GeneRAG: A Retrieval-Augmented Framework for Spatially Resolved Gene Expression Prediction

Hyeongsub Kim[1,2], Sihyun Kim[1], Minyoung Cho[1,2], Sanghyun Jo[1,3], Minhyeong Lee[1,4], and Kyungsu Kim[1⋆]

[1]Seoul National University, Korea [2]LG CNS, Korea [3]OGQ, Korea
[4]Asteromorph, Inc., Korea

Abstract. Spatial transcriptomics (ST) is pivotal for deciphering molecular organization, yet cross-modal variability challenges accurate H&E based profiling. Existing models struggle to generalize to unseen genes and lack clinical interpretability. We propose GeneRAG, a model-agnostic Retrieval-Augmented Generation framework with a Dual-Constrained Retrieval module. Unlike conventional black-box networks that rely solely on fixed parameters, GeneRAG explicitly decouples knowledge storage from model training. By optimizing an ElasticNet-based sparse sampling matrix, GeneRAG integrates morphological and biological constraints to fetch relevant samples from a pre-constructed bank. Leveraging conserved gene correlations, this enables accurate reconstruction of comprehensive profiles, including entirely unseen genes. On the HEST-1k dataset, GeneRAG seamlessly enhances state-of-the-art models in a plug-and-play manner, improving Stem's PCC-10 from 0.8322 to 0.8711(Breast dataset). For zero-shot generalization (5,000 genes), Stem+GeneRAG achieves a PCC-5000 of 0.5188, vastly outperforming DeepSpot (0.0748). GeneRAG provides robust, transparent predictions, highlighting its potential for clinical deployment.

Keywords: Spatial transcriptomics · Retrieval-augmented generation · Whole-transcriptome · Knowledge augmentation · Gene expression prediction

1 Introduction

Recent advancements in Spatial Transcriptomics (ST) have popularized virtual sequencing directly from H&E-stained images [4,6,7,13,15,18,21]. While probabilistic models like Stem [21] demonstrate remarkable performance, comprehensive profiling of large-scale multiplexed genes remains essential for understanding the tumor microenvironment [11,14]. However, most models are black-box networks failing to provide clinical rationales [1]. Methods like DeepSpot [15] attempt large-scale prediction but rely solely on internal capacity, suffering from

⋆ Corresponding authors: kyskim@snu.ac.kr

closed-set limitations that preclude predicting unseen genes. Other approaches using LLMs [20] lack morphological grounding, and retrieval-based methods like BLEEP [18] rely on rigid contrastive spaces, failing to fundamentally resolve the unseen gene problem. Crucially, these standalone models are tightly coupled to specific architectures.

To overcome these limitations, we propose GeneRAG, a universal framework for ST inspired by the Retrieval-Augmented Generation (RAG) paradigm [10]. Unlike models relying on fixed parameters, GeneRAG actively retrieves optimal samples from a pre-constructed Reference Bank by balancing morphological and gene expression constraints. By predicting final expressions via a weighted ensemble of retrieved reference samples, our framework effectively mitigates the inherent sparsity and noise of ST datasets. This flexible architecture seamlessly decouples knowledge storage from model training, fundamentally enhancing both the robustness and interpretability of existing Image-to-ST pipelines.

Our main contributions are three-fold: (1) We introduce GeneRAG, model-agnostic retrieval-augmented framework tailored for the spatial transcriptomics domain, effectively mitigating cross-modal variability. (2) We explicitly address the black-box nature of deep learning models by providing clear visual and transcriptomic rationales from retrieved reference data. (3) We demonstrate the framework's superior zero-shot generalization for unseen genes and validate its model-agnostic nature by seamlessly integrating it with various state-of-the-art architectures in a plug-and-play manner.

2 Method

2.1 Overall Architecture

Fig. 1 illustrates the overall inference pipeline of the proposed GeneRAG framework. Unlike conventional standalone approaches, GeneRAG is designed as a plug-and-play module that can be seamlessly integrated with any existing Image-to-ST prediction models without requiring structural modifications.

Given an input H&E image during the test phase, we first utilize the encoder of a pre-trained backbone model (e.g., a pathological foundation model), denoted as $\mathcal{E}$, to extract the morphological feature embedding f_{img}. Subsequently, a pre-trained gene expression decoder $\mathcal{D}$ processes f_{img} to generate an initial gene expression prediction $\hat{y}_{init}$. In a standard methodology, this $\hat{y}_{init}$ would serve as the final output, inheriting the closed-set limitations of the backbone.

In our framework, however, f_{img} and $\hat{y}_{init}$ are strictly utilized as anchor points for robust knowledge retrieval. Specifically, GeneRAG constructs a *Hybrid Query* to query the pre-constructed Reference Bank, which stores highly reliable multimodal information derived from the training set. This Hybrid Query operationalizes a dual-constraint search mechanism:

1. **Morphological Retrieval Query**: Utilizing f_{img} to find tissue patches with similar visual and cellular architectures.

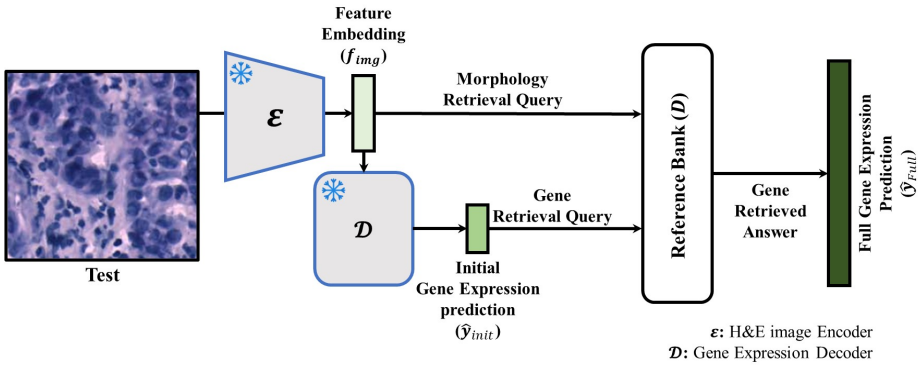

Fig. 1. Overall the GeneRAG framework: During test time, the model extracts a morphological feature embedding (f_{img}) and an initial gene expression prediction ($\hat{y}_{init}$) using a frozen encoder ($\mathcal{E}$) and decoder ($\mathcal{D}$). These are utilized as a hybrid query (morphological and gene retrieval queries) to fetch the most relevant answers from a pre-constructed Trainset Reference Bank, bypassing the closed-set limitation and yielding the full gene expression prediction ($\hat{y}_{full}$).

2. **Gene Retrieval Query**: Utilizing $\hat{y}_{init}$ as a functional constraint to ensure the retrieved patches share similar initial gene expression profiles.

By combining these two distinct semantic spaces, GeneRAG effectively queries the Reference Bank to fetch the most relevant Gene Retrieved Answers. Finally, these retrieved answers are aggregated to yield the full gene expression prediction $\hat{y}_{full}$. Through this retrieval-augmented process, GeneRAG bypasses the black-box prediction mechanism and successfully infers the expressions of a comprehensive panel of genes, including those completely unseen during the backbone's training phase.

2.2 Reference Bank Construction

To effectively retrieve multimodal knowledge, we pre-construct a Reference Bank $\mathcal{D}$ from the training dataset prior to inference (Fig. S1). $\mathcal{D}$ consists of two paired sub-banks: First, the **Morphology Bank (D_{img})** stores visual embeddings $f_{img}^{(i)}$ for N training H&E patches $x_{train}^{(i)}$. These are extracted using the frozen pre-trained encoder $\mathcal{E}$ to ensure perfect feature space alignment with test queries. Second, the **Full Gene Bank (D_{full})** stores the corresponding 1:1 ground-truth spatial transcriptomics profiles $y_{full}^{(i)}$. Crucially, D_{full} retains the complete gene expression panel, including completely unseen genes far beyond the backbone's initial training subset (y_{init}).

Consequently, the constructed Reference Bank $\mathcal{D} = \{(f_{img}^{(i)}, y_{full}^{(i)})\}_{i=1}^{N}$ functions as the essential knowledge base. During the test phase, it allows the Hybrid

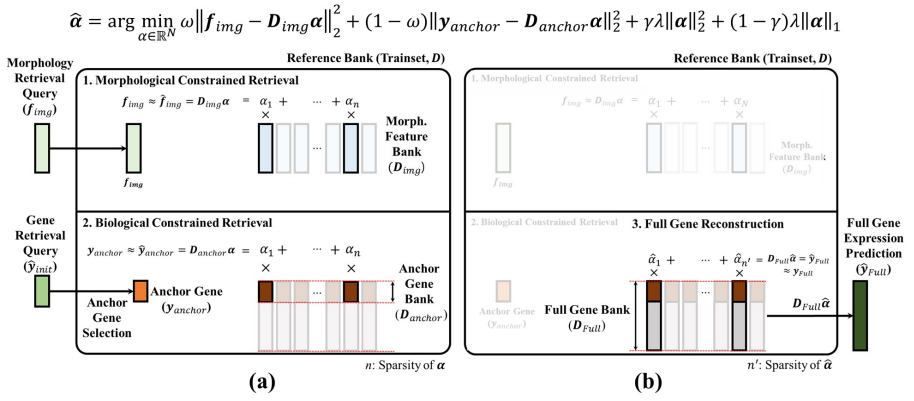

Fig. 2. GeneRAG Retrieval and Reconstruction Pipeline. (a) Dual-Constrained Retrieval: Hybrid query optimization is guided by morphological (f_{img}) and biological (y_{anchor}) constraints. An ElasticNet regression outputs a sparse sampling matrix $\hat{\alpha}$, selectively fetching optimal neighboring patches from the Reference Bank. **(b) Full Gene Expression Profiles Reconstruction:** The final prediction $\hat{y}_{full}$ is obtained by applying $\hat{\alpha}$ to the Full Gene Bank (D_{full}). Relying on robust gene-to-gene correlations, this linear combination accurately extrapolates the entire gene panel, including unseen genes.

Query to actively retrieve the optimal neighboring patches that satisfy both morphological and functional (gene expression) similarities.

2.3 Dual-Constrained Retrieval Module

Given the hybrid query $(f_{img}, \hat{y}_{init})$ and the Reference Bank $\mathcal{D}$, the objective of this module is to determine the optimal sampling weights $\alpha \in \mathbb{R}^N$ to retrieve the most relevant patches from the bank. Rather than relying on simple distance-based heuristics, we formulate the retrieval as an optimization problem that strictly enforces both morphological and biological constraints simultaneously (Fig. 2 (a)).

Prior to the optimization, we apply an anchor gene selection process to the gene retrieval query $\hat{y}_{init}$ to mitigate noise and improve retrieval efficiency. Instead of utilizing the entire gene panel, we extract an anchor gene vector y_{anchor}, representing the top-k Highly Variable Genes (HVG) or a highly co-expressed gene set. Correspondingly, we construct an Anchor Gene Bank, formally denoted as a matrix $D_{anchor} \in \mathbb{R}^{k \times N}$, as a subset of the Full Gene Bank (D_{full}) by selecting the identical anchor gene indices.

With the morphological query f_{img} and the refined biological query y_{anchor} prepared, GeneRAG solves the following ElasticNet-based dual-constrained optimization problem to obtain the optimal sampling matrix $\hat{\alpha}$:

$$\hat{\alpha} = \arg\min_{\alpha \in \mathbb{R}^N} \omega \| f_{img} - D_{img}\alpha \|_2^2 + (1-\omega) \| y_{anchor} - D_{anchor}\alpha \|_2^2 + \gamma\lambda \|\alpha\|_2^2 + (1-\gamma)\lambda \|\alpha\|_1. \tag{1}$$

Equation 1 is structurally composed of three main components. The first term minimizes the **morphological reconstruction error** between the Morphology Bank (formally defined as a matrix $D_{\text{img}} \in \mathbb{R}^{d \times N}$) and the input feature query. The second term minimizes the **biological reconstruction error** between the Anchor Gene Bank (defined as a matrix $D_{\text{anchor}} \in \mathbb{R}^{k \times N}$) and the initial gene prediction. The hyperparameter $\omega \in [0, 1]$ controls the balance between these two constraints.

The third and fourth terms represent the ElasticNet regularization. The ℓ_1 penalty enforces sparsity on α, ensuring that only a small, meaningful subset of patches from the vast Reference Bank is selected. Concurrently, the ℓ_2 penalty stabilizes the selection by grouping highly correlated patches. Consequently, the derived $\hat{\alpha}$ acts as the optimal sparse sampling matrix, indicating which reference patches best explain the query, and serves as the core component for the subsequent Full Gene Reconstruction. To ensure efficiency, Eq. (1) is solved as a multi-output ElasticNet problem using FISTA with batched matrix operations, completing the slide-level optimization within minutes.

2.4 Full Gene Expression Profiles Reconstruction

The optimal sparse sampling matrix $\hat{\alpha}$, derived from the dual-constrained retrieval module, identifies the most relevant reference patches and their corresponding weights (Fig. 2 (b)). In practice, $\hat{\alpha}$ activates approximately 50 reference samples on average per query. Consequently, the final prediction is computed as a weighted ensemble of these retrieved profiles, an approach optimally suited to mitigate the inherent sparsity and noise of spatial transcriptomics data. Mathematically, this reconstruction is formulated as a simple linear combination with the Full Gene Bank (D_{full}):

$$\hat{y}_{full} = D_{full} \cdot \hat{\alpha}. \tag{2}$$

Crucially, D_{full} encompasses comprehensive transcriptomic data extending far beyond the predictive scope of the backbone model. A natural question arises: how can $\hat{\alpha}$, optimized solely on a limited anchor gene set (y_{anchor}) and morphological features (f_{img}), accurately predict unseen genes? The fundamental rationale lies in biologically conserved gene-to-gene correlations [5,9]. Tissue patches exhibiting highly similar morphology and anchor gene environments inherently share broader expression profiles driven by intrinsic co-expression networks within individual spots. Thus, without requiring any structural modifications, GeneRAG achieves robust, near zero-shot predictions for unseen genes through this intuitive retrieval-augmented reconstruction.

3 Experiments

3.1 Experimental Setup

Datasets. To evaluate our framework, we utilized HEST-1k [8], a comprehensive spatial transcriptomics benchmark dataset, focusing on three major organs with high clinical significance: HER2ST (breast cancer), Prostate, and Kidney. To ensure rigorous evaluation and avoid arbitrary splits, we strictly adhered to the Leave-One-Slide-Out (LOSO) protocol of Stem [21]. This inherently evaluates two critical scenarios: intra-patient generalization (Prostate and HER2ST datasets) and strict inter-patient generalization (Kidney dataset; 23 slides from 22 patients). Testing on completely unseen patients robustly validates our framework against batch effects.

Evaluation Metrics. We employed PCC-k (Pearson Correlation Coefficient of top-k) to assess the linear correlation between predicted and ground-truth expression levels for the top-k highly correlated genes.

Baselines and Backbone Models. To rigorously validate the model-agnostic nature of GeneRAG, we selected representative models as baselines and backbones. We utilized Stem [21], a state-of-the-art diffusion generative model for spatial transcriptomics, as the representative conventional baseline. For the pathological foundation models, we selected UNI [2] to represent the class of models trained exclusively on H&E images (e.g., CONCH [12], GigaPath [19]). Additionally, we adopted EXAONE Path 2.5 [16] as the representative for the emerging class of multi-modal foundation models that incorporate genomic data alongside histology (e.g., Thread [17], OmiCLIP [3]). GeneRAG was flexibly integrated into these diverse models.

Evaluation Scenarios. To ensure fair comparisons by standardizing the population of evaluated genes, we divided the quantitative experiments into two distinct setups:

- **Core HVG Setup:** Evaluates the prediction performance within the targeted Highly Variable Genes (HVGs) included during the training phase (e.g., top 300 HVGs for Breast, and top 200 HVGs for Kidney and Prostate, strictly following the experimental design [21]). This scenario compares GeneRAG against conventional methods to verify its ability to calibrate and improve baseline accuracy.
- **Global HVG Setup:** Assesses the predictive capability when expanding to a large-scale panel of unseen genes (e.g., top 5,000 HVGs). In this setup, we introduced DeepSpot [15], a model designed for direct large-scale gene prediction via spatial context, as the primary baseline to benchmark against Stem, UNI, and EXAONE Path 2.5 augmented with GeneRAG.

Furthermore, for qualitative analysis, we designed visualizations for unseen gene predictions across varying hybrid retrieval weight ratios (ω). To validate interpretability, we also visualize the top-5 referenced patches from the Reference Bank alongside their top-5 gene expressions, retrieved based on the optimized sparse sampling matrix α.

3.2 Validation of Predictive Performance and Reference-based Interpretability

In-domain Calibration on Core HVG Table 1 presents a performance comparison (PCC-k) among conventional methods, foundation models, and their GeneRAG-augmented counterparts on the Core HVG, which was included during the training phase.

Table 1. Performance comparison on the Core HVG. The evaluation metric is PCC-k.

Model	GeneRAG	Breast			Kidney			Prostate		
		PCC-10↑	PCC-50↑	PCC-300↑	PCC-10↑	PCC-50↑	PCC-200↑	PCC-10↑	PCC-50↑	PCC-200↑
HisToGene [13]	✗	0.6812	0.6345	0.5250	0.4294	0.3503	0.0905	0.4035	0.3554	0.2235
BLEEP [18]	✗	0.7727	0.7141	0.5652	0.4998	0.4221	0.3143	0.5798	0.5102	0.3158
TRIPLEX [4]	✗	0.7907	0.7394	0.5766	0.4654	0.4105	0.3165	0.6173	0.4953	0.3601
Stem [21]	✗	0.8322	0.7696	0.6228	0.5768	0.4989	0.3354	0.6310	0.5546	0.3937
UNI [2]	✗	0.8301	0.7909	0.6024	0.4828	0.3888	0.2715	0.5548	0.4761	0.3076
CONCH [12]	✗	0.7799	0.7467	0.6043	0.3583	0.3109	0.2243	0.5660	0.4715	0.3171
EXAONE-Path 2.5 [16]	✗	0.8217	0.7850	0.6251	0.4584	0.4023	0.3009	0.6204	0.5313	0.3536
Stem [21]	✓	**0.8711**	**0.8275**	**0.7029**	**0.6068**	**0.5419**	**0.4055**	**0.7001**	**0.6693**	**0.5415**
UNI [2]	✓	0.8670	0.8257	0.7017	0.5529	0.4987	0.3525	0.6801	0.6322	0.5046
EXAONE-Path 2.5 [16]	✓	0.8589	0.8175	0.7002	0.5479	0.4886	0.3347	0.6911	0.6513	0.5371

The experimental results demonstrate that the proposed GeneRAG framework does not merely expand the predictive scope but significantly calibrates and enhances the intrinsic prediction performance of the backbone models. On the Breast dataset, the application of GeneRAG to Stem resulted in a substantial increase in PCC-10 from 0.8322 to 0.8711, with similar improvements observed when integrated with UNI. Notably, while performance improved across all organs, the most significant gains were observed in the Prostate dataset (e.g., UNI PCC-200 surged from 0.3076 to 0.5046). This is directly attributed to the Prostate tissue exhibiting the strongest inherent gene-to-gene correlation (mean $r = 0.5530$), providing the optimal biological context for GeneRAG to reconstruct unmeasured transcriptomic profiles. This indicates that GeneRAG's Dual-Constrained Retrieval Module effectively corrects the uncertainties inherent in the backbone's initial predictions ($\hat{y}_{init}$) by leveraging ground-truth data from the Reference Bank. Furthermore, GeneRAG consistently drives performance improvements regardless of the backbone's architectural characteristics—whether a pure vision model (UNI), a multimodal model (EXAONE-Path 2.5), or a generative model (Stem)—thereby proving the framework's robust model-agnosticism.

Zero-shot Extrapolation on Global HVG The true strength of GeneRAG becomes evident in the scenario of predicting Global HVG (HVG 5000) that the backbone models have never seen during training. Table 2 details the prediction results under the Global HVG setup.

When compared to DeepSpot, a baseline model explicitly trained to predict 5,000 genes directly utilizing spatial context, the GeneRAG-integrated frameworks recorded overwhelming performance margins. On the Breast dataset, DeepSpot

Table 2. Performance comparison on the Global HVG for Zero-shot Extrapolation.

Dataset	Model	GeneRAG	PCC-k ↑						
			10	50	300	1000	2000	3000	5000
Breast	DeepSpot [15]	✗	0.3932	0.3287	0.2469	0.1823	0.1413	0.1142	0.0748
	UNI [2]	✓	0.8716	**0.8393**	**0.7894**	**0.7297**	0.6721	0.6248	0.5465
	EXAONE-Path 2.5 [16]	✓	0.8619	0.8316	0.7825	0.7254	**0.6728**	**0.6301**	**0.5591**
	Stem [21]	✓	**0.8720**	0.8347	0.7801	0.7142	0.6505	0.6002	0.5188
Kidney	DeepSpot [15]	✗	0.2102	0.1685	0.1334	0.0889	0.0660	0.0510	0.0273
	UNI [2]	✓	0.5666	0.5182	0.4189	0.3148	0.2558	0.2236	0.1828
	EXAONE-Path 2.5 [16]	✓	0.5601	0.5069	0.4084	0.3107	0.2495	0.2149	0.1715
	Stem [21]	✓	**0.6158**	**0.5656**	**0.4778**	**0.3696**	**0.3060**	**0.2691**	**0.2221**
Prostate	DeepSpot [15]	✗	0.1362	0.1176	0.0868	0.0650	0.0503	0.0406	0.0265
	UNI [2]	✓	0.6840	0.6588	0.5875	0.4902	0.4157	0.3663	0.1980
	EXAONE-Path 2.5 [16]	✓	0.7051	0.6800	**0.6042**	**0.4977**	**0.4187**	**0.3669**	**0.2981**
	Stem [21]	✓	**0.7128**	**0.6846**	0.6000	0.4851	0.4015	0.3477	0.2781

achieved a PCC-10 of only 0.3932, whereas Stem + GeneRAG and UNI + GeneRAG achieved remarkable prediction accuracies of 0.8720 and 0.8716, respectively. This performance gap was maintained even under the extreme condition of evaluating all 5,000 genes (PCC-5000), where Stem + GeneRAG (0.5188) outperformed DeepSpot (0.0748) by approximately seven times. Furthermore, the Stem + GeneRAG combination demonstrated robust performance surpassing DeepSpot on the Kidney and Prostate datasets as well.

Furthermore, an interesting observation can be made between the foundation models on the Breast dataset. While the pure vision foundation model, UNI, exhibits slightly higher performance in predicting the top-tier genes (PCC-10 to PCC-1000), EXAONE-Path 2.5 surpasses it when evaluating a broader and more challenging range of genes (PCC-2000 and beyond). For instance, at PCC-5000, EXAONE-Path 2.5 achieved 0.5591 compared to UNI's 0.5465. This crossover highlights that EXAONE-Path 2.5's multimodal pre-training, which inherently incorporates both H&E images and genetic data, provides a more robust and comprehensive understanding when extrapolating to a massive scale of unseen genes.

Interpretability via Reference Bank Retrieval A significant advantage of GeneRAG is its inherent transparency of reference-based models. When predicting spatial gene expression, the model explicitly reveals the utilized reference samples and their weights. As visualization, examining the elements (α_i) of the optimized sampling matrix $\hat{\alpha}$ reveals the top retrieved patches. These patches not only share profound morphological resemblance with the target spot but also exhibit highly consistent transcriptomic profiles (e.g., prominent expressions of TMSB10, ERBB2). This validates that our dual-constrained mechanism fetches genuinely analogous samples, providing pathologists with transparent, biologically plausible rationales and significantly enhancing clinical reliability.

4 Conclusion

In this paper, we introduced GeneRAG, a novel, model-agnostic Retrieval-Augmented Generation framework for spatial transcriptomics. By synergistically integrating morphological features and biological anchor genes, GeneRAG actively retrieves contextually relevant knowledge from a pre-constructed bank. Extensive evaluations demonstrate that our framework consistently enhances the performance of state-of-the-art models and achieves remarkable zero-shot extrapolation for up to 5,000 unseen genes. Furthermore, GeneRAG addresses the black-box nature of deep learning by providing transparent, biologically plausible rationales, paving the way for reliable clinical decision-making.

Acknowledgements. This work was partly supported by the KHIDI grant funded by the Korean government (MOHW) [No.RS-2025-02307233], the NRF or IITP grants funded by the Korean government (MSIT) [No.RS-2026-25472075, No.RS-2026-25483206, No.RS-2025-02305581, No.RS-2025-25442338 (AI Star Fellowship-SNU), and No.RS-2021II211343 (SNU AI)], the ITIP grant funded by the Korean government (MOTIR) [No.RS-2026-25549946], the Advanced GPU Utilization and AI Computing Infrastructure Enhancement User Support Programs funded by the Korean government (MSIT) [No.05-26-04-0094], the Research grant from SNU, and the Strategic Hub grant for International Research Collaboration of SNU.

Kyungsu Kim is affiliated with the School of Transdisciplinary Innovations, Department of Biomedical Science, Interdisciplinary Program in Artificial Intelligence (IPAI), Medical Research Center, and AI Institute at SNU.

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

References

1. Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V., Madabhushi, A.: Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nature Reviews Clinical Oncology **16**(11), 703–715 (2019)
2. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**, 850–862 (2024)
3. Chen, W., Zhang, P., Tran, T.N., Xiao, Y., Li, S., Shah, V.V., Cheng, H., Brannan, K.W., Youker, K., Lai, L., et al.: A visual–omics foundation model to bridge histopathology with spatial transcriptomics. Nature Methods **22**(7), 1568–1582 (2025)
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. Elyanow, R., Dumitrascu, B., Engelhardt, B.E., Raphael, B.J.: netnmf-sc: leveraging gene–gene interactions for imputation and dimensionality reduction in single-cell expression analysis. Genome research **30**(2), 195–204 (2020)

6. 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)
7. 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)
8. 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)
9. Langfelder, P., Horvath, S.: Wgcna: an r package for weighted correlation network analysis. BMC bioinformatics **9**(1), 559 (2008)
10. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., et al.: Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems **33**, 9459–9474 (2020)
11. Lewis, S.M., Asselin-Labat, M.L., Nguyen, Q., Berthelet, J., Tan, X., Wimmer, V.C., Merino, D., Rogers, K.L., Naik, S.H.: Spatial omics and multiplexed imaging to explore cancer biology. Nature methods **18**(9), 997–1012 (2021)
12. Lu, M.Y., Chen, R.J., et al.: A visual-language foundation model for computational pathology. Nature Biotechnology (2024)
13. 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)
14. Rao, A., Barkley, D., França, G.S., Safarik, R.L., Thiagarajan, P.S., Salick, M.R., et al.: Exploring tissue architecture using spatial transcriptomics. Nature **596**(7871), 211–220 (2021)
15. Ratschlab: Deepspot: Multi-level spatial context for gene expression prediction. bioRxiv (2024)
16. Research, L.A.: Exaone path 2.5: Pathology foundation model with multi-omics alignment. arXiv preprint (2024)
17. Vaidya, A., Zhang, A., Jaume, G., Song, A.H., Ding, T., Wagner, S.J., Lu, M.Y., Doucet, P., Robertson, H., Almagro-Perez, C., et al.: Molecular-driven foundation model for oncologic pathology. arXiv preprint arXiv:2501.16652 (2025)
18. 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)
19. Xu, H., Usuyama, N., Bagga, J., et al.: A whole-slide foundation model for digital pathology from real-world data. Nature **630**, 181–188 (2024)
20. Zhou, Y., Lu, Y., Li, Q., Li, X., Wang, Y.: Deep association multimodal learning for zero-shot spatial transcriptomics prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 131–140. Springer (2025)
21. 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)