$	$0.486_{.003}$	$0.515_{.014}$	$0.482_{.019}$	$\underline{0.533}_{.007}$
		Colon (H)	$0.163_{.052}$	$0.161_{.040}$	$\underline{0.255}_{.007}$	$0.217_{.002}$	$0.254_{.007}$	$0.228_{.016}$	$0.145_{.030}$	$0.231_{.029}$	$\mathbf{0.298}_{\mathbf{.003}}$
	Cell-level Avg.		0.285	0.276	$\underline{0.376}$	0.331	0.353	0.339	0.268	0.325	**0.380**

6.2 EXTENSION TO ALL GENES

We next evaluate whether MANGO scales beyond restricted gene panels. Table 2 reports all-gene prediction on one representative dataset from each platform and resolution regime. The first three datasets are evaluated at spot level, whereas Colon (H) is evaluated at cell level. MANGO ranks first in all four settings, demonstrating scalability to substantially broader gene sets without modifying the framework or introducing platform- or resolution-specific components. Compared with the strongest baseline in each row, the absolute gains are 0.015, 0.035, 0.025, and 0.032, respectively; the advantage persists for panels exceeding 8,000 genes.

Table 2: **All-gene prediction across settings (PCC).**

Datasets	Genes	UNI	H0-mini	STFlow	MERGE	OURS
LUNG	197	$0.583_{.012}$	$0.585_{.017}$	$0.608_{.055}$	$0.519_{.033}$	$\mathbf{0.623}_{\mathbf{.024}}$
READ	8,828	$0.114_{.056}$	$0.112_{.073}$	$0.105_{.137}$	$0.151_{.047}$	$\mathbf{0.186}_{\mathbf{.045}}$
IDC	8,270	$0.213_{.106}$	$0.171_{.062}$	$0.216_{.109}$	$0.246_{.190}$	$\mathbf{0.271}_{\mathbf{.168}}$
Colon (H)	425	$0.095_{.000}$	$0.095_{.000}$	$0.107_{.006}$	$0.090_{.001}$	$\mathbf{0.139}_{\mathbf{.002}}$

6.3 EXTENSION TO SPATIAL MULTI-OMICS

Most histology-to-omics models focus exclusively on spatial transcriptomics. We extend MANGO to predict spatial metabolomics and proteomics, testing whether a single architecture transfers across molecular modalities, including metabolomics and proteomics.

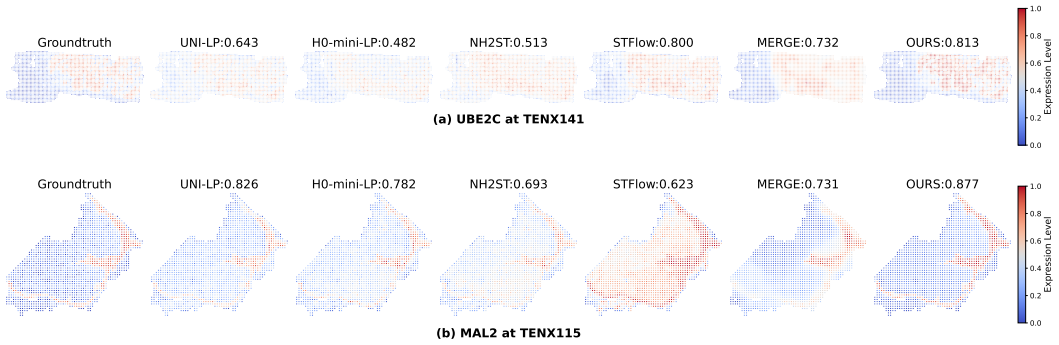


Figure 2: Comparison of spatial gene expression predictions across methods. Ground-truth and predicted expression maps for UBE2C at TENX141 (LUAD, top) and MAL2 at TENX115 (SKCM, bottom). PCC scores are shown for each method; MANGO achieves the closest match to ground truth in both samples.

As shown in Table 3, MANGO achieves PCC values of 0.409 on metabolomics and 0.292 on proteomics. STFlow is the strongest baseline on both tasks, with 0.233 and 0.204, leaving absolute improvements of 0.176 and 0.088. The advantage is therefore preserved beyond RNA expression and across molecular readouts with markedly different measurement processes. These results support the ability of the learned tissue representation and flow model to transfer across spatial-omics modalities within a unified framework.

Table 3: **Extension to other spatial omics**, including metabolomics and proteomics.

Methods	**Metabolomics** (Mouse PD)	**Proteomics** (69 Regions)
UNI	$0.188_{.037}$	$0.094_{.090}$
H0-mini	$0.147_{.047}$	$0.063_{.064}$
STFlow	$0.233_{.070}$	$0.204_{.098}$
MERGE	$0.155_{.079}$	$0.120_{.100}$
OURS	$\mathbf{0.409_{.066}}$	$\mathbf{0.292_{.050}}$

Table 4: **Ablation study of MANGO method**. Ablation of core components on COAD Visium.

Components	**UNI**	**H0-mini**
w/o Expression Node	$0.317_{.004}$	$0.313_{.017}$
w/o Gene Weights	$0.309_{.005}$	$0.291_{.017}$
w/o Dynamic Graph	$0.305_{.021}$	$0.315_{.006}$
w/o Multiplicative Edge	$0.301_{.003}$	$0.302_{.004}$
w/o Morph Edge	$0.307_{.006}$	$0.309_{.004}$
MANGO (Full)	$\mathbf{0.328_{.012}}$	$\mathbf{0.346_{.009}}$

6.4 Ablation Studies

Efficacy of Core Components. Table 4 evaluates the proposed components on COAD using both UNI and H0-mini. Every ablation reduces PCC under both encoders. Removing the multiplicative edge causes the largest drop with UNI, from 0.328 to 0.301, while removing gene weighting causes the largest drop with H0-mini, from 0.346 to 0.291. Disabling dynamic graph construction also lowers PCC to 0.305 and 0.315, respectively, and removing the expression node or morphological edge produces consistent degradation. These results show that the gains arise from the joint use of expression conditioning, dynamic topology, multi-view relations, and gene-balanced supervision rather than from a single dominant component.

Effect of Pathology Foundation Models. MANGO remains effective with both pathology encoders. On COAD, the complete model obtains 0.328 PCC with UNI and 0.346 with H0-mini, compared with 0.275 and 0.257 for their corresponding direct prediction baselines in Table 1. This translates into substantial absolute PCC gains of 0.053 and 0.089, respectively. Moreover, all component removals hurt both variants. The consistent behavior across encoders indicates that the gains primarily arise from the proposed modeling framework rather than the specific choice of pretrained pathology foundation model.

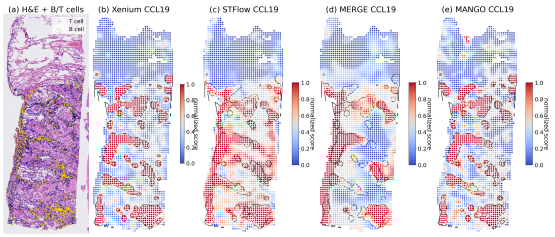


Figure 3: Spatial recovery of CCL19 on the TENX141 lung section. (a) H&E with aligned Xenium T and B cells; (b) measured CCL19; (c–e) predictions from STFlow, MERGE, and MANGO.

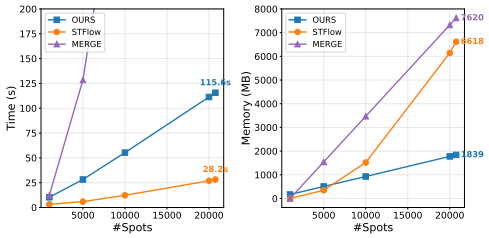


Figure 4: Inference time and peak GPU memory with increasing spot counts on TENX99, a 10x Genomics Xenium breast cancer sample from HEST benchmark IDC.

6.5 Biological Utility and Computational Efficiency

Recovery of spatially localized CCL19 expression. TLSs are spatially organized lymphocyte aggregates in which B cells play a central role, and their maturation is associated with prognosis in NSCLC (He et al., 2023). Beyond cell abundance, the spatial arrangement of B and T cells characterizes clinically relevant lung-cancer immune hotspots (Zhang et al., 2023); CCL19-expressing stromal and dendritic cells together with CCR7-expressing T cells are enriched in stem-immunity hubs (Chen et al., 2024a). Motivated by these findings, we evaluate recovery of this localized, low-abundance signal on TENX141 in Figure 3. The reference is derived independently from Xenium-measured CCL19 counts registered to H&E coordinates through HEST, rather than from model predictions. As complementary cellular context, Leiden clusters of normalized Xenium single-cell counts are annotated using canonical B-cell (CD19/MS4A1/CD79A) and T-cell (CD2/CD3D/CD3E/CCR7) markers. Measured and predicted CCL19 scores are standardized and smoothed identically; the top-20% contour is defined only from the measured reference and fixed across methods. MANGO most closely preserves the reference pattern and its correspondence with B/T-cell-rich regions. Compared with the strongest baseline STFlow, MANGO reduces MAE from 0.911 to 0.496, and raises the CCL19 PCC from 0.624 to 0.679.

Computational scalability. To evaluate efficiency, we compare total inference time and peak GPU memory across slide-based models as the number of spatial spots increases in Figure 4. Inference time is reported as the total duration of 100 repeated runs. For a fair comparison, visual feature extraction and the fixed memory required to load the pathology encoder are excluded for all methods. MANGO consistently uses substantially less GPU memory and remains faster than MERGE as the slide size increases, although STFlow has lower wall-clock time. This favorable memory scaling is enabled by restricting attention to dynamically selected local neighbors rather than constructing dense interactions over the entire slide. At approximately 20,000 spots, MANGO uses only 1,839 MB, versus 6,618 MB for STFlow and 7,620 MB for MERGE.

7 Conclusion

We presented MANGO, an expression-conditioned Riemannian graph flow for predicting spatial molecular profiles from routine histology. The framework combines multiplicative-geometry flow matching, clean-endpoint prediction, and a dynamic multi-view graph Transformer whose topology evolves with the inferred expression state. Its expression-aware objective further balances supervision across genes. Across 17 transcriptomics datasets, MANGO achieves the highest average PCC in every evaluation group and ranks first or second on every individual dataset. It also retains its advantage in representative all-gene experiments and transfers to spatial metabolomics and proteomics. Consistent ablation results under two pathology encoders further show that multiplicative relations, dynamic graph refinement, expression conditioning, and gene weighting contribute complementary gains. Overall, the results establish geometry-aware transport and adaptive tissue modeling as a unified approach to spatial-omics inference across platforms, resolutions, and molecular modalities.

AI USE STATEMENT

In this work, we did not use generative AI tools to generate synthetic datasets, develop theoretical models or conceptual frameworks, formulate mathematical claims or proofs, propose or refine hypotheses, design the research methodology or experiments, implement the proposed method, translate content, clean or reformat datasets, conduct qualitative or thematic data analysis, or interpret experimental results. We used generative AI tools for limited assistance in drafting and editing the manuscript, including improving readability, refining the wording of quantitative comparisons, and checking formatting consistency. All AI-assisted text and formatting recommendations were manually reviewed and verified by the authors for accuracy and consistency with the underlying methods and experimental results. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

We acknowledge that we have read and adhere to the ICLR Code of Ethics. All datasets used in this study are publicly available and were obtained from previously published works or official data repositories. Their sources, modalities, and exact sample identifiers are documented in Appendix C. We did not collect new human-subject data or interact directly with any participants, and we do not attempt to re-identify individuals represented in the datasets. The data were used solely for research purposes in accordance with their applicable licenses and terms of use. The proposed method is intended for scientific research rather than clinical decision-making; any downstream clinical application would require independent validation and appropriate ethical and regulatory review.

REPRODUCIBILITY STATEMENT

We have taken several steps to facilitate the reproducibility of this work. The problem formulation, model architecture, learning objectives, datasets, evaluation protocol, and experimental results are described in Sections 3–6. Appendix A defines the evaluation metrics, Appendix C documents the public data sources and exact sample identifiers, Appendix D provides the computing environment, implementation details, training settings, and optimization hyperparameters, and Appendix E reports the hyperparameter sensitivity analysis. Additional quantitative and qualitative results are provided in Appendices F and G. All datasets used in this study are publicly available through the cited sources. An anonymized repository containing the data-preprocessing, training, evaluation, and configuration files is available at https://anonymous.4open.science/r/MANGO-BFB3/.

REFERENCES

Jonathan H Chen, Linda T Nieman, Maxwell Spurrell, Vjola Jorgji, Liad Elmelech, Peter Richieri, Katherine H Xu, Roopa Madhu, Milan Parikh, Izabella Zamora, et al. Human lung cancer harbors spatially organized stem-immunity hubs associated with response to immunotherapy. *Nature immunology*, 25(4):644–658, 2024a.

Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3):850–862, 2024b.

Ricky TQ Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, volume 2024, pp. 47922–47945, 2024.

Donghai Fang, Yongheng Li, Zhen Wang, Yuansong Zeng, and Wenwen Min. Adapting a pre-trained single-cell foundation model to spatial gene expression generation from histology images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5720–5729, June 2026.

Alexandre Filiot, Nicolas Dop, Oussama Tchita, Auriane Riou, Rémy Dubois, Thomas Peeters, Daria Valter, Marin Scalbert, Charlie Saillard, Geneviève Robin, et al. Distilling foundation models for robust and efficient models in digital pathology. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 162–172. Springer, 2025.

Xiaohang Fu, Yue Cao, Beilei Bian, Chuhan Wang, Dinny Graham, Nirmala Pathmanathan, Ellis Patrick, Jinman Kim, and Jean Yee Hwa Yang. Spatial gene expression at single-cell resolution from histology using deep learning with ghist. *Nature methods*, 22(9):1900–1910, 2025.

Aniruddha Ganguly, Debolina Chatterjee, Wentao Huang, Jie Zhang, Alisa Yurovsky, Travis Steele Johnson, and Chao Chen. Merge: Multi-faceted hierarchical graph-based gnn for gene expression prediction from whole slide histopathology images. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15611–15620. IEEE, 2025.

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.

Miao He, Qihua He, Xiuyu Cai, Jun Liu, Hongshen Deng, Feng Li, Ran Zhong, Yi Lu, Haoxin Peng, Xiangrong Wu, et al. Intratumoral tertiary lymphoid structure (tls) maturation is influenced by draining lymph nodes of lung cancer. *Journal for Immunotherapy of Cancer*, 11(4):e005539, 2023.

Tinglin Huang, Tianyu Liu, Mehrtash Babadi, Wengong Jin, and Rex Ying. Scalable generation of spatial transcriptomics from histology images via whole-slide flow matching. In *International Conference on Machine Learning*, 2025.

Guillaume Jaume, Paul Doucet, Andrew H Song, Ming Y Lu, Cristina Almagro-Pérez, Sophia J Wagner, Anurag J Vaidya, Richard J Chen, Drew F Williamson, Ahrong Kim, et al. Hest-1k: A dataset for spatial transcriptomics and histology image analysis. *Advances in Neural Information Processing Systems*, 37:53798–53833, 2024.

Johanna Klughammer, Daniel L Abravanel, Åsa Segerstolpe, Timothy R Blosser, Yury Goltsev, Yi Cui, Daniel R Goodwin, Anubhav Sinha, Orr Ashenberg, Michal Slyper, et al. A multi-modal single-cell and spatial expression map of metastatic breast cancer biopsies across clinicopathological features. *Nature medicine*, 30(11):3236–3249, 2024.

Yonghao Liu, Chuyao Wang, Zhikang Wang, Liang Chen, Zhi Li, Jiangning Song, Qi Zou, Rui Gao, Bin-Zhi Qian, Xiaoyue Feng, et al. High-parameter spatial multi-omics through histology-anchored integration. *Nature Methods*, 23(2):373–386, 2026.

Minxing Pang, Kenong Su, and Mingyao Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *BioRxiv*, pp. 2021–11, 2021.

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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 574–584. Springer, 2025.

Qi Si, Penglei Wang, Yushuai Wu, Yifeng Jiao, Xuyang Liu, Xin Guo, Yuan Qi, and Yuan Cheng. Flag: Foundation model representation with latent diffusion alignment via graph for spatial gene expression prediction. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, 2026.

Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353 (6294):78–82, 2016.

Marco Vicari, Reza Mirzazadeh, Anna Nilsson, Reza Shariatgorji, Patrik Bjärterot, Ludvig Larsson, Hower Lee, Mats Nilsson, Julia Foyer, Markus Ekvall, et al. Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nature Biotechnology*, 42(7):1046–1050, 2024.

Chuhan Wang, Adam S Chan, Xiaohang Fu, Shila Ghazanfar, Jinman Kim, Ellis Patrick, and Jean YH Yang. Benchmarking the translational potential of spatial gene expression prediction from histology. *Nature Communications*, 16(1):1544, 2025.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12, 2019.

Yi Xiao and Dihua Yu. Tumor microenvironment as a therapeutic target in cancer. *Pharmacology & therapeutics*, 221:107753, 2021.

Hanyun Zhang, Khalid AbdulJabbar, David A Moore, Ayse Akarca, Katey SS Enfield, Mariam Jamal-Hanjani, Shan E Ahmed Raza, Selvaraju Veeriah, Roberto Salgado, Nicholas McGranahan, et al. Spatial positioning of immune hotspots reflects the interplay between b and t cells in lung squamous cell carcinoma. *Cancer Research*, 83(9):1410–1425, 2023.

Sichen Zhu, Yuchen Zhu, Molei Tao, and Peng Qiu. Diffusion generative modeling for spatially resolved gene expression inference from histology images. In *International Conference on Learning Representations*, volume 2025, pp. 19666–19690, 2025.

A METRIC COMPUTATION

This section details the evaluation metrics used to quantify the agreement between predicted and ground-truth spatial molecular profiles.

Let $y_{i,g}$ and $\hat{y}_{i,g}$ denote the ground-truth and predicted values of feature (gene/metabolite/protein) g at spatial spot/cell i, respectively. For each feature, we compute metrics over N valid spots/cells after $\log 1p$ normalization.

Pearson Correlation Coefficient (PCC). PCC measures the linear correlation between predicted and ground-truth profiles for each feature independently. It is computed as:

$$\mathrm{PCC}_g = \frac{\sum_{i=1}^{N}(y_{i,g} - \bar{y}_g)(\hat{y}_{i,g} - \bar{\hat{y}}_g)}{\sqrt{\sum_{i=1}^{N}(y_{i,g} - \bar{y}_g)^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_{i,g} - \bar{\hat{y}}_g)^2}}, \tag{12}$$

where $\bar{y}_g = \frac{1}{N}\sum_{i=1}^{N} y_{i,g}$ and $\bar{\hat{y}}_g = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_{i,g}$ are the mean ground-truth and predicted values for feature g, respectively. The reported PCC is averaged over the top 50 highly variable features.

Mean Squared Error (MSE). MSE measures the average squared difference between predictions and ground truths:

$$\mathrm{MSE}_g = \frac{1}{N}\sum_{i=1}^{N}(y_{i,g} - \hat{y}_{i,g})^2. \tag{13}$$

Lower MSE indicates better prediction accuracy.

Mean Absolute Error (MAE). MAE measures the average magnitude of prediction errors:

$$\mathrm{MAE}_g = \frac{1}{N}\sum_{i=1}^{N}\left|y_{i,g} - \hat{y}_{i,g}\right|. \tag{14}$$

Like MSE, lower MAE corresponds to more accurate predictions.

For the main results, we report the mean and standard deviation of each metric across three random seeds.

B OVERVIEW OF MODELING CAPABILITIES

Table 5: Comparison of modeling capabilities. Checkmarks indicate that a method explicitly incorporates the corresponding mechanism.

Capabilities	**UNI**	**HINGE**	**Spatial Ex**	**STFlow**	**NH^{2}ST**	**Ours**
Conditional generative modeling	×	✓	×	✓	×	✓
Slide-level contextual modeling	×	×	✓	✓	✓	✓
Multiplicative-geometry transport	×	×	×	×	×	✓
Expression-conditioned dynamic topology	×	×	×	×	×	✓
Multi-view edge-conditioned attention	×	×	×	×	×	✓
Gene-balanced supervision	×	×	×	×	×	✓

C DATASET DESCRIPTIONS

In this appendix, we provide detailed statistics and configurations for the datasets utilized in our experiments. Our benchmark comprises four distinct spatial omics dataset collections, spanning varying technologies, modalities, and organs. To ensure full reproducibility, the exact file names (sample IDs) used for each task are explicitly documented.

C.1 DATASET 1: HEST-BENCHMARK

Our first collection is based on the HEST-Benchmark, which encompasses 9 distinct spatial transcriptomics tasks covering various human organs. The spatial expression data is profiled using either 10x Genomics Xenium or Visium technologies. As detailed in Table 6, we process tasks for Breast (IDC), Prostate (PRAD), Pancreas (PAAD), Skin (SKCM), Colon (COAD), Rectum (READ), Kidney (ccRCC), Lung (LUAD), and Lymph Nodes (IDC_LN).

Specific sample filenames utilized include *TENX95, TENX99, NCBI783, NCBI785* for Task 1 (IDC), *TENX116, TENX126, TENX140* for Task 3 (PAAD), and *INT1* to *INT24* for Task 7 (ccRCC). Note that for the PRAD task, sample *MEND155* is explicitly excluded due to missing data in the original repository, resulting in 23 valid samples (*MEND139–MEND154, MEND156–MEND162*). A full list of sample IDs for all tasks is provided in Table 9.

Table 6: Dataset statistics of HEST-Bench.

	IDC	PRAD	PAAD	SKCM	COAD	READ	ccRCC	LUAD	IDC_LN
Organ	Breast	Prostate	Pancreas	Skin	Colon	Rectum	Kidney	Lung	Lymph Node
Technology	Xenium	Visium	Xenium	Xenium	Visium	Visium	Visium	Xenium	Visium
#Samples	4	23	3	2	4	4	24	2	4

C.2 DATASET 2: HEST-1K (VISIUMHD)

The second dataset collection is derived from HEST-1k, which focuses on sub-cellular and high-resolution spatial transcriptomics acquired via the VisiumHD technology. We construct four specific benchmark tasks corresponding to Breast Cancer (IDC), Lung Adenocarcinoma (LUAD), Colon Adenocarcinoma (COAD), and Serous Ovarian Cancer (SOC). The statistical breakdown is summarized in Table 7. The specific high-resolution slices used are meticulously selected, such as *TENX161, TENX162, TENX180* for IDC, and corresponding *TENX* identifiers for other organs (see Table 9).

Table 7: Dataset statistics of HEST-1k (VisiumHD).

	IDC	LUAD	COAD	SOC
Organ	Breast	Lung	Colon	Ovary
Technology	VisiumHD	VisiumHD	VisiumHD	VisiumHD
#Samples	3	5	6	4

C.3 DATASET 3: SPATIAL METABOLIC DATA

To evaluate model performance on spatial metabolic modalities, we utilize Mass Spectrometry Imaging (MSI) data from the Spatial Multimodal Analysis (SMA) repository. Specifically, we focus on the FMP-10 matrix applied to mouse brain tissues. To ensure robust model training, we strictly utilize high-quality capture areas: V11T16-085 A1, V11T16-085 B1, V11T16-085 C1, corresponding to three individual mice (mPD1, mPD3, mPD4). Sample *V11L12-109* was intentionally excluded because of its high sparsity rate ($\sim$75% zeros), which severely degrades model performance. These samples are grouped by individual mouse to perform a rigorous 3-fold leave-one-mouse-out cross-validation.

C.4 DATASET 4: SPATIAL PROTEIN DATA

Our fourth dataset incorporates spatial proteomics data targeting Metastatic Breast Cancer (MBC), sourced from the Human Tumor Atlas Pilot Project (HTAPP-MBC). This cohort utilizes CODEX multiplexed imaging technology for protein profiling. The sample inclusion relies on all currently available and matched patient metadata mapped within the repository. The configurations for Dataset 3 and Dataset 4 are jointly summarized in Table 8.

Table 8: Dataset statistics of spatial metabolic and protein data.

	FMP10	CODEX_Protein
Organ	Mouse Brain	Breast
Technology	FMP-10 (MSI)	CODEX
Modality	Metabolic	Protein
#Samples	3	13

C.5 Detailed Sample Identifiers

To facilitate straightforward dataset retrieval and reproducibility, Table 9 provides a comprehensive enumeration of all slice filenames (Sample IDs) assigned to each task in Dataset 1, Dataset 2, and Dataset 3.

Table 9: Detailed listing of exact sample filenames (Sample IDs) used across different benchmark tasks.

Dataset Collection	**Task Name**	**Included Sample IDs (Filenames)**
Dataset 1: HEST-Bench	Task1 IDC	TENX95, TENX99, NCBI783, NCBI785
	Task2 PRAD	MEND139 to MEND154, MEND156 to MEND162
	Task3 PAAD	TENX116, TENX126, TENX140
	Task4 SKCM	TENX115, TENX117
	Task5 COAD	TENX111, TENX147, TENX148, TENX149
	Task6 READ	ZEN36, ZEN40, ZEN48, ZEN49
	Task7 ccRCC	INT1, INT2, INT3, ..., INT24 *(inclusive from 1 to 24)*
	Task8 LUAD	TENX118, TENX141
	Task9 IDC_LN	NCBI681, NCBI682, NCBI683, NCBI684
Dataset 2: HEST-1k (VisiumHD)	Task1 IDC	TENX161, TENX162, TENX180
	Task2 LUAD	TENX163, TENX168, TENX169, TENX170, TENX171
	Task3 COAD	TENX128, TENX153, TENX154, TENX155, TENX156, TENX175
	Task4 SOC	TENX182, TENX183, TENX184, TENX185
Dataset 3:	Task1 FMP10	V11T16-085 A1, V11T16-085 B1, V11T16-085 C1
Dataset 4:	Task1 CODEX	HTAPP-213-SMP-6752, HTAPP-313-SMP-932, HTAPP-330-SMP-1082, HTAPP-364-SMP-1, HTAPP-514-SMP-6760, HTAPP-783-SMP-4081, HTAPP-812-SMP-8239, HTAPP-853MP-4381, HTAPP-878-SMP-7149, HTAPP-880-SMP-7179, HTAPP-917-SMP-4531, HTAPP-982-SMP-7629, HTAPP-997-SMP-7789

D Implementation Details

D.1 Running environment

The experiments are conducted on a Linux server equipped with two Intel Xeon Gold 6530 processors, 504GB RAM, and eight NVIDIA GeForce RTX 5090 GPUs. Our method is implemented on PyTorch 2.11.0 and Python 3.10.20.

D.2 Training details

The denoiser contains four dynamic multi-view graph-attention blocks with hidden dimension 128, pairwise edge dimension 128, and four heads. Both feature and attention dropout are set to 0.2. We set the spatial, morphology, and multiplicative graph weights to $w_s = w_p = w_m = 1$, retain $k = 3$ neighbors, and rebuild the graph after every block. The expression-aware loss uses $\gamma = 0.5$, stabilizer 10^{-3}, and maximum gene weight 5; all weights are renormalized to unit mean. The multiplicative geometry uses shift 1 and projection tolerance 10^{-6}.

All experiments use Adam with batch size 2, learning rate 5×10^{-4}, no weight decay, and gradient clipping at 1.0. We train for at most 100 epochs, evaluate after every epoch, and apply early stopping with patience 20 based on training PCC. Histology patches are augmented by random horizontal and vertical flips, 90° rotations, and mild brightness, contrast, and saturation jitter. Flow-matching times are sampled uniformly from $[0, 1]$, and the source distribution is a wrapped Gaussian with standard deviation 0.5. Inference uses a uniform five-point grid from 0.01 to 1.0 with Riemannian Euler updates, and we retain the checkpoint attaining the highest training PCC.

D.3 Model Parameters

The number of model parameters for all implemented methods is summarized in Table 10. The linear-probe baselines (UNI and H0-mini) require the fewest parameters, while NH2ST and Stem are the heaviest trainable models. Our main MANGO variants remain lightweight compared with MERGE and Stem.

Table 10: Number of parameters comparison for the implemented models.

	MANGO	MERGE	NH2ST	SpatialEx	STFlow	Stem	UNI	H0-mini
#Params (M)	1.747	18.575	51.692	1.208	1.706	35.581	0.279	0.212

E Sensitivity Sweep

We evaluate four key hyperparameters on two representative Xenium tasks. The SKCM and LUNG settings each contain two tissue sections with different numbers of spatial locations, as summarized in Table 11.

Table 11: Datasets used for sensitivity analysis. Spot counts are shown in parentheses.

Dataset	Slide 1	Slide 2
SKCM	TENX115 (1,830)	TENX117 (3,886)
LUNG	TENX118 (3,115)	TENX141 (4,390)

Table 12: Sensitivity of mean PCC to the gene-weighting exponent γ, retained-neighbor count k, patch-distance weight w_p, and manifold-distance weight w_m. Bold values denote the best result within each parameter sweep.

Hyperparameter	Value	SKCM	LUNG
Gene weighting γ	0.00	0.682	0.584
	0.25	0.693	0.598
	0.50	**0.701**	**0.618**
	1.00	0.685	0.579
Retained neighbors k	1	0.678	0.587
	2	0.693	0.593
	3	**0.701**	**0.618**
	5	0.682	0.585
	8	0.682	0.585
Patch-distance weight w_p	0.5	0.688	0.597
	1.0	**0.701**	**0.618**
	2.0	0.687	0.579
Manifold-distance weight w_m	0.5	0.686	0.587
	1.0	**0.701**	**0.618**
	2.0	0.687	0.586

Both datasets favor the same configuration: $\gamma = 0.5$, $k = 3$, $w_p = 1.0$, and $w_m = 1.0$. Performance is stable around these settings, while more aggressive gene reweighting or larger neighborhoods consistently reduces PCC.

F ADDITIONAL MSE AND MAE RESULTS

Table 13: The results of gene expression prediction (MSE). Datasets are systematically categorized by spatial resolution and sequencing platforms. **Bold** and underlined values indicate the best and second-best results, respectively.

			Discriminative Models					Generative Models			
			UNI	H0-mini	MERGE	NH^{2}ST	SpatialEx	STFlow	Stem	HINGE	**OURS**
Spot-level (Patch & Bin)	Xenium	IDC	$7.053_{3.447}$	$5.905_{2.043}$	$2.674_{.299}$	$2.267_{.671}$	$4.974_{.000}$	$\underline{2.062}_{.463}$	$5.109_{1.125}$	$2.824_{.629}$	$\mathbf{2.061}_{\mathbf{.301}}$
		PAAD	$3.282_{.497}$	$3.124_{.391}$	$0.989_{.191}$	$1.489_{.538}$	$2.644_{.557}$	$\mathbf{0.855}_{\mathbf{.146}}$	$2.652_{.258}$	$1.874_{.174}$	$\underline{0.938}_{.263}$
		SKCM	$4.127_{.887}$	$4.026_{.988}$	$1.832_{.376}$	$1.778_{.396}$	$4.335_{.872}$	$\underline{1.721}_{.352}$	$3.180_{.336}$	$2.086_{.308}$	$\mathbf{1.435}_{\mathbf{.007}}$
		LUNG	$5.679_{1.549}$	$5.392_{1.843}$	$2.047_{.304}$	$4.961_{1.089}$	$4.413_{.200}$	$\mathbf{1.572}_{\mathbf{.046}}$	$3.772_{.227}$	$2.022_{.091}$	$\underline{1.629}_{.065}$
	Xenium Avg.		5.035	4.612	1.886	2.624	4.092	$\underline{1.553}$	3.678	2.202	**1.516**
	Visium	PRAD	$1.212_{.288}$	$1.034_{.010}$	$0.550_{.118}$	$0.566_{.110}$	$0.911_{.265}$	$\underline{0.505}_{.145}$	$1.198_{.466}$	$0.698_{.225}$	$\mathbf{0.412}_{\mathbf{.204}}$
		COAD	$3.646_{1.826}$	$3.749_{1.430}$	$2.515_{.803}$	$2.176_{.726}$	$2.699_{1.049}$	$\underline{2.040}_{.509}$	$9.171_{6.427}$	$3.023_{1.988}$	$\mathbf{1.922}_{\mathbf{.783}}$
		READ	$0.401_{.039}$	$0.429_{.027}$	$0.570_{.171}$	$0.490_{.069}$	$0.412_{.017}$	$\underline{0.301}_{.081}$	$0.890_{.153}$	$0.548_{.099}$	$\mathbf{0.298}_{\mathbf{.075}}$
		CCRCC	$0.822_{.335}$	$0.814_{.296}$	$\underline{0.456}_{.144}$	$0.457_{.134}$	$0.609_{.180}$	$\mathbf{0.449}_{\mathbf{.107}}$	$1.051_{.263}$	$0.589_{.146}$	$0.495_{.178}$
		LYMPH	$0.766_{.451}$	$0.761_{.480}$	$0.743_{.249}$	$0.686_{.299}$	$\underline{0.657}_{.234}$	$0.658_{.288}$	$1.666_{.620}$	$0.710_{.371}$	$\mathbf{0.524}_{\mathbf{.268}}$
	Visium Avg.		1.369	1.357	0.967	0.875	1.058	$\underline{0.791}$	2.795	1.114	**0.730**
	Visium HD	Breast	$0.488_{.208}$	$0.464_{.256}$	$0.363_{.220}$	$0.425_{.178}$	$\underline{0.311}_{.228}$	$\mathbf{0.304}_{\mathbf{.201}}$	$0.485_{.215}$	$0.344_{.204}$	$0.322_{.150}$
		Lung	$0.321_{.021}$	$0.322_{.014}$	$0.240_{.054}$	$0.328_{.043}$	$0.421_{.038}$	$\mathbf{0.222}_{\mathbf{.016}}$	$0.335_{.019}$	$0.255_{.017}$	$\underline{0.235}_{.046}$
		Colon	$0.467_{.381}$	$0.466_{.366}$	$0.377_{.334}$	$\mathbf{0.266}_{\mathbf{.055}}$	$0.337_{.090}$	$0.343_{.313}$	$0.478_{.352}$	$0.367_{.323}$	$\underline{0.286}_{.085}$
		Ovary	$0.686_{.448}$	$0.692_{.455}$	$0.438_{.104}$	$\underline{0.398}_{.086}$	$0.619_{.339}$	$0.422_{.357}$	$0.708_{.448}$	$0.487_{.293}$	$\mathbf{0.374}_{\mathbf{.242}}$
	Visium HD Avg.		0.491	0.486	0.355	0.354	0.422	$\underline{0.323}$	0.502	0.363	**0.304**
Cell-level	Xenium	Breast	$0.529_{.000}$	$0.524_{.007}$	$0.322_{.012}$	$\underline{0.306}_{.015}$	$0.315_{.011}$	$0.468_{.044}$	$0.542_{.024}$	$0.426_{.064}$	$\mathbf{0.305}_{\mathbf{.010}}$
		Skin	$0.535_{.110}$	$0.551_{.127}$	$0.301_{.014}$	$0.318_{.016}$	$\underline{0.294}_{.013}$	$0.635_{.052}$	$0.584_{.061}$	$0.621_{.047}$	$\mathbf{0.288}_{\mathbf{.001}}$
		Colon (M)	$0.810_{.019}$	$0.805_{.018}$	$\underline{0.397}_{.015}$	$0.411_{.018}$	$0.421_{.016}$	$0.576_{.031}$	$0.618_{.038}$	$0.664_{.035}$	$\mathbf{0.390}_{\mathbf{.017}}$
		Colon (H)	$0.407_{.032}$	$0.410_{.034}$	$0.244_{.010}$	$\underline{0.232}_{.012}$	$0.253_{.011}$	$0.534_{.018}$	$0.589_{.021}$	$0.472_{.016}$	$\mathbf{0.225}_{\mathbf{.001}}$
	Cell-level Avg.		0.570	0.573	$\underline{0.316}$	0.317	0.321	0.553	0.583	0.546	**0.302**

Table 14: The results of gene expression prediction (MAE). Datasets are systematically categorized by spatial resolution and sequencing platforms. **Bold** and underlined values indicate the best and second-best results, respectively.

			Discriminative Models					Generative Models			
			UNI	H0-mini	MERGE	NH^{2}ST	SpatialEx	STFlow	Stem	HINGE	**OURS**
Spot-level (Patch & Bin)	Xenium	IDC	$2.125_{.592}$	$1.926_{.369}$	$1.240_{.091}$	$1.104_{.216}$	$1.640_{.000}$	$\mathbf{1.024}_{\mathbf{.156}}$	$1.611_{.283}$	$1.171_{.188}$	$\underline{1.072}_{.120}$
		PAAD	$1.376_{.147}$	$1.369_{.136}$	$0.752_{.071}$	$0.891_{.138}$	$1.116_{.168}$	$\mathbf{0.696}_{\mathbf{.055}}$	$1.200_{.082}$	$0.822_{.062}$	$\underline{0.720}_{.101}$
		SKCM	$1.486_{.244}$	$1.463_{.257}$	$0.973_{.099}$	$0.995_{.100}$	$1.320_{.230}$	$\underline{0.927}_{.052}$	$1.267_{.069}$	$1.012_{.056}$	$\mathbf{0.872}_{\mathbf{.005}}$
		LUNG	$1.940_{.343}$	$1.859_{.392}$	$1.054_{.096}$	$1.782_{.277}$	$1.563_{.049}$	$\mathbf{0.925}_{\mathbf{.026}}$	$1.440_{.043}$	$1.054_{.030}$	$\underline{0.956}_{.007}$
	Xenium Avg.		1.732	1.654	1.005	1.193	1.410	**0.893**	1.380	1.015	$\underline{0.905}$
	Visium	PRAD	$0.752_{.140}$	$0.683_{.041}$	$0.545_{.071}$	$0.562_{.064}$	$0.612_{.031}$	$\underline{0.518}_{.080}$	$0.773_{.168}$	$0.582_{.102}$	$\mathbf{0.481}_{\mathbf{.114}}$
		COAD	$1.360_{.329}$	$1.378_{.262}$	$1.104_{.126}$	$1.010_{.139}$	$1.049_{.254}$	$\underline{0.984}_{.078}$	$1.620_{.215}$	$1.143_{.112}$	$\mathbf{0.939}_{\mathbf{.117}}$
		READ	$0.435_{.003}$	$0.463_{.002}$	$0.542_{.074}$	$0.520_{.036}$	$\underline{0.383}_{.002}$	$0.397_{.059}$	$0.600_{.068}$	$0.448_{.061}$	$\mathbf{0.377}_{\mathbf{.045}}$
		CCRCC	$0.670_{.136}$	$0.669_{.113}$	$\underline{0.517}_{.060}$	$\underline{0.517}_{.063}$	$0.583_{.082}$	$\mathbf{0.516}_{\mathbf{.051}}$	$0.771_{.070}$	$0.580_{.056}$	$0.532_{.070}$
		LYMPH	$0.619_{.172}$	$0.613_{.185}$	$0.630_{.092}$	$0.595_{.156}$	$0.609_{.191}$	$\underline{0.588}_{.145}$	$0.886_{.211}$	$0.662_{.161}$	$\mathbf{0.517}_{\mathbf{.132}}$
	Visium Avg.		0.767	0.761	0.668	0.641	0.647	$\underline{0.601}$	0.930	0.683	**0.569**
	Visium HD	Breast	$0.470_{.050}$	$0.460_{.081}$	$0.407_{.100}$	$0.443_{.082}$	$\mathbf{0.301}_{\mathbf{.097}}$	$0.367_{.063}$	$0.472_{.076}$	$0.388_{.066}$	$\underline{0.360}_{.085}$
		Lung	$0.394_{.026}$	$0.397_{.019}$	$\underline{0.334}_{.045}$	$0.347_{.045}$	$0.337_{.032}$	$\mathbf{0.321}_{\mathbf{.033}}$	$0.405_{.021}$	$0.342_{.030}$	$\underline{0.364}_{.049}$
		Colon	$0.429_{.130}$	$0.431_{.122}$	$0.407_{.161}$	$0.350_{.044}$	$\underline{0.275}_{.058}$	$0.337_{.097}$	$0.442_{.118}$	$0.363_{.102}$	$\mathbf{0.209}_{\mathbf{.092}}$
		Ovary	$0.616_{.214}$	$0.617_{.214}$	$0.484_{.025}$	$0.461_{.025}$	$\underline{0.439}_{.172}$	$0.452_{.162}$	$0.632_{.210}$	$0.497_{.174}$	$\mathbf{0.429}_{\mathbf{.218}}$
	Visium HD Avg.		0.477	0.476	0.408	0.400	**0.338**	0.369	0.488	0.398	$\underline{0.341}$
Cell-level	Xenium	Breast	$0.450_{.000}$	$0.450_{.001}$	$0.379_{.012}$	$\underline{0.365}_{.014}$	$0.371_{.011}$	$0.432_{.018}$	$0.469_{.021}$	$0.414_{.017}$	$\mathbf{0.362}_{\mathbf{.009}}$
		Skin	$0.448_{.065}$	$0.459_{.074}$	$0.366_{.013}$	$0.382_{.015}$	$\underline{0.357}_{.014}$	$0.526_{.022}$	$0.489_{.028}$	$0.514_{.020}$	$\mathbf{0.350}_{\mathbf{.006}}$
		Colon (M)	$0.579_{.009}$	$0.580_{.008}$	$0.446_{.014}$	$0.458_{.016}$	$\underline{0.441}_{.015}$	$0.486_{.024}$	$0.512_{.029}$	$0.529_{.027}$	$\mathbf{0.437}_{\mathbf{.032}}$
		Colon (H)	$0.341_{.041}$	$0.341_{.045}$	$0.294_{.010}$	$\underline{0.283}_{.012}$	$0.301_{.011}$	$0.451_{.019}$	$0.482_{.023}$	$0.421_{.017}$	$\mathbf{0.277}_{\mathbf{.001}}$
	Cell-level Avg.		0.455	0.458	0.371	0.372	$\underline{0.368}$	0.474	0.488	0.470	**0.357**

Discussion. Tables 13 and 14 complement the correlation-based evaluation with pointwise errors. MANGO achieves the lowest average MSE in all four evaluation groups. For MAE, it ranks first on Visium and cell-level Xenium, second to STFlow on spot-level Xenium, and second to SpatialEx on Visium HD. These results show that the PCC gains generally coincide with accurate expression magnitude rather than only improved spatial ranking.

Table 15: The PCC results of gene expression prediction evaluated on top-10 HVGs. Bold and underlined values indicate the best and second-best results, respectively.

			Discriminative Models					Generative Models			
			UNI	H0-mini	MERGE	NH^{2}ST	SpatialEx	STFlow	Stem	HINGE	**OURS**
Spot-level (Patch & Bin)	Xenium	IDC	$0.521_{.103}$	$\underline{0.543}_{.094}$	$0.510_{.052}$	$0.523_{.105}$	$0.400_{.089}$	$\mathbf{0.555}_{\mathbf{.080}}$	0.496	$0.487_{.094}$	$0.529_{.134}$
		PAAD	$0.419_{.131}$	$0.421_{.090}$	$0.390_{.125}$	$0.343_{.143}$	$0.318_{.118}$	$0.433_{.077}$	0.374	$\underline{0.454}_{.054}$	$\mathbf{0.664}_{\mathbf{.128}}$
		SKCM	$0.614_{.016}$	$0.546_{.078}$	$0.607_{.050}$	$0.537_{.093}$	$0.591_{.013}$	$0.595_{.060}$	0.575	$\underline{0.633}$	$\mathbf{0.755}_{\mathbf{.236}}$
		LUNG	$0.472_{.005}$	$0.499_{.025}$	$0.418_{.067}$	$0.436_{.040}$	$0.431_{.013}$	$0.562_{.029}$	0.447	$\underline{0.629}_{.038}$	$\mathbf{0.702}_{\mathbf{.279}}$
	Xenium Avg.		0.507	0.502	0.481	0.460	0.435	0.536	0.473	$\underline{0.551}$	**0.663**
	Visium	PRAD	$0.282_{.082}$	$0.353_{.011}$	$0.333_{.030}$	$0.319_{.053}$	$0.298_{.063}$	$\underline{0.414}_{.017}$	0.313	$0.303_{.072}$	$\mathbf{0.541}_{\mathbf{.137}}$
		COAD	$0.216_{.026}$	$0.181_{.094}$	$0.223_{.095}$	$0.188_{.077}$	$0.282_{.052}$	$\mathbf{0.337}_{\mathbf{.031}}$	0.214	0.272	$\underline{0.328}_{.015}$
		READ	$0.168_{.048}$	$0.189_{.070}$	$0.173_{.088}$	$0.123_{.047}$	$0.162_{.018}$	$\underline{0.289}_{.054}$	0.159	$0.198_{.124}$	$\mathbf{0.514}_{\mathbf{.258}}$
		CCRCC	$0.242_{.099}$	$0.253_{.092}$	$0.248_{.073}$	$0.244_{.084}$	$0.263_{.089}$	$0.280_{.099}$	0.246	$\underline{0.304}$	$\mathbf{0.393}_{\mathbf{.029}}$
		LYMPH	$0.191_{.037}$	$0.192_{.043}$	$0.148_{.050}$	$0.188_{.045}$	$0.245_{.033}$	$0.220_{.035}$	0.189	$\underline{0.254}_{.071}$	$\mathbf{0.323}_{\mathbf{.550}}$
	Visium Avg.		0.220	0.234	0.225	0.212	0.250	$\underline{0.308}$	0.224	0.266	**0.420**
	Visium HD	Breast	$0.130_{.068}$	$0.147_{.078}$	$0.123_{.099}$	$0.138_{.092}$	$0.096_{.083}$	0.179	0.123	$\underline{0.181}$	$\mathbf{0.249}_{\mathbf{.022}}$
		Lung	$0.414_{.119}$	$0.426_{.112}$	$0.411_{.084}$	$0.436_{.093}$	0.421	$\underline{0.474}$	0.418	**0.476**	$0.462_{.037}$
		Colon	$0.358_{.048}$	$0.361_{.047}$	$0.313_{.063}$	$0.338_{.069}$	$0.377_{.068}$	$\underline{0.401}$	0.346	**0.403**	$0.380_{.307}$
		Ovary	$0.492_{.043}$	$0.474_{.044}$	$0.409_{.040}$	$0.498_{.050}$	$0.430_{.028}$	0.513	0.457	$\underline{0.515}$	**0.575**
	Visium HD Avg.		0.349	0.352	0.314	0.353	0.331	0.392	0.336	$\underline{0.394}$	**0.417**
Cell-level	Xenium	Breast	$0.372_{.010}$	$0.372_{.010}$	$0.373_{.009}$	$0.381_{.007}$	$0.371_{.013}$	$0.380_{.011}$	$0.324_{.008}$	$\underline{0.428}$	$\mathbf{0.575}_{\mathbf{.007}}$
		Skin	$0.041_{.045}$	$0.027_{.017}$	$\mathbf{0.248}_{\mathbf{.005}}$	$0.172_{.004}$	$0.195_{.005}$	$0.206_{.034}$	$\underline{0.232}_{.074}$	0.191	$0.202_{.000}$
		Colon (M)	$0.468_{.006}$	$0.481_{.015}$	$0.543_{.008}$	$0.473_{.001}$	$0.499_{.002}$	$0.448_{.004}$	$0.498_{.017}$	$\underline{0.547}$	$\mathbf{0.640}_{\mathbf{.003}}$
		Colon (H)	$0.158_{.051}$	$0.156_{.038}$	$\mathbf{0.261}_{\mathbf{.004}}$	$0.196_{.001}$	$0.224_{.006}$	$0.209_{.016}$	$0.133_{.026}$	$\underline{0.253}$	$0.239_{.012}$
	Cell-level Avg.		0.260	0.259	$\underline{0.356}$	0.306	0.322	0.311	0.297	0.355	**0.414**

G VISUALIZATION OF MARKER GENES

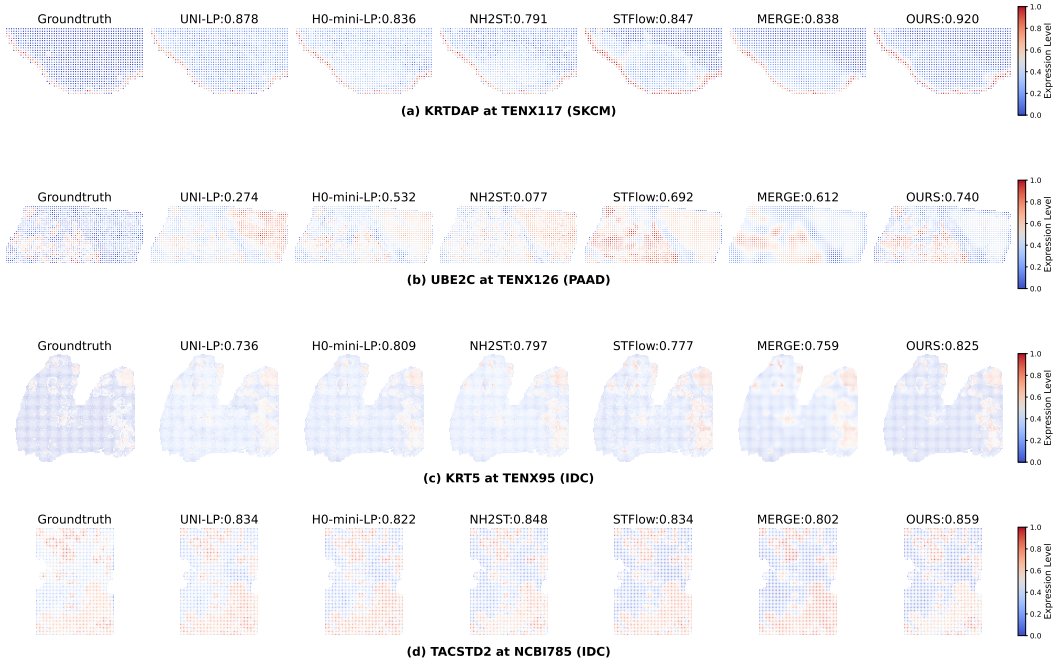


Figure 5: Qualitative marker-gene prediction on four representative Xenium sections. Each row shows measured expression followed by predictions from UNI-LP, H0-mini-LP, NH^{2}ST, STFlow, MERGE, and MANGO; titles report PCC.

Discussion. Figure 5 compares both localized and broadly distributed marker patterns across tissues. MANGO achieves the highest PCC in all four examples.