RESOLVE: RESIDUAL DECOMPOSITION WITH SINGLE-CELL PROTOTYPE GUIDANCE FOR SPATIAL GENE EXPRESSION PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting spatial gene expression from H&E-stained histology offers a promising low-cost alternative for reconstructing spatial transcriptomic maps. However, existing methods typically model all expression variation jointly, conflating stable morphology-associated patterns with slide-specific cell-state deviations and thereby limiting the recovery of signals that are not directly explained by histology. To this end, we propose RESOLVE, which reformulates spatial gene expression prediction as two complementary processes, a morphology-driven base prediction and a cell-state-driven residual correction. A lightweight base predictor learns the morphology-expression mapping shared across slides, capturing stable tissue-level expression patterns. A prototype-guided residual pathway builds cell-state prototypes with explicit biological meaning from a single-cell reference and uses prototype routing to correct the slide-specific deviations that the base prediction cannot adequately explain. A per-spot gate modulates the residual magnitude according to the local tissue state, and an orthogonality constraint encourages the two pathways to learn complementary and decoupled representations, explicitly separating stable morphological signals from slide-specific cell-state changes. Experiments on three public spatial transcriptomics datasets show that RESOLVE attains the best results on Her2ST and the best PCC on ST-Net and SCC. The gains of the residual pathway are interpretable. Genes that benefit most improve mainly by correcting slide-level differences and concentrate in interferon-α/γ and estrogen response programs. This pattern indicates that the residual corrects cell-state deviations rather than re-fitting what morphology already states.

1 INTRODUCTION

Spatial transcriptomics (ST) measures gene expression while preserving tissue spatial information, providing an important tool for dissecting tissue organization, cell states, and the microenvironment (Ståhl et al., 2016; Armingol et al., 2021; Zhang et al., 2022). However, its high experimental cost and technical barrier limit large-scale application. In contrast, H&E-stained slides are widely available in pathology research and clinical practice, making direct prediction of spatial gene expression from histology images an increasingly attractive alternative (He et al., 2020; Pang et al., 2021; Chung et al., 2024; Liu et al., 2024).

To address this task, existing research mainly improves spatial gene expression prediction by enriching the information available to histology models. Early methods directly regress gene expression at spatial spots from local image patches (He et al., 2020). Subsequent work gradually incorporates neighborhood information and multi-scale context to better characterize local tissue organization (Pang et al., 2021; Chung et al., 2024; Liu et al., 2024; Zhang et al., 2026a), and further employs graphs or hypergraphs to model broader spatial dependencies (Ganguly et al., 2025; Qu et al., 2026). Beyond spatial modeling, another line learns more complete conditional expression distributions through generative modeling to characterize expression uncertainty, or introduces single-cell reference data and other biological priors to characterize cell-state variability (Huang et al., 2025; Zhu et al., 2025; Nishimura et al., 2026; Fang et al., 2026; Li et al., 2025; Zhang et al., 2026b). Despite these advances in visual representation, spatial context modeling, and external biological priors, their

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.

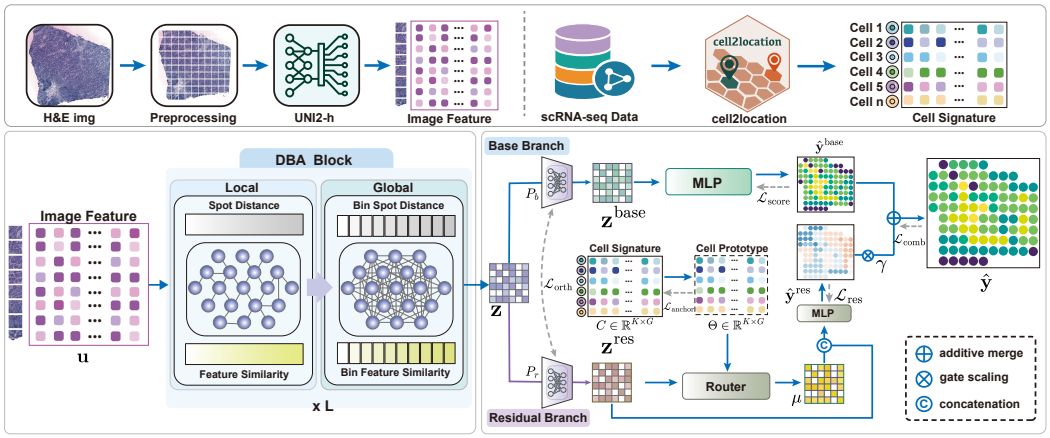

Figure 1: Overview of RESOLVE. A frozen UNI2-h encoder supplies spot features and cell2location supplies cell-type prototypes from a matched scRNA-seq reference. Both feed the DBA encoder, which drives a base branch and a prototype-routed residual branch.

basic prediction framework remains consistent: given a histology representation, directly predict the full gene expression profile at each spatial location.

However, such prediction frameworks implicitly assume that expression variation from different sources can be uniformly modeled by a single prediction mapping, without distinguishing their varying degrees of correspondence with tissue morphology. Prior work has shown that some spatial expression patterns are strongly associated with tissue morphology and local tissue organization (Schmauch et al., 2020). Meanwhile, cell-state transitions, disease-related responses, and microenvironmental perturbations also produce additional expression changes that may not be fully identifiable from H&E morphology (Barkley et al., 2022; Xie et al., 2026). When these two types of signals are compressed into a single prediction target, the model tends to prioritize fitting stable and morphologically predictable expression components, while state-related residual changes lack independent representation and modeling mechanisms. Incorporating spatial context or introducing single-cell priors can provide more information for prediction, but does not change this unified regression modeling approach. Therefore, a key limitation of current methods is not merely insufficient information, but the lack of structured decomposition of the prediction target: morphology-explainable expression components and slide-specific state-related deviations are still jointly modeled by the same pathway, limiting the model's ability to characterize the latter.

To this end, we propose RESOLVE, a framework that reformulates the modeling of gene expression prediction through expression decomposition (see Figure 1). The prediction target is decomposed into two parts: a tissue-wide, morphology-correlated expression pattern directly modeled by a lightweight base predictor, and a slide-level cell-state-specific deviation captured by a prototype-guided residual pathway. The residual pathway is conditioned on interpretable cell-state prototypes constructed from a single-cell atlas of the same tissue type (Li et al., 2025; Nishimura et al., 2026; Fang et al., 2026), and its contribution is scaled by per-spot learned weights, with an orthogonal constraint separating the representations of the two pathways. Through this decomposition, RESOLVE preserves the morphology-predictable component while providing an independent, interpretable modeling space for slide-level cell-state deviations.

Our main contributions are as follows:

- We propose an expression decomposition framework for gene expression prediction from histology images, separating the prediction target into a morphology-anchored base component and a residual component carrying slide-level cell-state-specific deviations.
- We design a prototype-guided residual pathway that conditions residual prediction on interpretable cell states constructed from a single-cell reference, complemented by per-spot gating that scales the residual contribution.

- On three public ST datasets, RESOLVE achieves seven best results: it is best on all metrics on Her2ST and ST-Net, and achieves the highest PCC on SCC.

2 METHODOLOGY

2.1 PROBLEM FORMULATION

Let $\mathcal{R}$ denote the set of real spots in a tissue slide. Each spot i corresponds to a histology image patch, from which we extract a feature vector $\mathbf{x}_i \in \mathbb{R}^D$ using a frozen pre-trained pathology foundation model, along with spatial coordinates $\mathbf{p}_i \in \mathbb{R}^2$ and a gene expression target $\mathbf{y}_i \in \mathbb{R}^G$, where D and G are the feature and gene dimensions. The task is to learn a mapping from $\mathbf{x}_i$ to $\mathbf{y}_i$. In addition, a matched single-cell reference provides a cell-type signature matrix, used only by the residual pathway (Section 2.3).

Existing approaches treat this task as deterministic regression, learning an end-to-end mapping $\mathbf{y}_i \approx f(\mathbf{x}_i)$. This regression formulation keeps the prediction target as a single output, leaving slide-level cell-state-specific deviations entangled with the tissue-wide expression pattern and without independent representation and modeling pathways.

RESOLVE instead reformulates the task through expression decomposition. For each spot i, the final prediction is

$$\hat{\mathbf{y}}_i = \hat{\mathbf{y}}_i^{\text{base}} + \gamma_i \hat{\mathbf{y}}_i^{\text{res}}, \tag{1}$$

where $\hat{\mathbf{y}}_i^{\text{base}}$ captures the tissue-wide expression pattern directly predictable from morphology, $\hat{\mathbf{y}}_i^{\text{res}}$ captures the slide-level cell-state-specific deviation, and $\gamma_i \in (0, 1)$ is a learned per-spot weight that scales the correction (Section 2.3). This decomposition assumes that spot expression can be separated into two components: one determined by local tissue context, and one determined by cell states shared across the slide, with the latter benefiting from separate modeling with single-cell prior information.

Three components instantiate this decomposition: a hierarchical DBA encoder that produces the shared context representation (Section 2.2), a two-pathway predictor that operates on it (Section 2.3), and a set of training objectives that supervise the two pathways (Section 2.4). Figure 1 shows the overall architecture, and Appendix B.2 summarizes the architectural dimensions and all hyperparameters.

2.2 HIERARCHICAL DBA ENCODER

The core of RESOLVE is to decompose the prediction target into a morphology-anchored base component and a slide-level cell-state deviation, and this decomposition requires both pathways to share a representation that jointly encodes local and global tissue context. However, gene expression at a spot is shaped by the local microenvironment as well as the global tissue architecture, and single-scale encoding cannot cover both. We therefore design **Dual-Bias Attention** (DBA) blocks, which build this shared representation through stacked local and global attention stages. Each block applies two bias terms to every attention logit, one for spatial distance and one for morphological feature similarity, so that attention is shaped by both the geometry and the morphology of the tissue rather than by content similarity alone. Figure 2 illustrates its internal structure.

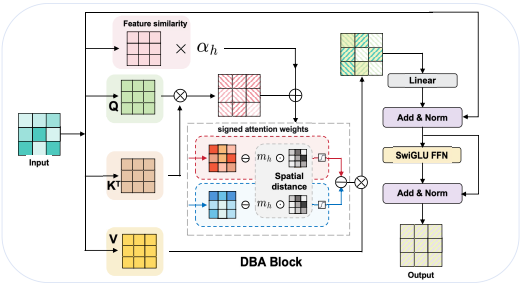

Figure 2: Architecture of the DBA block.

Each image feature $\mathbf{x}_i \in \mathbb{R}^D$ is first projected to the model dimension d through a two-layer MLP with feature dropout:

$$\mathbf{u}_i^{(0)} = W_2 \cdot \text{GELU}(W_1 \cdot \text{drop}(\mathbf{x}_i)) \in \mathbb{R}^d, \tag{2}$$

where $\text{GELU}(\cdot)$ is the activation and $\text{drop}(\cdot)$ denotes feature dropout.

Local DBA (L-DBA). The local microenvironment consists only of spatially adjacent spots, so L-DBA attends over the k spatial nearest neighbors $\mathcal{N}_k(i)$ of each spot i. However, spatial distance alone cannot distinguish neighbors that are morphologically similar but functionally different, so each logit additionally incorporates a learnable feature-similarity bias. For attention head h:

$$s_{ij}^{(h)} = \frac{\langle \mathbf{q}_i^{(h)}, \mathbf{k}_j^{(h)} \rangle}{\sqrt{d_h}} + \alpha_h \cdot \cos(\mathbf{u}_i, \mathbf{u}_j), \quad j \in \mathcal{N}_k(i), \tag{3}$$

where α_h is a per-head learnable gain initialized to zero, and the cosine term captures feature-level affinity not already expressed by spatial proximity.

Biological influence decays with spatial distance, so we further introduce a fixed ALiBi-style distance penalty (Press et al., 2021) $m_h \cdot \|\mathbf{p}_i - \mathbf{p}_j\|_2$, with head-specific slopes m_h that model this decay. Standard softmax attention, however, only produces non-negative weights and therefore cannot express inhibitory interactions, which are prevalent in the tissue microenvironment. To allow attention weights to be signed, we compute the final weight as

$$w_{ij}^{(h)} = \text{softmax}_j\left(s_{ij}^{(h)} - m_h d_{ij}\right) - \beta \cdot \text{softmax}_j\left(\frac{-s_{ij}^{(h)} - m_h d_{ij}}{\tau}\right), \tag{4}$$

where β and τ are fixed hyperparameters. The second term inverts the logit, so a spot whose features are anti-correlated with the query receives a negative contribution rather than merely a small positive one.

The attention output is aggregated over neighbors and passed through out-projection with a residual connection and LayerNorm, followed by a SwiGLU feed-forward network (Shazeer, 2020) that applies a nonlinear transformation at each spot:

$$\text{SwiGLU}(\mathbf{z}) = W_{\text{down}}(\text{SiLU}(W_{\text{gate}}\mathbf{z}) \odot W_{\text{up}}\mathbf{z}). \tag{5}$$

Global DBA (G-DBA). L-DBA is confined to spatial neighborhoods and cannot capture morphological similarities between spatially distant regions (e.g., scattered tumor foci), so G-DBA attends in a fully connected manner over all real spots $\mathcal{R}$. Fully connected attention lacks locality constraints, and directly using content scores for the bias would make it redundant with the scores it should regularize, so we follow Graphormer (Ying et al., 2021) and parameterize the bias as learnable **spatial and feature binning tables**:

$$s_{ij}^{(h)} = \frac{\langle \mathbf{q}_i^{(h)}, \mathbf{k}_j^{(h)} \rangle}{\sqrt{d_h}} + \mathbf{e}_{\text{sp}}^{(h)}[b_d(d_{ij})] + \mathbf{e}_{\text{ft}}^{(h)}[b_s(\cos(\mathbf{x}_i, \mathbf{x}_j))], \tag{6}$$

where $b_d(\cdot)$ and $b_s(\cdot)$ discretize pairwise distance and the cosine similarity of the raw UNI features into bins, respectively, and $\mathbf{e}_{\text{sp}}$ and $\mathbf{e}_{\text{ft}}$ are learned bias tables, with the spatial table warm-started from ALiBi-style values. Unlike L-DBA, G-DBA anchors feature affinity to the frozen UNI features, as its fully connected attention lacks the locality constraint that already regularizes L-DBA. The signed attention weights introduced in L-DBA are also used here, but different heads differ in their sensitivity to inhibitory relationships, so we adopt **per-head learnable** β_h and τ_h. The number of bins and their boundaries are listed in Appendix B.2.

After L blocks, each consisting of one L-DBA and one G-DBA stage, the hierarchical DBA encoder outputs a refined context feature $\mathbf{z}_i \in \mathbb{R}^d$ for each real spot, shared by the two pathways.

2.3 Two-Pathway Predictor

2.3.1 Feature Disentanglement Projection

The DBA encoder outputs a shared context feature $\mathbf{z}_i$ for each spot, and the two pathways of RESOLVE need to model a morphology-anchored base expression and a slide-level cell-state deviation on top of it. If both pathways directly share the same representation, the base prediction dominates the shared space and the residual pathway struggles to learn meaningful corrections. We therefore separate $\mathbf{z}_i$ through two independent linear projections:

$$\mathbf{z}_i^{\text{base}} = P_b \mathbf{z}_i, \qquad \mathbf{z}_i^{\text{res}} = P_r \mathbf{z}_i. \tag{7}$$

To reduce geometric redundancy between the two projected subspaces, we impose an orthogonality constraint on the cosine similarity between the projected features:

$$\mathcal{L}_{\text{orth}} = \frac{1}{|\mathcal{R}|} \sum_{i \in \mathcal{R}} \left(\frac{\langle \mathbf{z}_i^{\text{base}}, \mathbf{z}_i^{\text{res}} \rangle}{\|\mathbf{z}_i^{\text{base}}\| \|\mathbf{z}_i^{\text{res}}\| + \epsilon} \right)^2, \tag{8}$$

where ϵ is a small constant for numerical stability. This constraint only reduces geometric redundancy between the two subspaces. Biological non-overlap is further enforced by prototype conditioning and gradient isolation.

2.3.2 Base Predictor

The base pathway takes $\mathbf{z}_i^{\text{base}}$ as input and models the morphology-anchored, tissue-wide expression baseline. It is implemented as a lightweight MLP with deliberately constrained capacity:

$$\hat{\mathbf{y}}_i^{\text{base}} = f_b(\mathbf{z}_i^{\text{base}}). \tag{9}$$

The capacity constraint is intentional: the base pathway should fit only spatially smooth, tissue-wide expression patterns, not the slide-level deviations that belong to the residual pathway. This constraint alone, however, does not prevent the base pathway from fitting such deviations. The separation is enforced by gradient isolation in the residual loss and by the gating design (Section 2.3), while the capacity constraint only biases the base pathway toward smooth, tissue-wide patterns.

The base predictor is supervised on the full target $\mathbf{y}_i$ by a composite score loss with two terms. The first is a magnitude term, the mean squared error:

$$\mathcal{L}_{\text{base}} = \frac{1}{|\mathcal{R}|G} \sum_{i \in \mathcal{R}} \|\hat{\mathbf{y}}_i^{\text{base}} - \mathbf{y}_i\|_2^2. \tag{10}$$

However, downstream analyses of ST data, such as spatial domain identification and gene co-expression, rely mainly on the relative spatial pattern of each gene rather than its absolute magnitude, and MSE alone only weakly constrains this pattern. We therefore complement it with a scale-invariant pattern term that maximizes the per-gene Pearson correlation between base predictions and targets:

$$\mathcal{L}_{\text{corr}} = 1 - \frac{1}{G} \sum_{g=1}^{G} \rho_g\left(\hat{\mathbf{y}}_{\cdot g}^{\text{base}}, \mathbf{y}_{\cdot g}\right), \tag{11}$$

where $\rho_g(\cdot, \cdot)$ is the Pearson correlation computed across spots for gene g. The two terms are combined into the score loss of the base pathway:

$$\mathcal{L}_{\text{score}} = \mathcal{L}_{\text{base}} + \lambda_{\text{corr}} \mathcal{L}_{\text{corr}}, \tag{12}$$

where MSE is the primary term and the correlation term is auxiliary, so only the latter is weighted. This loss supervises both how strongly and in what spatial pattern each gene is expressed.

2.3.3 Prototype-Guided Residual Predictor

The residual pathway captures the slide-level, cell-state-specific deviations that the base prediction fails to cover. Learning such corrections from limited data is challenging: without guidance, the model is prone to fitting noise or producing biologically implausible corrections. To anchor residual learning toward biologically meaningful directions, we introduce a prototype routing mechanism that conditions residual prediction on interpretable cellular states derived from single-cell reference data.

Prototype construction. Let $C \in \mathbb{R}^{K \times G}$ denote the cell-type signature matrix inferred from single-cell references via cell2location (Kleshchevnikov et al., 2022), where K is the number of cell types (Appendix C.2 lists the reference datasets and cell-type labels). We maintain a learnable log-prototype matrix $\Theta \in \mathbb{R}^{K \times G}$ initialized as $\log C$. Since the single-cell reference already provides reliable cell-type signatures, Θ should only be fine-tuned rather than drift far from it, so we apply an anchoring loss:

$$\mathcal{L}_{\text{anchor}} = \frac{1}{KG} \|\log C - \Theta\|_2^2. \tag{13}$$

Routing. A spot is typically dominated by a small number of cell types, and allowing all types to participate uniformly would dilute the signal, so we compute a sparse routing vector over cell types for each spot i:

$$\boldsymbol{\pi}_i = \mathrm{TopK}_{K_\pi}\left(\mathrm{softmax}\left(\frac{W_r \mathbf{z}_i^{\mathrm{res}}}{\tau_\pi}\right)\right), \tag{14}$$

where TopK_{K_π} retains only the K_π most relevant cell types, enforcing sparsity and forcing the model to focus on a compact set of cellular states. The resulting biological condition vector is

$$\boldsymbol{\mu}_i = \boldsymbol{\kappa} \odot \left(\boldsymbol{\pi}_i^\top \exp(\Theta)\right) + \boldsymbol{\nu}, \tag{15}$$

where $\boldsymbol{\kappa}, \boldsymbol{\nu} \in \mathbb{R}^G$ are per-gene affine alignment parameters that adapt the prototype signatures to the target expression scale. This alignment is necessary because the single-cell reference and the ST data differ in technical scale.

Residual prediction. The residual MLP takes the concatenation of the residual feature and the biological condition vector:

$$\hat{\mathbf{y}}_i^{\mathrm{res}} = f_r([\mathbf{z}_i^{\mathrm{res}}; \boldsymbol{\mu}_i]). \tag{16}$$

The condition vector $\boldsymbol{\mu}_i$ does not directly provide expression values. It indicates which cellular state the spot is in, enabling the model to infer the systematic direction of base prediction error under that state. The output layer of the residual MLP applies no activation function, so $\hat{\mathbf{y}}_i^{\mathrm{res}}$ can take arbitrary real values and correct the base prediction in either direction.

The residual is trained against the detached target $\mathbf{y}_i - \mathrm{sg}(\hat{\mathbf{y}}_i^{\mathrm{base}})$, where $\mathrm{sg}(\cdot)$ denotes stop-gradient:

$$\mathcal{L}_{\mathrm{res}} = \frac{1}{|\mathcal{R}|G} \sum_{i \in \mathcal{R}} \|\hat{\mathbf{y}}_i^{\mathrm{res}} - (\mathbf{y}_i - \mathrm{sg}(\hat{\mathbf{y}}_i^{\mathrm{base}}))\|_2^2. \tag{17}$$

Without gradient isolation, the residual predictor would co-adapt with the base predictor, and the two could jointly reduce the total loss while breaking the semantics of the decomposition, in which the base pathway explains the morphology-interpretable component and the residual pathway explains the slide-level deviation. This loss ensures that the residual predictor focuses exclusively on correcting base prediction errors, while preventing the residual loss from influencing the base predictor. Propositions 1 and 2 formalize this isolation and characterize the optimal residual as the conditional mean of the base prediction error.

Finally, the learned residual scale need not match the scale of the true slide-level correction, so we scale it with a learned per-spot weight:

$$\gamma_i = \sigma\left(f_\gamma(\mathbf{z}_i^{\mathrm{base}})\right), \tag{18}$$

where σ is the sigmoid function, ensuring $\gamma_i \in (0, 1)$. The gate is read from the base feature because it should reflect how unreliable the base prediction is. If read from the residual feature, the gate would tend to amplify residuals that are already large rather than apply larger corrections where the base prediction is most uncertain. The residual-side representation is shaped only by $\mathcal{L}_{\mathrm{res}}$, and the direction of the correction is carried by the sign of $\hat{\mathbf{y}}_i^{\mathrm{res}}$. The weight is supervised by a combination loss on the final prediction, with the residual detached:

$$\mathcal{L}_{\mathrm{comb}} = \frac{1}{|\mathcal{R}|G} \sum_{i \in \mathcal{R}} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2^2, \tag{19}$$

where $\hat{\mathbf{y}}_i = \hat{\mathbf{y}}_i^{\mathrm{base}} + \gamma_i \cdot \mathrm{sg}(\hat{\mathbf{y}}_i^{\mathrm{res}})$. Here the residual is detached so that $\mathcal{L}_{\mathrm{comb}}$ only calibrates the gate scaling without affecting the residual direction. This objective calibrates the residual magnitude against the final prediction error, and Proposition 3 gives its per-spot optimum.

2.4 Overall Training Objectives

The full training objective combines the component-wise losses:

$$\mathcal{L} = \mathcal{L}_{\mathrm{score}} + \lambda_r \mathcal{L}_{\mathrm{res}} + \lambda_c \mathcal{L}_{\mathrm{comb}} + \lambda_a \mathcal{L}_{\mathrm{anchor}} + \lambda_o \mathcal{L}_{\mathrm{orth}}, \tag{20}$$

where $\lambda_r, \lambda_c, \lambda_a, \lambda_o > 0$ are loss weights given in Appendix B.2. Except for the gradient-isolated paths described in Section 2.3, all parameters are optimized jointly. Inference is a single deterministic forward pass.

Table 1: Cross-validation performance on ST-Net, Her2ST, and SCC. Best in bold, second best underlined.

Method	ST-Net			Her2ST			SCC		
	PCC↑	MSE↓	MAE↓	PCC↑	MSE↓	MAE↓	PCC↑	MSE↓	MAE↓
ResNet+FCN	0.5221	0.1999	0.3448	0.4629	0.6623	0.6385	0.4619	0.6103	0.6290
BLEEP	0.0784	0.3756	0.4736	0.2747	0.7426	0.6591	0.4176	0.6079	0.6013
HisToGene	0.1211	0.3054	0.4336	0.2062	0.9452	0.7739	0.1225	**0.3095**	**0.4367**
Hist2ST	0.1525	0.3811	0.4822	0.2479	0.7843	0.7286	0.3003	1.0190	0.7639
THItoGene	0.3666	0.2925	0.4111	0.3445	0.8436	0.7069	0.3897	0.6798	0.6442
TRIPLEX	0.2320	0.1472	0.2943	0.3927	0.8982	0.6946	0.5416	0.4891	0.5356
MERGE	<u>0.6795</u>	0.1347	0.2834	0.5037	0.6422	0.6255	0.5512	0.5353	0.5838
CPNN	−0.0310	0.3055	0.4360	<u>0.5372</u>	<u>0.5910</u>	0.5932	<u>0.5747</u>	0.4956	0.5505
DKAN	0.2406	<u>0.1315</u>	<u>0.2780</u>	0.4126	0.5925	<u>0.5768</u>	0.5578	<u>0.4734</u>	0.5326
HyperST	0.2220	0.1469	0.2902	0.3965	0.6352	0.6013	0.5488	0.4965	0.5317
RESOLVE (Ours)	**0.7179**	**0.1178**	**0.2613**	**0.5649**	**0.5751**	**0.5738**	**0.5954**	0.4799	<u>0.5219</u>

3 Experimental Results

Datasets and Preprocessing. We use three public datasets: two breast cancer datasets (ST-Net (He et al., 2020) and Her2ST (Andersson et al., 2021)) and one skin cancer dataset (squamous cell carcinoma, SCC (Ji et al., 2020)). For each dataset, we select the 250 most highly expressed genes and smooth the targets with SPCS (Liu et al., 2022), so all reported results are trained and evaluated on SPCS-smoothed data. Appendix C.1 gives the dataset statistics, the patch extraction protocol, and the full preprocessing details.

Evaluation Metrics. We assess generalization with 8-fold cross-validation, with folds split by slide, and report averaged MSE, MAE, and PCC.

Baselines. We compare RESOLVE against ten baselines: ResNet18 with a fully connected layer (FCN) (He et al., 2016), BLEEP (Xie et al., 2023), HisToGene (Pang et al., 2021), Hist2ST (Zeng et al., 2022), THItoGene (Jia et al., 2024), TRIPLEX (Chung et al., 2024), MERGE (Ganguly et al., 2025), CPNN (Nishimura et al., 2026), DKAN (Zhang et al., 2026b), and HyperST (Zhang et al., 2026a). All methods are trained and evaluated under the same data splits and preprocessing as MERGE. CPNN uses prototypes from the same single-cell references as RESOLVE (Appendix C.2), so that the comparison isolates how the prior is used rather than differences in the reference.

3.1 Quantitative Results

We evaluate prediction accuracy via 8-fold cross-validation (Table 1). RESOLVE achieves the best PCC on all three datasets: 0.5649 on Her2ST, 0.7179 on ST-Net, and 0.5954 on SCC. In MAE, RESOLVE is best on Her2ST and ST-Net and second to HisToGene on SCC. In MSE, it is best on Her2ST and ST-Net and third on SCC, behind HisToGene and DKAN. On SCC, HisToGene attains the best MSE and MAE, but its PCC is near zero. Such a near-zero PCC indicates that the prediction is almost uncorrelated with true expression, which can occur when the model outputs a near-constant value. This yields a moderate MSE and MAE but no predictive power, a degenerate solution rather than predictive skill. We therefore treat PCC as the primary metric and the others as secondary. Under this criterion, the three datasets consistently test the RESOLVE design. Among these comparisons, CPNN builds prototypes from the same single-cell reference as RESOLVE, but predicts directly as a weighted combination of prototype profiles. The difference is not the reference itself, which is identical, but the absence of a fallback estimate: when the prototype combination is unreliable for a given cohort, CPNN has no base prediction to fall back on. RESOLVE decomposes the prediction, so the base pathway provides a stable estimate even when the prototype-conditioned residual is unreliable. This is why RESOLVE remains at 0.7179 on ST-Net while CPNN collapses. On Her2ST and SCC, RESOLVE also leads, and these results consistently reflect the effectiveness of the decomposition and the residual pathway. We next examine the source of this contribution at the gene level.

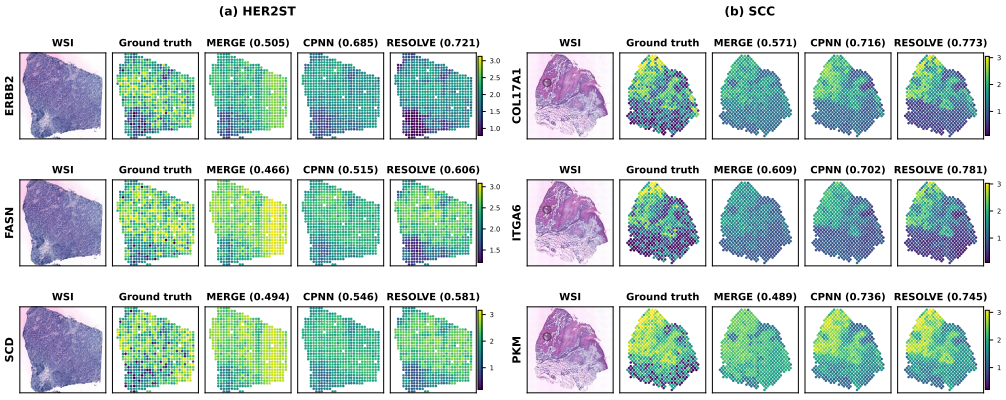

Figure 3: Predictions of MERGE, CPNN, and RESOLVE against ground truth on one validation slide per dataset (a: Her2ST, b: SCC). Columns from left to right: H&E, ground truth, MERGE, CPNN, RESOLVE. All panels of a gene share a color scale, and slide-level single-gene PCC is reported above each prediction.

3.2 Qualitative Results

Figure 3 compares predictions of RESOLVE against two representative baselines, MERGE and CPNN, on one validation slide per dataset, using tumor-enriched markers identified by differential expression analysis (Fig. 3a: Her2ST, ERBB2/FASN/SCD and Fig. 3b: SCC, COL17A1/ITGA6/PKM). MERGE is the strongest graph-based baseline on Her2ST and SCC, and CPNN is the closest prior-based counterpart to RESOLVE, as it builds prototypes from the same single-cell reference. This selection allows us to examine both spatial modeling and prior usage. All spots are rendered at their full-resolution pixel coordinates, so the predicted heatmaps align geometrically with the underlying tissue. On both datasets, RESOLVE recovers the spatial extent and boundaries of tumor-enriched high-expression regions more faithfully than MERGE and CPNN, whose predictions appear blurred or mislocalized at tumor boundaries. The single-gene PCC values confirm this pattern: on the Her2ST slide, RESOLVE achieves 0.721 versus 0.685 for CPNN and 0.505 for MERGE on ERBB2, 0.606 versus 0.515 and 0.466 on FASN, and 0.581 versus 0.546 and 0.494 on SCD. The SCC slide shows the same trend, with RESOLVE reaching 0.773 versus 0.716 and 0.571 on COL17A1, 0.781 versus 0.702 and 0.609 on ITGA6, and 0.745 versus 0.736 and 0.489 on PKM. Across all six markers, RESOLVE attains the highest PCC, indicating that the advantage observed in the quantitative analysis translates to the gene level, with improvements in both spatial-structure fidelity and per-gene correlation.

3.3 Ablation Study

The qualitative results confirm that the gains are real, but not which component produces them. We therefore conduct ablation experiments on Her2ST, with results in Tables 3 and 4 in Appendix B.4.

Key Components. Removing the residual branch costs only 0.007 PCC, whereas removing the prototype router and the gate costs 0.022 and 0.014, far more than the branch itself. The branch therefore contributes only when it is conditioned on a meaningful prior. Without the router, the predictor retains its residual features but loses the biological condition vector, so it cannot distinguish genuine slide-level deviations from noise and fits the latter. This is the failure mode that prototype guidance is designed to prevent. The two encoder stages have effects comparable to the residual branch, with G-DBA at 0.007 and L-DBA at 0.005, indicating that both the local and global stages contribute to accuracy.

Key Loss Terms. The same pattern holds among the loss terms. $\mathcal{L}_{\text{corr}}$, the per-gene pattern term on the base prediction defined in Eq. 11, has the largest effect, reducing PCC by 0.019 and raising MSE from 0.5751 to 0.5876. This term directly optimizes the per-gene spatial pattern that PCC measures, so its removal degrades the primary evaluation metric. Next, $\mathcal{L}_{\text{comb}}$, which calibrates the residual magnitude against the final prediction, costs 0.006. The two structural constraints, $\mathcal{L}_{\text{anchor}}$

and $\mathcal{L}_{\text{orth}}$, are nearly free, with $\Delta \leq 0.002$, which fits their role in stabilizing the base-residual separation rather than directly improving accuracy.

Prototype Prior Control. To further examine the role of the cell-type prototypes, we replace them with a random initialization while keeping all other settings unchanged. PCC drops from 0.5649 to 0.4503, the residual correction becomes harmful, and the full model falls below the no-residual baseline across the gene panel, while the Hallmark enrichment of residual-gain genes disappears, with permutation $p = 1.0$ against 0.001–0.011 for the true prior. Therefore, under the same architecture and parameter count, a meaningless prior is worse than omitting the branch entirely. This shows that RESOLVE's gain comes from conditioning the residual on genuine biological content, as detailed in Appendix B.5.

3.4 Interpretability Analysis

The ablation study shows that the gain depends on a genuine biological prior, but does not specify what this prior corrects. We find that the residual gain is modest but broad. On Her2ST, it improves 208 of 250 genes, or 83%, with sign test $p = 1.3 \times 10^{-27}$, and the largest gains are on cancer programs that vary most across slides, while the largest losses fall on housekeeping genes, as detailed in Appendix D.1. This directional distribution indicates that the residual is not fitting noise randomly, but selectively acting on genes with specific biological properties. To characterize the functional identity of these beneficiary genes, we submit the top 15 genes ranked by residual gain to Hallmark over-representation analysis (Subramanian et al., 2005; Liberzon et al., 2015), shown in Figure 6 in Appendix D.3. These genes are significantly enriched for interferon-α response with FDR $= 0.013$, interferon-γ response with FDR $= 0.015$, and estrogen response early with FDR $= 0.038$. These programs are expressed by immune cells and luminal tumor cells and vary substantially across patients (Kyalwazi et al., 2023; Dunphy et al., 2020), consistent with a patient-level program shared across the slide rather than expression directly explained by local morphology. The top-15 gain genes do not overlap the DEG-defined tumor-cell gene set, further indicating that the residual is not re-predicting tumor-core expression. Together, these findings point to the residual acting on immune- and luminal-related program deviations rather than on tumor-core expression.

This interpretation is further supported by the source and the spatial distribution of the gain. A component-level decomposition shows that for the top-15 gain genes, the improvement comes almost entirely from the between-slide component, with $+0.0321$ versus $+0.0021$ PCC within slides, as detailed in Appendix D.4, confirming that the residual corrects a slide-level rather than an intra-slide signal. Spatial visualization further shows that the gated residual concentrates at the stromal-immune interface in breast cancer and at tumor cell-state transition regions in skin SCC, shown in Figure 5 in Appendix D.2. These regions are where local morphology provides the least information, and the residual concentrates there, indicating that it is not repeating what morphology already expresses. In summary, the base branch carries the dominant, highly expressed and spatially smooth structure of the expression profile, while the residual branch intervenes where local morphology is least informative, correcting slide-level cell states rather than re-fitting what morphology already expresses.

4 Conclusion and Future Work

We proposed RESOLVE, which disentangles histology-based gene expression prediction into a morphology-anchored base component and a prototype-guided residual component. Across three public ST datasets, RESOLVE attains seven best results, and its residual pathway provides interpretable gains, correcting mainly slide-level differences and enriching for interferon and estrogen response programs. The current design depends on a fixed single-cell reference and a deterministic predictor. Future work will address these limitations through more expressive single-cell integration and generative uncertainty quantification.

AI use statement

In this work, we used generative AI tools to polish the language of the manuscript and to check the consistency of the reported results. We have not used generative AI tools for any of the other tasks requiring disclosure, including the formulation of mathematical claims or proofs, the design

or implementation of methods and experiments, the generation of synthetic data, the cleaning or reformatting of datasets, or the interpretation of results. We have reviewed all AI-assisted content and take full responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

All three datasets (ST-Net, Her2ST, and skin SCC) are previously published collections of de-identified human tissue, released under the consent and ethics approvals of their original studies. No new human subjects data were collected, and no personally identifiable information is used. Our method operates only on histology images and gene expression profiles, and we are not aware of any specific harmful application it enables, although as with any spatial gene expression predictor its outputs should not be used for clinical decision-making without independent validation.

Reproducibility statement

All experiments use three publicly available spatial transcriptomics datasets: ST-Net, Her2ST, and skin SCC. No new data were collected, and all tissue profiles come from previously published, de-identified human specimens. Image and expression preprocessing, including the SPCS-smoothed targets, follows the protocol of MERGE, described in Appendix C.1, and the single-cell references used to build the prototypes are listed in Appendix C.2. All architectural and training hyperparameters are reported in Table 2. Generalization is assessed by 8-fold cross-validation with folds split by slide, and every reported MSE, MAE, and PCC value is averaged over the folds. The code is available in anonymized form at `https://anonymous.4open.science/r/RESOLVE-DBD0` and will be publicly released upon acceptance.

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.

Erick Armingol, Adam Officer, Olivier Harismendy, and Nathan E Lewis. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics*, 22(2):71–88, 2021.

Dalia Barkley, Reuben Moncada, Maayan Pour, Deborah A Liberman, Ian Dryg, Gregor Werba, Wei Wang, Maayan Baron, Anjali Rao, Bo Xia, et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nature genetics*, 54(8):1192–1201, 2022.

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, 2024.

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.

Ozan Ciga, Tony Xu, and Anne Louise Martel. Self supervised contrastive learning for digital histopathology. *Machine learning with applications*, 7:100198, 2022.

Karen A Dunphy, Amye L Black, Amy L Roberts, Aman Sharma, Zida Li, Sneha Suresh, Eva P Browne, Kathleen F Arcaro, Jennifer Ser-Dolansky, Carol Bigelow, et al. Inter-individual variation in response to estrogen in human breast explants. *Journal of mammary gland biology and neoplasia*, 25(1):51–68, 2020.

Donghai Fang, Yongheng Li, Zhen Wang, Yuansong Zeng, and Wenwen Min. Adapting a pre-trained single-cell foundation model to spatial gene expression generation from histology images. *arXiv preprint arXiv:2603.19766*, 2026.

Alexandre Filiot, Ridouane Ghermi, Antoine Olivier, Paul Jacob, Lucas Fidon, Axel Camara, Alice Mac Kain, Charlie Saillard, and Jean-Baptiste Schiratti. Scaling self-supervised learning for histopathology with masked image modeling. *MedRxiv*, pp. 2023–07, 2023.

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

Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering*, 4(8):827–834, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Mo Huang, Jingshu Wang, Eduardo Torre, Hannah Dueck, Sydney Shaffer, Roberto Bonasio, John I Murray, Arjun Raj, Mingyao Li, and Nancy R Zhang. Saver: gene expression recovery for single-cell rna sequencing. *Nature methods*, 15(7):539–542, 2018.

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*, 2025.

Andrew L Ji, Adam J Rubin, Kim Thrane, Sizun Jiang, David L Reynolds, Robin M Meyers, Margaret G Guo, Benson M George, Annelie Mollbrink, Joseph Bergenstråhle, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *cell*, 182(2): 497–514, 2020.

Yuran Jia, Junliang Liu, Li Chen, Tianyi Zhao, and Yadong Wang. Thitogene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics*, 25(1):bbad464, 2024.

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.

Beverly Kyalwazi, Christina Yau, Michael J Campbell, Toshio F Yoshimatsu, A Jo Chien, Anne M Wallace, Andres Forero-Torres, Lajos Pusztai, Erin D Ellis, Kathy S Albain, et al. Race, gene expression signatures, and clinical outcomes of patients with high-risk early breast cancer. *JAMA Network Open*, 6(12):e2349646, 2023.

Yixin Li, Fan Zhong, and Lei Liu. Scist: single-cell reference-informed spatial gene expression prediction from pathological images. *Briefings in Bioinformatics*, 26(6):bbaf613, 2025.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6): 417–425, 2015.

Anqi Liu, Yue Zhao, Hui Shen, Zhengming Ding, and Hong-Wen Deng. Ressat: Enhancing spatial transcriptomics prediction from h&e-stained histology images with interactive spot transformer. *Res Sq*, pp. rs–3, 2024.

Yusong Liu, Tongxin Wang, Ben Duggan, Michael Sharpnack, Kun Huang, Jie Zhang, Xiufen Ye, and Travis S Johnson. Spcs: a spatial and pattern combined smoothing method for spatial transcriptomic expression. *Briefings in Bioinformatics*, 23(3):bbac116, 2022.

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.

Tongxuan Lv, Ying Zhang, Mei Li, Qiang Kang, Shuangsang Fang, Yong Zhang, Susanne Brix, and Xun Xu. Eags: efficient and adaptive gaussian smoothing applied to high-resolved spatial transcriptomics. *GigaScience*, 13:giad097, 2024.

Kazuya Nishimura, Ryoma Bise, Shinnosuke Matsuo, Haruka Hirose, and Yasuhiro Kojima. Cell-type prototype-informed neural network for gene expression estimation from pathology images. *arXiv preprint arXiv:2603.18461*, 2026.

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

Ofir Press, Noah A Smith, and Mike Lewis. Train short, test long: Attention with linear biases enables input length extrapolation. *arXiv preprint arXiv:2108.12409*, 2021.

Mingcheng Qu, Yuchuan Zhao, Guang Yang, Donglin Di, Xiu Su, Hongyan Xu, Yang Song, and Lei Fan. Stag: Biologically guided spatial transcriptomics prediction via hypergraph learning. *Medical Image Analysis*, pp. 104206, 2026.

Benoît Schmauch, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro, Aurélie Kamoun, Meriem Sefta, Sylvain Toldo, Mikhail Zaslavskiy, et al. A deep learning model to predict rna-seq expression of tumours from whole slide images. *Nature communications*, 11(1):3877, 2020.

Noam Shazeer. Glu variants improve transformer. *arXiv preprint arXiv:2002.05202*, 2020.

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

Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the national academy of sciences*, 102(43):15545–15550, 2005.

David Van Dijk, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J Carr, Cassandra Burdziak, Kevin R Moon, Christine L Chaffer, Diwakar Pattabiraman, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3):716–729, 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature medicine*, 30(10): 2924–2935, 2024.

Sunny Z Wu, Ghamdan Al-Eryani, Daniel Lee Roden, Simon Junankar, Kate Harvey, Alma Andersson, Aatish Thennavan, Chenfei Wang, James R Torpy, Nenad Bartonicek, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics*, 53(9):1334–1347, 2021.

Jacky Xie, Xiao Tan, Yuanhao Jiang, Hao Nguyen, Nan Ye, and Quan Nguyen. Performance of deep learning methods for spatial gene expression prediction from histology images: a comprehensive assessment. *Bioinformatics Advances*, pp. vbag202, 2026.

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.

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.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34:28877–28888, 2021.

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.

Chen Zhang, Yilu An, Ying Chen, Hao Li, Xitong Ling, Lihao Liu, Junjun He, Yuxiang Lin, Zihui Wang, and Rongshan Yu. Hyperst: Hierarchical hyperbolic learning for spatial transcriptomics prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5730–5739, 2026a.

Linlin Zhang, Dongsheng Chen, Dongli Song, Xiaoxia Liu, Yanan Zhang, Xun Xu, and Xiangdong Wang. Clinical and translational values of spatial transcriptomics. *Signal transduction and targeted therapy*, 7(1):111, 2022.

Wei Zhang, Jiajun Chu, Xinci Liu, Chen Tong, and Xinyue Li. Dual-path knowledge-augmented contrastive alignment network for spatially resolved transcriptomics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12807–12815, 2026b.

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 Related Work

A.1 Gene Expression Prediction from Histology Images

Early methods predict gene expression at each spot directly from local histology images. For example, ST-Net (He et al., 2020) uses a convolutional neural network (CNN) (He et al., 2016) to encode each image patch independently and regresses gene expression values through a fully connected layer. Because each spot is processed in isolation, these methods struggle to model spatial dependencies between neighboring regions and the local tissue microenvironment. To address this, subsequent work incorporates spatial context and multi-scale tissue organization into the prediction process. HisToGene (Pang et al., 2021) introduces a Transformer (Vaswani et al., 2017) to spatial gene expression prediction, modeling global dependencies among spots through self-attention with spatial position encodings. TRIPLEX (Chung et al., 2024) further adopts multi-resolution representations, jointly fusing visual information at three scales: the target spot, its local neighborhood, and the whole slide. ResSAT (Liu et al., 2024) jointly models image representations and spatial locations through an interactive spot Transformer to better exploit local spatial context. Going further, HyperST (Zhang et al., 2026a) exploits the inherent hierarchy of ST data, extracting spot-level and niche-level image and gene representations and aligning them in hyperbolic space.

Beyond attention-based context modeling, graph structures are widely used to characterize more complex tissue microenvironments. MERGE (Ganguly et al., 2025) organizes spots by spatial proximity and morphological similarity, constructing both intra- and inter-cluster connections to link spatially distant but morphologically similar regions. STAG (Qu et al., 2026) adopts a dual-branch hypergraph to capture higher-order spatial dependencies that pairwise graphs cannot represent. Meanwhile, generative modeling offers an alternative to deterministic regression for this task. STFlow (Huang et al., 2025) learns the slide-level joint expression distribution via flow matching, while Stem (Zhu et al., 2025) performs generative inference of spatial gene expression with a conditional diffusion model. These methods improve the characterization of expression distributions and their uncertainty, but their prediction target remains the recovery of the full gene expression profile given histology information.

Another line of work injects external biological priors into the predictor. CPNN (Nishimura et al., 2026) summarizes a matched single-cell reference as cell-type prototypes whose composition weights are estimated from the image, so that regression becomes a weighted combination of prototype profiles. HINGE (Fang et al., 2026) adapts a pretrained single-cell foundation model into the histology-conditioned generator itself. SciSt (Li et al., 2025) converts cell segmentation and reference weighting into an initial expression estimate that prompts the predictor. DKAN (Zhang et al., 2026b) introduces gene semantic knowledge retrieved from external databases and uses it to align image and expression representations.

Overall, existing methods continue to strengthen the mapping from histology images to gene expression through visual representations, spatial context, graph structures, generative modeling, and external biological priors. However, these methods still regress the full expression profile under a unified objective, without explicitly distinguishing which components are explainable by morphology and which come from beyond morphology. As a result, signals beyond morphology can only remain entangled with the morphology-driven prediction in the same regression target, lacking an independent representation and modeling pathway that could be inspected or corrected separately.

A.2 Foundation Models for Computational Pathology

Gene expression prediction from histology images depends on an encoder strong enough to expose the morphological patterns that carry expression information. This requirement has motivated a separate line of work on general-purpose histology representations, where large-scale self-supervised pretraining has produced pathology foundation models (Ciga et al., 2022; Filiot et al., 2023; Vorontsov et al., 2024; Xu et al., 2024; Chen et al., 2024; Lu et al., 2024) whose embeddings encode morphological patterns of tissue cells. These embeddings provide a transferable morphological representation and have been shown to carry information relevant to gene expression, making them a natural feature backbone for this task. In this study, we adopt UNI2-h (Chen et al., 2024) as the frozen feature extractor. Appendix B.3 details the backbone and its alignment with our patch extraction protocol.

B Architecture and Training Details

B.1 Analysis of the Two-Pathway Objective

This section formalizes three properties of the two-pathway predictor (Section 2.3): the stop-gradient supervision keeps the gradients of the two predictors disjoint, the residual predictor is optimal when it predicts the conditional mean of the base prediction error, and the gate is optimal when it rescales that prediction by its regression coefficient on the unexplained part of the target.

B.1.1 Gradient Isolation

Proposition 1. *Consider the full objective in Eq. 20. Let θ_b denote the parameters of the base predictor f_b, and θ_r those of the residual predictor, comprising its router and its MLP f_r. Then*

$$\frac{\partial \mathcal{L}_{res}}{\partial \theta_b} = 0 \qquad \textit{and} \qquad \frac{\partial \mathcal{L}_{comb}}{\partial \theta_r} = 0.$$

Neither predictor is penalized for the behaviour of the other.

Proof. By Eq. 17, the loss $\mathcal{L}_{\text{res}}$ depends on $\hat{\mathbf{y}}_i^{\text{base}}$ only through $\text{sg}(\hat{\mathbf{y}}_i^{\text{base}})$, which is held constant during differentiation, so no gradient reaches θ_b through $\mathcal{L}_{\text{res}}$. By the definition following Eq. 19, the combined prediction depends on $\hat{\mathbf{y}}_i^{\text{res}}$ only through $\gamma_i \cdot \text{sg}(\hat{\mathbf{y}}_i^{\text{res}})$, so no gradient reaches θ_r through $\mathcal{L}_{\text{comb}}$. The remaining losses separate cleanly: $\mathcal{L}_{\text{score}}$ depends on $\hat{\mathbf{y}}_i^{\text{base}}$ but not on $\hat{\mathbf{y}}_i^{\text{res}}$, and neither $\mathcal{L}_{\text{anchor}}$ nor $\mathcal{L}_{\text{orth}}$ involves either predictor.

Remark 1. *The isolation holds for the two predictors, not for the model as a whole. The base predictor is supervised by $\mathcal{L}_{score}$ on the full target $\mathbf{y}_i$ and by $\mathcal{L}_{comb}$ through the combined prediction. The residual predictor receives gradients only through $\mathcal{L}_{res}$. The prototypes receive them through $\mathcal{L}_{res}$ and additionally through the anchoring loss $\mathcal{L}_{anchor}$, which involves no other module. The gate f_γ is supervised only by $\mathcal{L}_{comb}$. The shared encoder remains a common conduit and receives gradients from every loss except $\mathcal{L}_{anchor}$, through the two projections.*

B.1.2 Optimality of the Residual Predictor

Proposition 2. *Fix the encoder, the projections, and the base predictor, and write $Z_i = [\mathbf{z}_i^{res}; \boldsymbol{\mu}_i]$ for the input of the residual predictor. Let $\mathcal{F}$ be the class of measurable functions of Z_i. The population risk underlying Eq. 17 is minimized over $\mathcal{F}$, coordinate-wise per gene, by*

$$f_r^*(Z_i) = \mathbb{E}\big[\mathbf{y}_i - \hat{\mathbf{y}}_i^{base} \,\big|\, Z_i\big]\,. \tag{21}$$

Proof. The loss separates over the G gene coordinates. For a fixed spot, write Z for its residual input $[\mathbf{z}_i^{\text{res}}; \boldsymbol{\mu}_i]$ and $X = y_g - \hat{y}_g^{\text{base}}$ for one gene coordinate of the unexplained target. For any measurable h,

$$\mathbb{E}\big[(X - h(Z))^2\big] = \mathbb{E}\big[\mathbb{E}\big[(X - h(Z))^2 \,\big|\, Z\big]\big]\,,$$

and conditionally on Z the value $h(Z)$ is a constant, so

$$\mathbb{E}\big[(X - h(Z))^2 \,\big|\, Z\big] = \text{Var}(X \mid Z) + (h(Z) - \mathbb{E}[X \mid Z])^2 \geq \text{Var}(X \mid Z),$$

with equality if and only if $h(Z) = \mathbb{E}[X \mid Z]$ almost surely.

B.1.3 Calibrating the Residual Scale

Proposition 3. *Fix the encoder, the projections, the base and residual predictors, and the prototypes, and let $\mathbf{r}_i = sg(\hat{\mathbf{y}}_i^{res})$ denote the detached residual prediction. The per-spot minimizer of $\mathcal{L}_{comb}$ over $\gamma_i \in [0, 1]$ is, for $\mathbf{r}_i \neq \mathbf{0}$,*

$$\gamma_i^\star = \text{clip}\left(\frac{\langle \mathbf{y}_i - \hat{\mathbf{y}}_i^{base},\ \mathbf{r}_i \rangle}{\|\mathbf{r}_i\|_2^2},\ 0,\ 1\right). \tag{22}$$

Proof. With the remaining modules fixed, $\mathbf{r}_i$ is constant, so $\mathcal{L}_{\text{comb}}$ separates over spots and each γ_i appears in a single term. Writing $\boldsymbol{\delta}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i^{\text{base}}$,

$$\|\hat{\mathbf{y}}_i^{\text{base}} + \gamma_i \mathbf{r}_i - \mathbf{y}_i\|_2^2 = \gamma_i^2 \|\mathbf{r}_i\|_2^2 - 2\gamma_i \langle \boldsymbol{\delta}_i, \mathbf{r}_i \rangle + \|\boldsymbol{\delta}_i\|_2^2,$$

a convex quadratic in γ_i with positive leading coefficient whenever $\mathbf{r}_i \neq \mathbf{0}$. Its unconstrained minimizer is $\langle \boldsymbol{\delta}_i, \mathbf{r}_i \rangle / \|\mathbf{r}_i\|_2^2$, and clipping to $[0, 1]$ yields the constrained minimizer.

Remark 2. *Propositions 1, 2, and 3 give the decomposition its operational meaning. The base predictor is regressed on the full target $\mathbf{y}_i$, so it learns what its input supports, and by Proposition 1 it is never penalized for what the residual predictor does. By Proposition 2, the residual predictor does not re-predict the target: it estimates the conditional mean of whatever the base prediction leaves unexplained, and Eq. 21 identifies that estimate with the conditional mean of the residual target. By Proposition 3, the gate scales this estimate by its regression coefficient on the unexplained target, so the final correction is the portion of the residual direction that the combined loss can use. Three boundaries are worth stating. The characterization concerns the population optimum of each predictor with the encoder, the projections, and the remaining modules fixed, not the joint global optimum of Eq. 20. In the model the gate is not a free per-spot variable but $\gamma_i = \sigma(f_\gamma(\mathbf{z}_i^{base}))$, so Eq. 22 is the per-spot target that f_γ regresses toward under the constraint $\gamma_i \in (0, 1)$. And gradient isolation is a property of the gradient paths rather than of the representations: $\mathcal{L}_{orth}$ is the only term coupling the two projections, so the disentanglement of the two feature spaces is driven by that objective rather than by the stop-gradient.*

B.2 Hyperparameters and Model Configuration

Table 2 summarizes the architectural dimensions and hyperparameters of RESOLVE.

B.3 The Pathology Backbone

RESOLVE uses UNI2-h (Chen et al., 2024) as its feature extractor. UNI2-h is a vision foundation model for computational pathology built on a ViT-H/14 backbone and pretrained with a DINOv2-based self-supervised objective that combines self-distillation with masked image modeling. Its pretraining corpus comprises more than 200 million H&E and IHC tiles sampled from over 350,000 whole slide images, roughly twice the tile count of its predecessor UNI (Chen et al., 2024), which used a ViT-L backbone and about 100 million tiles from over 100,000 slides spanning more than 20 major tissue types.

The encoder is kept frozen throughout training, and its 1536-dimensional patch embeddings serve as the spot features $\mathbf{x}_i$ ($D = 1536$ in Table 2). Two properties make UNI2-h well suited to this task. First, its pretraining tiles are extracted at 256×256 pixels and $20\times$ magnification, exactly matching our patch extraction protocol, so the encoder operates on the same image statistics it was pretrained on. Second, because cell morphology reflects cell type and state, the learned embeddings carry information relevant to spot-level gene expression, on which both pathways build.

B.4 Full Ablation Tables

The complete component and loss ablations on Her2ST discussed in Section 3.3 are listed in Tables 3 and 4.

B.5 Random-Prior Control

To isolate what the single-cell prior contributes, we replace the prototype initialization $\log C$ with a random matrix of the same shape, drawn uniformly in log space, and keep every other setting identical. Because the no-residual ablation never reads the prototypes, its trained model is shared by both configurations, so per-gene gains are computed against the identical baseline.

With the true prior, the top-15 gain genes improve by $+0.020$ to $+0.035$ and enrich for interferon-α, interferon-γ, and estrogen-response programs (permutation $p = 0.001$, 0.0015, and 0.0105). With random prototypes, per-gene gains are negative across the panel (-0.009 to -0.036 for the fifteen largest), the full model falls below the no-residual baseline, and no pathway survives permutation

Table 2: Architectural dimensions and hyperparameters of RESOLVE.

Module	Hyperparameter	Value
Hierarchical DBA encoder		
Input projection	input dimension D (UNI2-h)	1536
	model dimension d	768
L-DBA	nearest neighbors k	32
	†ST-Net	48
	ALiBi slopes m_h	$2^{-(h+1)}$
	β, τ (fixed)	1.5, 0.6
G-DBA	distance bins	8 (cut points $\{1, 2, 4, 8, 16, 32, 48\}$)
	similarity bins	6 (cut points $\{0.3, 0.5, 0.7, 0.85, 0.95\}$)
	β_h, τ_h initialization	1.5, 0.6
Encoder	stacked blocks L	3
	†ST-Net	4
Two-pathway predictor		
Base predictor f_b	MLP dimensions	$768 \to 512 \to 512 \to 250$
Gate f_γ	MLP dimensions	$768 \to 128 \to 1$
Residual predictor f_r	MLP dimensions	$1018 \to 512 \to 512 \to 250$
Prototype routing		
Cell types K	breast / skin	29 / 35
Target genes G	most highly expressed genes	250
Routing	temperature τ_π	1
	routed types K_π	15
	†ST-Net	10
Optimization		
Training	optimizer	Adam
	epochs	1500
	learning rate	3×10^{-4}
	†ST-Net	2×10^{-4}
	weight decay	10^{-4}
	cosine annealing $T_{\max}$	100
	gradient clipping	1.0
	loss weights λ_r, λ_c, λ_a, λ_o	0.5, 1.0, 1.0, 0.1
	†ST-Net	0.5, 3.0, 2.0, 0.1
	†SCC	2.0, 2.0, 2.0, 0.1
	$\mathcal{L}_{\text{score}}$ weight λ_{corr}	2.0
	†ST-Net	0.5

†Value used for that dataset; values without a marker apply to all three datasets.

Table 3: Ablation of key components on Her2ST (8-fold cross-validation).

Settings	PCC	MAE	MSE
FULL	**0.5649**	**0.5738**	**0.5751**
w/o L-DBA	0.5597	0.5752	0.5792
w/o G-DBA	0.5577	0.5787	0.5813
w/o residual branch	0.5583	0.5796	0.5822
w/o prototype router	0.5425	0.5889	0.5897
w/o gate	0.5512	0.5835	0.5842

Table 4: Ablation of loss terms on Her2ST (8-fold cross-validation).

Settings	PCC	MAE	MSE
FULL	**0.5649**	**0.5738**	**0.5751**
w/o $\mathcal{L}_{\text{corr}}$	0.5460	0.5840	0.5876
w/o $\mathcal{L}_{\text{comb}}$	0.5592	0.5767	0.5793
w/o $\mathcal{L}_{\text{anchor}}$	0.5632	0.5740	0.5775
w/o $\mathcal{L}_{\text{orth}}$	0.5634	0.5740	0.5781

testing ($p = 1.0$, with zero overlap with the interferon gene sets). The prior therefore supplies both the direction and the biological content of the correction, and anchors the aggregate result.

Table 5: Random-prior control on Her2ST (8-fold cross-validation).

Settings	PCC	MAE	MSE
Cell-type prototypes	**0.5649**	**0.5738**	**0.5751**
Random prototypes	0.4503	0.6201	0.6451

C Data and Preprocessing Protocol

C.1 Preprocessing Details

This section details the preprocessing protocol behind the experiments in Section 3. The protocol follows MERGE (Ganguly et al., 2025) so that RESOLVE and all baselines share identical inputs, targets, and data splits, with one adaptation to our feature backbone noted below.

Datasets. ST-Net comprises 68 samples from 23 patients, Her2ST 36 samples from 8 patients, and skin SCC 12 samples from 4 patients, averaging approximately 450, 378, and 723 in-tissue spots per sample, respectively.

Image patch extraction. The three datasets are profiled on the original ST platform (Ståhl et al., 2016), whose spots have a diameter of $100\,\mu\mathrm{m}$ and sit on a square grid with $200\,\mu\mathrm{m}$ center-to-center spacing. For each spot we extract a 256×256 pixel patch from the $20\times$ magnification H&E image, centered at the spot coordinate. This patch size matches the pretraining protocol of UNI2-h (Appendix B.3), whereas MERGE extracts 224×224 patches for its ResNet18 encoder. To recover morphological context that grid-centered patches leave out, we further extract additional patches from intermediate locations between adjacent spot coordinates. These additional patches participate only in L-DBA (Section 2.2) as spatial context and are excluded from all losses and evaluation metrics.

Gene selection. For each dataset we use the fixed subset of the 250 most highly expressed genes introduced by TRIPLEX (Chung et al., 2024) and adopted by MERGE, so the prediction target $\mathbf{y}_i \in \mathbb{R}^G$ with $G = 250$ is identical across all compared methods.

SPCS smoothing. Transcriptomic quantification is prone to dropouts, in which subsets of genes are missed during measurement and artificial zeros enter the expression matrix. Smoothing is well studied for scRNA-seq (Huang et al., 2018; Van Dijk et al., 2018), but these methods take no spatial information into account. Earlier work on ST suppressed the artifacts by averaging each spot with its eight grid neighbors, but such spatial averaging further shrinks the scale of an already sparse matrix and ignores gene-level information. We therefore smooth the targets with two-factor methods such as SPCS (Liu et al., 2022) and EAGS (Lv et al., 2024). SPCS replaces the expression of each spot with a weighted combination of all other spots in the same sample. Each weight combines a spatial factor, inversely proportional to the Manhattan distance between spot coordinates, and a pattern factor computed by projecting the expression profiles into a 10-dimensional PCA space and exponentiating the Pearson correlation distance between the projected profiles. Following MERGE, all methods in Table 1, including the baselines, are trained and evaluated on the SPCS-smoothed targets.

C.2 Single-Cell References for Prototype Construction

The prototypes used by the residual pathway (Section 2.3) are derived from external single-cell RNA-seq references of matched tissue. For each ST dataset, the reference is first restricted to the 250 target genes, cell2location (Kleshchevnikov et al., 2022) is then fitted on that restricted reference, and the posterior mean expression of each cell type yields the signature matrix $C \in \mathbb{R}^{K\times G}$. The matrix is affinely aligned to the target expression scale of the corresponding ST dataset, and all entries are clipped to a small positive value before the logarithm. Genes absent from the reference receive the same small positive value. Table 6 lists the reference datasets and the cell-type labels used in each case.

Her2ST and ST-Net share the same 29 cell types from the breast atlas (Wu et al., 2021): Cancer Basal SC, Cancer Cycling, Cancer Her2 SC, Cancer LumA SC, Cancer LumB SC, Luminal Progenitors,

Table 6: Single-cell references used for prototype construction.

ST dataset	Single-cell reference	Accession	Label	K
Her2ST	Breast cancer atlas, 26 patients	GSE176078	`celltype_minor`	29
ST-Net	Same as Her2ST	GSE176078	`celltype_minor`	29
SCC	Cutaneous SCC atlas	GSE144236	`level3`	35

Mature Luminal, Myoepithelial, CAFs MSC iCAF-like, CAFs myCAF-like, Cycling PVL, PVL Differentiated, PVL Immature, B cells Memory, B cells Naive, Cycling T-cells, Cycling_Myeloid, DCs, Macrophage, Monocyte, NK cells, NKT cells, Plasmablasts, T cells CD4+, T cells CD8+, Endothelial ACKR1, Endothelial CXCL12, Lymphatic LYVE1, Endothelial RGS5.

The SCC prototypes use 35 cell types from the cutaneous SCC atlas (Ji et al., 2020), label `level3`: ASDC, B Cell, CD1C, CD4_Exh, CD4_Naive, CD4_Pre_Exh, CD4_RGCC, CD8_EM, CD8_EMRA, CD8_Exh, CD8_Naive, CLEC9A, Eccrine, Endothelial Cell, Fibroblast, Keratinocyte, LC, MDSC, Mac, Melanocyte, Mig_CD1C, Mig_CLEC9A, Mig_LC, Multiplet, NK, Normal_KC_Basal, Normal_KC_Cyc, Normal_KC_Diff, PDC, Pilosebaceous, TSK, Treg, Tumor_KC_Basal, Tumor_KC_Cyc, Tumor_KC_Diff.

D Analysis of the Residual Gain

D.1 Per-Gene Structure of the Residual Gain

We first characterize how the residual gain is distributed across genes on Her2ST, using the per-gene PCC difference between the full model and the retrained no-residual ablation (Figure 4).

Panel (a) shows that the improvement is broad rather than concentrated. Of the 250 genes, 208 (83%) improve under the residual branch ($p = 1.3 \times 10^{-27}$, sign test), the largest single-gene gain exceeds the largest loss in magnitude, and the mean gain of $+0.007$ is carried by many small corrections in the same direction rather than by a few extreme genes. The aggregate difference between the full model and the ablation (Table 3) is thus a broad, one-directional shift across the gene panel.

Panel (b) relates the gain to expression level. The per-gene gain is negatively correlated with mean expression ($r = -0.27$), indicating that the residual pathway skews toward low-abundance genes. The labeled beneficiaries concentrate in the cancer programs that vary most across slides: luminal and HER2 amplicon genes (MUC1, TFF3, PPP1R1B, IGFBP2, LGALS3BP), interferon response genes (ISG15, B2M), stromal markers (AEBP1, MYL9), and the immunoglobulin gene IGLC2, consistent with the Hallmark enrichment in Appendix D.3. At the high-expression end, the largest losses are confined to housekeeping genes (FTL, TMSB10, UBA52), which vary little across slides and carry little additional signal for the residual pathway to model. The two ends point the same way: the residual draws its gain from cross-slide variation, which the highest-expressed housekeeping genes lack.

D.2 Spot-Level Decomposition of the Residual Gain

We next localize where the corrections of the residual pathway occur. For one representative validation slide per dataset, we render the base component $\hat{y}_{\text{base}}$, the gated residual $\gamma \cdot \hat{y}_{\text{res}}$, and their sum spot by spot in the H&E pixel coordinate system (Figure 5).

D.3 Hallmark Enrichment of Residual-Gain Genes

We then characterize the functional identity of the gain through Hallmark over-representation analysis (Subramanian et al., 2005; Liberzon et al., 2015) of the top-15 residual-gain genes on Her2ST (Figure 6), where the residual gain of a gene is defined as its per-gene PCC under the full model minus that under the retrained no-residual-branch ablation (Section 3.4).

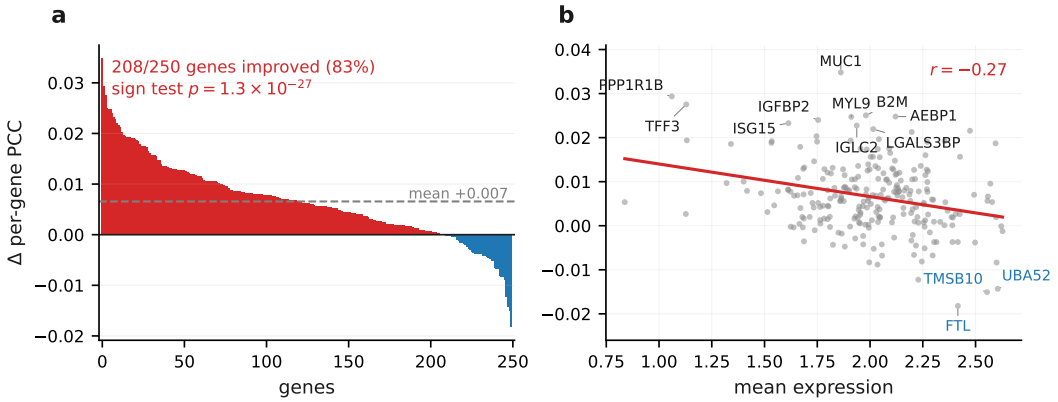

Figure 4: Per-gene structure of the residual gain on Her2ST. (a) Per-gene PCC difference between the full model and the retrained no-residual ablation, rank-sorted. 208 of the 250 genes improve ($p = 1.3 \times 10^{-27}$, sign test), the dashed line marks the mean gain ($+0.007$). (b) The same per-gene gain against mean expression, the red line is the linear fit ($r = -0.27$). Black labels mark the ten largest beneficiaries, blue labels the three largest losses.

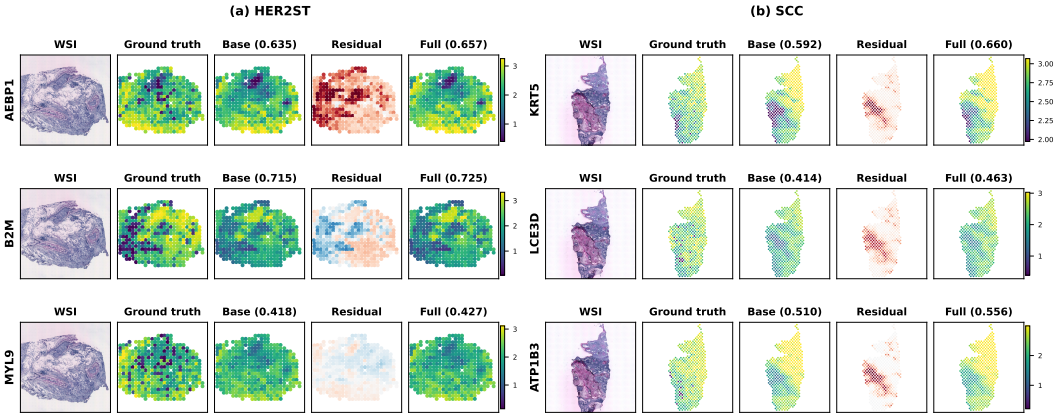

Figure 5: Decomposition of predictions on one representative validation slide per dataset (a: Her2ST, slide G1, and b: skin SCC, slide P5_rep2). Columns from left to right: H&E, ground truth, base branch $\hat{y}_{\text{base}}$, gated residual $\gamma \cdot \hat{y}_{\text{res}}$, and full prediction $\hat{y}_{\text{base}} + \gamma \cdot \hat{y}_{\text{res}}$. Ground truth, base, and full panels share a color scale per gene, the residual uses a diverging scale (red: up, blue: down), and slide-level PCC is reported per column. Genes are the largest residual gains per dataset (AEBP1/B2M/MYL9 on Her2ST, KRT5/LCE3D/ATP1B3 on skin SCC).

D.4 Component Decomposition of the Residual Gain

Finally, we ask whether the gain lands in the between-slide component rather than in within-slide refinement. The two-pathway design targets the slide-level deviation specifically, and we test this by recomputing the per-gene residual gain, the per-gene PCC difference averaged over 8 folds between the full model and the retrained no-residual ablation, under two restricted variants in addition to the pooled one used elsewhere.

The **within-slide** variant evaluates each validation slide separately, computing the spot-level PCC within that slide and averaging over the 36 validation slides. It measures how well the model refines the spatial expression pattern inside a slide. The **between-slide** variant reduces every slide to its mean prediction and mean ground truth and computes the correlation across the 36 slides. It measures how well the model recognizes which expression level, and therefore which cell-state program, is active in each slide.

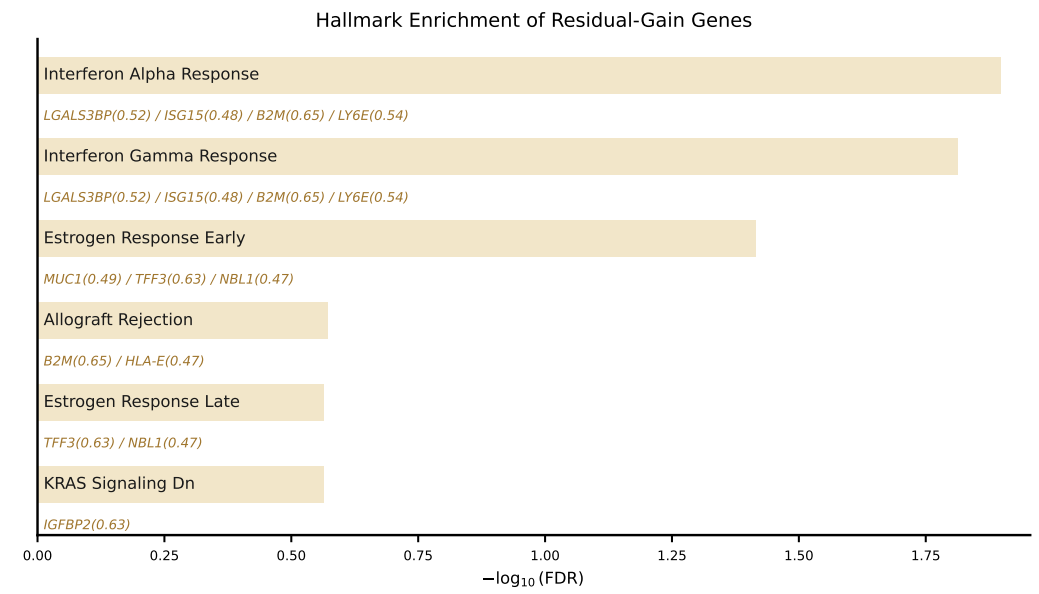

Figure 6: Hallmark over-representation analysis of the top-15 residual-gain genes on Her2ST (background: the 250 measured genes). Bars show $-\log_{10}(\mathrm{FDR})$, hit genes are listed with per-gene PCC in parentheses.

For the top-15 gain genes, the pooled gain averages $+0.0240$ PCC, and the two restricted variants give $+0.0321$ between slides and $+0.0021$ within slides (Figure 7a): the between-slide component carries fifteen times the within-slide gain. The per-gene view (Figure 7b) confirms the same ordering for most of the panel, although the share varies across genes. The residual pathway therefore corrects the slide-level cell-state-specific deviation that the base pathway does not represent, rather than refining the within-slide spatial pattern.

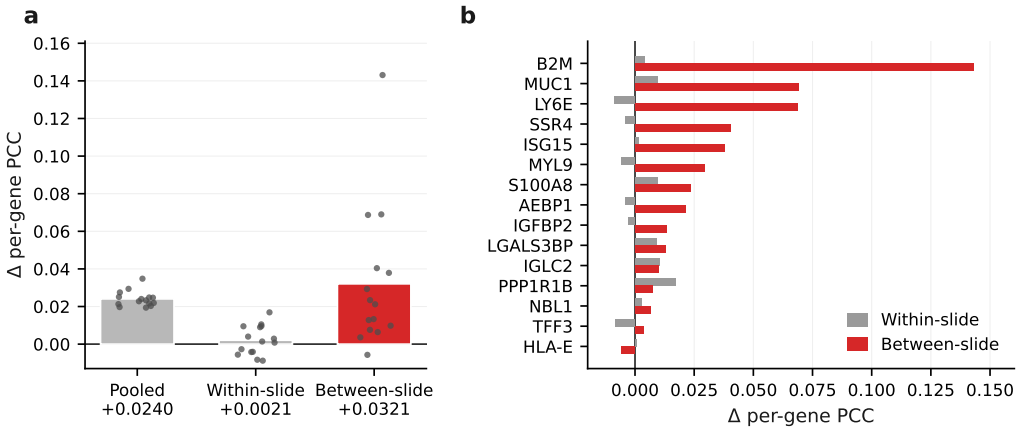

Figure 7: Decomposition of the residual gain on Her2ST. (a) Mean per-gene PCC gain of the top-15 gain genes under the pooled, within-slide, and between-slide variants, with dots showing individual genes. (b) Per-gene within-slide and between-slide gains for the same genes.