GLoDe: Global–Local Decomposition for Histology-based Spatial Gene Expression Prediction

Anonymous authors
Paper under double-blind review

Abstract

The high cost of spatial transcriptomics motivates scalable prediction of spatial gene expression from readily available hematoxylin and eosin (H&E) images. However, existing methods typically predict absolute gene expression without explicitly disentangling slide-level molecular shifts from within-slide spatial variation, which may cause models to capture slide-specific biases rather than transferable spatial patterns, limiting their generalization across tissue sections. To address this limitation, we propose GLoDe, a global–local decomposition framework that decomposes gene expression into a slide-shared molecular baseline and a spot-specific residual, explicitly accounting for their distinct sources of variation. The global component captures tissue-level molecular characteristics shared across spots within each tissue section, while the local component employs a gene-conditioned residual decoder to integrate gene priors, local morphology, and spatial context for predicting gene-specific residual expression. Extensive experiments on HEST-1K and HER2ST demonstrate consistent performance gains, including a 16.9% relative improvement in average Pearson correlation under out-of-fold evaluation over the strongest evaluated baseline on HEST-1K.

1 Introduction

Spatial transcriptomics (ST) enables genome-wide gene expression profiling while preserving tissue spatial organization, offering a powerful tool for studying tissue architecture, cellular heterogeneity, and disease microenvironments (Ståhl et al., 2016; Chen et al., 2024a). However, its expensive and specialized experimental procedures limit scalability to large clinical cohorts (Liang et al., 2023). In contrast, hematoxylin and eosin (H&E) slides are routinely available in clinical pathology and contain rich morphological information, motivating the task of predicting spatial gene expression directly from histology (Wang et al., 2025; He et al., 2020).

Early histology-to-spatial-transcriptomics (HE-to-ST) prediction methods directly predict multi-gene expression from local image patches centered at individual spots (He et al., 2020). Since local morphology alone may not fully capture the tissue context relevant to spatial gene expression, subsequent methods incorporate neighboring spots, graph structures, and multi-resolution tissue regions to model broader spatial dependencies (Pang et al., 2021; Zeng et al., 2022; Chung et al., 2024). To further capture uncertainty in morphology-to-expression mapping and expression dependencies across locations, generative approaches model conditional expression distributions or slide-level joint expression patterns (Huang et al., 2025a). Meanwhile, gene-aware methods incorporate gene information to capture gene-specific morphology–expression relationships (Xiong et al., 2024; Liu et al., 2026), while cross-slide approaches establish correspondences between tissue sections or draw on reference slides to address distribution shifts (Shi et al., 2026; Wang et al., 2026).

Despite these advances, HE-to-ST prediction still faces two interconnected modeling challenges. First, most existing methods directly predict absolute spot-level expression without

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.

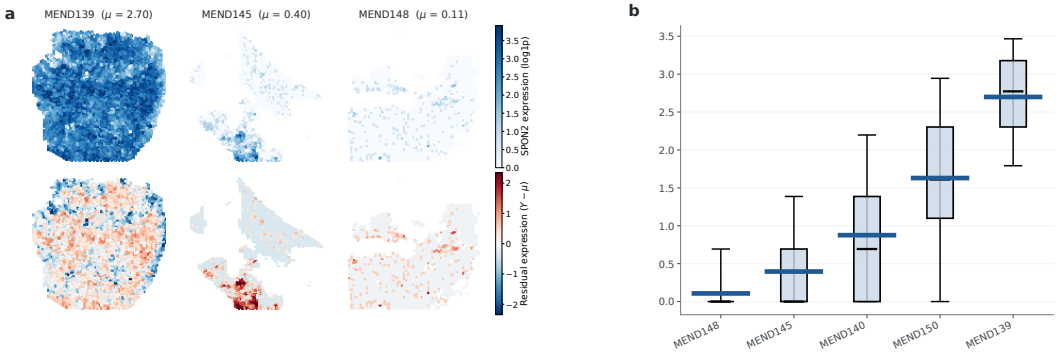


Figure 1: Slide-level expression shifts motivate global–local decomposition. (a) Observed SPON2 expression (top) and residual expression after subtracting the corresponding slide mean μ (bottom) for three representative PRAD slides. (b) SPON2 expression distributions across five PRAD slides, ordered by increasing slide mean.

explicitly distinguishing slide-shared expression levels from within-slide spatial variation. As illustrated in Figure 1, gene-wise mean expression exhibits pronounced shifts across tissue sections, while structured spatial variation remains within individual slides after subtracting the corresponding slide-level mean. Variance decomposition across multiple cancer cohorts further indicates that between-slide variation accounts for a non-negligible fraction of the total expression variance, supporting separate modeling of slide-level molecular baselines and within-slide spatial variation (Appendix C.1). Second, within-slide expression variation in different genes may correspond to distinct cellular and morphological patterns that a shared spot representation may not fully capture. This motivates integrating gene priors with local morphology to construct gene-specific visual (GCV) features for within-slide residual prediction.

To address these limitations, we propose GLoDe, a global–local decomposition framework for histology-based spatial gene expression prediction. GLoDe models the hierarchical structure and gene specificity of expression through two complementary designs. First, it decomposes expression at each spot into a slide-shared molecular baseline and a spot-specific residual, separately modeling between-slide expression differences and within-slide spatial variation. Second, it introduces a Gene-Conditioned Residual Decoder (GCRD) that integrates gene priors, local morphology, and spatial context into GCV features and predicts gene-specific residuals through a parameter-shared decoder. Extensive experiments on HEST-1K (Jaume et al., 2024) and HER2ST (Andersson et al., 2021) demonstrate consistent performance improvements across cancer types, gene panels, and evaluation settings, supporting its applicability across cancer cohorts and its ability to generalize to unseen patients.

In summary, our main contributions are as follows:

1. **Global–local decomposition for spatial gene expression prediction.** We propose GLoDe, which explicitly models the hierarchical structure of spatial transcriptomic data by decomposing spot-level gene expression into a slide-specific molecular baseline and within-slide residual variation.
2. **Gene-conditioned residual decoding.** We introduce a GCRD that integrates gene priors, local morphology, and spatial context to construct GCV features for each target gene at each spot. A parameter-shared decoder maps these features to gene-specific residual expression predictions.
3. **Extensive evaluation across benchmarks and settings.** Experiments on HEST-1K and HER2ST demonstrate consistent improvements across multiple cancer types, gene panels, and evaluation settings, including a 16.9% relative improvement on HEST-1K over the strongest baseline; ablation studies and model analyses further support the proposed design.

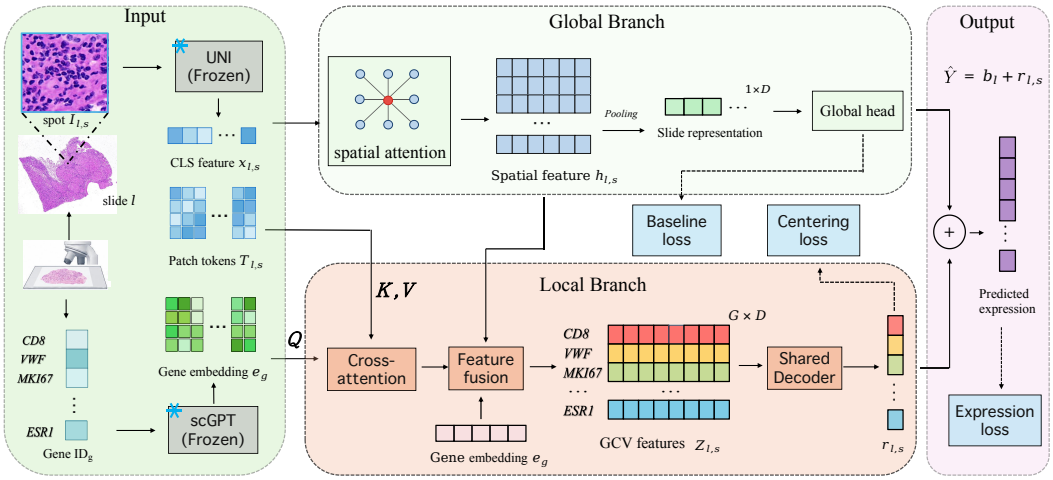


Figure 2: Overview of GLoDe for HE-to-ST prediction. The global branch aggregates spatially contextualized CLS features to predict a slide-shared molecular baseline. The GCRD integrates gene priors, local morphology, and spatial context into GCV features and predicts spot-specific residuals through a parameter-shared decoder. The baseline and residual predictions are added to obtain the final expression estimate.

2 Related Work

2.1 Histology-Based Spatial Gene Expression Prediction

Early HE-to-ST methods formulate spatial gene expression prediction as a direct regression problem, using local histological image patches to predict the expression profiles of individual spots (He et al., 2020; Dawood et al., 2021; Monjo et al., 2022). However, local morphology alone may not fully capture the tissue context relevant to spatial gene expression. Subsequent studies therefore incorporate neighboring spot information, graph structures, and multi-scale tissue representations to model within-slide spatial dependencies. For example, HisToGene and Hist2ST incorporate spatial positional information and employ Transformers or graph structures to model relationships among spots (Pang et al., 2021; Zeng et al., 2022); TRIPLEX and DeepSpot further integrate neighborhood and multi-scale tissue information to enrich morphological and contextual representations for expression prediction (Chung et al., 2024; Nonchev et al., 2025).

Beyond spatial context modeling, recent studies improve HE-to-ST prediction through cross-modal representation learning and expression distribution modeling. Methods such as BLEEP and OmiCLIP learn associations between histological images and gene expression through cross-modal alignment (Xie et al., 2023; Chen et al., 2025). Generative approaches further model conditional distributions of expression given morphology: STFlow jointly models expression across an entire tissue section to capture dependencies across spatial locations (Huang et al., 2025a; Zhu et al., 2025); STPath learns transferable histology–expression associations through large-scale generative pretraining (Huang et al., 2025b). Notably, recent studies have begun to consider the impact of between-slide differences on spatial gene expression prediction. SpaHGC establishes connections between target and reference slides to facilitate knowledge transfer and improve cross-slide prediction robustness (Shi et al., 2026).

2.2 Foundation Models and Gene-Conditioned Representation Learning

Although spatial context modeling improves the representation of tissue organization, different genes may depend on distinct morphological cues. A shared spot-level visual representation may therefore be insufficient to fully capture gene-specific morphology–expression relationships. Recent studies have consequently begun to incorporate gene information into

multimodal learning frameworks. Gene foundation models such as scGPT and Geneformer learn transferable molecular representations from large-scale transcriptomic data, providing gene-level priors for expression prediction (Cui et al., 2024; Theodoris et al., 2023). Meanwhile, pathology foundation models such as UNI and GigaPath learn rich morphological features through large-scale histological image pretraining (Chen et al., 2024b; Xu et al., 2024). These visual and molecular representations provide a foundation for integrating tissue morphology and gene priors in HE-to-ST prediction.

Building on these advances, HAGE and HistoGPA incorporate pretrained gene embeddings as priors and enhance spot-level morphological representations through cross-modal attention to improve spatial gene expression prediction (Dang et al., 2025; Liu et al., 2026). These methods focus on incorporating gene information into representation learning, while their prediction formulations still directly output spot-level expression without explicitly decomposing it into a slide-shared molecular baseline and a spot-specific residual component.

3 Method

In this section, we introduce GLoDe, a global–local decomposition framework for spatial gene expression prediction. We first formulate the HE-to-ST prediction task (Section 3.1) and provide a framework overview (Section 3.2). We then describe the global branch for slide-level baseline modeling (Section 3.3) and the local branch for gene-conditioned residual prediction (Section 3.4), followed by the training objective (Section 3.5).

3.1 Preliminaries

We consider the standard HE-to-ST prediction task. H&E staining and ST profiling provide paired histological images and spatially resolved gene expression measurements for a tissue section. For slide l, the paired data are represented as $\mathcal{D}_l = \{(p_{l,s}, I_{l,s}, y_{l,s})\}_{s=1}^{N_l}$, where N_l is the number of spots. Each spot s is associated with a spatial coordinate $p_{l,s} \in \mathbb{R}^2$, a spot-centered H&E image patch $I_{l,s} \in \mathbb{R}^{H\times W\times 3}$, and a gene expression vector $y_{l,s} \in \mathbb{R}^G$. Here, H and W denote the patch height and width, and G is the number of target genes. The HE-to-ST task aims to predict gene expression from histology and spatial coordinates:

$$f_\theta : \{(I_{l,s}, p_{l,s})\}_{s=1}^{N_l} \mapsto \{y_{l,s}\}_{s=1}^{N_l}, \tag{1}$$

where θ denotes the trainable model parameters.

3.2 Framework Overview

Figure 2 provides an overview of GLoDe, which explicitly models the slide–spot hierarchy through global–local decomposition. The global branch aggregates spatially contextualized spot features into a slide-level representation to predict a molecular baseline b_l shared across spots within each slide. The local branch integrates gene priors, gene-specific local morphological features, and spatial context into gene-conditioned visual (GCV) features, and predicts spot-specific residuals $r_{l,s}$ through a parameter-shared decoder. The two predictions are added to obtain the final expression estimate:

$$y_{l,s} = b_l + r_{l,s}. \tag{2}$$

3.3 Global Branch

In the global branch, we first use a pretrained pathology foundation model to extract visual features from each spot-centered H&E image patch:

$$(x_{l,s}, T_{l,s}) = f_{\mathrm{PFM}}(I_{l,s}), \tag{3}$$

where $x_{l,s} \in \mathbb{R}^d$ is the class (CLS) token feature summarizing the overall morphology of each spot image, and $T_{l,s} \in \mathbb{R}^{N_t\times d_t}$ contains N_t patch-token features preserving local morphological details. CLS features support slide-level modeling in the global branch, while patch-token features provide fine-grained morphological information for the local branch.

Specifically, we use spatial attention to model spatial relationships among spots, obtaining spot representations $h_{l,s}$ that incorporate neighborhood context. We then apply mean pooling across all spots within the slide to form a slide-level representation and use a multilayer perceptron to predict the slide-shared molecular baseline:

$$h_{l,s} = \mathrm{SpatialAttn}\left(W_x x_{l,s}, \{W_x x_{l,j} : j \in \mathcal{N}(s)\}\right) \in \mathbb{R}^D, \tag{4}$$

$$b_l = \mathrm{MLP}\left(\frac{1}{N_l}\sum_{s=1}^{N_l} h_{l,s}\right) \in \mathbb{R}^G. \tag{5}$$

Here, $W_x \in \mathbb{R}^{D\times d}$ is a learnable linear projection matrix that maps CLS features into a D-dimensional latent space, and $\mathcal{N}(s)$ denotes the K nearest neighbors of spot s determined by spatial coordinates. SpatialAttn computes attention weights from the features of the central spot and its neighbors and aggregates neighborhood information through a weighted sum. MLP maps the pooled slide-level representation to a G-dimensional molecular baseline vector.

3.4 Local Branch

In the local branch, we first retrieve gene embeddings from a pretrained molecular foundation model using the vocabulary identifier ID_g of each target gene:

$$e_g = f_{\mathrm{MFM}}(\mathrm{ID}_g), \tag{6}$$

where $e_g \in \mathbb{R}^{d_g}$. The embeddings of all target genes form the matrix $E = [e_1, \ldots, e_G]^\top \in \mathbb{R}^{G\times d_g}$.

Gene embeddings provide gene identity and potential molecular associations shared across spots, while patch tokens preserve local morphological details specific to each spot. To combine these complementary sources of information, we project gene embeddings into queries and use cross-attention to extract local morphological features associated with each gene. Specifically, we define the queries as $Q = \mathrm{LN}(EW_e^\top) \in \mathbb{R}^{G\times D}$ and the keys and values as $K_{l,s} = V_{l,s} = T_{l,s}W_t \in \mathbb{R}^{N_t\times D}$. Here, W_e and W_t are learnable projection matrices, and LN denotes layer normalization. The gene-specific morphological features are computed as:

$$M_{l,s} = \mathrm{CrossAttn}\left(Q, K_{l,s}, V_{l,s}\right) \in \mathbb{R}^{G\times D}. \tag{7}$$

Each row of $M_{l,s}$ corresponds to a target gene and contains morphological information retrieved from patch tokens through its gene query. We then fuse the morphological features $M_{l,s}$, gene priors Q, and spatial context $h_{l,s}$ through adaptive gating to obtain the GCV features:

$$A_{l,s} = f_{\mathrm{gate}}\left([Q; M_{l,s}; h_{l,s}]\right), \tag{8}$$

$$Z_{l,s} = \mathrm{LN}\left(Q + M_{l,s} + A_{l,s} \odot (W_h h_{l,s})^\top\right) \in \mathbb{R}^{G\times D}. \tag{9}$$

Here, $[\,;\,]$ denotes concatenation along the feature dimension, with $h_{l,s}$ broadcast along the gene dimension to align with Q and $M_{l,s}$. The gating network f_{gate} operates row-wise to generate weights $A_{l,s} \in \mathbb{R}^{G\times D}$, adaptively controlling the contribution of spatial context to each gene. W_h is a learnable projection matrix, and $\odot$ denotes element-wise multiplication, with the projected context vector likewise shared across genes.

Since the GCV features already encode gene-specific information, we use a parameter-shared decoder f_{shared} to learn a common predictive mapping across genes, yielding the spot-level residual expression vector:

$$r_{l,s} = f_{\mathrm{shared}}(Z_{l,s}) \in \mathbb{R}^G. \tag{10}$$

3.5 Training Objective

We jointly use an expression reconstruction loss, a baseline loss, and a centering loss to supervise expression prediction and guide global–local decomposition:

$$\mathcal{L} = \mathcal{L}_{\mathrm{expr}} + \lambda_{\mathrm{base}}\mathcal{L}_{\mathrm{base}} + \lambda_{\mathrm{center}}\mathcal{L}_{\mathrm{center}}, \tag{11}$$

where λ_{base} and λ_{center} weight the baseline and centering losses, respectively.

The expression reconstruction loss supervises the final prediction $\hat{y}_{l,s} = b_l + r_{l,s}$ using SmoothL1, enabling smooth optimization while reducing the influence of extreme errors:

$$\mathcal{L}_{\text{expr}} = \frac{1}{N_l G} \sum_{s=1}^{N_l} \sum_{g=1}^{G} \begin{cases} \dfrac{(\hat{y}_{l,s,g} - y_{l,s,g})^2}{2\beta}, & |\hat{y}_{l,s,g} - y_{l,s,g}| < \beta, \\ |\hat{y}_{l,s,g} - y_{l,s,g}| - \dfrac{\beta}{2}, & \text{otherwise}, \end{cases} \tag{12}$$

where $\beta > 0$ controls the transition between the quadratic and linear regions. Throughout this subsection, the subscript g indexes the corresponding gene component of each vector.

The baseline loss supervises the global branch using the slide-wise mean expression $\bar{y}_{l,g} = \frac{1}{N_l} \sum_{s=1}^{N_l} y_{l,s,g}$:

$$\mathcal{L}_{\text{base}} = \frac{1}{G} \sum_{g=1}^{G} \begin{cases} \dfrac{(b_{l,g} - \bar{y}_{l,g})^2}{2\beta}, & |b_{l,g} - \bar{y}_{l,g}| < \beta, \\ |b_{l,g} - \bar{y}_{l,g}| - \dfrac{\beta}{2}, & \text{otherwise}. \end{cases} \tag{13}$$

The centering loss encourages the slide-wise mean residual of each gene to remain close to zero, limiting the absorption of overall expression shifts by the local branch:

$$\mathcal{L}_{\text{center}} = \frac{1}{G} \sum_{g=1}^{G} \left(\frac{1}{N_l} \sum_{s=1}^{N_l} r_{l,s,g} \right)^2. \tag{14}$$

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate GLoDe on two public spatial transcriptomics datasets paired with H&E-stained histology images: HEST-1K (Jaume et al., 2024) and HER2ST (Andersson et al., 2021). HEST-1K brings together paired data from diverse tissue sources and technological platforms. Its diversity in tissue morphology and expression distributions poses challenges for prediction, enabling evaluation across different cancer cohorts. HER2ST focuses on HER2+ breast cancer, allowing us to assess cross-patient generalization within the same cancer setting amid morphological and molecular differences between patients. Using these two datasets, we conduct evaluations across 11 cancer cohorts, comprising 108 tissue sections and 250,115 spatial locations generated using three spatial transcriptomics technologies: Visium, Xenium, and ST. Detailed statistics are provided in Appendix A.

Evaluation Protocol. For HEST-1K, we perform patient-level k-fold cross-validation separately within each cancer cohort. For HER2ST, we use leave-one-patient-out cross-validation. Both settings ensure that tissue sections from the same patient do not appear in both the training and test sets. We evaluate prediction performance using the gene-wise Pearson correlation coefficient (PCC):

$$\text{PCC}_g = \frac{\text{Cov}(y_g, \hat{y}_g)}{\sqrt{\text{Var}(y_g)\,\text{Var}(\hat{y}_g)}}. \tag{15}$$

Here, y_g and $\hat{y}_g$ denote the observed and predicted expression vectors of gene g across the evaluation samples, respectively. Cov and Var denote covariance and variance. We report mean-fold (MF) PCC and out-of-fold (OOF) PCC. MF PCC is computed by calculating gene-wise PCC within each test fold and then averaging across genes and folds. OOF PCC pools held-out predictions from all test folds within the same cohort, computes gene-wise PCC, and averages across genes. These metrics assess average test-fold performance and overall cross-slide predictive correlation at the cohort level, respectively.

Baselines. We compare GLoDe with BLEEP (Xie et al., 2023), TRIPLEX (Chung et al., 2024), and STFlow (Huang et al., 2025a), representing cross-modal alignment, multi-

Table 1: Performance comparison on HEST-1K across ten cancer cohorts. Best and second-best mean scores are highlighted in bold and underlined. All trainable methods are evaluated with three random seeds; standard deviations across runs are shown as subscripts. † STPath is evaluated directly using its released pretrained model without fine-tuning; OOF PCC is not reported.

Cancer	MF PCC					OOF PCC			
	STFlow	BLEEP	TRIPLEX	STPath†	GLoDe	STFlow	BLEEP	TRIPLEX	GLoDe
CCRCC	$\underline{0.229}_{.011}$	$0.110_{.023}$	$0.113_{.013}$	0.143	$\mathbf{0.254}_{.015}$	$\underline{0.241}_{.003}$	$0.113_{.034}$	$0.102_{.048}$	$\mathbf{0.250}_{.005}$
COAD	$0.299_{.012}$	$0.158_{.035}$	$0.240_{.015}$	**0.400**	$\underline{0.347}_{.025}$	$\underline{0.052}_{.013}$	$-0.031_{.032}$	$0.010_{.026}$	$\mathbf{0.116}_{.038}$
HCC	$0.085_{.003}$	$0.060_{.015}$	$0.035_{.015}$	$\underline{0.098}$	$\mathbf{0.115}_{.008}$	$\underline{-0.040}_{.006}$	$-0.133_{.042}$	$-0.148_{.046}$	$\mathbf{-0.003}_{.020}$
IDC	$0.569_{.006}$	$0.364_{.016}$	$0.477_{.018}$	**0.661**	$\underline{0.593}_{.005}$	**0.338**$_{.023}$	$0.178_{.047}$	$0.233_{.046}$	$\underline{0.269}_{.030}$
LUNG	$0.579_{.003}$	$0.512_{.015}$	$0.572_{.006}$	**0.640**	$\underline{0.610}_{.004}$	$\underline{0.354}_{.003}$	$0.279_{.008}$	$0.300_{.004}$	$\mathbf{0.442}_{.014}$
LYMPH	$\underline{0.269}_{.007}$	$0.178_{.014}$	$0.188_{.007}$	0.214	$\mathbf{0.291}_{.006}$	$\underline{0.026}_{.014}$	$-0.022_{.014}$	$-0.074_{.047}$	$\mathbf{0.065}_{.048}$
PAAD	$0.453_{.013}$	$0.361_{.016}$	$0.384_{.005}$	**0.477**	$\underline{0.465}_{.005}$	**0.320**$_{.008}$	$0.207_{.041}$	$0.165_{.002}$	$\underline{0.289}_{.046}$
PRAD	$0.209_{.032}$	$0.125_{.014}$	$0.229_{.005}$	$\underline{0.275}$	$\mathbf{0.298}_{.034}$	$\underline{0.222}_{.028}$	$0.090_{.027}$	$0.213_{.003}$	$\mathbf{0.304}_{.019}$
READ	$0.200_{.014}$	$0.099_{.034}$	$0.108_{.011}$	**0.292**	$\underline{0.243}_{.008}$	$\underline{0.193}_{.013}$	$0.078_{.051}$	$-0.028_{.014}$	$\mathbf{0.290}_{.026}$
SKCM	$\underline{0.653}_{.023}$	$0.374_{.082}$	$0.547_{.012}$	0.606	$\mathbf{0.704}_{.010}$	$\underline{0.604}_{.022}$	$0.355_{.013}$	$0.470_{.010}$	$\mathbf{0.678}_{.010}$
Average	$0.354_{.007}$	$0.234_{.010}$	$0.289_{.004}$	$\underline{0.381}$	$\mathbf{0.392}_{.005}$	$\underline{0.231}_{.004}$	$0.111_{.009}$	$0.124_{.003}$	$\mathbf{0.270}_{.012}$

Table 2: Performance comparison on HER2ST across the top 50, 100, and 250 highly variable genes. All trainable methods are evaluated with three random seeds; standard deviations across runs are shown as subscripts.

Dataset	Panel	MF PCC				OOF PCC			
		BLEEP	STFlow	TRIPLEX	GLoDe	BLEEP	STFlow	TRIPLEX	GLoDe
HER2ST	Top-50	$0.0906_{.0018}$	$0.1060_{.0051}$	$\underline{0.1261}_{.0008}$	$\mathbf{0.1332}_{.0014}$	$\underline{0.0605}_{.0209}$	$0.0564_{.0073}$	$0.0566_{.0097}$	$\mathbf{0.0611}_{.0058}$
	Top-100	$0.0903_{.0015}$	$0.1124_{.0051}$	$\underline{0.1264}_{.0023}$	$\mathbf{0.1350}_{.0009}$	$0.0767_{.0228}$	$\underline{0.0818}_{.0093}$	$0.0738_{.0012}$	$\mathbf{0.0869}_{.0125}$
	Top-250	$0.0961_{.0007}$	$0.1207_{.0041}$	$\underline{0.1343}_{.0018}$	$\mathbf{0.1419}_{.0012}$	$0.0993_{.0248}$	$\underline{0.1100}_{.0144}$	$\underline{0.1100}_{.0065}$	$\mathbf{0.1312}_{.0094}$

resolution spatial modeling, and whole-slide generative modeling, respectively. These methods use matched gene panels, expression preprocessing, and patient-level splits, with checkpoint selection based on the corresponding validation data. We additionally evaluate STPath (Huang et al., 2025b) using its released pretrained checkpoint without fine-tuning. Implementation details of GLoDe are provided in Appendix B.

4.2 Results

GLoDe demonstrates strong spatial gene expression prediction performance across diverse cancer cohorts. Table 1 presents the MF and OOF PCC of GLoDe and competing methods across ten cancer cohorts in HEST-1K. GLoDe achieves the highest average PCC under both evaluation settings, with relative improvements of 2.9% and 16.9% over the respective best-performing baselines, demonstrating its overall predictive advantage across multiple cancer cohorts. Qualitative comparisons of predicted and observed spatial expression patterns are provided in Appendix Figure 5. These results support the broad applicability of GLoDe across diverse tissue morphologies and molecular expression patterns.

GLoDe also demonstrates strong cross-patient generalization. Table 2 presents prediction results for gene panels comprising the top 50, 100, and 250 highly variable genes under leave-one-patient-out cross-validation on HER2ST. GLoDe outperforms the compared methods in both MF and OOF PCC across all panel sizes, indicating its ability to predict spatial gene expression in unseen patients while maintaining a consistent performance advantage as the number of target genes increases. Together with the HEST-1K results, these findings support the broad applicability of GLoDe across cancer cohorts and its cross-patient generalization ability.

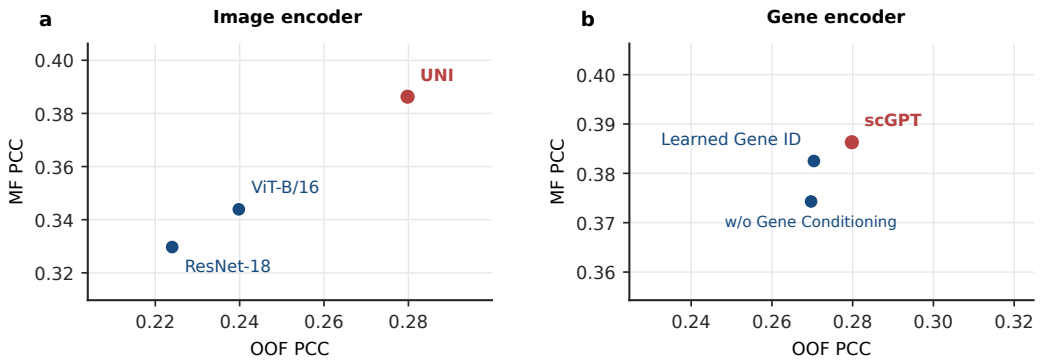


Figure 3: Ablation study of the image encoder and gene encoder.

GLoDe demonstrates good accuracy in predicting the spatial expression of disease-associated biomarkers. We further evaluate GLoDe on five representative biomarkers across eight cancer–gene settings. As shown in Appendix Table 8, GLoDe achieves the highest Macro PCC in seven of the eight settings, with particularly notable improvements for CD8A and VWF. These results demonstrate its ability to predict representative molecular markers across different cancer cohorts.

4.3 Further Analysis

We further examine the contributions of global–local decomposition, pretrained encoders, and GCV features through ablation studies and representation analysis.

Table 3: Ablation Study on Global–Local Decomposition.

Model Configuration	Global Branch	Decomposition Supervision	MF PCC	OOF PCC
w/o Global Branch	×	×	0.376	0.244
w/o Supervision	✓	×	0.383	0.264
GLoDe	✓	✓	**0.386**	**0.280**

Ablation Study on Global–Local Decomposition. To test our central hypothesis that explicitly separating a slide-shared baseline from spot-specific residuals benefits expression prediction, we examine the contributions of the global–local architecture and decomposition supervision through two ablated configurations. The first configuration, w/o Global Branch, removes the global branch and decomposition supervision, directly predicting absolute expression with the same gene-conditioned predictor. The second configuration, w/o Supervision, retains both branches but removes slide-mean supervision and residual centering, using only the final expression prediction loss. The comparison results are presented in Table 3. Adding the global branch improves both MF and OOF PCC, while introducing decomposition supervision further improves both metrics. These results show that the global–local architecture and decomposition supervision both contribute to improved prediction performance.

Ablation Study on Pretrained Encoders. We examine the effects of image and gene encoder selection. For image encoding, we compare ImageNet-1K-pretrained ResNet18 and ViT-B/16 with the pathology-pretrained UNI. As shown in Figure 3(a), UNI achieves the highest MF and OOF PCC, supporting its use as the default image encoder. For gene encoding, we compare no gene conditioning, randomly initialized learnable gene-ID embeddings, and frozen scGPT embeddings within GCRD. Figure 3(b) shows that scGPT achieves the highest MF and OOF PCC, exceeding learnable gene-ID embeddings by 0.0038 and 0.0094,

respectively. These results support the use of pretrained gene embeddings as molecular priors in GCRD.

Table 4: Gene-Specific Representation Analysis.

Expression Model	MF PCC		OOF PCC	
	Shared Feature	GCV Feature	Shared Feature	GCV Feature
w/o Global Branch	0.371	0.375	0.237	0.244
GLoDe	0.379	**0.386**	0.274	**0.280**

Effect of Gene-Conditioned Visual Features. To examine the contribution of gene-specific morphological information to expression prediction, we compare two feature construction strategies: Shared Feature, which averages the cross-attention output $M_{l,s}$ along the gene dimension before subsequent fusion for each gene; and GCV Feature, which retains the morphological features associated with individual genes and fuses them separately. We compare these strategies under global–local decomposition (GLoDe) and without the global branch (w/o Global Branch), keeping the remaining architecture unchanged. As shown in Table 4, GCV features improve both MF PCC and OOF PCC in both settings. These results support the contribution of gene-specific morphological information to expression prediction and its value as a complementary component for within-slide residual modeling.

We further analyze the learned representations of GCRD. As shown in Appendix Table 9, GCV features better capture gene-associated within-slide expression variation than shared spot representations.

5 Conclusion

In this work, we proposed GLoDe, a framework that decomposes gene expression into a slide-shared molecular baseline and spot-specific residuals, explicitly accounting for their distinct sources of variation. Furthermore, we incorporate gene-conditioned visual representations as a complementary component for within-slide residual modeling, capturing gene-specific local morphological cues and their associated expression variation. Experiments on two datasets and biomarker prediction demonstrate that GLoDe achieves strong predictive performance and cross-patient generalization across different cancer cohorts and gene panels, providing an effective modeling approach for histology-based spatial gene expression prediction.

Limitations. Our experiments demonstrate the effectiveness of global–local decomposition for expression prediction, but its practical value in clinical pathology assessment requires further validation in independent clinical cohorts and studies targeting specific clinical tasks. Moreover, our experiments focus on predefined target-gene panels and do not evaluate prediction for genes unseen during training. Extending gene coverage through gene priors is a promising direction for future work. We hope that our empirical findings help researchers understand how between-slide expression differences affect spatial expression prediction and inform future research.

References

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.

Julia Chen, Ludvig Larsson, Alexander Swarbrick, and Joakim Lundeberg. Spatial landscapes of cancers: insights and opportunities. *Nature reviews Clinical oncology*, 21(9): 660–674, 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.

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

Youngmin Chung, Ji Hun Ha, Kyeong Chan Im, and Joo Sang Lee. Accurate spatial gene expression prediction by integrating multi-resolution features. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11591–11600. IEEE, 2024.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024.

Thao M Dang, Haiqing Li, Yuzhi Guo, Hehuan Ma, Feng Jiang, Yuwei Miao, Qifeng Zhou, Jean Gao, and Junzhou Huang. Hage: hierarchical alignment gene-enhanced pathology representation learning with spatial transcriptomics. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 228–238. Springer, 2025.

Muhammad Dawood, Kim Branson, Nasir M Rajpoot, and Fayyaz ul Amir Afsar Minhas. All you need is color: image based spatial gene expression prediction using neural stain learning. In *Joint European Conference on machine learning and knowledge discovery in databases*, pp. 437–450. Springer, 2021.

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.

Tinglin Huang, Tianyu Liu, Mehrtash Babadi, Wengong Jin, and Rex Ying. Scalable generation of spatial transcriptomics from histology images via whole-slide flow matching. *arXiv preprint arXiv:2506.05361*, 2025a.

Tinglin Huang, Tianyu Liu, Mehrtash Babadi, Rex Ying, and Wengong Jin. Stpath: a generative foundation model for integrating spatial transcriptomics and whole-slide images. *NPJ Digital Medicine*, 8(1):659, 2025b.

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.

Guohao Liang, Hong Yin, and Fangyuan Ding. Technical advances and applications of spatial transcriptomics. *GEN biotechnology*, 2(5):384–398, 2023.

Ziang Liu, Xinhai Chen, Yigui Feng, Shuai Li, Qingyang Zhang, and Jie Liu. Histogpa: A context-conditioned gene-prior attention framework for histology-based spatial gene expression prediction. *arXiv preprint arXiv:2607.24364*, 2026.

Taku Monjo, Masaru Koido, Satoi Nagasawa, Yutaka Suzuki, and Yoichiro Kamatani. Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. *Scientific reports*, 12(1):4133, 2022.

Kalin Nonchev, Sebastian Dawo, Karina Silina, Holger Moch, Sonali Andani, Tumor Profiler Consortium, Viktor H Koelzer, and Gunnar Rätsch. Deepspot: Leveraging spatial context for enhanced spatial transcriptomics prediction from h&e images. *MedRxiv*, pp. 2025–02, 2025.

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.

Zhiceng Shi, Changmiao Wang, Jun Wan, and Wenwen Min. Cross-slice knowledge transfer via masked multi-modal heterogeneous graph contrastive learning for spatial gene expression inference. *arXiv preprint arXiv:2603.22821*, 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.

Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

Changfan Wang, Xinran Wang, Donghai Liu, Fei Su, Lulu Sun, Zhicheng Zhao, and Zhu Meng. Chrep: Cross-modal histology representation and post-hoc calibration for spatial gene expression prediction. *arXiv preprint arXiv:2604.21573*, 2026.

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.

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: 70626–70637, 2023.

Ying Xiong, Linjing Liu, Yufei Cui, Shangyu Wu, Xue Liu, Antoni B Chan, and Chun Jason Xue. Genequery: A general qa-based framework for spatial gene expression predictions from histology images. *arXiv preprint arXiv:2411.18391*, 2024.

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, et al. A whole-slide foundation model for digital pathology from real-world data. *Nature*, 630(8015):181–188, 2024.

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.

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 Datasets

A.1 Dataset Statistics

Table 5 summarizes the datasets used in our experiments.

Table 5: Dataset and cohort statistics used in our experiments.

Dataset	Cancer cohort	Slides	Spots	Platform
HEST-1K	CCRCC	24	74,220	Visium
	COAD	4	15,651	Xenium
	HCC	2	4,196	Visium
	IDC	4	35,536	Xenium
	LUNG	2	5,206	Xenium
	LYMPH-IDC	4	19,964	Visium
	PAAD	3	7,571	Xenium
	PRAD	23	62,710	Visium
	READ	4	8,407	Visium
	SKCM	2	3,034	Xenium
	Subtotal	**72**	**236,495**	–
HER2ST	HER2+ breast cancer	36	13,620	ST
Total		**108**	**250,115**	–

A.2 Preprocessing and Gene Selection

HEST-Bench. We use the official fold-specific panels of 50 target genes provided by HEST-Bench without additional gene selection. Raw expression counts are transformed using $\log(1 + x)$ to obtain the prediction targets.

HER2ST. We select the top 250 highly variable genes as the target gene panel. For each spot, expression counts within this panel are normalized to a total of 10^4 and subsequently transformed using $\log_{10}(1 + x)$ to obtain the prediction targets. The gene symbols included in the Top-50, Top-100, and Top-250 panels are provided in Table 6.

B Implementation

Running Environment. All experiments were conducted on a Linux server equipped with two NVIDIA A800 GPUs, each with 80 GB of memory. Each training run used a single GPU, with independent runs executed in parallel across the two GPUs. Our implementation used Python 3.10, PyTorch 2.2.2, CUDA 12.1, and cuDNN 8.9.2.

Feature Extraction. For HEST-1K, we use the provided spot-aligned RGB patches of size 224×224, normalized using ImageNet statistics. A frozen UNI-v1 encoder (ViT-L/16) extracts 1024-dimensional CLS features and 196×1024 patch-token features.

For each spot, we cache the original image and three augmented views generated using random rotations, horizontal and vertical flips, and color perturbations of up to 10%. During training, one view is randomly selected per batch, while validation and testing use only the original view.

Gene embeddings are obtained from pretrained scGPT through gene-symbol matching and remain frozen during training. The 512-dimensional embeddings are projected into the shared 128-dimensional latent space, with trainable fallback queries used for genes absent from the scGPT vocabulary.

Table 6: HER2ST descriptive gene panels. The Top-100 panel comprises the Top-50 genes and 50 additional genes; the Top-250 panel comprises the Top-100 genes and 150 additional genes.

Panel	Gene symbols
Top-50	CORO2A, TJP3, LARP7, C19orf52, AGPAT2, ANGPTL4, RHOB, ZNF787, KMT2E, TFF3, IGLC7, ANKRD12, NR4A1, IFI44L, SFSWAP, OASL, MRPL21, AGR2, LUC7L3, PNISR, QSOX1, SEC14L2, GATA3, CCL19, DDB2, GPX3, RAB11FIP1, SCD, JCHAIN, SOX4, FARP1, ISG15, ARID5A, CCL2, IER5L, LTF, GSTM3, HELZ2, CCL21, KRT86, LINC00176, AQP1, PPP1R1B, ARID1B, S100A9, CCDC80, GBP5, HAGHL, RNF170, EXOSC10.
Top-100 (additional genes)	KRT17, GBP2, POSTN, GPRC5A, FLNA, ADAM15, PTGDS, SPSB1, OSBPL8, CXCL10, CLASRP, FARSB, INAFM1, CADM4, DLG1, DOCK6, SPAG9, ILKAP, OPTN, GBP1, HOOK3, TNNT1, ZNF91, IGKC, IGLC2, IGLC3, C5orf38, FAH, TRIM22, PTPRC, CD3D, CHCHD6, PLSCR1, STK17A, ELN, AKAP9, ARID3A, MUC1, GLTSCR2, R3HDM2, NFATC4, MID1IP1, MUCL1, TRAP1, ITGB4, NEBL, CCDC88B, FGFR1, CHD6, THBS1.
Top-250 (additional genes)	NFIC, GMFG, MFAP4, EFHD2, COL3A1, IGHA1, CD74, CRABP2, NIT2, HLA-DPB1, VWF, COL4A1, DPYSL3, LUM, NCKAP1L, IGHG4, ABHD14A, SLC43A1, CARS2, IGFBP4, SLC2A1, NELFE, CX3CL1, WDFY3, GNL2, CRACR2B, SLPI, XPO6, YIPF4, SMCHD1, UXT, MICAL1, A2M, KIAA1468, PITX1, CES2, FNBP4, STAT5A, UFD1L, PIH1D1, MMP9, PHF21A, CCT4, NDUFB9, ZNF83, MZT2A, MED25, TAX1BP1, NRP1, EXD3, FN1, WDR13, DPP9, SAMM50, ECHDC2, TNC, PWP1, CALML5, RAMP2, KDM4B, TXNDC17, TCEA2, MPND, PLK2, RAB3IL1, TBXAS1, SDHB, YTHDC1, MICALL2, RSRP1, LST1, CXXC5, AGO2, INTS8, PPIH, EIF3B, JAK3, S100A14, DNAJC7, MAVS, METTL16, PER1, DNAJC1, CTIF, ATF6B, TCEA3, PTRF, JAK1, PDCD5, LRRC8A, UBR2, HCST, SNF8, ATG12, TCF25, PEX14, ZNF644, LMF1, ZFP36, WIPF1, MPST, ZG16B, HYI, TBCA, ZMYND8, NR1H2, IDUA, C10orf10, NDUFB3, DENND1A, CLU, PPP1R1A, SURF1, FAM118A, C12orf45, VAV1, NDRG1, SOD3, PROSER1, GYPC, CHD1, CERS5, MAP3K6, IFI27, GAS6, MAP4K2, SRPX2, ARID5B, TEX10, LGALS1, ARHGEF40, NUP214, MLLT4, SPARC, ITGB6, PNMT, FUS, VARS, GLG1, TMEM101, HSPB1, ECM1, MAP3K14, BST2, TNFRSF4, AKR1A1, CFDP1, TIGD5, SLC25A24, MYO1G.

Model Configuration. GLoDe uses a shared latent dimension of $D = 128$. The spatial encoder consists of two attention layers with four heads each, operating on an 8-nearest-neighbor graph. The resulting spot representations are averaged across each slide and mapped to gene-wise molecular baselines through an MLP.

The gene-conditioned residual branch employs an eight-head gene-to-patch cross-attention layer and a 128-dimensional adaptive gate to integrate spatial context. A parameter-shared token refinement network and decoder predict gene-specific residual expression. Dropout is set to 0.3 throughout the attention and feed-forward modules.

Training and Inference. We train GLoDe using AdamW with a learning rate of 5×10^{-5} and a weight decay of 10^{-4} for up to 30 epochs. Mixed-precision training is employed, with gradient clipping at a maximum norm of 1.0. The training batch sizes are set to 64 and 32 for HEST-Bench and HER2ST, respectively.

During training, GLoDe combines full-slide spatial encoding with center-spot local prediction. For each batch, the spatial encoder processes all CLS features and the complete spatial graph of the current slide to capture slide-level context and predict the global molecular

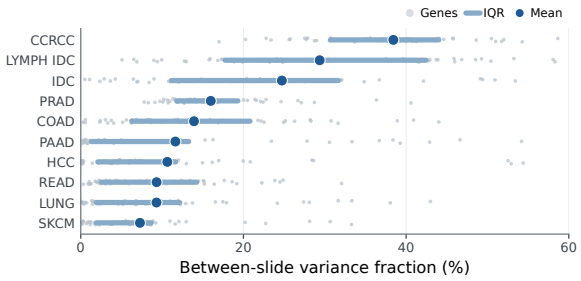


Figure 4: Gene-wise between-slide variance fractions across cancer cohorts.

Cancer cohort	Slides	Mean ρ_g (%)
CCRCC	24	38.43
LYMPH-IDC	4	29.38
IDC	4	24.73
PRAD	23	15.99
COAD	4	13.92
PAAD	3	11.63
HCC	2	10.64
READ	4	9.32
LUNG	2	9.31
SKCM	2	7.28

Table 7: Gene-averaged between-slide variance fractions across HEST-Bench cohorts.

baseline. The local branch performs patch-token cross-attention and residual prediction only for the selected center spots, while neighboring spots contribute through their CLS features in the spatial encoder.

The expression reconstruction loss is computed over the selected center spots, while baseline supervision uses the ground-truth mean expression across the entire training slide. The residual centering loss is computed from the mean predicted residual of the current center-spot batch rather than the full slide. The weights of the three loss terms are set to 1, 0.1, and 0.01, respectively.

For folds with validation slides, we evaluate pooled gene-wise PCC after each epoch and select the checkpoint with the highest mean PCC across genes, using early stopping with a patience of eight epochs. For HEST-Bench folds without validation slides, we train for all 30 epochs and use the final checkpoint for testing. Test data are never used for checkpoint selection.

During inference, the spatial encoder and global branch are evaluated once per slide to obtain the slide-specific molecular baseline. The local branch then predicts spot-specific expression residuals in batches, which are added to the slide-level baseline to obtain the final expression predictions. The decoder residuals are used directly without additional slide-wise mean subtraction.

C Additional Results and Analyses

C.1 Between-Slide Expression Variation

We quantify between-slide expression variation to further examine the motivation for global–local decomposition. For a cohort containing L slides, let $N = \sum_{l=1}^{L} N_l$ denote the total number of spots. For a given gene, the slide-wise and cohort-wide mean expression levels are defined as

$$\bar{y}_l = \frac{1}{N_l} \sum_{s=1}^{N_l} y_{l,s}, \qquad \bar{y} = \frac{1}{N} \sum_{l=1}^{L} N_l \bar{y}_l. \tag{16}$$

We define the between-slide and within-slide variances as

$$V_{\text{between}} = \frac{1}{N} \sum_{l=1}^{L} N_l (\bar{y}_l - \bar{y})^2, \tag{17}$$

$$V_{\text{within}} = \frac{1}{N} \sum_{l=1}^{L} \sum_{s=1}^{N_l} (y_{l,s} - \bar{y}_l)^2. \tag{18}$$

Table 8: Biomarker prediction performance across cancer cohorts, measured by Macro PCC. Standard deviations are shown as subscripts.

Cancer	Gene	MF PCC			
		BLEEP	STFlow	TRIPLEX	GLoDe
IDC	ERBB2	$0.662_{.032}$	$\underline{0.852}_{.008}$	$0.738_{.036}$	$\mathbf{0.855}_{.005}$
	ESR1	$0.564_{.062}$	$0.612_{.015}$	$\mathbf{0.647}_{.010}$	$\underline{0.637}_{.022}$
	VWF	$0.259_{.011}$	$\underline{0.581}_{.030}$	$0.306_{.027}$	$\mathbf{0.665}_{.008}$
LUNG	MKI67	$0.696_{.044}$	$0.714_{.027}$	$\underline{0.768}_{.008}$	$\mathbf{0.785}_{.002}$
PAAD	ESR1	$0.230_{.035}$	$\underline{0.299}_{.043}$	$0.254_{.030}$	$\mathbf{0.319}_{.011}$
	MKI67	$0.419_{.054}$	$\underline{0.508}_{.048}$	$0.329_{.043}$	$\mathbf{0.528}_{.031}$
	CD8A	$0.518_{.007}$	$\underline{0.676}_{.005}$	$0.524_{.024}$	$\mathbf{0.698}_{.007}$
SKCM	VWF	$0.408_{.017}$	$\underline{0.522}_{.038}$	$0.507_{.034}$	$\mathbf{0.746}_{.015}$

We further define the fraction of total variance attributable to between-slide variation as

$$\rho = \frac{V_{\text{between}}}{V_{\text{between}} + V_{\text{within}}}. \tag{19}$$

Figure 4 presents the distribution of gene-wise between-slide variance fractions across ten cancer cohorts, and Table 7 reports the corresponding gene-averaged values. The mean between-slide variance fraction ranges from 7.28% to 38.43%, indicating that the contribution of slide-level expression differences to total expression variation varies across cancer cohorts.

These results indicate that between-slide differences account for a non-negligible proportion of total expression variation, motivating the explicit modeling of slide-shared molecular baselines alongside spot-specific residuals.

C.2 Biomarker Prediction

We provide detailed biomarker prediction results across eight cancer–gene settings involving five representative genes: ERBB2, ESR1, MKI67, CD8A, and VWF. Table 8 reports the gene-wise Macro PCC of GLoDe and the compared methods across different cancer cohorts.

C.3 Gene-Specific Representation Analysis.

We further examine whether GCV features capture within-slide expression variation. Specifically, we compare the shared context-aware spot representation $h_{l,s}$ with the GCV feature $z_{l,s}$. We quantify the proportion of observed expression residual variance explained by representation-derived clusters using η^2. For each slide–gene pair, we independently cluster the representations using K-means with $K = 3$, applying the same clustering configuration to both Shared and GCV features.

For each slide–gene case, we omit the gene index for simplicity and define $u_s = y_{l,s} - \bar{y}_l$ as the observed expression residual at spot s. Let $\mathcal{C}_1, \ldots, \mathcal{C}_K$ denote the resulting clusters. We compute

$$\eta^2 = \frac{\sum_{k=1}^{K} |\mathcal{C}_k| \left(\bar{u}_k - \bar{u}\right)^2}{\sum_{s=1}^{N_l} \left(u_s - \bar{u}\right)^2}, \tag{20}$$

where $y_{l,s}$ denotes the observed expression of the selected gene, $\bar{y}_l$ is its slide-wise mean, $\bar{u}_k$ is the mean residual within cluster $\mathcal{C}_k$, and $\bar{u}$ is the mean residual across all evaluated spots. A larger η^2 indicates that the representation-derived clusters explain a greater proportion of within-slide expression variation.



Figure 5: Spatial gene expression prediction in IDC Xenium samples. Columns compare the ground truth with predictions from GLoDe and baseline methods. Subfigure titles report PCC between observed and predicted expression.

C.4 Spatial Expression Visualization

We visualize predicted gene expression in IDC Xenium samples to examine the spatial expression patterns underlying the quantitative results. Figure 5 compares the observed expression of ERBB2, MYH11, VWF, and CEACAM6 with predictions from GLoDe and other baseline methods.

Table 9: Gene-Specific Representation Analysis.

Cancer	**Gene**	**Shared Feature η^2**	**GCV Feature η^2**
IDC	ERBB2	0.547	**0.678**
	ESR1	0.436	**0.452**
LUNG	MKI67	**0.579**	0.552
PAAD	CD8A	0.225	**0.409**
	ESR1	0.107	**0.148**
	MKI67	0.363	**0.381**
Overall	–	0.378	**0.444**