

DUET: Dual-Paradigm Adaptive Expert Triage with Single-cell Inductive Prior for Spatial Transcriptomics Prediction

Junchao Zhu¹, Ruining Deng², Junlin Guo¹, Tianyuan Yao¹, Chongyu Qu¹,
Juming Xiong¹, Zhengyi Lu¹, Yanfan Zhu¹, Marilyn Lions¹, Yuechen Yang¹,
Yu Wang³, Shilin Zhao³, Haichun Yang³, Yuankai Huo^{1*}

¹ Vanderbilt University, Tennessee, USA

² Weill Cornell Medicine, New York, USA

³ Vanderbilt University Medical Center, Tennessee, USA

Abstract. Inferring spatially resolved gene expression from histology images offers a cost-effective complement to spatial transcriptomics (ST). However, existing methods reduce this task to a simple morphology-to-expression mapping, where visual similarity does not guarantee molecular consistency. Meanwhile, single-cell data has amassed rich resources far surpassing the scale of ST data, yet it remains underexplored in vision-omics modeling. Furthermore, current approaches commit to a monolithic paradigm with bottlenecks, unable to balance expressive flexibility with biological fidelity. To bridge these gaps, we propose DUET, a novel dual-paradigm framework that synergizes parametric prediction and memory-based retrieval under cellular inductive priors. DUET implements a parallel regression-retrieval paradigm, adaptively reconciling the outputs of its complementary pathways. To mitigate aleatoric vision ambiguity, we incorporate large-scale single-cell references to impose molecular states as biological constraints for faithful learning. Building upon structural refinement, we further design a lightweight adapter to dynamically assign branch preference across spatial contexts to achieve optimal performance. Extensive experiments on three public datasets across varied gene scales demonstrate that DUET achieves SOTA performance, with consistent gains contributed by each proposed component. Code is available at <https://github.com/hrlblab/DUET>.

Keywords: Computational Pathology · Spatial Transcriptomics · scRNA-sequencing Genomics.

1 Introduction

Spatial transcriptomics (ST) links tissue morphology with spatial molecular programs [1–3]. However, its high cost constrains data scale, while sequencing noise and insufficient read depth further weaken the available supervisory signal [4].

* Email: yuankai.huo@vanderbilt.edu

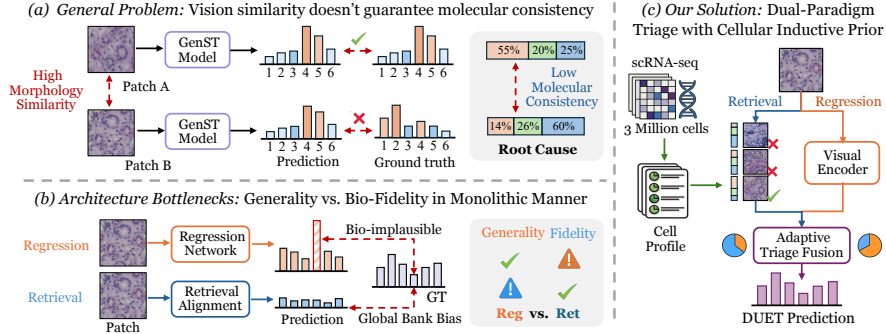


Fig. 1. Paradigm limitations in vision-omics modeling. (a) Morphological visual similarity does not guarantee molecular consistency. (b) Monolithic learning manner fails to reconcile flexibility with biological reliability. (c) Proposed DUET introduces a dual-paradigm framework with cellular priors and adaptive fusion, jointly modeling parametric prediction and memory-based inference for biologically reliable estimation.

Against this backdrop, inferring spatial expression profiles from histology images has offered a cost-effective strategy [5, 6]. Yet, existing methods simplify the task into a vision-to-expression mapping [7, 8]; however, high visual similarity does not guarantee molecular consistency, leading to aleatoric morphology ambiguity in molecular modeling, as presented in Fig. 1a. Meanwhile, single-cell RNA sequencing (scRNA-seq) has amassed cell-type and state references far beyond the scale of ST [9, 10], but remains underutilized in vision-omics modeling.

Moreover, current vision-driven methods commit to a monolithic paradigm [11–13]. As shown in Fig. 1b, regression-based models [5, 8, 14] provide expressive flexibility through parametric generalization, yet may yield biologically implausible outputs. Retrieval-based methods [7, 15], by anchoring predictions to reference expression distributions, ensure biological reliability but limit extrapolation beyond the empirical support. Monolithic learning manner has each bottleneck, while an adaptive fusion of both complementary paradigms could trade off expressive freedom and biological fidelity across spatial contexts.

To achieve this, we introduce DUET, a novel dual-paradigm framework that integrates cellular inductive priors and adaptive expert triage to synergize flexibility and biological reliability. DUET establishes a coordinated regression-retrieval architecture, where parametric prediction and memory-based inference operate in parallel and are adaptively fused, as illustrated in Fig. 2. Moreover, we leverage large-scale scRNA-seq data to infer spot cell-type composition, incorporating cellular inductive signals to filter out biologically incompatible neighbors and reinforce coherence. Building upon structure-aware refinement, we further introduce a lightweight adapter trained on a small held-out set to perform expert triage, dynamically learning fusion weights that assign branch preference for each spatial location. Our contributions can be summarized as threefold:

- We propose a pioneering dual-paradigm framework DUET that transcends monolithic modeling by unifying parametric prediction and memory-based inference to reconcile flexibility and biological fidelity.
- We curate large-scale scRNA-seq references to impose cellular inductive priors as biological constraints, alleviating morphological ambiguity with plausible molecular states for faithful vision-omics modeling.

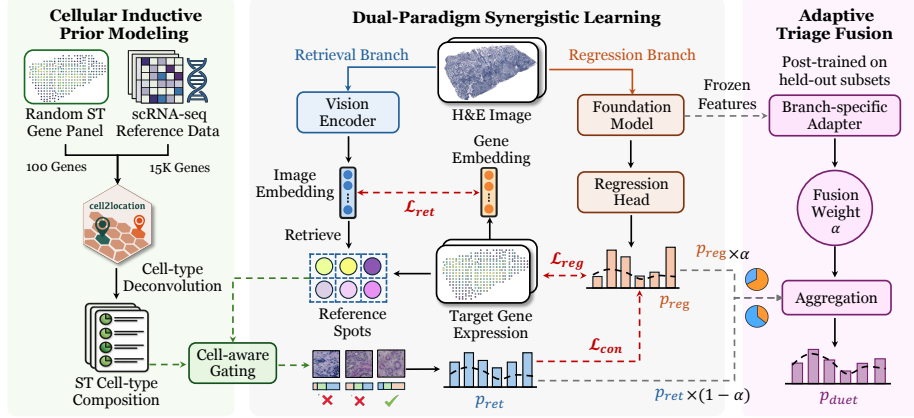


Fig. 2. Overview of proposed DUET framework. DUET comprises three key components. We first leverage large-scale scRNA-seq references to perform cell-type deconvolution and derive cellular inductive priors, enabling a gating mechanism to filter biologically incompatible candidates. We then conduct unified dual-paradigm joint training and facilitate cross-branch knowledge transfer through a retrieval-guided consistency loss. Finally, a lightweight adapter is introduced for adaptive triage fusion, dynamically routing each spot to its more reliable paradigm for the final output.

- We design an adaptive expert triage module equipped with a lightweight adapter to dynamically modulate the contribution of regression and retrieval pathways for optimal expression prediction.

2 Method

2.1 Cellular Inductive Prior Modeling

Expression profile of ST spot consists of mixed signals from multiple cell types [16]. Accurate estimation of cell-type composition is therefore essential for reliable modeling. To address this, we construct cellular inductive priors by integrating large-scale single-cell references. Given multi-batch scRNA-seq datasets $\mathcal{D}_{sc} = \{(\mathbf{x}_c, t_c, b_c)\}_{c=1}^C$, where $\mathbf{x}_c \in \mathbb{Z}_{\geq 0}^G$ denotes the raw UMI counts of cell c , $t_c \in \{1, \dots, T\}$ represents cell type, and b_c indicates batch label. We adopt a nega-

tive binomial regression model [17] to learn batch-robust cell-type signatures:

$$x_{cg} \sim \text{NB}(\mu_{cg}, \theta_g), \quad \mu_{cg} = l_c \cdot \exp\left(m_g^{(b_c)}\right) \cdot \mu_{t_c, g} \quad (1)$$

where l_c denotes scaling factor, $m_g^{(b_c)}$ represents batch effect, and $\mu_{t, g}$ is the average expression of gene g in cell type t . After optimization via variational inference, we obtain the signature matrix $\mathbf{M} \in \mathcal{R}_{>0}^{G \times T}$, where $M_{gt} = \hat{\mu}_{t, g}$ encodes the characteristic pattern of cell type. Based on the signature matrix $\mathbf{M}$, we then perform cell-type deconvolution on ST data. Specifically, we randomly sample 100 genes non-overlapped with target genes to construct a deconvolution gene panel G_{de} . We then conduct Bayesian deconvolution [17] for each ST slide:

$$y_{sg} \sim \text{NB}(\mu_{sg}, \alpha_g), \quad \mu_{sg} = d_s \cdot \sum_{t=1}^T w_{st} \cdot M_{gt} \quad (2)$$

where $w_{st} \geq 0$ denotes the abundance of cell type t in spot s , and d_s represents detection efficiency. The 5% quantile $\hat{w}_{st}^{q05}$ is then extracted and normalized. We then apply CellViT [18] to perform cell count n_s for each spot to construct a cellular gating signal as $g_{st} = n_s \cdot \hat{w}_{st}^{q05}$ for the following learning.

2.2 Dual-Paradigm Synergistic Learning

Cross-Modality Alignment. We adopt DenseNet-121 [19] as visual encoder, followed by a projection head that maps and ℓ_2 -normalizes image features into embeddings $v_i \in \mathcal{R}^d$. A gene encoder with similar projects expression vector y_i into $h_i \in \mathcal{R}^d$. Following [7], we align these two modalities via a symmetric InfoNCE loss $\mathcal{L}_{\text{ret}}$ with temperature $\tau = 0.07$, defined as:

$$\mathcal{L}_{\text{ret}} = -\frac{1}{2N} \sum_{i=1}^N \left[\log \frac{\exp(v_i^\top h_i / \tau)}{\sum_{j=1}^N \exp(v_i^\top h_j / \tau)} + \log \frac{\exp(h_i^\top v_i / \tau)}{\sum_{j=1}^N \exp(h_i^\top v_j / \tau)} \right] \quad (3)$$

Cell-Aware Gating Retrieval. After training, we encode all training expression vectors into an embedding database $\{h_j\}_{j=1}^N$. For a query spot s , we compute similarity $\phi_{sj} = v_s^\top h_j$ and select top-150 entries to construct candidate set. However, relying solely on embedding similarity may introduce false-positive references with mismatched molecular compositions. We leverage the gating signal g_s to perform cell-aware filtering on candidates. Let $\text{sim}_{sj} = \cos(g_s, g_j)$ composition similarity. For query spot s and candidate spot j , we retain only those candidates that satisfy cell count deviation and composition similarity constraints:

$$m_{sj} = 1 \left[\frac{|\sum_t g_{st} - \sum_t g_{jt}|}{\max(\sum_t g_{st}, \sum_t g_{jt})} \leq \tau_c \right] \cdot 1[\text{sim}_{sj} \geq \tau_p] \quad (4)$$

where $\tau_c=0.5$ and $\tau_p=0.3$ are cell count deviation and composition similarity threshold, respectively. Then the final retrieval score is defined as :

$$r_{sj} = (1-\beta) \phi_{sj} + \beta \text{sim}_{sj}, \quad \text{s.t. } m_{sj}=1 \quad (5)$$

where $\beta = 0.3$ controls the trade-off between the two similarities. We rank candidates by r_{sj} and select top- k (or all remaining). The retrieval prediction p_{ret} is obtained as a softmax-weighted average of the retrieved expression vectors.

Regression and Dynamic Soft Consistency. To enhance our framework’s extrapolation capability, we employ the pathology foundation model CONCH [20] to extract features f_i , which are then passed through a multi-layer regression head to produce p_{reg} . However, under limited samples and high-dimensional output, purely supervised regression may produce biologically implausible predictions. To reduce estimation uncertainty, we introduce the structured predictions from retrieval as a soft constraint $\mathcal{L}_{con} = \lambda(e) \cdot \text{MSE}(p_{reg}, p_{ret})$. The joint optimization objective of the regression branch is formulated as:

$$\mathcal{L}_{reg} = \text{MSE}(p_{reg}, y) + \lambda(e) \cdot \text{MSE}(p_{reg}, p_{ret}) \quad (6)$$

where guidance weight $\lambda(e) = \frac{\lambda_0}{2} (1 + \cos \frac{\pi e}{E_d})$ cosine-anneals from $\lambda_0 = 1$ to 0 over first $E_d = 30$ epochs. Thereby, it constrains the output to the biological expression structure in retrieval at early stage, while this constraint is released to allow free optimization later. Embedding database $\{h_j\}_{j=1}^N$ will be fresh each epoch to ensure soft guidance signal improves in tandem with embedding quality.

2.3 Adaptive Expert Triage Fusion

We design a lightweight adaptive fusion module to learn spot-level preference for route predictions. We first take frozen features $f_s \in \mathcal{R}^D$ from the vision encoder in the regression branch and use them as input to a lightweight MLP that predicts a spot-level fusion weight α_s . Specifically, the fusion weight is defined as: $\alpha_s = \frac{1}{2} + \frac{1}{2} \tanh(\text{MLP}(f_s))$, where the tanh function naturally constrains $\alpha_s \in (0, 1)$. The final fused prediction y_{duet} is then formulated as:

$$y_{duet} = \alpha_s \cdot y_{ret} + (1 - \alpha_s) \cdot y_{reg} \quad (7)$$

This module is trained in a post-hoc stage with both branches fully frozen, using a small set of held-out slides from the training set. The optimization objective is defined as $\mathcal{L}_{fuse} = \text{MSE}(y_{duet}, y) + \|\delta\|_2^2$, where $\delta = \alpha_s - 0.5$ is the residual deviation from equal weighting, and the regularization term $\|\delta\|_2^2$ prevents overfitting on the small-scale held-out data, enabling the adapter to dynamically modulate pathway contributions for each spot and yield final estimation of y_{duet} .

3 Data and Experiments

Datasets and Preprocessing. All methods are fairly assessed on three publicly available ST datasets, including HER2 [21], Breast Cancer [5], and Kidney [22]. We extract a 224×224 image patch centered at each spot’s coordinates as the model input. As prediction targets, we consider the top 100, 300, and 500 high-variance genes (HVGs). Consistent with BLEEP [7], raw gene expression counts are transformed using $\log(1+x)$. For external single-cell references, we adopt two

Table 1. Quantitative comparisons on expression prediction. Best performance is highlighted in **orange**, where DUET outperforms SOTAs across datasets and HVGs.

HVG Model	Breast Cancer [5]			HER2 [21]			Kidney [22]			
	MSE ↓	MAE ↓	PCC ↑	MSE ↓	MAE ↓	PCC ↑	MSE ↓	MAE ↓	PCC ↑	
100	ST-Net [5]	0.4126	0.5179	0.1270	0.6083	0.6157	0.2260	0.5412	0.5925	0.1672
	His2ST [27]	0.4517	0.5525	0.0822	0.6533	0.6588	0.1413	0.5231	0.5879	0.0778
	EGN [8]	0.4294	0.5297	0.1216	0.5866	0.6031	0.2312	0.5179	0.5291	0.1182
	BLEEP [7]	0.3738	0.4794	0.1975	0.6810	0.6183	0.2223	0.5923	0.6028	0.1898
	mclSTExp [12]	0.5141	0.5333	0.0950	0.6467	0.6125	0.2168	0.5804	0.6026	0.1942
	OmiCLIP [15]	0.3945	0.5006	0.2432	0.6033	0.6028	0.2649	0.4865	0.5639	0.2227
	UMPIRE [14]	0.4209	0.5213	0.1905	0.5802	0.6028	0.2385	0.4529	0.5379	0.2496
	DUET (Ours)	<u>0.2886</u>	<u>0.4117</u>	<u>0.3583</u>	<u>0.5726</u>	<u>0.5906</u>	<u>0.3038</u>	<u>0.4210</u>	<u>0.5286</u>	<u>0.3099</u>
300	ST-Net [5]	0.4745	0.5525	0.1133	0.6201	0.6018	0.2240	0.5038	0.5899	0.1466
	His2ST [27]	0.4399	0.5489	0.0501	0.6345	0.6496	0.1339	0.5212	0.5889	0.0539
	EGN [8]	0.4164	0.5277	0.1214	0.6771	0.6809	0.1549	0.5566	0.6047	0.1061
	BLEEP [7]	0.3593	0.4778	0.1860	0.6311	0.5968	0.2195	0.5428	0.5481	0.2473
	mclSTExp [12]	0.4307	0.5046	0.1142	0.6661	0.6189	0.1958	0.6412	0.6247	0.1736
	OmiCLIP [15]	0.3903	0.5045	0.1843	0.5672	0.5884	0.2382	0.4710	0.5576	0.2127
	UMPIRE [14]	0.4208	0.5309	0.2169	0.5838	0.6104	0.2500	0.4586	0.5492	0.2748
	DUET (Ours)	<u>0.2959</u>	<u>0.4321</u>	<u>0.3082</u>	<u>0.5297</u>	<u>0.5668</u>	<u>0.3158</u>	<u>0.4067</u>	<u>0.5196</u>	<u>0.3105</u>
500	ST-Net [5]	0.4849	0.5796	0.1094	0.6485	0.6459	0.1901	0.4957	0.5678	0.1359
	His2ST [27]	0.4419	0.5496	0.0799	0.6359	0.6508	0.0769	0.5274	0.5881	0.0400
	EGN [8]	0.4232	0.5323	0.1155	0.6519	0.6648	0.2135	0.4937	0.5833	0.0960
	BLEEP [7]	0.3539	0.4747	0.2255	0.6335	0.6042	0.2037	0.5262	0.5755	0.1930
	mclSTExp [12]	0.4073	0.4975	0.1461	0.6416	0.6083	0.2100	0.5672	0.5939	0.1367
	OmiCLIP [15]	0.4007	0.5139	0.2522	0.5540	0.5826	0.2551	0.4873	0.5632	0.2088
	UMPIRE [14]	0.4249	0.5336	0.1739	0.5745	0.6042	0.2416	0.4751	0.5565	0.2583
	DUET (Ours)	<u>0.3095</u>	<u>0.4481</u>	<u>0.3008</u>	<u>0.5248</u>	<u>0.5654</u>	<u>0.3150</u>	<u>0.4099</u>	<u>0.5078</u>	<u>0.3153</u>

million cells from [23] for the Kidney dataset, and approximately three million cells collected from [24–26] for the Breast Cancer and HER2 datasets.

Baseline and Evaluation Metrics. We benchmark our DUET against SOTA methods, covering regression-based ST-Net [5], EGN [8] and His2ST [27]; retrieval-based BLEEP [7] and mclSTExp [12]; and foundation models OmiCLIP [15] and UMPIRE [14]. All methods were implemented under the same protocols. Model performance was evaluated using Pearson correlation coefficient (PCC), mean squared error (MSE), and mean absolute error (MAE).

Implementation Details. All experiments were performed on two NVIDIA RTX A6000 GPUs. We adopted SGD optimizer with a momentum of 0.9 and a weight decay of 10^{-4} . The learning rate was set to $lr_0 = 10^{-4}$, and the batch size was set to 128. For retrieval, we set $\tau_c = 0.5$, $\tau_p = 0.15$, and top- $k = 100$.

4 Results

Empirical Cross-validation on Expression Prediction. We evaluate DUET against SOTAs using four-fold sample-level cross-validation. Table 1 summarizes

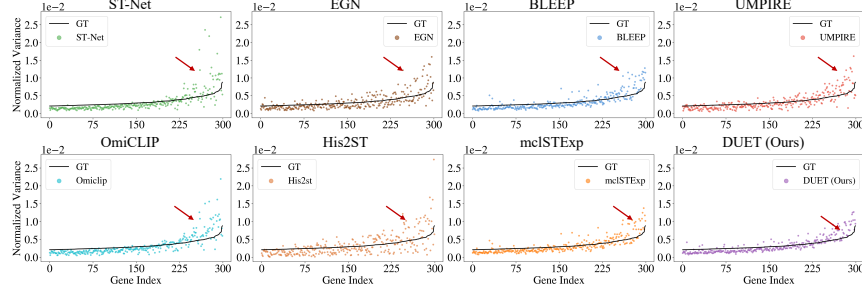


Fig. 3. Normalized variance of predictions on Kidney (300 HVGs). Red arrows highlight the failure of current methods to preserve the bio-fidelity of inter-gene variance pattern.

quantitative comparisons across cohorts and HVG settings, where DUET consistently achieves best performance on all metrics. For HER2 with 300 HVGs, DUET reaches 0.3158 ± 0.135 of PCC and 0.5297 ± 0.141 of MSE, substantially outperforming single-manner models BLEEP ($0.2195 \pm 0.175 / 0.6311 \pm 0.167$) and OmiCLIP ($0.2500 \pm 0.185 / 0.5838 \pm 0.207$). This advantage is consistently maintained as the number of HVGs varies, indicating its stronger adaptability.

Following [28], we further evaluate prediction fidelity by computing the normalized variance and sorting in ascending order of truth variance (Fig. 3). Regression-based methods (e.g., His2ST, UMPIRE) markedly overestimate highly variable genes, leading to unstable predictions. In contrast, DUET closely follows the ground-truth curve across the entire variance spectrum, avoiding over-smoothing in low-variance genes and noise amplification in high-variance ones. These results suggest that our dual-paradigm design effectively regularizes the output distribution while preserving bio-meaningful gene variability patterns.

Pivotal Gene Expression Prediction. We identified α -actinin-4 (ACTN4) [29, 30], a key breast-cancer-associated marker, to assess models’ practical applicability. Fig. 4 visualizes the spatial pattern of ACTN4 along with the PCC across different models. Our DUET achieved the best PCC of 0.729. In contrast, monolithic-paradigm methods such as EGN (0.341) or BLEEP (-0.024) failed to reliably recover ACTN4 due to limited ability to capture cross-sample spatial heterogeneity. By incorporating a cellular prior and dual-paradigm adaptive design, DUET demonstrates strong utility in real-world clinical scenarios.

Ablation Study. We conducted a detailed ablation study with 300 HVGs to evaluate the effectiveness of each functional block, as is summarized in Table 2. Results for DUET_{Reg} and DUET_{Ret} demonstrated that incorporation of cellular prior leads to clear improvement in either regression or retrieval manner. Sensitivity analysis of gating hyperparameters τ_c and τ_p further demonstrates the impact of the cellular signal. A low threshold admits many mismatched references and adds noise, whereas a high threshold discards informative signals; therefore, a moderate threshold yields the best performance. Our adaptive fusion strategy further improves performance over either single branch, reducing MSE

Table 2. Ablation study for the functional blocks in DUET with 300 HVGs, where best performance is highlighted in **orange** and the second is highlighted in **blue**.

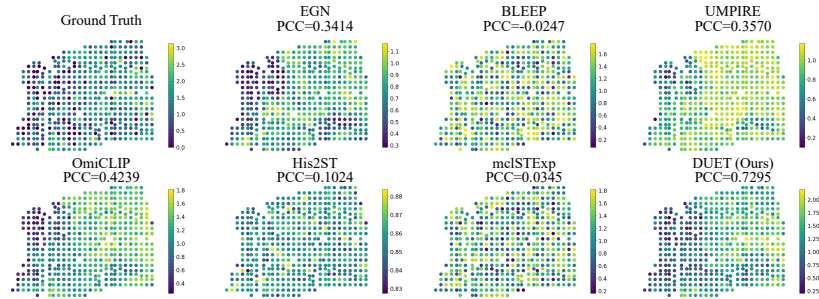
Setting	HER2 [21]			Kidney [22]		
	MSE ↓	MAE ↓	PCC ↑	MSE ↓	MAE ↓	PCC ↑
w.o. $\mathcal{L}_{con}$	0.589 $\pm$.192	0.608 $\pm$.084	0.255 $\pm$.175	0.471 $\pm$.045	0.556 $\pm$.022	0.267 $\pm$.024
DUET _{Reg}	0.574 $\pm$.184	0.581 $\pm$.104	0.281 $\pm$.147	0.478 $\pm$.031	0.549 $\pm$.013	0.273 $\pm$.037
w.o. Gating	0.631 $\pm$.158	0.597 $\pm$.083	0.220 $\pm$.128	0.543 $\pm$.026	0.548 $\pm$.015	0.249 $\pm$.029
$\tau_c, \tau_p=0.1, 0.1$	0.598 $\pm$.118	0.589 $\pm$.066	0.239 $\pm$.133	0.481 $\pm$.036	0.533 $\pm$.019	0.251 $\pm$.022
$\tau_c, \tau_p=0.7, 0.5$	0.581 $\pm$.143	0.590 $\pm$.077	0.242 $\pm$.129	0.490 $\pm$.024	0.525 $\pm$.013	0.273 $\pm$.030
DUET _{Ret}	0.569 $\pm$.134	0.586 $\pm$.074	0.266 $\pm$.134	0.488 $\pm$.027	0.526 $\pm$.015	0.276 $\pm$.027
$G_{de}=300$	0.529$\pm$.131	0.570$\pm$.083	0.315$\pm$.135	0.416 $\pm$.025	0.524 $\pm$.015	0.304 $\pm$.029
$G_{de}=1000$	0.517$\pm$.154	0.578 $\pm$.085	0.312 $\pm$.138	0.405$\pm$.021	0.511$\pm$.012	0.304 $\pm$.027
Avg $\alpha_s=0.5$	0.544 $\pm$.151	0.570 $\pm$.085	0.303 $\pm$.145	0.447 $\pm$.023	0.545 $\pm$.014	0.294 $\pm$.033
DUET(Ours)	0.530$\pm$.141	0.567$\pm$.081	0.316$\pm$.135	0.407$\pm$.023	0.520$\pm$.013	0.311$\pm$.030

from 0.478 ± 0.031 to 0.407 ± 0.023 . The gains from each module are orthogonal, and DUET achieves optimal performance with all blocks integrated.

We further vary the size of gene panels g_{de} for cell type deconvolution and observe that only a small set of prior genes is sufficient to achieve robust bio-guidance. Moreover, DUET contains 97M parameters, fewer than or similar to OmiCLIP (ViT-L/16, 304M) and UMPIRE (ViT-B/16, 87M), yet achieves superior performance, highlighting its parameter efficiency and effective architecture.

5 Conclusion

We present a novel cellular-empowered dual-paradigm framework that moves beyond monolithic vision-to-expression modeling in ST prediction. By orchestrating parametric regression and memory-based retrieval under cellular inductive priors, DUET reconciles expressive flexibility with biological fidelity. Embedded large-scale scRNA-seq knowledge then anchors predictions to biologically

**Fig. 4.** Predicted spatial expression distribution of breast-cancer-related gene ACTN4.

grounded molecular states, promoting principled and faithful inference. An adaptive expert triage mechanism further dynamically modulates the preference of dual pathways across spatial contexts, thereby yielding a more reliable estimation. Extensive experiments on three public datasets validate our framework’s superior performance against diverse prediction architectures, with consistent gains under various settings. By bridging expressive modeling with molecular reliability, DUET establishes a unified paradigm for spatial molecular modeling and paves the way toward scalable deployment in real-world omics analysis.

Acknowledgments. This research was supported by NIH R01DK135597 (Huo), DoD HT9425-23-1-0003 (HCY), NSF 2434229 (Huo), and KPMP Glue Grant. This work was also supported by Vanderbilt Seed Success Grant, Vanderbilt Discovery Grant, and VISE Seed Grant. This project was supported by The Leona M. and Harry B. Helm-sley Charitable Trust grant G-1903-03793 and G-2103-05128. This research was also supported by NIH grants R01EB033385, R01DK132338, REB017230, R01MH125931, and NSF 2040462. We extend gratitude to NVIDIA for their support by means of the NVIDIA hardware grant. This work was also supported by NSF NAIRR Pilot Award NAIRR240055.

References

1. Darren J Burgess. Spatial transcriptomics coming of age. *Nature Reviews Genetics*, 20(6):317–317, 2019.
2. Michaela Asp, Stefania Giacomello, Ludvig Larsson, Chenglin Wu, Daniel Fürth, Xiaoyan Qian, Eva Wärdell, Joaquin Custodio, Johan Reimegård, Fredrik Salmén, et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell*, 179(7):1647–1660, 2019.
3. Michaela Asp, Joseph Bergenstråhle, and Joakim Lundeberg. Spatially resolved transcriptomes—next generation tools for tissue exploration. *Bioessays*, 42(10):1900221, 2020.
4. Kyongho Choe, Unil Pak, Yu Pang, Wanjun Hao, and Xiuqin Yang. Advances and challenges in spatial transcriptomics for developmental biology. *Biomolecules*, 13(1):156, 2023.
5. 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.
6. Junchao Zhu, Ruining Deng, Tianyuan Yao, Juming Xiong, Chongyu Qu, Junlin Guo, Siqi Lu, Mengmeng Yin, Yu Wang, Shilin Zhao, et al. Assign: an anatomy-aware spatial imputation graphic network for 3d spatial transcriptomics. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 30829–30838, 2025.
7. Ronald Xie, Kuan Pang, Sai Chung, Catia Perciani, Sonya MacParland, Bo Wang, and Gary Bader. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. *Advances in Neural Information Processing Systems*, 36, 2024.
8. Yan Yang, Md Zakir Hossain, Eric A Stone, and Shafin Rahman. 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*, pages 5039–5048, 2023.

9. Aviv Regev, Sarah A Teichmann, Eric S Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, Peter Campbell, Piero Carninci, Menna Clatworthy, et al. The human cell atlas. *elife*, 6:e27041, 2017.
10. Junchao Zhu, Ruining Deng, Junlin Guo, Tianyuan Yao, Chongyu Qu, Juming Xiong, Siqu Lu, Zhengyi Lu, Yanfan Zhu, Marilyn Lions, et al. Scr2-st: Combine single cell with spatial transcriptomics for efficient active sampling via reinforcement learning. *arXiv preprint arXiv:2512.13635*, 2025.
11. Junchao Zhu, Ruining Deng, Junlin Guo, Tianyuan Yao, Siqu Lu, Chongyu Qu, Juming Xiong, Yanfan Zhu, Zhengyi Lu, Yuechen Yang, et al. Computer vision methods for spatial transcriptomics: A survey. *bioRxiv*, pages 2025–10, 2025.
12. Wenwen Min, Zhiceng Shi, Jun Zhang, Jun Wan, and Changmiao Wang. Multi-modal contrastive learning for spatial gene expression prediction using histology images. *Briefings in Bioinformatics*, 25(6):bbae551, 2024.
13. Minxing Pang, Kenong Su, and Mingyao Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *BioRxiv*, pages 2021–11, 2021.
14. Minghao Han, Dingkan Yang, Jiabei Cheng, Xukun Zhang, Zizhi Chen, Haopeng Kuang, and Lihua Zhang. Towards unified molecule-enhanced pathology image representation learning via integrating spatial transcriptomics. *Pattern Recognition*, page 112458, 2025.
15. Weiqing Chen, Pengzhi Zhang, Tu N Tran, Yiwei Xiao, Shengyu Li, Vrutant V Shah, Hao Cheng, Kristopher W Brannan, Keith Youker, Li Lai, et al. A visual-omics foundation model to bridge histopathology with spatial transcriptomics. *Nature Methods*, pages 1–15, 2025.
16. Lucie C Gaspard-Boulin, Luca Gortana, Thomas Walter, Emmanuel Barillot, and Florence MG Cavalli. Cell-type deconvolution methods for spatial transcriptomics. *Nature Reviews Genetics*, 26(12):828–846, 2025.
17. Vitalii Kleshchevnikov, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W King, Tong Li, Rasa Elmentaite, Artem Lomakin, Veronika Kedlian, Adam Gayoso, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature biotechnology*, 40(5):661–671, 2022.
18. Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical image analysis*, 94:103143, 2024.
19. Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
20. Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature medicine*, 30(3):863–874, 2024.
21. Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, et al. Spatial deconvolution of her2-positive breast cancer delineates tumor-associated cell type interactions. *Nature communications*, 12(1):6012, 2021.
22. Blue B Lake, Rajasree Menon, Seth Winfree, Qiwen Hu, Ricardo Melo Ferreira, Kian Kalhor, Daria Barwinska, Edgar A Otto, Michael Ferkowicz, Dinh Diep, et al. An atlas of healthy and injured cell states and niches in the human kidney. *Nature*, 619(7970):585–594, 2023.

23. Blue B Lake, Ricardo Melo Ferreira, Jens Hansen, Rajasree Menon, Jeannine Basta, Heather Thiessen Philbrook, Stephanie Reinert, Robin Fallegger, Asmita K Lagwankar, Xi Chen, et al. Cellular and spatial drivers of unresolved injury and functional decline in the human kidney. *BioRxiv*, pages 2025–09, 2025.
24. Andrew Chen, Lina Kroehling, Christina S Ennis, Gerald V Denis, and Stefano Monti. A highly resolved integrated transcriptomic atlas of human breast cancers. *bioRxiv*, pages 2025–03, 2025.
25. Austin D Reed, Sara Pensa, Adi Steif, Jack Stenning, Daniel J Kunz, Linsey J Porter, Kui Hua, Peng He, Alecia-Jane Twigger, Abigail JQ Siu, et al. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. *Nature genetics*, 56(4):652–662, 2024.
26. 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.
27. Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5):bbac297, 2022.
28. Sichen Zhu, Yuchen Zhu, Molei Tao, and Peng Qiu. Diffusion generative modeling for spatially resolved gene expression inference from histology images. *arXiv preprint arXiv:2501.15598*, 2025.
29. Neng Wang, Qi Wang, Hailin Tang, Fengxue Zhang, Yifeng Zheng, Shengqi Wang, Jin Zhang, Zhiyu Wang, and Xiaoming Xie. Direct inhibition of actn4 by ellagic acid limits breast cancer metastasis via regulation of β -catenin stabilization in cancer stem cells. *Journal of Experimental & Clinical Cancer Research*, 36(1):172, 2017.
30. Dmitri Tentler, Ekaterina Lomert, Ksenia Novitskaya, and Nikolai A Barlev. Role of actn4 in tumorigenesis, metastasis, and emt. *Cells*, 8(11):1427, 2019.