BRIDGING HISTOLOGY AND SPATIAL GENE EXPRESSION WITH LATENT CO-EXPRESSION PROGRAMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting spatial gene expression from hematoxylin and eosin (H&E)-stained histology could extend routinely acquired pathology images toward spatial molecular characterization. Existing methods primarily map local morphology to a selected target-gene panel, yet the same visual appearance can accompany different target-gene levels, and similar target expression can occur in morphologically distinct regions. Paired training spatial transcriptomics profiles also measure non-target genes whose coordinated expression provides complementary molecular context for these cases. To use this training signal, we introduce BRIDGE, a histology-based spatial gene expression prediction framework that learns cohort-adaptive co-expression programs from a disjoint support-gene set and converts them into target-specific visual predictors. The framework comprises three complementary modules: the Local Morphology Encoding Module (LME) extracts target-gene-conditioned morphology features; the Co-expression Program Learning Module (CPL) converts support-gene programs into queries for program-specific morphology aggregation; and the Gene-Adaptive Decoding Module (GAD) combines pooled and program-specific morphology with target-gene features through gene-adaptive experts. Across patient-disjoint evaluations, BRIDGE achieves the best MAE and MSE on all three datasets and the highest PCC on ANDERSSON and LUAD. On ANDERSSON, its PCC is 0.734 versus 0.710 for SpaHGC, while its MSE is 0.399 versus 0.497 for MERGE. Controlled comparisons, cross-cohort transfer, and ablations assess the roles of support-derived supervision, program-guided morphology aggregation, and gene-adaptive decoding.

1 INTRODUCTION

Spatial transcriptomics (ST) measures gene expression within tissue sections, providing spatially resolved molecular maps of cellular states, tissue organization, and microenvironmental interactions Ståhl et al. (2016). H&E-stained histology, in contrast, is routinely acquired in clinical pathology and captures tissue architecture, cellular morphology, and regional heterogeneity Bera et al. (2019). Inferring spatial gene expression from H&E images could therefore extend routinely available pathology slides from morphological assessment toward spatial molecular characterization Schmauch et al. (2020); Kather et al. (2020).

H&E-to-ST prediction is challenging because ST profiles contain high-dimensional dependencies across genes, whereas histology provides only indirect morphological evidence of the underlying molecular state Zeng et al. (2022b); Chelebian et al. (2025). Existing methods have strengthened this mapping through local visual encoding He et al. (2020), spatial-context modeling Zeng et al. (2022a), cross-modal alignment Xie et al. (2023), and generative learning Zhu et al. (2025).

However, spatial expression can vary substantially across regions within the same tumor Andersson et al. (2021), and these molecular differences may not be fully apparent from morphology alone. As illustrated in Fig. 1(a), two regions with similar H&E morphology may exhibit divergent target-gene expression, causing a morphology-driven model to underestimate molecular differences. Conversely, Fig. 1(b) shows that regions with distinct morphology may share similar expression of the same target gene, causing the model to overinterpret visual differences. Together, these complementary observations suggest that local morphology alone may not capture variation in target-gene expression.

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.

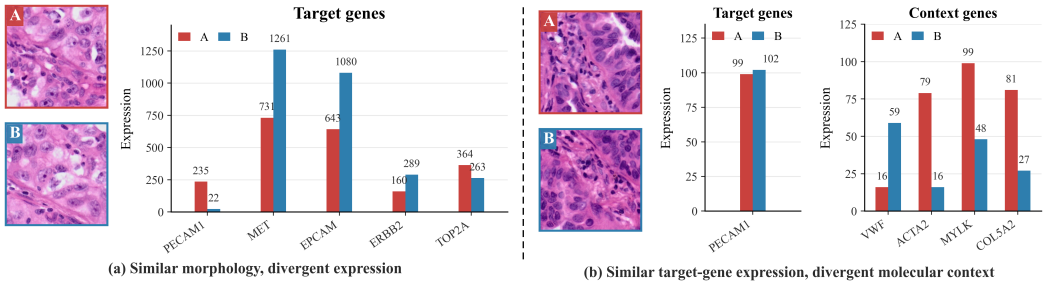

Figure 1: Motivation of BRIDGE using two LUAD slides: (a) similar morphology with divergent target-gene expression (TENX141); (b) different morphology with similar target-gene expression but distinct context-gene patterns (TENX118).

Modern spatial transcriptomics platforms can measure genome-scale expression profiles at spatially resolved spots Jaume et al. (2024). In standard H&E-to-ST benchmarks, methods evaluate a selected target-gene panel, often comprising genes chosen for high expression or variability Yang et al. (2023); Xie et al. (2023); Chung et al. (2024). In the H&E-to-ST methods considered here, prediction is typically formulated around a selected target-gene panel, whereas additional measured non-target genes are less commonly used as explicit training-time supervision to organize the visual representation for target-gene prediction. For example, the context-gene profiles in the right panel of Fig. 1(b) show that similar expression of the target gene can occur within distinct cross-gene contexts. Together with the fact that functionally related genes can participate in shared cellular processes and exhibit coordinated expression changes Subramanian et al. (2005), this motivates a more specific question: *can training-time supervision from non-target support genes help organize histological features into more informative, target-specific visual representations*?

Representative pathway-level methods, such as DeepPathway Ahsan et al. (2026), construct pathway-level targets for H&E-to-ST prediction by aggregating expression within predefined gene sets such as the MSigDB Hallmark collection Liberzon et al. (2015). However, the fixed gene memberships in these collections may miss cohort-specific co-expression patterns that vary with tissue type, cellular composition, or microenvironmental state Sonawane et al. (2017); Langfelder & Horvath (2008). For example, several immune-related transcripts may increase together in a tumor region with greater immune-cell abundance, while being distributed across multiple entries in a fixed collection such as MSigDB Hallmark Andersson et al. (2021); Liberzon et al. (2015). A representation based on a single predefined set would capture only part of this coordinated change, motivating a data-adaptive alternative for incorporating cross-gene context into H&E-to-ST prediction.

Building on this observation, we propose the Bridging Histology and Spatial Gene Expression with Latent Co-Expression Programs framework, referred to as BRIDGE. BRIDGE partitions the measured genes into a target set and a disjoint non-target support set, then derives cohort-adaptive latent programs using support-gene expression from the training slides only. First, the Local Morphology Encoding Module (LME) maps local morphology tokens and target-gene embeddings into a shared latent space, retaining token-level detail for program queries while providing a pooled morphology summary for decoding. Next, the Co-expression Program Learning Module (CPL) derives latent programs from support-gene expression and turns their gene loadings into queries. These queries aggregate program-relevant morphology, with auxiliary supervision aligning the result to program activity. Finally, the Gene-Adaptive Decoding Module (GAD) combines pooled morphology, program-specific morphology, and the target-gene embedding through adaptively weighted program experts. Support genes are training-only; inference uses histology and target genes.

Across patient-disjoint evaluations on ANDERSSON, LYMPH_IDC, and LUAD, BRIDGE achieved the lowest MAE and MSE on all three datasets and the highest PCC on ANDERSSON and LUAD. On ANDERSSON, PCC improved from 0.710 to 0.734 and MSE decreased from 0.497 to 0.399 relative to the strongest baseline; in the LUAD reciprocal evaluation, PCC improved from 0.705 to 0.760 and MSE decreased from 3.176 to 2.674. Cross-cohort transfer and controlled ablations further evaluate generalization and component contributions.

The main contributions of this work are summarized as follows:

- BRIDGE introduces a training-only support-gene paradigm that uses disjoint non-target expression measurements to inform histology-based target-panel prediction.
- BRIDGE converts support-gene co-expression into visual queries for program-relevant morphology and target-gene-adaptive expert routing under program activity supervision.
- Three patient-disjoint evaluations, cross-cohort transfer, and controlled ablations assess BRIDGE's predictive accuracy, transfer behavior, and component contributions.

2 Related Work

2.1 Histology-Based Spatial Gene Expression Prediction

Histology-based spatial gene expression prediction infers spatially resolved molecular measurements from H&E images. StNet established local patch-to-expression regression He et al. (2020), while later methods introduced exemplar guidance, contrastive pretraining, multi-resolution integration, diffusion modeling, and hierarchical graphs Yang et al. (2023); Xie et al. (2023); Chung et al. (2024); Zhu et al. (2025); Ganguly et al. (2025). These approaches improve image-to-expression prediction for selected target genes, but do not explicitly use expression from a disjoint non-target gene set as training-only information for constructing the predictor.

2.2 Spatial Context and Cross-Modal Representation Learning

Spatially aware methods model local neighborhoods, exemplars, and multi-scale graph relations Hu et al. (2021); Zhao et al. (2021); Dries et al. (2021); Yang et al. (2023); Shi et al. (2026); Ganguly et al. (2025). Contrastive methods align histology and expression to learn transferable morphology representations Xie et al. (2023), whereas STEM uses a histology-conditioned diffusion model to generate gene expression Zhu et al. (2025). These approaches organize context around patches, spots, or expression endpoints, whereas BRIDGE derives support-gene programs to guide morphology aggregation and target-gene decoding.

2.3 Gene-Aware and Cross-Gene Structure Modeling

DeepPathway incorporates predefined pathway priors into H&E-to-ST prediction Ahsan et al. (2026), whereas other recent approaches use pretrained transcriptomic or cross-modal gene representations Lee et al. (2024); Fang et al. (2026); Cui et al. (2024); Theodoris et al. (2023). These fixed or pretrained priors encode general gene relationships, but do not derive co-expression structure from non-target genes in the current training cohort. BRIDGE instead uses expression from a disjoint set of non-target genes on each training split to derive cohort-adaptive latent programs. This support-gene expression supplies training-only molecular context: it shapes the program queries and auxiliary supervision, but is not an inference input.

3 Methodology

3.1 Overview

We introduce BRIDGE, a framework for histology-based spatial gene expression prediction that converts training-only information from non-target genes into gene-specific visual predictors. Given a spot-centered H&E region and a predefined target-gene panel, BRIDGE predicts expression for each target gene. During training, a disjoint support-gene set is used to derive latent co-expression programs; support genes are training-only, and inference uses histology and target genes. As illustrated in Fig. 2, BRIDGE consists of three modules: the Local Morphology Encoding Module (LME), the Co-expression Program Learning Module (CPL), and the Gene-Adaptive Decoding Module (GAD), denoted by Φ_{LME}, Φ_{CPL}, and Φ_{GAD}, respectively. LME encodes local morphology and target-gene representations, CPL converts support-derived programs into program-specific morphology queries, and GAD combines pooled and program-specific morphology with target-gene features through adaptive program experts, with auxiliary supervision applied to program activity.

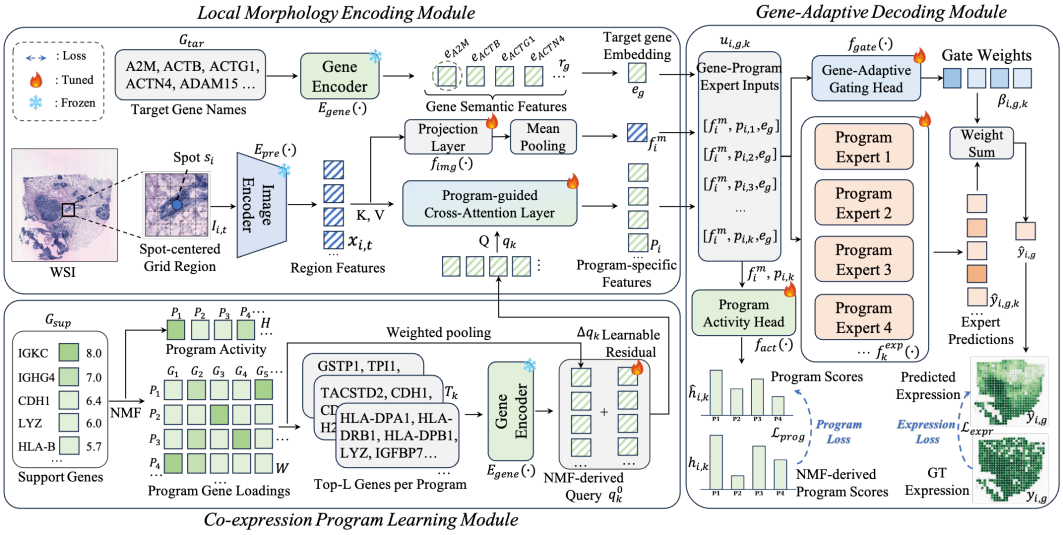

Figure 2: Overall framework of BRIDGE. The model encodes spot-centered local H&E morphology and target-gene representations, derives NMF-based program queries from support gene expression in the training slides, aggregates program-specific morphology evidence through cross-attention, and performs gene-adaptive decoding for target gene expression prediction.

3.2 Problem Formulation

Let $\mathcal{D} = \{(\mathcal{I}_m, \mathcal{S}_m, \mathbf{Y}_m)\}_{m=1}^{M}$ denote a dataset of M paired H&E whole-slide images and spatial transcriptomics measurements, where $\mathcal{I}_m$ is the m-th H&E whole-slide image, $\mathcal{S}_m = \{s_{m,i}\}_{i=1}^{N_m}$ is its set of registered spatial spots, and $\mathbf{Y}_m \in \mathbb{R}^{N_m \times G}$ contains the measured expression values of G target genes after preprocessing. For each spot $s_{m,i}$, we crop a spot-centered histology region $\mathcal{R}_{m,i}$ from $\mathcal{I}_m$ at its registered coordinate. For notational simplicity, we subsequently flatten the slide and within-slide spot indices and write s_i, $\mathbf{y}_i$, and $\mathcal{R}_i$ for the corresponding spot, target-gene expression vector, and spot-centered histology region, respectively, where $i = 1, \ldots, N$ and $N = \sum_{m=1}^{M} N_m$. The region $\mathcal{R}_i$ yields T morphology tokens $\mathcal{X}_i = \{\mathbf{x}_{i,t}\}_{t=1}^{T}$, with $\mathbf{x}_{i,t} \in \mathbb{R}^{D_{\text{img}}}$.

Let $\mathcal{G}_{\text{tar}}$ denote the target-gene set, with $G = |\mathcal{G}_{\text{tar}}|$. In current H&E-to-ST prediction, supervision is typically defined only on these target genes. BRIDGE augments this supervision with a disjoint set $\mathcal{G}_{\text{sup}}$ of non-target measured genes, termed support genes, with $G_s = |\mathcal{G}_{\text{sup}}|$. Their expression values across the N_{tr} training spots are collected in $\mathbf{X}_{\text{tr}}^{\text{sup}} \in \mathbb{R}_{+}^{N_{\text{tr}} \times G_s}$ and used during training to construct the program-query matrix $\mathbf{Q}$. Thus, the target genes define the prediction task, whereas the support genes provide training-only molecular information and are not inputs to the prediction pathway for held-out slides. The goal of BRIDGE is to learn a gene-adaptive prediction function that estimates each target gene through three coupled modules:

$$(\mathbf{Z}_i, \mathbf{f}_i^{\text{m}}, \mathbf{e}_g) = \Phi_{\text{LME}}(\mathcal{X}_i, \mathbf{r}_g),\ \mathbf{P}_i = \Phi_{\text{CPL}}(\mathbf{Z}_i, \mathbf{Q}),\ \hat{y}_{i,g} = \Phi_{\text{GAD}}(\mathbf{f}_i^{\text{m}}, \mathbf{P}_i, \mathbf{e}_g), \quad g = 1, \ldots, G. \tag{1}$$

Here, $\mathbf{r}_g$ denotes the pretrained representation of target gene g. The intermediate variables $\mathbf{Z}_i$, $\mathbf{f}_i^{\text{m}}$, $\mathbf{e}_g$, and $\mathbf{P}_i$ denote the projected morphology tokens, pooled morphology feature, target-gene embedding, and program-specific morphology representations, respectively. Collecting the predictions over the target panel gives $\hat{\mathbf{y}}_i = [\hat{y}_{i,1}, \ldots, \hat{y}_{i,G}]^{\top} \in \mathbb{R}^{G}$.

3.3 Local Morphology Encoding Module

For each spot s_i, we crop a local H&E region of context size ρ at its registered coordinate and partition it into a regular grid of $p \times p$-pixel tokens $I_{i,t}$, where $p = 224$. A frozen pretrained histology encoder then extracts each morphology feature as $\mathbf{x}_{i,t} = E_{\text{pre}}(I_{i,t})$.

Each morphology token is projected into a shared hidden space using a layer-normalized projection network. The projected token matrix and its mean-pooled spot representation are

$$\mathbf{z}_{i,t} = f_{\text{img}}(\mathbf{x}_{i,t}) \in \mathbb{R}^{D}, \qquad \mathbf{Z}_i = [\mathbf{z}_{i,1}, \ldots, \mathbf{z}_{i,T}] \in \mathbb{R}^{T \times D}, \qquad \mathbf{f}_i^{\text{m}} = \text{MeanPool}(\mathbf{Z}_i) \in \mathbb{R}^{D}. \tag{2}$$

The pooled feature $\mathbf{f}_i^{\mathrm{m}}$ provides a compact representation of the local morphology, while the full token set $\mathbf{Z}_i$ is retained for program-guided aggregation.

In parallel, each target gene g is represented by $\mathbf{r}_g = E_{\mathrm{gene}}(\mathrm{name}(g))$ and mapped into the shared hidden space as $\mathbf{e}_g = f_{\mathrm{gene}}(\mathbf{r}_g) \in \mathbb{R}^D$ for gene-specific prediction. Here, $\mathrm{name}(\cdot)$ maps a gene identifier to its vocabulary entry, and $E_{\mathrm{gene}}(\cdot)$ denotes lookup in the pretrained embedding table.

3.4 Co-expression Program Learning Module

To transform training-only non-target gene expression into a reusable predictor, BRIDGE constructs latent programs from the support-gene matrix. Let $\mathcal{G}_{\mathrm{all}}$ denote the measured gene set. Using the target-gene set $\mathcal{G}_{\mathrm{tar}}$ introduced above, we first select a top-ranked candidate gene panel $\mathcal{G}_{\mathrm{cand}}$ using a predefined gene selection criterion, such as expression variability or abundance, and then construct the support-gene set by excluding target genes:

$$\mathcal{G}_{\mathrm{cand}} = \mathrm{Top}_{G_c}(\mathcal{G}_{\mathrm{all}}; \psi), \qquad \mathcal{G}_{\mathrm{sup}} = \mathcal{G}_{\mathrm{cand}} \setminus \mathcal{G}_{\mathrm{tar}}. \tag{3}$$

where ψ denotes the dataset-specific gene selection criterion and $G_c = |\mathcal{G}_{\mathrm{cand}}|$ is the size of the candidate gene panel. By construction, $\mathcal{G}_{\mathrm{sup}} \cap \mathcal{G}_{\mathrm{tar}} = \emptyset$, preventing target-gene expression from being directly used in program construction. To summarize support-gene co-expression in a low-dimensional, non-negative form, we use non-negative matrix factorization (NMF) to derive data-adaptive latent programs, represented by spot-level program scores and program-level gene loadings Lee & Seung (1999). For each training split, NMF is fitted using support-gene expression from training slides only. The matrix $\mathbf{X}_{\mathrm{tr}}^{\mathrm{sup}} \in \mathbb{R}_+^{N_{\mathrm{tr}} \times G_s}$ introduced above collects the expression values of these support genes across all training spots, with rows corresponding to spots and columns corresponding to support genes. NMF factorizes this matrix as

$$\mathbf{X}_{\mathrm{tr}}^{\mathrm{sup}} \approx \mathbf{H}\mathbf{W}, \tag{4}$$

where $\mathbf{H} \in \mathbb{R}_+^{N_{\mathrm{tr}} \times K}$ is the spot-by-program score matrix, with each entry indicating the score of a latent co-expression program in a training spot. $\mathbf{W} \in \mathbb{R}_+^{K \times G_s}$ is the program-by-support-gene loading matrix, where each row represents the support-gene signature of one latent co-expression program, and K is the number of latent programs. For each program k, the top m support genes by NMF loading are encoded with the frozen gene encoder:

$$\mathcal{T}_k = \mathrm{Top}_m\{g \in \mathcal{G}_{\mathrm{sup}} \mid W_{k,g}\}, \qquad \mathbf{s}_g = E_{\mathrm{gene}}(\mathrm{name}(g)), \quad g \in \mathcal{T}_k. \tag{5}$$

The base program query is a loading-weighted pooling of these gene features, with weights normalized within the selected gene set:

$$\alpha_{k,g} = \frac{W_{k,g}}{\sum_{g' \in \mathcal{T}_k} W_{k,g'}}, \qquad \mathbf{q}_k^0 = \sum_{g \in \mathcal{T}_k} \alpha_{k,g} \mathbf{s}_g. \tag{6}$$

The NMF-derived query preserves co-expression structure but is not optimized for the histology feature space. We therefore add a learnable residual for task-specific alignment:

$$\mathbf{q}_k = \mathbf{q}_k^0 + \Delta\mathbf{q}_k, \qquad \mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_K]^\top \in \mathbb{R}^{K \times D}. \tag{7}$$

The Co-expression Program Learning Module Φ_{CPL} then projects these queries and aggregates program-specific morphology from the local histology tokens $\mathbf{Z}_i$:

$$\widetilde{\mathbf{Q}} = f_{\mathrm{qry}}(\mathbf{Q}), \qquad \mathbf{P}_i = \mathrm{MHA}(\widetilde{\mathbf{Q}}, \mathbf{Z}_i, \mathbf{Z}_i). \tag{8}$$

Here, $f_{\mathrm{qry}}(\cdot)$ is a trainable linear projection followed by layer normalization, and $\mathrm{MHA}(\cdot)$ denotes multi-head attention. The output $\mathbf{P}_i = [\mathbf{p}_{i,1}, \ldots, \mathbf{p}_{i,K}]^\top \in \mathbb{R}^{K \times D}$ contains the local morphology evidence aggregated under each latent co-expression program.

3.5 Gene-Adaptive Decoding Module

The latent programs capture distinct co-expression patterns and their associated morphological evidence. Because their contributions can vary across target genes and spots, the Gene-Adaptive Decoding Module assigns a separate prediction expert to each program and learns how strongly each expert should contribute to a given target. For target gene g and program k at spot i, the expert

Table 1: In-domain patient-disjoint evaluation (a)–(c) and cross-dataset transfer (d). Results are reported as mean ± standard deviation across outer evaluations. Best and second-best results are shown in bold and underline, respectively.

(a) ANDERSSON

Method	PCC ↑	MAE ↓	MSE ↓
StNet	0.563 ± 0.046	0.631 ± 0.089	0.696 ± 0.191
EGN	0.528 ± 0.050	0.792 ± 0.059	1.279 ± 0.135
BLEEP	0.565 ± 0.062	0.624 ± 0.078	0.738 ± 0.162
TRIPLEX	0.652 ± 0.049	0.589 ± 0.048	0.602 ± 0.094
STEM	0.583 ± 0.045	0.619 ± 0.082	0.726 ± 0.176
MERGE	0.684 ± 0.031	0.532 ± 0.070	0.497 ± 0.126
SpaHGC	0.710 ± 0.040	0.574 ± 0.076	0.558 ± 0.142
BRIDGE	**0.734 ± 0.033**	**0.472 ± 0.064**	**0.399 ± 0.112**

(b) LYMPH_IDC

Method	PCC ↑	MAE ↓	MSE ↓
StNet	0.421 ± 0.031	1.049 ± 0.125	2.331 ± 0.320
EGN	0.508 ± 0.102	1.101 ± 0.059	2.549 ± 0.182
BLEEP	0.424 ± 0.077	1.140 ± 0.098	2.711 ± 0.273
TRIPLEX	0.535 ± 0.095	1.091 ± 0.101	2.537 ± 0.566
STEM	0.308 ± 0.063	1.103 ± 0.177	2.736 ± 0.789
MERGE	0.427 ± 0.081	1.042 ± 0.174	2.372 ± 0.706
SpaHGC	**0.574 ± 0.123**	1.072 ± 0.074	2.199 ± 0.197
BRIDGE	0.569 ± 0.120	**0.956 ± 0.048**	**1.937 ± 0.116**

(c) LUAD

Method	PCC ↑	MAE ↓	MSE ↓
StNet	0.633 ± 0.021	1.353 ± 0.041	3.176 ± 0.121
EGN	0.579 ± 0.058	1.481 ± 0.082	3.869 ± 0.295
BLEEP	0.517 ± 0.041	1.463 ± 0.024	3.628 ± 0.192
TRIPLEX	0.705 ± 0.017	1.358 ± 0.057	3.218 ± 0.090
STEM	0.558 ± 0.027	1.386 ± 0.024	3.315 ± 0.055
MERGE	0.703 ± 0.012	1.339 ± 0.121	3.327 ± 0.737
SpaHGC	0.659 ± 0.020	1.362 ± 0.013	3.223 ± 0.106
BRIDGE	**0.760 ± 0.007**	**1.256 ± 0.037**	**2.674 ± 0.108**

(d) ANDERSSON → LYMPH_IDC

Method	PCC ↑	MAE ↓	MSE ↓
StNet	0.224 ± 0.054	1.199 ± 0.029	2.319 ± 0.112
EGN	0.109 ± 0.061	1.309 ± 0.051	2.804 ± 0.216
BLEEP	0.154 ± 0.032	1.328 ± 0.024	2.838 ± 0.081
TRIPLEX	0.154 ± 0.049	1.258 ± 0.012	2.500 ± 0.034
STEM	0.105 ± 0.014	1.363 ± 0.022	3.200 ± 0.097
MERGE	0.219 ± 0.065	1.201 ± 0.045	2.358 ± 0.121
SpaHGC	0.222 ± 0.089	1.307 ± 0.072	2.785 ± 0.256
BRIDGE	**0.226 ± 0.038**	**1.146 ± 0.019**	**2.144 ± 0.057**

input concatenates the pooled morphology, program-specific morphology, and target-gene features as $\mathbf{u}_{i,g,k} = [\mathbf{f}_i^{\mathrm{m}}, \mathbf{p}_{i,k}, \mathbf{e}_g] \in \mathbb{R}^{3D}$.

The gene-adaptive prediction head uses K independently parameterized MLP experts $\{f_k^{\exp}\}_{k=1}^{K}$ shared across target genes. Given gene-specific inputs, the gate normalizes their program-wise estimates to form the final prediction:

$$\begin{aligned} \hat{y}_{i,g,k} &= f_k^{\exp}(\mathbf{u}_{i,g,k}), \qquad a_{i,g,k} = f_{\mathrm{gate}}(\mathbf{u}_{i,g,k}), \\ \beta_{i,g,k} &= \frac{\exp(a_{i,g,k})}{\sum_{k'=1}^{K} \exp(a_{i,g,k'})}, \qquad \hat{y}_{i,g} = \sum_{k=1}^{K} \beta_{i,g,k} \hat{y}_{i,g,k}. \end{aligned} \tag{9}$$

The experts specialize by latent program and are shared across target genes, while the gate produces spot- and gene-specific weights through $\mathbf{u}_{i,g,k}$. Their weighted combination therefore adapts the contribution of each co-expression program to the target being predicted.

To preserve the program information in the retrieved morphology features, a shared Program Activity Head predicts $\hat{h}_{i,k} = f_{\mathrm{act}}([\mathbf{p}_{i,k}, \mathbf{f}_i^{\mathrm{m}}])$ for each program k. The head consists of LayerNorm, a linear layer, GELU, dropout, and a final linear projection to a scalar score. The resulting vector $\hat{\mathbf{h}}_i \in \mathbb{R}^K$ is supervised by the corresponding NMF-derived spot-level program scores in $\mathbf{H}$.

3.6 Optimization Objective

The model is optimized with a primary target-gene expression loss and an auxiliary NMF-derived program score loss:

$$\mathcal{L}_{\mathrm{expr}} = \frac{1}{N_{\mathrm{tr}} G} \sum_{i=1}^{N_{\mathrm{tr}}} \sum_{g=1}^{G} (\hat{y}_{i,g} - y_{i,g})^2, \qquad \mathcal{L}_{\mathrm{prog}} = \frac{1}{N_{\mathrm{tr}} K} \sum_{i=1}^{N_{\mathrm{tr}}} \sum_{k=1}^{K} (\hat{h}_{i,k} - h_{i,k})^2. \tag{10}$$

Here, $h_{i,k}$ is the NMF-derived program score pseudo-label, and both losses are averaged over the N_{tr} training spots in the current split. We jointly optimize target-gene prediction and program-activity estimation through $\mathcal{L} = \mathcal{L}_{\mathrm{expr}} + \lambda_{\mathrm{prog}} \mathcal{L}_{\mathrm{prog}}$, where λ_{prog} controls the contribution of the auxiliary program objective. The auxiliary loss preserves train-derived cross-gene structure in program-guided morphology features, while the primary loss supervises target-gene prediction.

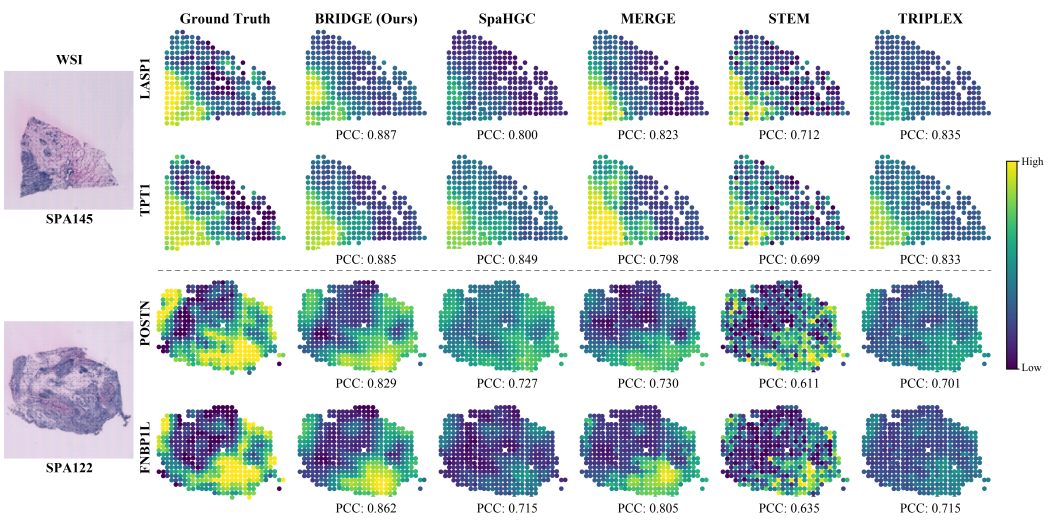

Figure 3: Spatial prediction visualization on ANDERSSON outer test slides. WSI thumbnails are shown on the left, followed by ground truth, BRIDGE, and baseline predictions. Rows correspond to selected target genes, with PCC reported below each predicted map.

4 Experiments and Results

4.1 Experimental Setup

Datasets. We evaluated BRIDGE on three HEST-1K cohorts Jaume et al. (2024) pairing H&E images with spatial transcriptomics profiles: ANDERSSON, LYMPH_IDC, and LUAD. Dataset statistics are reported in Appendix Table A1. To ensure comparability with established benchmarks, we used one fixed target-gene panel per dataset, shared by all methods and patient-disjoint splits Yang et al. (2023); Xie et al. (2023); Chung et al. (2024); Ganguly et al. (2025). Specifically, *ANDERSSON* contains 36 slides from eight patients and was evaluated with five-fold patient-grouped cross-validation; across folds, 18–24 slides were used for training, 6–9 for validation, and 6–9 for testing because patients contribute unequal numbers of slides. *LYMPH_IDC* contains four slides from four patients and was evaluated in four patient-disjoint runs, each using two patients for training, one for validation, and one for testing, with the test patient rotated so that each patient was evaluated once. *LUAD* contains two slides from two patients and was included as a small-cohort reciprocal evaluation, with each slide serving once for training and once for testing, and approximately 20% of the training-slide spots reserved as a contiguous validation block for checkpoint selection.

Implementation. BRIDGE used frozen UNI v1 features extracted from 224×224-pixel patches Chen et al. (2024) and frozen scGPT gene embeddings Cui et al. (2024). Each spot was represented by a 5×5 local grid, with hidden dimension $D = 512$ and $K = 12$ latent programs initialized from the top 100 support genes per program. Training used AdamW with a learning rate of 1×10^{-4}, batch size 128, and program loss weight $\lambda_{\text{prog}} = 0.1$. Checkpoints were selected by validation MAE. Additional preprocessing, NMF solver, network, training-schedule, early-stopping, and hardware details are provided in Appendix A.2.

Evaluation metrics. Within each split, predictions and ground-truth expression values from all outer test slides were combined before computing the metrics. For each target gene, PCC was computed over the combined test spots, and the resulting gene-wise PCC values were averaged. MAE and MSE were computed over all spot–gene pairs in the combined test set. We then calculated the mean and standard deviation of each metric across the patient-level outer folds or splits. For domain transfer, each ANDERSSON source-fold checkpoint was evaluated separately on each of the four LYMPH_IDC target slides using the matched transfer-gene panel. The four target-slide results were averaged for each source fold, and the five source-fold averages were then summarized. Higher PCC and lower errors indicate better performance.

Table 2: Module-subtraction ablations on ANDERSSON and LYMPH_IDC. All remaining settings and matched outer folds are fixed. Results are mean ± standard deviation; higher PCC and lower MAE and MSE indicate better performance, and best values are bolded.

	ANDERSSON			LYMPH_IDC		
Setting	PCC ↑	MAE ↓	MSE ↓	PCC ↑	MAE ↓	MSE ↓
w/o NMF program prior	0.719 ± 0.035	0.487 ± 0.066	0.427 ± 0.118	0.536 ± 0.131	0.989 ± 0.057	2.061 ± 0.151
w/o program-guided aggregation	0.722 ± 0.034	0.484 ± 0.063	0.422 ± 0.115	0.548 ± 0.126	0.975 ± 0.052	2.005 ± 0.134
w/o program activity supervision	0.728 ± 0.033	0.478 ± 0.065	0.413 ± 0.114	0.552 ± 0.123	0.970 ± 0.051	1.991 ± 0.129
w/o gene-adaptive decoding	0.707 ± 0.037	0.496 ± 0.068	0.446 ± 0.124	0.521 ± 0.138	1.006 ± 0.063	2.127 ± 0.178
BRIDGE	**0.734 ± 0.033**	**0.472 ± 0.064**	**0.399 ± 0.112**	**0.569 ± 0.120**	**0.956 ± 0.048**	**1.937 ± 0.116**

4.2 Comparisons With State-of-the-Art Methods

We compared BRIDGE with representative state-of-the-art methods for histology-based spatial gene expression prediction, including StNet He et al. (2020), EGN Yang et al. (2023), BLEEP Xie et al. (2023), TRIPLEX Chung et al. (2024), STEM Zhu et al. (2025), MERGE Ganguly et al. (2025), and SpaHGC Shi et al. (2026). All methods were evaluated using the same patient-disjoint partitions, dataset-level target-gene panels, and split-wise evaluation protocol. Baselines were run from their official implementations or reimplemented according to the settings described in the original papers, with only dataset-specific adaptations required by the input format. BRIDGE achieved the lowest MAE and MSE on all three datasets and the highest PCC on ANDERSSON and LUAD. On ANDERSSON, PCC increased from 0.710 for SpaHGC to 0.734, while MAE and MSE decreased from 0.532 and 0.497 for the corresponding best baselines to 0.472 and 0.399. On LYMPH_IDC, SpaHGC had slightly higher PCC, whereas BRIDGE achieved the lowest errors despite comparable spatial correlation; the paired PCC comparison based on the four patient-level outer runs was not significant under a paired t-test ($p = 0.678$). In the LUAD reciprocal evaluation, BRIDGE improved PCC by 0.055 and reduced MAE and MSE by 0.083 and 0.502 relative to the best baselines.

4.3 Domain Transfer Experiments

We tested domain transfer by applying models trained on ANDERSSON directly to LYMPH_IDC, without fine-tuning or parameter updates. Both cohorts come from breast-cancer studies but represent different tissue environments, providing a focused test of cross-cohort transfer. At inference, BRIDGE uses only histology and target-gene identities, enabling direct application to the target cohort. Each checkpoint from the five ANDERSSON folds was evaluated on all four LYMPH_IDC slides using the fixed 196-gene transfer panel. This panel contains the genes from the 200-gene ANDERSSON target panel that could be matched to LYMPH_IDC by exact gene name. It was not used to train or select the ANDERSSON checkpoints. We first averaged the four target-slide results for each source fold and then summarized the five source-fold averages. As reported in Table 1(d), BRIDGE achieved a PCC comparable to that of the strongest baseline (0.226 versus 0.224), while reducing MAE from 1.199 to 1.146 and MSE from 2.319 to 2.144. The modest PCC values across methods indicate that transfer remains difficult because cohort-specific tissue composition, imaging conditions, and morphology-expression associations can shift between the source and target cohorts.

4.4 Visualization and Case Studies

Figure 3 presents two held-out ANDERSSON cases with different spatial expression patterns. On SPA145, BRIDGE reproduced the left-to-right expression gradients of LASP1 and TPT1, achieving PCCs of 0.887 and 0.885, respectively. On SPA122, POSTN and FNBP1L exhibited more heterogeneous spatial patterns, with high-expression regions distributed along the tissue periphery and lower portion. BRIDGE preserved these major high- and low-expression regions and achieved PCCs of 0.829 and 0.862, exceeding the strongest displayed baseline for each gene. Although some fine-scale spot variation was attenuated, the predictions retained the main gene-specific spatial organization observed in the ground truth. The main visualization focuses on target-gene prediction maps in ANDERSSON. As a complementary view of the learned program representations, Appendix A.3 visualizes their predicted spot-level scores $\hat{h}_{i,k}$ on LYMPH_IDC.

Table 3: Query-design and image-encoder controls on ANDERSSON and LYMPH_IDC. Results are reported as mean ± standard deviation across matched patient-level outer folds or splits. Higher PCC and lower MAE and MSE indicate better performance; the reference setting in each control block is bolded.

	ANDERSSON			LYMPH_IDC		
Setting	PCC ↑	MAE ↓	MSE ↓	PCC ↑	MAE ↓	MSE ↓
Query-design controls						
Fixed random queries	0.721 ± 0.035	0.487 ± 0.062	0.421 ± 0.109	0.542 ± 0.125	0.983 ± 0.055	2.040 ± 0.148
Shuffled NMF mapping	0.718 ± 0.036	0.490 ± 0.066	0.426 ± 0.115	0.531 ± 0.128	1.000 ± 0.058	2.082 ± 0.155
NMF-derived queries	**0.734 ± 0.033**	**0.472 ± 0.064**	**0.399 ± 0.112**	**0.569 ± 0.120**	**0.956 ± 0.048**	**1.937 ± 0.116**
Image-encoder controls						
ResNet50	0.706 ± 0.037	0.498 ± 0.054	0.448 ± 0.102	0.530 ± 0.132	0.995 ± 0.060	2.084 ± 0.160
CONCH	0.714 ± 0.033	0.486 ± 0.061	0.426 ± 0.117	0.548 ± 0.126	0.974 ± 0.054	2.012 ± 0.142
UNI v1	**0.734 ± 0.033**	**0.472 ± 0.064**	**0.399 ± 0.112**	**0.569 ± 0.120**	**0.956 ± 0.048**	**1.937 ± 0.116**

4.5 Ablation Studies

Table 2 reports module-subtraction results, while Table 3 reports query-design and image-encoder controls. Additional controls and sensitivity results are provided in the Appendix.

4.5.1 Effect of Each Module in BRIDGE

We removed one component at a time while keeping the remaining settings unchanged. In Table 2, the first ablation row, w/o NMF program prior, replaces the NMF-derived queries with learnable queries of the same size. The second row, w/o program-guided aggregation, replaces program-guided cross-attention with shared mean-pooled morphology. The third row, w/o program activity supervision, sets $\lambda_{\text{prog}} = 0$. The fourth row, w/o gene-adaptive decoding, replaces the K program-specific experts and their gene-conditioned gate with a single shared decoder that uses the pooled program context together with pooled morphology and target-gene features. For each of the four variants, removing the corresponding component reduced PCC and increased both prediction errors on both datasets (Table 2). Relative to the full model, the largest PCC decreases were 0.027 on ANDERSSON and 0.048 on LYMPH_IDC, while the largest MSE increases were 0.047 and 0.190, respectively. These consistent directional changes support the role of each component.

4.5.2 Effect of Query Design and Image-Encoder Choice

Query-design controls. We compared the NMF-derived queries with two alternatives of the same number and dimensionality. Fixed random queries were initialized once and kept unchanged, while shuffled NMF mapping permuted the correspondence between NMF loadings and support-gene embeddings. The remaining BRIDGE architecture was retained in all variants. Table 3 shows that both alternatives reduced PCC and increased prediction errors on both datasets. PCC decreased by 0.013–0.016 on ANDERSSON and by 0.027–0.038 on LYMPH_IDC. The shuffled NMF mapping performed worst on LYMPH_IDC, indicating the importance of preserving the NMF-derived correspondence between program loadings and support-gene embeddings.

Image-encoder controls. We replaced UNI v1 Chen et al. (2024) with ResNet50 He et al. (2016) or CONCH Lu et al. (2024). On ANDERSSON, UNI v1 achieved a PCC of 0.734, compared with 0.706 for ResNet50 and 0.714 for CONCH. The corresponding PCC values on LYMPH_IDC were 0.569, 0.530, and 0.548, respectively, with UNI v1 also yielding the lowest MAE and MSE on both datasets. These results show that encoder choice affects absolute performance and that UNI v1 provides the strongest representation among the tested encoders.

5 Conclusion

BRIDGE provides a histology-to-ST framework that uses non-target support-gene measurements during training to learn latent co-expression programs for morphology aggregation and target-gene-adaptive decoding. Across the patient-disjoint evaluations, BRIDGE achieved the lowest MAE and MSE on all three datasets and the highest PCC on ANDERSSON and LUAD. In cross-cohort trans-

fer, it achieved comparable PCC and lower MAE and MSE than the strongest baseline. Module ablations consistently reduced performance, supporting the contribution of the support-derived program pathway and gene-adaptive decoding. Together, these results demonstrate that training-time cross-gene context provides a useful complement to histology for target-gene prediction.

AI USE STATEMENT

The authors did not use generative AI tools in conducting the research or preparing this manuscript.

REFERENCES

Muhammad Ahtazaz Ahsan, Karen Piper Hanley, Martin Fergie, Claire O'Leary, Gerben Borst, Federico Roncaroli, Fayyaz Minhas, Magnus Rattray, Mudassar Iqbal, and Syed Murtuza Baker. DeepPathway: Predicting pathway expression from histopathology images. *Bioinformatics*, pp. btag643, 2026.

Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z. Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, Jonas Frisén, Camilla Engblom, and Joakim Lundeberg. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nature Communications*, 12:6012, 2021.

Kaustav Bera, Kurt A. Schalper, David L. Rimm, Vamsidhar Velcheti, and Anant Madabhushi. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, 16(11):703–715, 2019.

Christos Boutsidis and Efstratios Gallopoulos. Svd based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4):1350–1362, 2008.

Eduard Chelebian, Christophe Avenel, and Carolina Wählby. Combining spatial transcriptomics with tissue morphology. *Nature Communications*, 16:4452, 2025.

Richard J. Chen, Tong Ding, Ming Y. Lu, Drew F. K. Williamson, Guillaume Jaume, Andrew H. Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, Mane Williams, Lukas Oldenburg, Luca L. Weishaupt, Judy J. Wang, Anurag Vaidya, Long Phi Le, Georg Gerber, Sharifa Sahai, Walt Williams, and Faisal Mahmood. 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11591–11600, 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*, 21:1470–1480, 2024.

Ruben Dries, Qian Zhu, Rui Dong, Chee-Huat Linus Eng, Huipeng Li, Kan Liu, Yuntian Fu, Tianxiao Zhao, Arpan Sarkar, Feng Bao, Rani E. George, Nico Pierson, Long Cai, and Guo-Cheng Yuan. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biology*, 22(1):78, 2021.

Donghai Fang, Yongheng Li, Zhen Wang, Yuansong Zeng, and Wenwen Min. Adapting a pretrained single-cell foundation model to spatial gene expression generation from histology images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5720–5729, June 2026.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15611–15620, June 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*, 4:827–834, 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.

Jian Hu, Xiangjie Li, Kyle Coleman, Amelia Schroeder, Nan Ma, David J. Irwin, Edward B. Lee, Russell T. Shinohara, and Mingyao Li. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods*, 18(11):1342–1351, 2021.

Guillaume Jaume, Paul Doucet, Andrew H. Song, Ming Y. Lu, Cristina Almagro-Pérez, Sophia J. Wagner, Anurag J. Vaidya, Richard J. Chen, Drew F. K. Williamson, Ahrong Kim, and Faisal Mahmood. HEST-1k: A dataset for spatial transcriptomics and histology image analysis. *arXiv preprint arXiv:2406.16192*, 2024.

Jakob Nikolas Kather, Lara R. Heij, Heike I. Grabsch, Chiara Loeffler, Amelie Echle, Hannah Sophie Muti, Jeremias Krause, Jan M. Niehues, Kai A. J. Sommer, Peter Bankhead, Loes F. S. Kooreman, Jefree J. Schulte, Nicole A. Cipriani, Roman D. Buelow, Peter Boor, Nadina Ortiz-Brüchle, Andrew M. Hanby, Valerie Speirs, Sara Kochanny, Akash Patnaik, Andrew Srisuwananukorn, Hermann Brenner, Michael Hoffmeister, Piet A. van den Brandt, Dirk Jäger, Christian Trautwein, Alexander T. Pearson, and Tom Luedde. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer*, 1(8):789–799, 2020.

Peter Langfelder and Steve Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9:559, 2008.

Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

Yongju Lee, Xinhao Liu, Minsheng Hao, Tianyu Liu, and Aviv Regev. PathOmCLIP: Connecting tumor histology with spatial gene expression via locally enhanced contrastive learning of pathology and single-cell foundation model. *bioRxiv*, 2024.

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.

Ming Y. Lu, Bowen Chen, Drew F. K. Williamson, Richard J. Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, Anil V. Parwani, Andrew Zhang, and Faisal Mahmood. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024.

Benoît Schmauch, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro, Aurélie Kamoun, Meriem Sefta, Sylvain Toldo, Mikhail Zaslavskiy, Thomas Clozel, Matahi Moarii, Pierre Courtiol, and Gilles Wainrib. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nature Communications*, 11(1):3877, 2020.

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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5710–5719, June 2026.

Abhijeet R. Sonawane, John Platig, Maud Fagny, Chao-Yu Chen, Joseph N. Paulson, Camila M. Lopes-Ramos, Dawn L. DeMeo, John Quackenbush, Kimberly Glass, and Marieke L. Kuijjer. Understanding tissue-specific gene regulation. *Cell Reports*, 21(4):1077–1088, 2017.

Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O. Westholm, Mikael Huss, Annelie Mollbrink, Sten Linnarsson, Simone Codeluppi, Åke Borg, Fredrik Pontén, Paul Igor Costea, Pelin Sahlén, Jan Mulder, Olaf Bergmann, Joakim Lundeberg, and Jonas Frisén. 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, and Jill P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550, 2005.

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, and Patrick T. Ellinor. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

Ronald Xie, Kuan Pang, Sai W. Chung, Catia T. Perciani, Sonya A. MacParland, Bo Wang, and Gary D. Bader. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 70626–70637, 2023.

Yan Yang, Md Zakir Hossain, Eric A. Stone, and Shafin Rahman. Exemplar guided deep neural network for spatial transcriptomics analysis of gene expression prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5039–5048, 2023.

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, 2022a.

Zexian Zeng, Yawei Li, Yiming Li, and Yuan Luo. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biology*, 23(1):83, 2022b.

Edward Zhao, Matthew R. Stone, Xing Ren, Jamie Guenthoer, Kimberly S. Smythe, Thomas Pulliam, Stephen R. Williams, Cedric R. Uytingco, Sarah E. B. Taylor, Paul Nghiem, Jason H. Bielas, and Raphael Gottardo. Spatial transcriptomics at subspot resolution with BayesSpace. *Nature Biotechnology*, 39(11):1375–1384, 2021.

Sichen Zhu, Yuchen Zhu, Molei Tao, and Peng Qiu. Diffusion generative modeling for spatially resolved gene expression inference from histology images. In *The Thirteenth International Conference on Learning Representations*, 2025.

A Additional Experimental Details and Analyses

A.1 Dataset Statistics

The three HEST-1K cohorts Jaume et al. (2024) differ in cohort size, spot count, and gene coverage. ANDERSSON has the most slides, LYMPH_IDC has the most spots and the broadest measured gene set, and LUAD comprises two slides measured with a targeted Xenium panel. Table A1 reports the original spot-level expression matrices before dataset-level target and support gene selection; the range for ANDERSSON reflects differences in measured gene coverage across slides.

Table A1: Dataset statistics used in the experiments. Measured genes denote the number of genes in the original spot-level expression matrix before target gene selection. Support-gene counts denote the genes used by BRIDGE after target-gene selection.

Dataset	Slides	Spots	Measured genes	Target genes	Support genes
ANDERSSON	36	13,612	14,861–15,842	200 HVGs	800
LYMPH_IDC	4	19,964	33,931	50 mean genes	950
LUAD	2	7,505	541	50 mean genes	148

The target-gene counts denote the prediction endpoints rather than all molecular measurements available during training. We followed the dataset-level panel convention used in prior H&E-to-ST benchmarks: a target-gene panel is selected once for each dataset and then kept fixed across methods and patient-level splits Yang et al. (2023); Xie et al. (2023); Chung et al. (2024); Ganguly et al. (2025). BRIDGE retains this convention for fair comparison and additionally forms a disjoint support-gene panel from the remaining genes in the corresponding dataset-level candidate panel.

For ANDERSSON, the main experiments used a dataset-level panel of 1,000 highly expressed and variable genes; the top 200 formed the target panel and the remaining 800 formed the support panel. For LYMPH_IDC, genes were ranked by mean expression across all four slides to form a fixed 1,000-gene panel; the top 50 formed the target panel and the remaining 950 formed the support panel. For LUAD, filtering non-biological Xenium features yielded 198 eligible biological genes; the 50 genes in the mean-expression target panel were used as target genes and the remaining 148 as support genes. Target and support gene sets were disjoint in all datasets, i.e., $\mathcal{G}_{\text{tar}} \cap \mathcal{G}_{\text{sup}} = \varnothing$. For cross-cohort transfer, the 200-gene ANDERSSON panel was matched to LYMPH_IDC by exact gene name, yielding a fixed 196-gene panel for all source–target pairs.

A.2 Implementation Details

Preprocessing and input representation. Following a commonly used benchmark protocol in histology-based ST prediction Chung et al. (2024), target-gene expression values were normalized to a total of 10^4 per spot and then log-transformed. After the data split was fixed, target-gene expression was smoothed over the available 3×3 neighborhood separately within each subset, without crossing subset boundaries. The same target-gene preprocessing and prediction endpoints were used for BRIDGE and all compared methods within each split. Support-gene expression was separately normalized to 10^4 per spot and log-transformed, without spatial smoothing, before NMF-based program construction. UNI features Chen et al. (2024) were deterministically precomputed at the native slide resolution, producing morphology tokens with $D_{\text{img}} = 1024$. Each local grid contained at most $T = 25$ tokens, forming $\mathcal{X}_i = \{\mathbf{x}_{i,t}\}_{t=1}^{T}$, and the image projection network mapped them to $\mathbf{Z}_i \in \mathbb{R}^{T \times D}$. Target genes used scGPT whole-human embeddings Cui et al. (2024).

Program construction and model setting. For each data split, NMF was fitted only on support-gene expression from the training observations, excluding all validation and outer-test observations. For LUAD, the contiguous validation block was excluded before NMF fitting. NMF used NNDSVDa initialization Boutsidis & Gallopoulos (2008), a coordinate-descent solver, a maximum of 1,000 iterations, a convergence tolerance of 10^{-4}, and a random seed of 2021. The image projection consisted of LayerNorm, a linear layer mapping 1024-dimensional UNI features to the $D = 512$ hidden space, GELU, and dropout. Target-gene and program-query projections used a linear layer

followed by layer normalization, and the cross-attention layer used 8 attention heads. The Program Activity Head took the concatenation of a program-specific morphology feature and the pooled morphology feature, with input dimension $2D$, and applied LayerNorm, a linear layer mapping $2D$ inputs to D hidden units, GELU, dropout, and a final linear layer mapping the hidden representation to one output. Each of the $K = 12$ program experts received the $3D$-dimensional concatenation of pooled morphology, program-specific morphology, and target-gene features. Each expert used LayerNorm, a linear layer mapping $3D$ inputs to D hidden units, GELU, dropout, and a final projection to one expression value. The gene-adaptive gate used the same input and hidden dimensions and produced one scalar routing logit for each program. For baseline methods, we followed the official implementations and the training procedures specified in the corresponding papers whenever available. Dataset-specific adaptations were limited to the input format, target-gene dimensionality, and fixed patient-level splits. Baselines were evaluated under a common protocol while retaining method-specific model settings.

Training configuration. The dropout rate was set to 0.1, non-negative output clamping was enabled, and mixed-precision training was used. The maximum number of epochs was 1,000 for ANDERSSON and 200 for LYMPH_IDC and LUAD. Training stopped when validation MAE did not improve for 20 consecutive epochs. A ReduceLROnPlateau scheduler was used with a reduction factor of 0.1 and patience of 10 epochs for the ANDERSSON dataset and 5 epochs for LYMPH_IDC and LUAD. All models were implemented in PyTorch and trained on a single NVIDIA A800 GPU.

A.3 Predicted Latent Co-Expression Program Scores

We visualized the predicted program scores $\hat{h}_{i,k}$ on slide NCBI681 from LYMPH_IDC, where $\hat{h}_{i,k}$ denotes the morphology-derived score estimate of latent co-expression program k at spot i. This visualization was used to examine the spatial behavior of the auxiliary program activity branch.

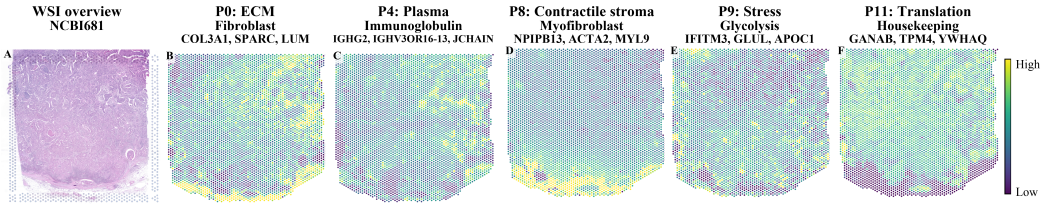

Figure A1: Predicted program-score maps on the NCBI681 slide. Panel A shows the WSI overview; panels B–F show programs P0, P4, P8, P9, and P11 on a shared Low–High scale, with representative high-loading support genes listed above each map.

As shown in Fig. A1, different programs exhibited distinct spatial distributions rather than a shared global intensity pattern. For example, P0 showed elevated activity in tissue-boundary regions and regions displaying stromal-like morphology, with high-loading genes including the extracellular-matrix- and fibroblast-associated genes COL3A1, SPARC, and LUM. In contrast, P4, whose high-loading genes included the immune-associated genes IGHG2 and JCHAIN, exhibited a more spatially localized activity pattern. Other programs, such as P9 and P11, displayed scattered or tissue-wide distributions, respectively. These observations suggest that the learned program scores capture spatially heterogeneous latent cross-gene patterns from H&E morphology, providing qualitative support for using train-derived programs as intermediate constraints for target gene prediction.

A.4 Hyper-Parameter Sensitivity

To study the sensitivity of BRIDGE to key hyper-parameters, we varied the number of latent co-expression programs K, the visual context size $s \times s$, the program activity loss weight λ_{prog}, the support gene count G_s, and the number of top support genes m used for query initialization. The support-gene count was varied by changing the size of the candidate panel while keeping the target panel fixed. All default hyper-parameters were selected exclusively using the ANDERSSON validation subsets, without access to the corresponding outer-test results. After the default config-

uration was frozen, the sensitivity variants were evaluated on the corresponding outer-test folds for descriptive reporting. The results are shown in Fig. A2.

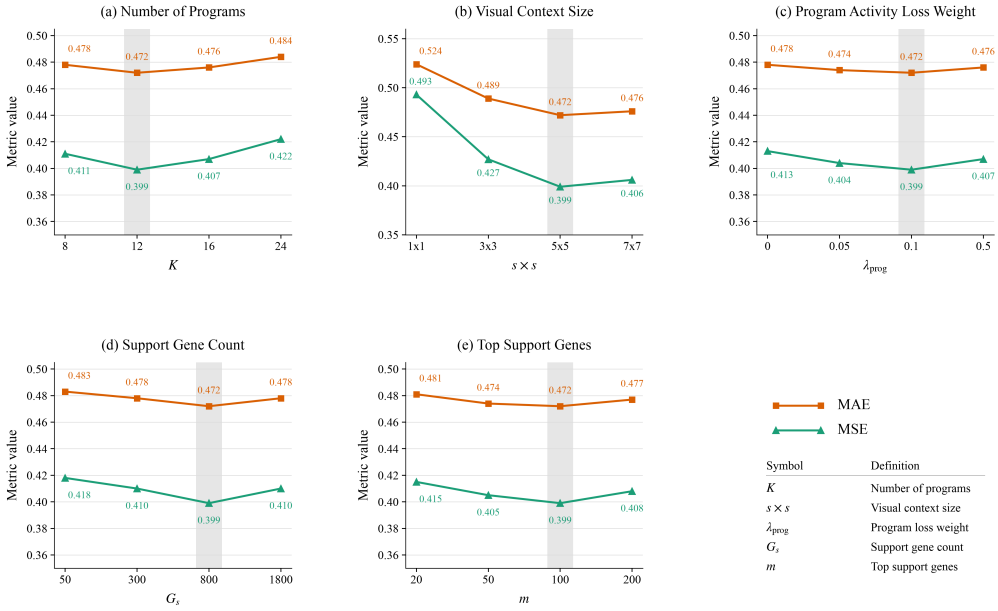

Figure A2: Post-selection parameter sensitivity analysis on the ANDERSSON outer-test folds. The five panels evaluate the number of latent co-expression programs K, visual context size $s \times s$, program activity loss weight λ_{prog}, support gene count G_s, and the number of top support genes m used for query initialization.

The validation-selected default setting, with $K = 12$, a 5×5 visual context, $\lambda_{\text{prog}} = 0.1$, $G_s = 800$, and $m = 100$, also achieved the lowest reported outer-test MAE and MSE among the evaluated settings. Among these factors, the visual context size had the most visible effect. Using only the central spot led to substantially higher errors, while enlarging the context to 5×5 improved both MAE and MSE. Increasing the context to 7×7 did not further improve performance, suggesting that overly large neighborhoods may introduce less relevant morphology tokens. The model was relatively stable with respect to K and λ_{prog}, whereas G_s and m showed moderate sensitivity, with intermediate values producing the lowest prediction errors.

A.5 Effectiveness of Pretrained Gene Representations

To assess the effectiveness of gene representations derived from single-cell foundation models, we compared the full scGPT Cui et al. (2024) target-gene embeddings with fixed random embeddings of the same dimensionality and with low-rank reconstructions of the scGPT embeddings. The random embeddings were initialized once and kept fixed during training, thereby preserving distinct gene identities without using the pretrained information encoded by scGPT. Only target-gene representations for gene-adaptive decoding were replaced; all other model components remained unchanged.

Table A2: Target-gene representation results on ANDERSSON. Results are reported as mean ± standard deviation. Higher PCC and lower MAE and MSE indicate better performance; the best results are bolded.

Setting	PCC ↑	MAE ↓	MSE ↓
Fixed random	0.705 ± 0.026	0.504 ± 0.048	0.446 ± 0.112
scGPT Rank-4	0.713 ± 0.038	0.497 ± 0.058	0.437 ± 0.103
scGPT Rank-8	0.724 ± 0.035	0.481 ± 0.059	0.414 ± 0.104
Full scGPT (Ours)	**0.734 ± 0.033**	**0.472 ± 0.064**	**0.399 ± 0.112**

As shown in Table A2, fixed random embeddings produced the lowest PCC and the highest prediction errors. Performance improved consistently from the fixed random representation to the Rank-4 and Rank-8 scGPT representations, while the full scGPT embeddings achieved the best results. Compared with fixed random embeddings, the full scGPT representation increased PCC by 0.029 and reduced MAE and MSE by 0.032 and 0.047, respectively. These results show that scGPT-derived pretrained representations provide more effective target-gene conditioning than fixed random gene identities, while retaining a richer embedding representation further improves prediction.

A.6 Architecture-Controlled Comparison Design

To isolate the effects of support-derived supervision and the program-guided prediction pathway, we compared BRIDGE with three controlled variants: a direct gene-conditioned predictor, an NMF auxiliary-only predictor, and a learnable gene-weighted query variant (Table A3). These controls separately test the value of the program pathway, the support-derived supervision alone, and the train-derived NMF structure beyond generic learnable gene weighting. All variants share the same data partitions and dataset-specific gene panels.

Direct gene-conditioned predictor. This variant removes the NMF-derived program queries, program activity loss, and program-expert decoder. Each target-gene embedding instead directly attends to the local morphology tokens. This baseline tests whether an intermediate program representation provides additional value beyond gene-conditioned morphology aggregation.

NMF auxiliary-only predictor. This variant uses the direct gene-conditioned prediction pathway above and adds an auxiliary head that predicts the NMF-derived score vector from the pooled morphology representation. The NMF scores provide auxiliary supervision but are not used as attention queries or decoder experts. This control distinguishes the effect of auxiliary program-score prediction from the effect of program-conditioned morphology aggregation and gene-adaptive decoding.

Learnable gene-weighted program queries. This variant preserves the support-gene-weighted query construction and the downstream program-guided aggregation and gene-adaptive decoder, but replaces the train-derived NMF loading matrix with end-to-end learnable gene weights:

$$A_{k,g} = \frac{\exp(\widetilde{A}_{k,g})}{\sum_{g' \in \mathcal{G}_{\text{sup}}} \exp(\widetilde{A}_{k,g'})}, \qquad \mathbf{q}_k^{\text{learn}} = \sum_{g \in \mathcal{G}_{\text{sup}}} A_{k,g} \mathbf{s}_g. \tag{A1}$$

The weights $\widetilde{A}$ are initialized randomly and learned using the target-gene expression objective. Because this variant does not use NMF-derived program scores, the program activity loss is disabled. Together with the existing fixed-random and shuffled-NMF controls, this comparison tests whether train-derived co-expression loadings outperform generic learned gene weights.

Table A3: Architecture-controlled comparison results on ANDERSSON. All settings use the same data partitions, frozen encoders, gene panels, optimization schedule, checkpoint-selection rule, and split-wise evaluation protocol. Results are reported as mean ± standard deviation across matched runs; higher PCC and lower MAE and MSE indicate better performance.

Setting	PCC ↑	MAE ↓	MSE ↓
Direct gene-conditioned	0.713 ± 0.036	0.493 ± 0.067	0.438 ± 0.120
NMF auxiliary-only	0.720 ± 0.035	0.486 ± 0.065	0.424 ± 0.117
Learnable gene-weighted queries	0.725 ± 0.034	0.480 ± 0.064	0.415 ± 0.114
BRIDGE	**0.734 ± 0.033**	**0.472 ± 0.064**	**0.399 ± 0.112**

The controlled comparison separates access to training-only support-gene information from the way BRIDGE uses that information. Relative to the direct target-only predictor, the NMF auxiliary-only variant increased PCC from 0.713 to 0.720, showing that support-derived program supervision provides a useful auxiliary signal even without program queries or experts. BRIDGE further increased PCC to 0.734, outperforming both the NMF auxiliary-only predictor and the learnable gene-weighted query variant (0.725). These results indicate that the gain is not explained solely by access to additional support-derived supervision or by the capacity to learn gene-weighted queries; the full program-query aggregation and gene-adaptive decoding provide further gains.

A.7 MATCHED-BACKBONE BENCHMARK

To examine whether the comparison is affected by the choice of visual encoder, we define an additional matched-backbone protocol in which all methods use the same frozen UNI v1 features. UNI features are extracted deterministically from 224×224-pixel patches at the native slide resolution. Each baseline retains its method-specific image aggregation, graph construction, decoder, and loss functions. The same patient-level outer splits, fixed dataset-specific target-gene panels, checkpoint-selection rule, and evaluation script are used for all methods.

Table A4: Matched-backbone results on ANDERSSON. All methods use the same frozen UNI v1 features while retaining their method-specific prediction architectures. Results are reported as mean ± standard deviation across the same five patient-level outer folds; best results are bolded.

Method	PCC ↑	MAE ↓	MSE ↓
TRIPLEX	0.680 ± 0.043	0.552 ± 0.066	0.526 ± 0.121
MERGE	0.703 ± 0.035	0.510 ± 0.065	0.461 ± 0.118
SpaHGC	0.716 ± 0.037	0.543 ± 0.071	0.515 ± 0.134
BRIDGE	**0.734 ± 0.033**	**0.472 ± 0.064**	**0.399 ± 0.112**

With the visual feature source held fixed, BRIDGE achieved the highest PCC and the lowest MAE and MSE among the four methods. Its PCC was 0.734, compared with 0.716 for SpaHGC, 0.703 for MERGE, and 0.680 for TRIPLEX. BRIDGE ranked first on all three metrics, with an MAE of 0.472 and an MSE of 0.399, both lower than those of every baseline. Thus, the performance advantage persisted after controlling the visual encoder.

B DISCUSSION

BRIDGE's results support the use of training-time support-gene information as molecular context for histology-based target-gene prediction. The ablation results are consistent with complementary contributions from support-derived supervision and program-guided, gene-adaptive decoding. The cross-cohort evaluation further shows that this advantage remains evident under cohort shift, although the lower absolute correlations indicate that morphology–expression associations remain sensitive to changes in tissue composition and acquisition conditions. The program-score maps provide qualitative insight into the spatial variation of the learned representations, but do not establish a specific biological mechanism. The current study is limited to three datasets and fixed target-gene panels; evaluation on broader cohorts, gene panels, and independent molecular measurements will be important for assessing generality and biological validity.