Gene-Selective Morphological Conditioning with Heteroscedastic Flow Matching for Spatial Transcriptomics

Hyun Namgung[1], Soopil Kim[2], and Sang Hyun Park[1*]

[1] Pohang University of Science & Technology (POSTECH), Republic of Korea
[2] Daegu Gyeongbuk Institute of Science & Technology (DGIST), Republic of Korea
{hyun.namgung,sanghyunpark}@postech.ac.kr

[*] Corresponding author.

Abstract. Predicting spatially resolved gene expression from histology images requires capturing how tissue morphology relates to each gene, but the strength of this coupling differs across genes. Current methods assign every gene the same morphological representation, overlooking this heterogeneity. We introduce gene binding modules that decompose morphological features into functional sub-spaces, letting each gene attend selectively to relevant programs rather than a shared global representation. To capture inter-gene dependencies, compact regulator tokens compress and redistribute information across gene groups, modeling coordinated transcriptional programs. The framework further learns per-gene heteroscedastic uncertainty that focuses model capacity on morphologically predictable genes during training, and modulates stochastic sampling at inference: tightly coupled genes receive near-deterministic generation while loosely coupled genes are sampled more broadly. Experiments on four spatial transcriptomics datasets show consistent improvements over both deterministic and generative baselines. The learned modules recover biologically coherent gene groupings without supervision, validated through pathway enrichment analysis, demonstrating that the framework provides interpretable biological insights beyond prediction accuracy.

Keywords: Spatial Transcriptomics · Flow Matching · Gene Expression Prediction · Uncertainty Estimation

1 Introduction

Spatial Transcriptomics (ST) captures gene expression within native tissue contexts [19, 2, 17], revealing biological processes invisible to bulk RNA sequencing [15]. However, measuring spatially resolved transcriptomes remains expensive and technically demanding [21], motivating computational prediction of gene expression from routinely available hematoxylin and eosin (H&E) histology images. Existing approaches span deterministic mappings [5, 16], spatial context modeling [23, 22, 4, 24, 20, 9], and generative frameworks [25, 6], yet three biological realities remain unaddressed. First, genes differ in how tightly their expression

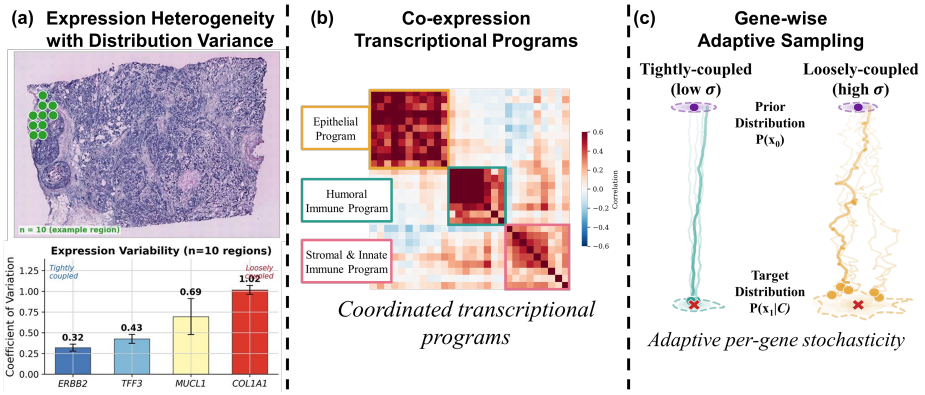

Fig. 1. Biological motivations. (a) Among morphologically similar spots, expression variability ranges from tightly coupled genes (e.g., *ERBB2*) to highly variable ones (e.g., *COL1A1*). Error bars: std over 10 regions. **(b)** Gene–gene correlation shows block-diagonal structure reflecting co-regulated transcriptional programs. **(c)** Heteroscedastic flow matching adapts stochasticity per gene: tightly-coupled genes follow near-deterministic trajectories (left) while loosely-coupled genes are sampled broadly (right).

is coupled with tissue morphology (Fig. 1(a)): among morphologically similar spots, some genes such as *ERBB2* exhibit narrow expression distributions while others such as *COL1A1* vary widely. Every existing method, however, feeds all genes the same global morphological representation. Second, genes are organized into coordinated transcriptional programs such as epithelial, humoral immune, and stromal/innate immune (Fig. 1(b)), yet no current method explicitly models these inter-gene dependencies. Third, a well-calibrated generative model should modulate its stochasticity per gene: tightly-coupled genes warrant near-deterministic generation, while loosely-coupled genes require broader sampling (Fig. 1(c)). Existing generative methods instead apply uniform noise across all genes. We propose a flow matching framework that addresses these three gaps through **gene binding modules** that decompose morphological features into functional sub-spaces for gene-selective attention, **regulator tokens** that capture inter-gene coordination via latent bottleneck compression, and **heteroscedastic flow matching** that learns per-gene uncertainty to guide adaptive SDE sampling. Experiments on four ST datasets show consistent improvements over both deterministic and generative baselines, with learned modules recovering biologically coherent gene groupings without supervision.

2 Method

Overview. Given a histology image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$ and spatial coordinates $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N]$, we aim to predict the gene expression matrix $\mathbf{X} \in \mathbb{R}^{N\times G}$, where N is the number of spatial spots and G is the number of target genes.

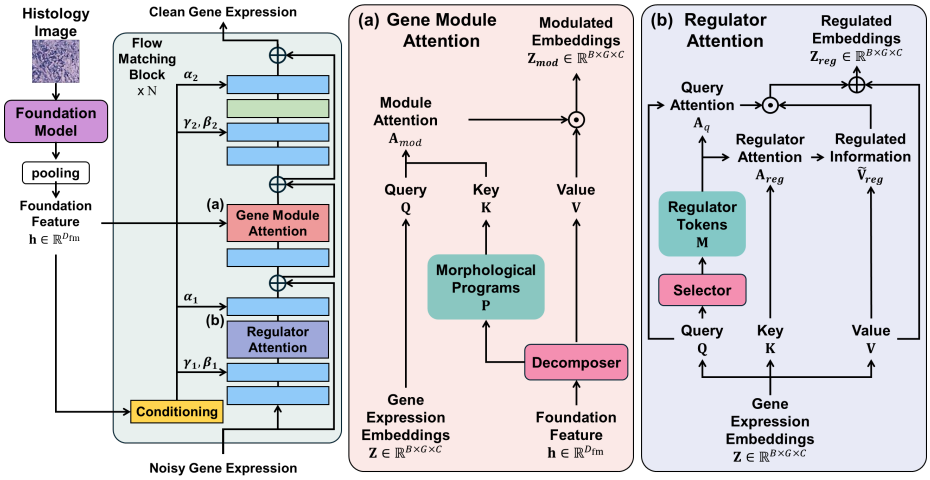

Fig. 2. Overview of the proposed architecture. (Left) Each flow matching block consists of Gene Module Attention, Regulator Attention, and AdaLN layers. (a) **Gene Module Attention** decomposes the foundation feature into K morphological programs and lets each gene selectively attend to relevant programs. (b) **Regulator Attention** compresses gene-gene interactions through m learnable regulator tokens.

Rather than learning a deterministic mapping, we model the conditional distribution $p(\mathbf{X}|\mathbf{I}, \mathbf{S})$ using flow matching. For each spot, we extract morphological features from pre-trained pathology foundation models, producing a feature vector $\mathbf{h} \in \mathbb{R}^{D_{\text{fm}}}$. The gene expression vector $\mathbf{x} \in \mathbb{R}^G$ is embedded into a sequence of G token representations, each corresponding to one gene. A transformer backbone of L stacked blocks processes these gene tokens, where each block contains regulator token attention, gene binding module attention, and a feed-forward network with adaptive layer normalization conditioned on the diffusion timestep. The model jointly predicts velocity and per-gene uncertainty via a heteroscedastic loss, enabling uncertainty-aware SDE sampling at inference.

Regulator Attention. To capture inter-gene regulatory dependencies, we introduce m learnable regulator tokens that encode latent regulatory states. Given gene token representations $\mathbf{Z} \in \mathbb{R}^{B\times G\times C}$, where B is the batch size, G is the number of target genes, and C denotes the embedding dimension, we obtain $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ via linear projections ($d_h = C/n_{\text{heads}}$). We generate m regulator tokens, where m controls the bottleneck capacity through which inter-gene dependencies are compressed and redistributed (see Table 2 for ablation), from queries via a learned projection: $\mathbf{M} = \text{MLP}(\mathbf{Q}^\top)^\top \in \mathbb{R}^{B\times m\times C}$. The mechanism operates in two stages. First, regulator-to-gene attention aggregates global information into each regulator:

$$\widetilde{\mathbf{V}}_{\text{reg}} = \text{softmax}\left(\frac{\mathbf{M}\mathbf{K}^\top}{\sqrt{d_h}}\right)\mathbf{V} \tag{1}$$

Second, gene-to-regulator attention allows each gene to query the compressed representations:

$$\mathbf{Z}_{\text{reg}} = \text{softmax}\left(\frac{\mathbf{Q}\widetilde{\mathbf{V}}_{\text{reg}}^{\top}}{\sqrt{d_h}}\right)\widetilde{\mathbf{V}}_{\text{reg}} \tag{2}$$

This two-stage bottleneck mediates information flow across gene groups.

Gene Binding Module Attention. The regulated output $\mathbf{Z}_{\text{reg}}$ then serves as input to gene binding module attention. To enable gene-selective morphological conditioning, we decompose the pooled foundation feature $\mathbf{h} \in \mathbb{R}^{D_{\text{fm}}}$ into K morphological programs, each representing a distinct biological process. Here K controls the number of functional sub-spaces; we set $K = 6$ based on the ablation in Table 2, and the learned modules recover biologically coherent gene groupings without supervision (Fig. 4): $\mathbf{P} = \text{Reshape}(\text{MLP}(\mathbf{h})) \in \mathbb{R}^{K\times C}$, where $\mathbf{P}$ is normalized via LayerNorm. Each gene binds to these programs via multi-head cross-attention, producing a gene-module association matrix:

$$\mathbf{A}_{\text{mod}} = \text{softmax}\left(\frac{(\mathbf{Z}_{\text{reg}}\mathbf{W}_Q)(\mathbf{P}\mathbf{W}_K)^{\top}}{\sqrt{d_h}}\right) \in \mathbb{R}^{B\times n_{\text{heads}}\times G\times K} \tag{3}$$

The modulated embeddings are computed as $\mathbf{Z}_{\text{mod}} = \mathbf{A}_{\text{mod}}(\mathbf{P}\mathbf{W}_V)$ and projected back to the original dimension. This soft assignment enables each gene to receive a tailored morphological representation rather than relying on a single shared feature, directly addressing the heterogeneous morphology–expression coupling across genes.

Uncertainty-Aware Flow Matching. Flow matching [13] learns a time-dependent velocity field $\mathbf{v}_\theta(\mathbf{x}_t, t, \mathbf{h})$ that transports a Gaussian prior $p_0 = \mathcal{N}(\mathbf{0}, \mathbf{I})$ to the data distribution p_1 along the linear path $\mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_0$ for $t \in [0, 1]$, with target velocity $\mathbf{u} = \mathbf{x}_1 - \mathbf{x}_0$. Rather than predicting velocity alone, our model jointly outputs velocity and per-gene log-variance, which serves as adaptive loss weighting during training and controls per-gene generation stochasticity at inference:

$$[\mathbf{v}_\theta, \log \boldsymbol{\sigma}_\theta^2] = f_\theta(\mathbf{x}_t, t, \mathbf{h}) \tag{4}$$

where $\boldsymbol{\sigma}_\theta \in \mathbb{R}^G$ represents gene-wise uncertainty, clamped to $[-6.9, 0.0]$ in log-space for numerical stability.

Training. We sample $t \sim \mathcal{U}(0, 1)$, construct $\mathbf{x}_t$, and minimize the heteroscedastic Gaussian negative log-likelihood loss of Kendall and Gal [10], applied here to the velocity regression of flow matching:

$$\mathcal{L} = \frac{1}{G}\sum_{g=1}^{G}\left[\frac{(v_g - u_g)^2}{2\sigma_g^2} + \frac{1}{2}\log\sigma_g^2\right] \tag{5}$$

The first term acts as an adaptive weighting mechanism: genes with large prediction error encourage the model to increase σ_g, while the second term regularizes

against trivially large variance. At optimality, σ_g^2 converges toward the actual squared prediction error, providing a theoretical basis for calibrated uncertainty.
Inference. Distinct from the standard use of heteroscedastic NLL purely as a training-time loss [10], our learned $\boldsymbol{\sigma}_\theta$ doubles as the per-gene SDE diffusion coefficient [18] at inference:

$$d\mathbf{x} = \mathbf{v}_\theta(\mathbf{x}_t, t, \mathbf{h})\, dt + \boldsymbol{\sigma}_\theta(\mathbf{x}_t, t, \mathbf{h})\, d\mathbf{w} \tag{6}$$

where $d\mathbf{w}$ is a Wiener process. Each gene g receives a different magnitude of stochastic perturbation governed by $\sigma_{\theta,g}$: genes confidently predicted from morphology yield near-deterministic trajectories, while genes with high biological ambiguity produce diverse plausible profiles. We integrate Eq. (6) with a second-order Heun SDE solver over 50 steps, applying a Tweedie correction at the final step. For each spot, 20 samples are drawn and averaged for the mean prediction, while their dispersion provides empirical uncertainty intervals.

3 Experiments

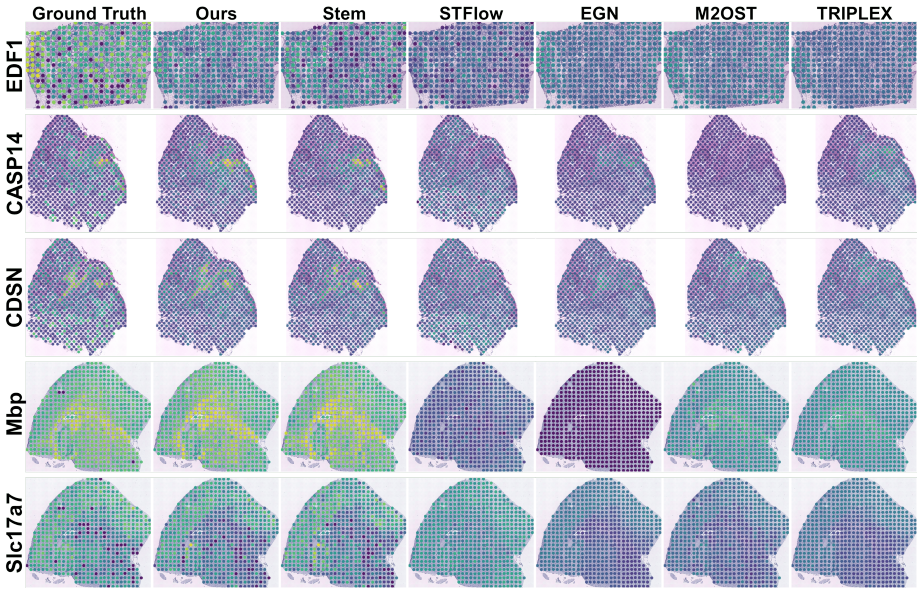

Fig. 3. Qualitative comparison of spatial gene expression prediction. Each row shows a representative gene from different datasets. Our method most recovers localized expression patterns where baselines produce overly smoothed predictions.

We curate four 10x Visium evaluation datasets from HEST-1k [7]: (1) **Breast Cancer I** (BC-I) [5], invasive ductal carcinoma with heterogeneous architectures (invasive carcinoma, DCIS, tumor-infiltrating lymphocytes); (2) **Breast**

Table 1. Quantitative comparison across four diverse tissue types. Best results are shown in **bold** and second-best are underlined. † denotes deterministic flow matching where lower MSE/MAE reflects point estimation rather than distribution modeling.

Method	Breast Cancer I					Squamous Cell Carcinoma				
	$\rho_{10}\uparrow$	$\rho_{50}\uparrow$	$\rho_{200}\uparrow$	MSE↓	MAE↓	$\rho_{10}\uparrow$	$\rho_{50}\uparrow$	$\rho_{200}\uparrow$	MSE↓	MAE↓
ST-Net [5]	.3969	.2673	.1235	.7898	.6647	.2928	.1689	.0464	2.2187	1.1277
EGN [23]	.5494	.4364	.2474	1.2404	.7535	.7403	.6875	.5303	1.9452	1.0075
BLEEP [22]	.3837	.2700	.1379	.8216	.6630	.3100	.2058	.0753	2.5041	1.0576
HisToGene [16]	.4296	.3181	.1567	.8098	**.6615**	.4637	.4049	.2842	2.0999	1.0762
TRIPLEX [4]	.5344	.4074	.2137	.7576	.6478	.5058	.4552	.3273	2.0117	1.0492
Hist2ST [24]	.4401	.3263	.1346	.9022	.7229	.4924	.4263	.2105	2.9221	1.2459
M2OST [20]	.5434	.4191	.2217	.9128	.6818	.3795	.3361	.2497	2.0901	1.0702
THItoGene [9]	.3412	.2506	.1202	1.4361	.9305	.5406	.5025	.3996	2.2358	1.1091
Stem [25]	.4689	.3252	.1689	1.7856	1.0295	.8159	.7689	.6901	1.3121	.8385
STFlow† [6]	.3809	.2490	.1150	**.7309**	.6669	.5924	.4650	.2953	**.6555**	**.6217**
Ours	**.5546**	**.4237**	**.3521**	1.4667	.9325	**.8786**	**.8312**	**.7943**	.9781	.7518

Method	Breast Cancer II					Mouse Brain				
	$\rho_{10}\uparrow$	$\rho_{50}\uparrow$	$\rho_{200}\uparrow$	MSE↓	MAE↓	$\rho_{10}\uparrow$	$\rho_{50}\uparrow$	$\rho_{200}\uparrow$	MSE↓	MAE↓
ST-Net [5]	.5937	.4880	.3204	1.0313	.7677	.2371	.1828	.0911	1.4279	.9519
EGN [23]	.6658	.5448	.3587	1.0768	.7720	.6414	**.5911**	.3943	1.9944	1.0459
BLEEP [22]	.6318	.5291	.3826	1.0158	.7530	.2055	.0974	.0517	1.8150	1.0593
HisToGene [16]	.5113	.4394	.2557	1.5872	.9444	.6167	.5109	.3190	1.4118	.9417
TRIPLEX [4]	.6444	.5680	.4134	.9902	.7430	.6348	.5390	.3713	1.2893	.9074
Hist2ST [24]	.7127	**.6514**	.4963	1.0437	.7975	.5848	.4808	.2577	1.3057	.9098
M2OST [20]	.5829	.5041	.3792	1.0316	.7550	.5415	.4700	.3338	1.4084	.9422
THItoGene [9]	.6811	.5609	.3067	1.9446	1.0824	.6245	.5185	.3496	1.3173	.9175
Stem [25]	.6938	.5799	.4187	1.2769	.8342	.6269	.5354	.3737	.9049	.7208
STFlow† [6]	.3244	.2042	.0948	**.6559**	**.6289**	.6097	.4667	.2892	**.4152**	**.4907**
Ours	**.7142**	.6287	**.5452**	1.0642	.7764	**.6418**	.5544	**.4995**	.7999	.6779

Cancer II (BC-II) [1], an independent HER2-positive cohort offering a homogeneous molecular context; (3) **Skin Cancer** (Squamous Cell Carcinoma) [8], cutaneous samples with keratinization patterns and tumor-stromal-immune organization; and (4) **Mouse Brain**, neural tissue with cortical structure for cross-species validation. For each spot, we extract patch-level features from UNI [3] and CONCH [14] and concatenate them as the morphological input $\mathbf{h} \in \mathbb{R}^{D_{\text{fm}}}$. Gene expressions are normalized via $y = \log(x/\sum x + 1)$ and the top 200 highly variable genes are selected per dataset; all baselines are trained and evaluated on this identical gene panel for fair comparison. Following [4, 25], we use leave-one-out cross-validation across tissue sections.

Baselines. We compare against both deterministic and generative methods: ST-Net [5], EGN [23], BLEEP [22], HisToGene [16], TRIPLEX [4], Hist2ST [24], M2OST [20], THItoGene [9], Stem [25], and STFlow [6]. All baselines are reproduced using their author-provided codebases with their native encoder configurations; notably, Stem [25] uses the same UNI+CONCH combination as our method, isolating the contribution of our architectural design from encoder choice. We report Pearson Correlation Coefficient (PCC) averaged over the top-k highly variable genes, denoted ρ_k for $k \in \{10, 50, 200\}$, along with Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Table 2. Hyperparameter analysis on the number of regulator tokens M and morphological programs K.

		$M=4$	$M=8$	$M=16$	$M=32$
BC-I	$\rho_{50}\uparrow$	.4235	.4065	**.4237**	.4194
	$\rho_{200}\uparrow$	.3519	.3373	**.3521**	.3503
BC-II	$\rho_{50}\uparrow$	.6160	.6074	**.6287**	.5966
	$\rho_{200}\uparrow$	.5610	.5488	**.5776**	.5452
Mouse	$\rho_{50}\uparrow$	.5526	.5476	**.5544**	.5461
	$\rho_{200}\uparrow$	.4973	.4935	**.4995**	.4918
		$K=2$	$K=4$	$K=6$	$K=8$
BC-I	$\rho_{50}\uparrow$	.4219	.4102	**.4237**	.4194
	$\rho_{200}\uparrow$	.3472	.3457	**.3521**	.3503
BC-II	$\rho_{50}\uparrow$	.6195	.6090	**.6287**	.5982
	$\rho_{200}\uparrow$	.5630	.5558	**.5776**	.5435
Mouse	$\rho_{50}\uparrow$	.5505	**.5585**	.5544	.5477
	$\rho_{200}\uparrow$	.4962	**.5024**	.4995	.5022

Table 3. Component ablation study. B: baseline model with MSE loss (corresponds to Stem), M: regulator tokens, P: morphological programs. Ours includes heteroscedastic loss.

	Method	$\rho_{10}\uparrow$	$\rho_{50}\uparrow$	$\rho_{200}\uparrow$	MSE↓	MAE↓
BC-I	B	.4689	.3252	.1689	1.786	1.030
	B+M	.5241	.4008	.3261	**1.459**	**0.925**
	B+P	.5338	.4191	.3497	1.479	0.943
	Ours	**.5646**	**.4237**	**.3521**	1.467	0.933
BC-II	B	.6938	.5799	.4187	1.277	0.834
	B+M	.7012	.6027	.5409	1.098	0.798
	B+P	.7053	.6038	.5419	**1.038**	0.792
	Ours	**.7142**	**.6287**	**.5776**	1.064	**0.776**
Mouse	B	.6269	.5354	.3737	0.905	0.721
	B+M	.6408	**.5571**	.4972	0.812	0.679
	B+P	.6368	.5493	.4961	0.817	0.687
	Ours	**.6418**	.5544	**.4995**	**0.800**	**0.678**

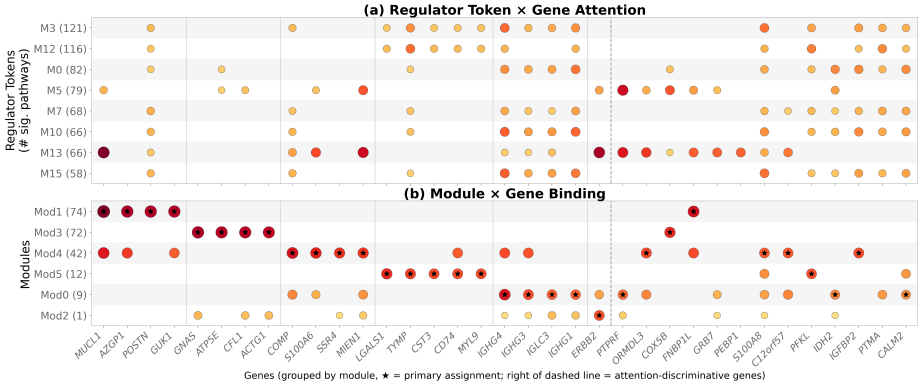

Fig. 4. Interpretability analysis of learned attention patterns on a breast cancer sample. (Top) Regulator token attention weights, ranked by number of significantly enriched pathways. (Bottom) Gene binding module scores (⋆=primary assignment). Genes left of the dashed line are top-scoring per module; those to the right exhibit high attention variance across regulator tokens. Both panels share a gene axis to contrast block-diagonal binding with broadly distributed regulator attention.

Main Results. Table 1 presents quantitative comparisons across four datasets. Our method achieves the best or competitive performance across all correlation metrics, with the most consistent gains on ρ_{200}: improvements of 42%, 10%, 15%, and 27% over the respective best baselines on Breast Cancer I, II, SCC, and Mouse Brain. Since ρ_k averages per-gene PCC over the top-k best-predicted genes, ρ_{10} is dominated by genes whose expression is strongly readable from morphology alone, where all methods perform comparably. As k increases to 200, genes with weaker morphology-expression coupling enter the evaluation,

and the advantage of gene-selective conditioning becomes evident. Deterministic methods are particularly affected, as they rely on a single global representation that cannot adapt to per-gene morphological relevance. Among generative baselines, Stem [25] achieves competitive ρ_{10} but falls behind on ρ_{200}, indicating that uncertainty-aware generation and gene-selective conditioning provide complementary benefits. Our method, along with other generative approaches, reports higher MSE/MAE than some deterministic baselines. This is expected: deterministic regression learns the conditional mean, which minimizes squared error by definition but collapses per-gene expression diversity. STFlow† similarly achieves low MSE/MAE through deterministic flow matching, yet its substantially lower PCC confirms that low reconstruction error does not imply faithful gene-level prediction. Fig. 3 corroborates this, where low-MSE methods produce spatially smoothed predictions that lose localized expression patterns, whereas our method most closely recovers the ground truth spatial structure.

Component Ablation. Table 3 isolates the contribution of each component. Starting from the baseline (B), adding regulator tokens (B+M) yields substantial ρ_{200} gains across all datasets, with the largest improvement on Breast Cancer I. Gene binding modules (B+P) provide comparable PCC improvements, confirming that both inter-gene modeling and gene-selective conditioning independently contribute. Combining all components with heteroscedastic loss (Ours) consistently achieves the best ρ_{200}, indicating that the three components address complementary aspects. Gains are most pronounced on breast cancer datasets where tumor heterogeneity demands gene-selective conditioning, while Mouse Brain shows smaller but consistent improvements due to its anatomically distinct regions providing strong morphological cues from a single global feature.

Hyperparameter Analysis. Table 2 examines the number of regulator tokens M and morphological programs K. Both exhibit a clear optimum ($M = 16$, $K = 6$): smaller values lack representational capacity while larger values degrade performance due to redundancy.

Biological Coherence of Learned Modules. Fig. 4 visualizes both attention mechanisms on a shared gene axis. Gene binding modules (bottom) exhibit block-diagonal structure, grouping biologically related genes without any supervision: secretory/adhesion genes in Mod1, cytoskeletal/metabolic genes in Mod3, extracellular matrix/signaling genes in Mod4, stromal/immune regulators in Mod5, and immunoglobulins in Mod0. Enrichment analysis via Enrichr [11] confirms that these groupings correspond to established pathways including epithelial mesenchymal transition [12], antigen processing, and neutrophil degranulation. In contrast, regulator token attention (top) is broadly distributed, with top-ranked tokens (M3, M12) attending to genes across multiple modules rather than specializing in any one. The attention-discriminative genes (right of the dashed line) illustrate this complementarity most clearly: they receive strong attention from multiple tokens yet lack concentrated binding to any single module.

4 Conclusion

We introduced gene binding modules and regulator tokens that let each gene selectively access morphological features, combined with heteroscedastic flow matching for per-gene adaptive sampling. Beyond consistent quantitative gains, the unsupervised emergence of biologically coherent gene modules suggests that morphological features encode richer gene-level information than previously exploited. This opens the possibility of using learned module structures as a discovery tool for identifying novel gene-morphology associations.

Acknowledgements This work was supported by Smart HealthCare Program funded by the Korean National Police Agency(KNPA) (No. 220222M01), National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2025-00516124), and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2025-25431811).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Andersson, A., Larsson, L., Stenbeck, L., Salmu00e9n, F., Ehinger, A., Wu, S.Z., Al-Eryani, G., Roden, D., Swarbrick, A., u00c5. Borg, Frisu00e9n, J., Engblom, C., Lundeberg, J.: Spatial deconvolution of her2-positive breast tumors reveals novel intercellular relationships. In: bioRxiv (2020). https://doi.org/10.1101/2020.07.14.200600
2. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., Zhuang, X.: Spatially resolved, highly multiplexed RNA profiling in single cells. Science **348**(6233), aaa6090 (2015)
3. Chen, R.J., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine (2024)
4. Chung, Y., Ha, J.H., Im, K.C., Lee, J.S.: Accurate spatial gene expression prediction by integrating multi-resolution features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11591–11600 (2024)
5. He, B., Bergenstråhle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å., Maaskola, J., Lundeberg, J., Zou, J.: Integrating spatial gene expression and breast tumour morphology via deep learning. Nature Biomedical Engineering **4**(8), 827–834 (2020)
6. Huang, T., Liu, T., Babadi, M., Jin, W., Ying, R.: Scalable generation of spatial transcriptomics from histology images via whole-slide flow matching. In: International Conference on Machine Learning (2025)
7. Jaume, G., Doucet, P., Song, A., Lu, M.Y., Almagro Pérez, C., Wagner, S., Vaidya, A., Chen, R., Williamson, D., Kim, A., et al.: Hest-1k: A dataset for spatial transcriptomics and histology image analysis. Advances in Neural Information Processing Systems **37**, 53798–53833 (2024)

8. Ji, A., Rubin, A.J., Thrane, K., Jiang, S., Reynolds, D., Meyers, R., Guo, M., George, B., Mollbrink, A., Bergenstru00e5hle, J., Larsson, L., Bai, Y., Zhu, B., Bhaduri, A., Meyers, J.M., Rovira-Clavu00e9, X., Hollmig, S., Aasi, S., Nolan, G.P., Lundeberg, J., Khavari, P.: Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. In: Cell (2020). https://doi.org/10.1016/j.cell.2020.05.039
9. Jia, Y., Liu, J., Chen, L., Zhao, T., Wang, Y.: Thitogene: a deep learning method for predicting spatial transcriptomics from histological images. Briefings in Bioinformatics **25**(1) (2023)
10. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems. vol. 30 (2017)
11. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al.: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research **44**(W1), W90–W97 (2016)
12. Liberzon, A., et al.: The molecular signatures database hallmark gene set collection. Cell systems **1**(6), 417–425 (2015)
13. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. ICLR (2023)
14. Lu, M.Y., et al.: A visual-language foundation model for computational pathology. Nature Medicine (2024)
15. Marx, V.: Method of the Year: spatially resolved transcriptomics. Nature Methods **18**(1), 9–14 (2021)
16. Pang, M., Su, K., Li, M.: Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. BioRxiv pp. 2021–11 (2021)
17. Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., Macosko, E.Z.: Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science **363**(6434), 1463–1467 (2019)
18. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021)
19. Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P.I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., Frisén, J.: Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science **353**(6294), 78–82 (2016)
20. Wang, H., Du, X., Liu, J., Ouyang, S., Chen, Y.W., Lin, L.: M2ost: Many-to-one regression for predicting spatial transcriptomics from digital pathology images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 7709–7717 (2025)
21. Williams, C.G., Lee, H.J., Asatsuma, T., Vento-Tormo, R., Haque, A.: An introduction to spatial transcriptomics for biomedical research. Genome medicine **14**(1), 68 (2022)
22. Xie, R., Pang, K., Chung, S., Perciani, C., MacParland, S., Wang, B., Bader, G.: Spatially resolved gene expression prediction from histology images via bimodal contrastive learning. Advances in Neural Information Processing Systems **36**, 70626–70637 (2023)

23. Yang, Y., Hossain, M.Z., Stone, E.A., Rahman, S.: Exemplar guided deep neural network for spatial transcriptomics analysis of gene expression prediction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 5039–5048 (2023)
24. Zeng, Y., Wei, Z., Yu, W., Yin, R., Yuan, Y., Li, B., Tang, Z., Lu, Y., Yang, Y.: Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. Briefings in Bioinformatics **23**(5) (2022)
25. Zhu, S., Zhu, Y., Tao, M., Qiu, P.: Diffusion generative modeling for spatially resolved gene expression inference from histology images. arXiv preprint arXiv:2501.15598 (2025)