# TISSUEFIELD: RECONSTRUCTING SPATIAL GENE EXPRESSION WITH REUSABLE CONDITIONAL NEURAL FIELDS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Spatial transcriptomics samples gene expression at discrete locations, creating a need to reconstruct spatial expression patterns from partial and irregular observations. Implicit neural representations offer a continuous model of these patterns, but fitting a separate field for each tissue section limits the reuse of learned structure. We present TissueField, a conditional neural field that learns from source sections to reconstruct expression on a new section from partial observations. TissueField constructs local context–query episodes and uses a Transformer hypernetwork to infer window-specific modulation of a shared coordinate network. The network predicts compact expression embeddings, which a shared decoder maps to gene expression. On a new section, conditioning and reconstruction require only forward computation, with modulation reused across queries within each window. Across Stereo-seq MOSTA, MERFISH, Visium, and Slide-seqV2, within-platform cross-section evaluations show consistent reconstruction gains over directly transferred baselines. Spatial analyses show preservation of characteristic expression patterns and tissue organization. These findings support reusing learned reconstruction structure to recover expression on new sections without repeated model fitting.

## 1 INTRODUCTION

Spatial transcriptomics (ST) links gene-expression measurements to positions in tissue, making it possible to study molecular variation together with tissue organization (Ståhl et al., 2016; Marx, 2021). Measurements nevertheless cover a finite set of locations, with coverage and sampling density varying across technologies and tissue sections. Recovering a more complete expression field from partial observations is therefore an important computational task. The difficulty is both spatial and statistical: observations can be irregularly distributed, expression profiles are high-dimensional, and a reconstruction must retain localized variation as well as broad tissue structure. A useful model should predict expression at unobserved coordinates while making effective use of the measurements that are available.

Existing approaches exploit spatial relationships and shared expression structure in several ways. Graph-based methods learn representations from neighboring spots and support spatial-domain analysis (Hu et al., 2021; Dong & Zhang, 2022; Long et al., 2023). Histology-guided methods use tissue morphology to predict expression (He et al., 2020; Zeng et al., 2022; Chung et al., 2024). For spatial refinement, BayesSpace infers subspot expression structure using a Bayesian spatial model (Zhao et al., 2021), while STAGE learns a coordinate-to-expression mapping for higher-density prediction (Li et al., 2024). These methods establish the value of spatial information for reconstruction. For representations defined on measured spots, querying between measurements requires an additional mapping from coordinates to expression. Coordinate-based methods provide this mapping, while the high dimensionality and sparsity of expression remain important modeling challenges.

Implicit neural representations (INRs) represent a signal as a continuous function of spatial coordinates, allowing predictions at locations beyond the original measurements. Fourier features and periodic activations allow coordinate networks to represent spatial variation at different frequencies (Tancik et al., 2020; Sitzmann et al., 2020). SUICA adapts this approach to high-dimensional, sparse



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.

ST data by learning a coordinate mapping in a compact latent space and decoding the resulting embeddings into gene-expression profiles (Zhu et al., 2025). This separates spatial modeling from high-dimensional decoding while retaining the ability to reconstruct expression at arbitrary query positions.

A continuous expression field fitted to an individual section remains tied to that section's coordinate–expression relationship. Applying it directly to another section does not account for changes in shape, composition, or local expression, while fitting a new field repeats the optimization procedure. Source sections may already contain useful reconstruction structure, but this structure alone does not describe the local expression state of a new section. Partial observations from that section provide information about this state. These limitations motivate a conditional reconstruction mapping that combines a learned reconstruction prior with target observations to construct an expression field without target-specific parameter optimization. We focus on transfer between sections from the same measurement platform with a shared gene representation.

We introduce **TissueField**, which learns this conditional reconstruction mapping through local windows. During training, each window provides observed context spots and disjoint query spots. A Transformer hypernetwork encodes the context and predicts lightweight, layer-specific modulation for a shared INR. The modulated network maps query coordinates to latent expression embeddings, which are decoded into gene profiles. At inference, the model covers a new section with overlapping windows, infers modulation once for each window, and reuses it for all queries in that window. Local observations determine how the shared field is adapted, and the learned parameters remain fixed throughout target-section reconstruction.

We make three contributions:

- We introduce TissueField, a reusable reconstruction framework that maps partial observations from a new tissue section to a conditional expression field without target-specific parameter optimization.
- We develop a local-window conditioning method in which context–query training teaches a Transformer hypernetwork to modulate a shared INR from local observations. A shared decoder converts the resulting latent field into gene-expression predictions.
- Experiments on Stereo-seq MOSTA, MERFISH, Visium, and Slide-seqV2 show improved reconstruction over the evaluated baselines under same-platform cross-section transfer. Spatial analyses support preservation of expression patterns and tissue organization in the evaluated examples, while timing results show reduced target-section processing time relative to section-specific fitting.

## 2 Related Work

**Spatial transcriptomics reconstruction and expression completion.** Spatially informed representations help recover expression structure and identify tissue domains. SpaGCN combines expression, spatial location, and histology in a graph model (Hu et al., 2021); STAGATE learns spatial representations with graph attention (Dong & Zhang, 2022); and GraphST addresses clustering, integration, and deconvolution (Long et al., 2023). These methods provide relevant ways to encode spatial dependence, although their primary tasks differ from coordinate-query reconstruction. A complementary line predicts expression from histology, including ST-Net, Hist2ST, and TRIPLEX (He et al., 2020; Zeng et al., 2022; Chung et al., 2024). BayesSpace performs subspot spatial refinement, and STAGE generates higher-density expression through a learned relationship between spatial coordinates and expression (Zhao et al., 2021; Li et al., 2024). TissueField uses coordinates and observed expression as its inputs and focuses on reusing a learned reconstruction model across sections.

**Implicit representations for biological spatial data.** INRs represent signals as coordinate-dependent functions. Fourier feature mappings (Tancik et al., 2020) and sinusoidal networks (Sitzmann et al., 2020) are two established parameterizations used in our comparisons. In ST, the output dimension and sparsity make direct coordinate-to-expression regression difficult. SUICA addresses these properties through a graph-augmented autoencoder, latent spatial mapping, and sparsity-aware supervision (Zhu et al., 2025). TissueField builds on the latent-field formulation and the expression reconstruction objective. Its additional design concerns conditional reuse: observations from a target

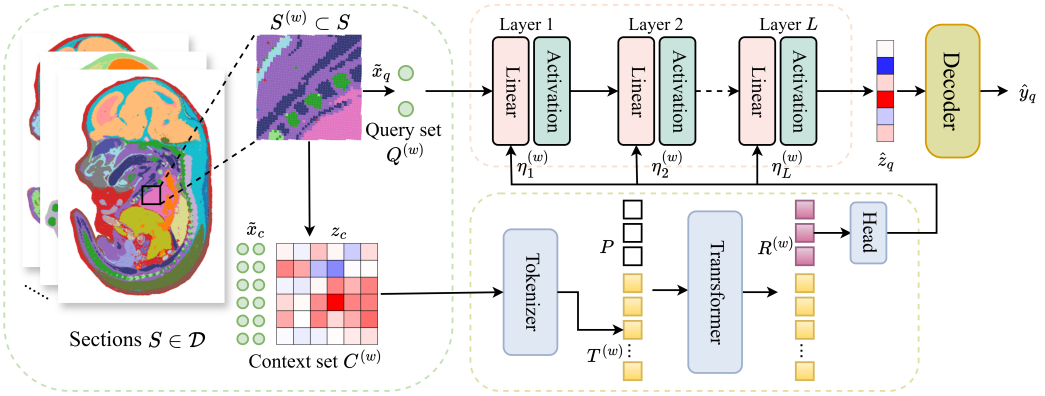


Figure 1: **TissueField overview.** During training, a local window $w$ supplies disjoint context and query sets $C^{(w)}$ and $Q^{(w)}$. Context coordinates and expression embeddings are tokenized and processed by a Transformer together with learnable parameter tokens $P$. The parameter-token outputs $R^{(w)}$ pass through prediction heads to produce window-specific modulation $\eta^{(w)}$ of a shared INR. The field predicts a latent embedding at each query coordinate, and a shared decoder recovers gene expression. At inference, context observations from each window on the target section determine its modulation, which is reused for all queries in that window while the learned parameters remain fixed.

window determine modulation of a field whose parameters were learned on source sections. The experiments distinguish this setting from fitting a field independently to each section.

**Hypernetworks and conditional neural fields.** Hypernetworks generate parameters for another network from an input representation (Ha et al., 2017). FiLM adapts shared features through condition-dependent affine transformations (Perez et al., 2018). For neural fields, Chen & Wang (2022) use Transformers to infer INR parameters from observation sets, with a weight-grouping strategy that modulates shared weight vectors. TissueField uses this observation-conditioned construction within local ST windows. Its context tokens encode coordinates and latent expression, while one parameter token per modulated layer yields row-wise weight scales and bias shifts for a shared INR. A shared expression decoder then maps the field's latent outputs to high-dimensional gene profiles.

## 3 Method

### 3.1 Problem Formulation

Let $\mathcal{D}_{\text{src}} = \{S_1, \ldots, S_M\}$ be source sections from one ST platform. A section $S = \{(x_i, y_i)\}_{i=1}^{N_S}$ contains coordinates $x_i \in \mathbb{R}^2$ and expression profiles $y_i \in \mathbb{R}^G$. A section-specific mask $m_S \in \{0,1\}^G$ identifies the genes available for loss computation and evaluation. On a new target section $S \notin \mathcal{D}_{\text{src}}$ with spatial domain $\Omega_S$, an observed set $C \subset S$ supplies expression measurements. We learn from the source sections to map the observed context $C$ to a conditional expression field

$$F_C : \Omega_S \longrightarrow \mathbb{R}^G, \qquad x \longmapsto \hat{y}(x \mid C), \tag{1}$$

with the learned model parameters held fixed on the target section. Observed target expression enters through $C$ to condition predictions at query coordinates $x$.

Reconstruction and densification use the same function but have different evaluation support. Reconstruction predicts at held-out measured locations, where the withheld expression provides a reference. Densification queries additional coordinates to produce a more densely sampled map. Quantitative evaluation therefore uses held-out measured locations shared across methods, while predictions at additional coordinates support visualization and spatial analysis.

### 3.2 LOCAL-WINDOW EPISODES

Tissue sections can differ in global geometry and anatomical composition. TissueField therefore decomposes section-level reconstruction into local episodes (Figure 1). For an episode sampled from a source section, let $S^{(w)}$ denote the local measured spots associated with window $w$. We sample context and query sets of valid measured spots

$$C^{(w)}, Q^{(w)} \subset S^{(w)}, \qquad C^{(w)} \cap Q^{(w)} = \emptyset, \qquad |C^{(w)}| \le N_c, \quad |Q^{(w)}| \le N_q. \tag{2}$$

The context contains observed coordinate–expression pairs. Query coordinates are inputs to the field, and query expression is used as a training or evaluation target. In particular, query responses are excluded from the hypernetwork input. Context and query arrays have fixed sizes $N_c$ and $N_q$, with padding used when a sparse window supplies fewer valid spots. Padded context entries are masked in attention, and padded query entries are excluded from the reconstruction loss. The experiments vary the context budget to assess how many observations are needed for useful conditioning.

Coordinates are translated relative to the window center and scaled by the window size, giving a local coordinate $\tilde{x}$. This removes dependence on the absolute placement of a window within a section. The model can then learn a relationship between local expression context and local spatial variation, while the shared parameters carry information across episodes. The assumption is that source and target sections contain useful similarities at the neighborhood scale. Local conditioning still allows neighborhoods to differ in their observed expression.

Two windows with similar local coordinates can therefore produce different fields when their expression observations differ. The window size determines the spatial extent over which this context is gathered, connecting the representation to the scale of local tissue variation.

### 3.3 CONTEXT ENCODING AND HYPERNETWORK

Following the latent-expression formulation of SUICA (Zhu et al., 2025), a pretrained expression encoder provides a compact embedding $z_i \in \mathbb{R}^{d_z}$ for each observed spot. The expression encoder is trained exclusively on the training sections, without using target-section expression. A context token combines learned projections of this embedding and the local coordinate with a shared spot-type embedding:

$$t_i = \mathrm{LayerNorm}(\pi_x(\tilde{x}_i) + \pi_z(z_i) + e_{\mathrm{spot}}), \qquad T^{(w)} = [t_i]_{i \in C^{(w)}}. \tag{3}$$

The projections map coordinates and expression embeddings to $d_h = 256$, with $d_z = 32$. Layer normalization is applied to the combined token before it enters the Transformer. Spatial position is carried by the coordinate projection, and expression is carried by the latent projection. The equations use valid context tokens; padding slots are masked in Transformer attention.

We append $L$ learnable parameter tokens $P = [p_1, \ldots, p_L]$, one for each modulated INR layer. A Transformer encoder $h_\varphi$ jointly processes the context and parameter tokens:

$$[T^{(w)\prime}; R^{(w)}] = h_\varphi([T^{(w)}; P]), \qquad \eta_\ell^{(w)} = \mathrm{Head}_\ell(r_\ell^{(w)}). \tag{4}$$

The output states $r_\ell^{(w)}$ at parameter-token positions summarize the context for their respective layers. Small prediction heads convert these states into modulation parameters. Thus, the hypernetwork determines how the field should behave in a window; the coordinate network performs the pointwise prediction. Conditioning is computed once and can be shared by any number of queries in that window. This separation also keeps the conditioning sequence independent of the requested query-grid density: requesting a denser map adds field evaluations without changing the observed context or requiring a new set of learned parameters.

### 3.4 SHARED MODULATED NEURAL FIELD

The shared field is a SIREN backbone with periodic activations (Sitzmann et al., 2020). It maps a query coordinate to a latent prediction, followed by expression decoding:

$$\hat{z}_q = f_{\theta,\eta^{(w)}}(\tilde{x}_q), \qquad \hat{y}_q = g_\psi(\hat{z}_q). \tag{5}$$

The INR parameters $\theta$ and decoder parameters $\psi$ are shared across windows and sections. Only $\eta^{(w)}$ depends on the target context. We evaluate reconstruction on the decoded gene-expression profiles.

TissueField uses weight-row-wise modulation. For a modulated layer with shared matrix $W^{(\ell)}$ and bias $b^{(\ell)}$, the hypernetwork outputs $\eta_\ell^{(w)} = (\alpha^{(w,\ell)}, \Delta b^{(w,\ell)})$, where both vectors have the layer's output width. The effective parameters are

$$\widetilde{W}^{(w,\ell)} = \mathrm{diag}(\mathbf{1} + \alpha^{(w,\ell)})W^{(\ell)}, \qquad \widetilde{b}^{(w,\ell)} = b^{(\ell)} + \Delta b^{(w,\ell)}. \tag{6}$$

Row scaling and bias shifts adapt a common coordinate mapping without generating a full weight matrix for every window. The modulation heads are initialized near the identity, with $\alpha \approx 0$ and $\Delta b \approx 0$. At initialization, the effective weights and biases are therefore close to their shared values. Expression predictions follow Equation 5 directly.

For a layer of output width $d_\ell$, the window-specific output comprises $2d_\ell$ scalars, one scale and one bias shift per unit. The shared matrix retains the learned combinations of incoming features, while the predicted scales and shifts control their response to local context. Thus, the hypernetwork's output size depends on the modulated layer widths rather than the number of queried locations. The expression decoder remains shared as well: each window adapts spatial prediction in the latent space without generating a separate gene-expression decoder.

### 3.5 Training and Inference

Training samples episodes from source sections and supervises predictions at their query locations. It follows two stages. First, latent-space warmup trains the conditional field against the pretrained encoder's query embeddings:

$$\mathcal{L}_{\mathrm{latent}}^{(w)} = \frac{1}{|Q^{(w)}|} \sum_{q \in Q^{(w)}} \|\hat{z}_q - z_q\|_2^2. \tag{7}$$

The decoder is fixed during this stage. Second, expression-space fine-tuning jointly optimizes the hypernetwork, INR, and decoder. The reconstruction objective combines squared error on positively expressed entries, absolute error over available entries including zeros, and a Dice term for expression support:

$$\mathcal{L}_{\mathrm{expr}}^{(w)} = \mathcal{L}_{\mathrm{MSE},+}^{(w)} + \mathcal{L}_{\mathrm{MAE}}^{(w)} + \lambda \mathcal{L}_{\mathrm{Dice}}^{(w)}. \tag{8}$$

Both stages evaluate losses only on valid query spots in $Q^{(w)}$, excluding padding slots. Each expression-space term is additionally restricted to available genes by $m_S$. This follows the sparsity-aware reconstruction objective of SUICA (Zhu et al., 2025). Appendix B gives the training settings. Episode supervision teaches the field to use context observations to predict other locations, rather than simply reconstructing the context supplied to it.

The two stages address different parts of this mapping. Latent warmup learns to recover compact expression structure from coordinates and context. Expression-space fine-tuning then aligns those predictions with the decoded quantities used for evaluation. Its loss terms provide supervision for expression magnitude and support, including the many zero entries in ST profiles. In both stages, query errors train the shared conditioning mechanism across episodes; window-specific modulation is an output of that mechanism, rather than an independently optimized variable for each training window.

On a new section, the spatial domain is covered by overlapping windows. Each window draws its context from available target observations, computes $\eta^{(w)}$, and predicts expression at the requested coordinates it covers. If $\mathcal{W}(x)$ is the set of covering windows, the final prediction is

$$\hat{y}(x \mid C) = \frac{1}{|\mathcal{W}(x)|} \sum_{w \in \mathcal{W}(x)} g_\psi\Big(f_{\theta,\eta^{(w)}}(\tilde{x}^{(w)})\Big). \tag{9}$$

Here $\tilde{x}^{(w)}$ denotes the local coordinate of $x$ in window $w$. We use 50% window overlap. Averaging combines the local predictions where windows meet. The learned parameters remain fixed: adaptation to the target consists of context encoding and modulation prediction. Additional coordinate queries within a window reuse its existing modulation.

Table 1: **Cross-section reconstruction on MOSTA and MERFISH.** MOSTA uses E15.5 as the single source and other developmental stages as targets. MERFISH averages the three one-source/two-target splits. All learned model parameters are fixed on the target; TissueField uses target observations as context. Metrics are evaluated at held-out measured locations. Values are mean (standard deviation) over test sections. Bold and underlining indicate the best and second-best displayed values within each dataset.

| Method | Stereo-seq MOSTA | | | | MERFISH | | | |
|---|---|---|---|---|---|---|---|---|
| | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
| FFN | 0.082(0.026) | 0.024(0.016) | 0.586(0.048) | 0.575(0.063) | 0.776(0.038) | 0.963(0.086) | 0.769(0.005) | 0.568(0.027) |
| SIREN | 0.088(0.026) | 0.026(0.016) | 0.504(0.049) | 0.521(0.064) | 0.812(0.050) | 1.065(0.121) | 0.727(0.022) | 0.509(0.055) |
| STAGE | 0.073(0.002) | 0.023(0.008) | 0.678(0.056) | 0.535(0.021) | 0.857(0.064) | 1.072(0.018) | 0.830(0.008) | 0.506(0.020) |
| SUICA-FFN | 0.072(0.022) | 0.021(0.014) | 0.691(0.042) | 0.613(0.059) | 0.665(0.036) | 0.765(0.074) | 0.815(0.018) | 0.505(0.054) |
| SUICA-SIREN | 0.074(0.022) | 0.021(0.014) | 0.693(0.041) | 0.611(0.058) | 0.715(0.069) | 0.879(0.154) | 0.791(0.033) | 0.474(0.082) |
| **TissueField** | **0.065(0.017)** | **0.019(0.009)** | **0.835(0.042)** | **0.763(0.067)** | **0.436(0.021)** | **0.335(0.027)** | **0.929(0.002)** | **0.762(0.005)** |

# 4 EXPERIMENTS

## 4.1 EXPERIMENTAL SETUP

**Data and preprocessing.** We evaluate on four platforms: 10x Visium brain sections (Palla et al., 2022), Slide-seqV2 mouse brain (Stickels et al., 2021), Stereo-seq MOSTA mouse embryos (Chen et al., 2022), and MERFISH human heart (Farah et al., 2024). They cover regular spot arrays, irregular bead sampling, developmental tissue sections, and imaging-based targeted panels. Within each dataset, methods use the same section-level coordinate–expression matrices, quality control, normalization, and gene filtering. Gene masks restrict losses and metrics to available genes. Appendix A describes the sections and distinguishes their evaluation splits. Models are trained separately within each platform.

**Protocols and baselines.** The main comparison evaluates cross-section reconstruction without target-section optimization. MOSTA uses E15.5 as the single source section and evaluates other developmental stages. MERFISH uses three one-source/two-target splits: each of three heart sections serves as the source once, and the other two are targets. This is distinct from training on two sections and testing on one. Visium transfers between human brain sections, and Slide-seqV2 transfers from hippocampus to somatosensory cortex. We compare FFN (Tancik et al., 2020), SIREN (Sitzmann et al., 2020), SUICA-FFN, SUICA-SIREN (Zhu et al., 2025), and STAGE (Li et al., 2024). Learned parameters are fixed on the target in this protocol. TissueField additionally receives the available target expression as context, excluding held-out query responses. The comparison therefore tests context-conditioned reconstruction against direct transfer of the fitted baselines. Separate in-section fitting and multi-section training results appear in Appendix C.

**Metrics and configuration.** We report MAE, MSE, cosine similarity, and Pearson correlation on decoded expression at the same held-out measured locations. Lower errors and higher similarities indicate better reconstruction. For the biological-organization analysis, the same downstream clustering pipeline is applied to each reconstruction and evaluated by adjusted Rand index (ARI) against reference labels. TissueField is implemented in PyTorch and trained with AdamW, weight decay 0.01, cosine learning-rate decay, and gradient clipping at norm 5. Latent warmup uses 20,000 steps at learning rate $5 \times 10^{-4}$, followed by 2,000 expression-space steps at $2 \times 10^{-5}$. The reported default batch comprises 256 episodes, with context and query budgets of $N_c = 512$ and $N_q = 2048$, respectively. Further configuration details are in Appendix B.

## 4.2 CROSS-SECTION RECONSTRUCTION

Table 1 tests whether source-trained models reconstruct expression on other sections. On MOSTA, TissueField improves cosine similarity from the best baseline value of 0.693 to 0.835 and Pearson correlation from 0.613 to 0.763. Its MAE and MSE are also the lowest among the compared methods. The target sections vary in developmental stage and morphology. Conditioning on observations from each target supplies information about its local expression state while reusing the learned field.

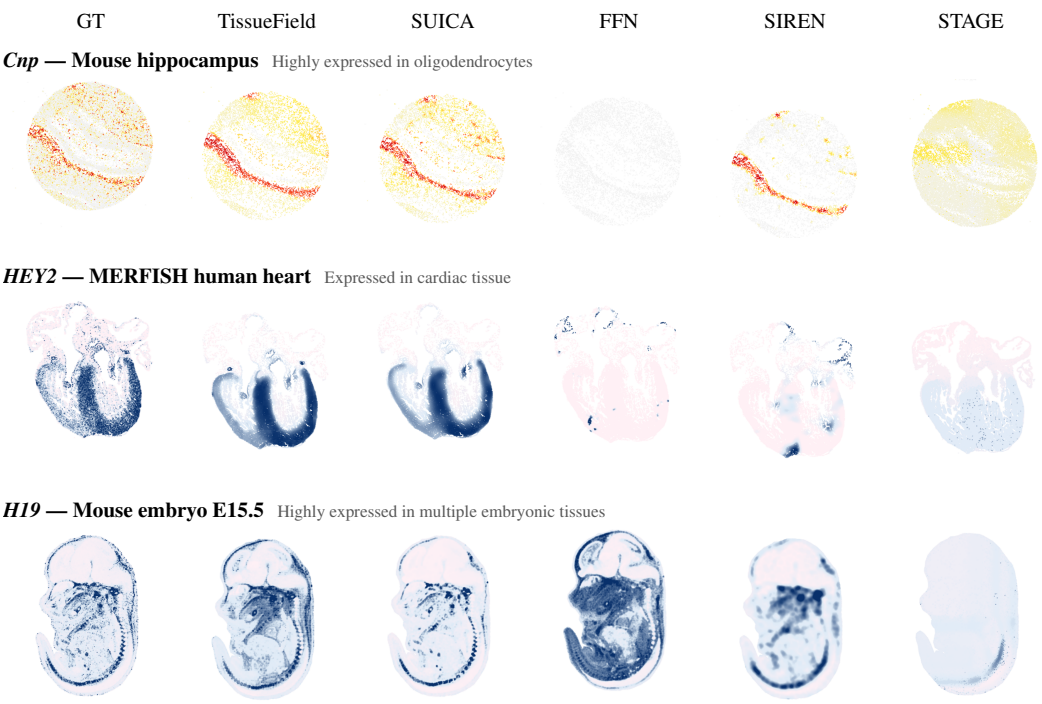


Figure 2: **Spatial expression patterns under cross-section reuse and in-section fitting.** Rows show *Cnp* in mouse hippocampus, *HEY2* in MERFISH human heart, and *H19* in mouse embryo E15.5. TissueField performs cross-section reconstruction; SUICA-SIREN (labeled SUICA), FFN, SIREN, and STAGE are fitted within the displayed section. Columns show measured expression (GT), TissueField, SUICA, FFN, SIREN, and STAGE, from left to right. These spatial examples use a different comparison setting from Table 1.

The same pattern appears on MERFISH. TissueField improves cosine similarity from 0.830 to 0.929 and Pearson correlation from 0.568 to 0.762 relative to the strongest baseline on each respective metric. Its MAE is 0.436 and MSE is 0.335, compared with 0.665 and 0.765 for SUICA-FFN, the strongest error-based baseline. The improvements span both expression error and pattern similarity under the one-source/two-target protocol.

The complete Visium and Slide-seqV2 comparisons show the same direction under their cross-section protocols (Appendix C.2). TissueField reaches cosine similarity 0.872 on the Visium human brain target and 0.725 on the Slide-seqV2 somatosensory cortex target. The corresponding best baseline values are 0.765 and 0.566. Appendix C also retains in-section reconstruction results, where the relative ordering can differ by metric. Multi-section MOSTA training is reported separately: TissueField obtains MAE 0.060, MSE 0.015, cosine similarity 0.844, and Pearson correlation 0.773. This is a different training regime from the single-source comparison in Table 1.

### 4.3 Spatial Gene-expression Patterns

Figure 2 examines spatial patterns that aggregate expression metrics alone cannot show. Here, TissueField reconstructs across sections, whereas the baselines are fitted within the displayed section. In mouse hippocampus, TissueField retains the curved high-expression band of *Cnp*, an oligodendrocyte-associated gene (Patel et al., 2010). SUICA-SIREN and SIREN also recover parts of this band, while FFN has weak contrast and STAGE distributes expression more broadly. For *HEY2*, whose role in ventricular myocardial maturation has been demonstrated in mice (Koibuchi & Chin, 2007), TissueField and SUICA-SIREN preserve the main localized expression pattern in the MERFISH heart section. FFN and SIREN recover only part of this structure. For *H19*, an imprinted long non-coding RNA involved in embryonic growth regulation (Gabory et al., 2009), TissueField preserves the embryo outline and internal expression variation.

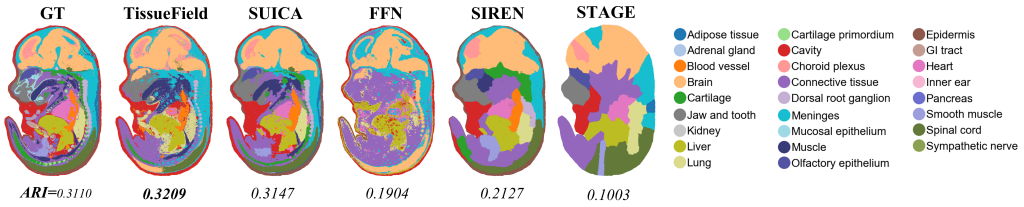


Figure 3: **Spatial organization in the MOSTA clustering example.** TissueField uses cross-section reconstruction; SUICA-SIREN (labeled SUICA), FFN, SIREN, and STAGE use in-section fitting. The same downstream analysis is applied to measured expression (GT) and reconstructed expression. ARI measures agreement with tissue-domain annotations. GT is the measured-expression analysis, not an ARI upper bound. Appendix F contains the full panel across platforms.

These examples show that a reused TissueField can retain recognizable spatial patterns when compared with models fitted to each target section. These maps complement the held-out evaluations by showing the spatial distribution of predicted expression.

### 4.4 Preservation of Biological Organization

We next examine whether reconstructed expression retains organization relevant to downstream analysis. Figure 3 compares expression-derived clusters against MOSTA tissue-domain annotations. Cross-section TissueField obtains ARI 0.3209, slightly above the in-section SUICA-SIREN result of 0.3147 and above FFN, SIREN, and STAGE. The measured-expression analysis gives 0.3110. The spatial maps retain major domains, while the absolute ARI values indicate only partial agreement with the reference annotation. These results support preservation of broad organization under model reuse.

The full panel in Appendix F adds MERFISH, Slide-seqV2 hippocampus, and Visium mouse brain. It supports the preservation of organization in these displayed examples, with annotation-based references for MOSTA and MERFISH and reference clusterings for the brain datasets. This analysis complements expression-level metrics by measuring the agreement between clusters derived from reconstructed expression and the dataset-specific reference labels.

In the MERFISH example, TissueField yields ARI 0.6676 compared with 0.3092 for the displayed SUICA reconstruction. The corresponding values are 0.5395 versus 0.3172 in hippocampus and 0.6198 versus 0.3809 in Visium mouse brain. The separation from SUICA is larger in these examples than in the MOSTA panel. These comparisons extend the spatial-organization analysis beyond the embryo example, using the reference labeling appropriate to each dataset.

### 4.5 Ablation and Robustness

Across both platforms in Table 2, removing the hypernetwork causes the largest deterioration on all four metrics. Cosine similarity falls from 0.854 to 0.660 on MOSTA and from 0.875 to 0.444 on Visium. This consistent pattern supports context-dependent modulation for adapting the shared field to new sections.

Replacing local training episodes with whole-section episodes worsens all four metrics on both platforms. Replacing local-window inference with one section-wide conditioning window has a smaller effect: cosine similarity decreases to 0.851 on MOSTA and 0.862 on Visium. The inference ablation retains target observations as context but shares one modulation across the entire section and removes overlap aggregation. The hypernetwork ablation instead uses a shared SIREN without context-dependent modulation. This distinction separates the use of target context from the spatial granularity of conditioning: removing context-dependent modulation produces the largest deterioration, while local training episodes and local-window inference provide further improvements in these comparisons. All variants keep source-trained parameters fixed during target evaluation, so the comparison concerns training and conditioning rather than differences in target-specific optimization. Appendix D defines the three variants.

Table 2: **Component ablation under cross-section transfer.** MOSTA variants are trained on E15.5 and evaluated on E13.5; Visium variants are trained on Brain1 and evaluated on Brain2. Each variant removes one component from TissueField. In-section results are provided in Appendix D.

| Variant | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
|---|---|---|---|---|
| **MOSTA: E15.5 → E13.5** | | | | |
| TissueField | **0.041** | **0.008** | **0.854** | **0.812** |
| w/o local-window training | 0.046 | 0.010 | 0.815 | 0.784 |
| w/o hypernetwork | 0.053 | 0.013 | 0.660 | 0.526 |
| w/o local-window inference | 0.042 | 0.009 | 0.851 | 0.807 |
| **Visium: Brain1 → Brain2** | | | | |
| TissueField | **0.043** | **0.005** | **0.875** | **0.837** |
| w/o local-window training | 0.046 | 0.006 | 0.856 | 0.816 |
| w/o hypernetwork | 0.103 | 0.017 | 0.444 | 0.708 |
| w/o local-window inference | 0.044 | 0.006 | 0.862 | 0.817 |

On MERFISH, increasing the context budget from 64 to 512 improves cross-section Pearson correlation from 0.725 to 0.752; at 768, it is 0.748 (Appendix D.2). Increasing the held-out fraction from 40% to 80% reduces available observations: TissueField's cross-section cosine similarity changes from 0.925 to 0.914 and Pearson correlation from 0.734 to 0.724. These results support robustness to reduced observations within the tested context budgets and held-out fractions.

### 4.6 Computational Efficiency

TissueField moves optimization to source training and reconstructs each target section through context-conditioned forward computation. On MERFISH, target-section processing takes approximately 9 seconds for TissueField, compared with 53.83–58.56 minutes for section-specific fitting of the INR baselines and 41.26 minutes for STAGE (Appendix E). This timing comparison measures the cost of processing new sections with the respective workflows; it is separate from the fixed-parameter transfer evaluation in Table 1.

Source training is shared across subsequent target sections and is excluded from these measurements, which quantify target-section processing rather than total training-and-inference cost. Within each target window, modulation is computed once and reused across query coordinates, enabling reconstruction without fitting a new field.

## 5 Discussion and Conclusion

TissueField learns how local observations condition a reusable expression field. A shared coordinate network and decoder retain source-trained reconstruction structure, while a Transformer hypernetwork adapts the field to each observed neighborhood. This connects the continuous, compact representation of ST expression to a practical reuse mechanism: a new section supplies context, and its field is constructed through forward computation. The experiments show improved cross-section reconstruction over the compared baselines and preservation of spatial expression patterns and broad tissue organization in the displayed examples. Component ablations support the role of context-dependent modulation, while the runtime comparison shows shorter target-section processing than per-section fitting.

The scope is same-platform transfer under the reported section splits. Performance can depend on the similarity of source and target neighborhoods, local observation coverage, and the ability of the expression representation to capture relevant variation. Heterogeneous boundaries and very sparse context may limit recovery. Reconstruction accuracy is evaluated at held-out measured locations. Broader cohorts, uncertainty estimates, and additional modalities are useful directions for evaluating these limits. Within the present scope, TissueField reconstructs spatial expression from partial observations by conditioning a shared neural field, avoiding repeated fitting on new sections.

## AI USE STATEMENT

Generative AI was used only to assist manuscript writing and English–Chinese translation, and to help interpret the existing experimental results. Numerical results and scientific figure panels were taken from the existing experimental materials. The authors are responsible for the accuracy of the text, citations, interpretations, and reported results.

## ETHICS STATEMENT

This study analyzes publicly available, previously published spatial transcriptomics datasets. It does not involve new recruitment of human participants or new animal experiments. The source publications describe the original data collection and associated approvals.

## REPRODUCIBILITY STATEMENT

Section 3 describes the conditional field, modulation, objective, and inference procedure. Section 4.1 and Appendices A–B describe the evaluation protocols and implementation settings. Appendices C–E give the complete quantitative comparisons, ablations, sensitivity analyses, and target-processing times. The spatial examples and their evaluation settings are described in Appendix F.

## REFERENCES

Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

Yinbo Chen and Xiaolong Wang. Transformers as meta-learners for implicit neural representations. In *European Conference on Computer Vision*, pp. 170–187. Springer, 2022.

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.

Kangning Dong and Shihua Zhang. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*, 13(1):1739, 2022.

Elie N Farah, Robert K Hu, Colin Kern, Qingquan Zhang, Ting-Yu Lu, Qixuan Ma, Shaina Tran, Bo Zhang, Daniel Carlin, Alexander Monell, et al. Spatially organized cellular communities form the developing human heart. *Nature*, 627(8005):854–864, 2024.

Anne Gabory, Marie-Anne Ripoche, Anne Le Digarcher, Françoise Watrin, Ahmed Ziyyat, Thierry Forné, Hélène Jammes, Justin FX Ainscough, M Azim Surani, Laurent Journot, et al. H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development*, 136(20): 3413–3421, 2009.

David Ha, Andrew M. Dai, and Quoc V. Le. HyperNetworks. International Conference on Learning Representations, poster, 2017. URL https://openreview.net/forum?id=rkpACe1lx.

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.

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.

Nobutaka Koibuchi and Michael T Chin. Chf1/hey2 plays a pivotal role in left ventricular maturation through suppression of ectopic atrial gene expression. *Circulation Research*, 100(6):850–855, 2007.

Shang Li, Kuo Gai, Kangning Dong, Yiyang Zhang, and Shihua Zhang. High-density generation of spatial transcriptomics with STAGE. *Nucleic Acids Research*, 52(9):4843–4856, 2024.

Yahui Long, Kok Siong Ang, Mengwei Li, Kian Long Kelvin Chong, Raman Sethi, Chengwei Zhong, Hang Xu, Zhiwei Ong, Karishma Sachaphibulkij, Ao Chen, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications*, 14(1):1155, 2023.

Vivien Marx. Method of the year: spatially resolved transcriptomics. *Nature Methods*, 18(1):9–14, 2021.

Giovanni Palla, Hannah Spitzer, Michal Klein, David Fischer, Anna Christina Schaar, Louis Benedikt Kuemmerle, Sergei Rybakov, Ignacio L Ibarra, Olle Holmberg, Isaac Virshup, et al. Squidpy: a scalable framework for spatial omics analysis. *Nature Methods*, 19(2):171–178, 2022.

Jigisha R Patel, Erin E McCandless, Denise Dorsey, and Robyn S Klein. Cxcr4 promotes differentiation of oligodendrocyte progenitors and remyelination. *Proceedings of the National Academy of Sciences*, 107(24):11062–11067, 2010.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020.

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

Robert R Stickels, Evan Murray, Pawan Kumar, Jilong Li, Jamie L Marshall, Daniela J Di Bella, Paola Arlotta, Evan Z Macosko, and Fei Chen. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology*, 39(3):313–319, 2021.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020.

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), 2022.

Edward Zhao, Matthew R Stone, Xing Ren, Jamie Guenthoer, Kimberly S Smythe, Thomas Pulliam, Stephen R Williams, Cedric R Uytingco, Sarah EB Taylor, Paul Nghiem, et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nature Biotechnology*, 39(11):1375–1384, 2021.

Qingtian Zhu, Yumin Zheng, Yuling Sang, Yifan Zhan, Ziyan Zhu, Jun Ding, and Yinqiang Zheng. SUICA: Learning super-high dimensional sparse implicit neural representations for spatial transcriptomics. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 80448–80462. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/zhu25af.html.

# A Datasets and Evaluation Protocols

**10x Visium.** The Visium data comprise two mouse brain sagittal posterior sections and two human brain sections (Palla et al., 2022). The mouse sections contain 3,342 spots with 19,099 genes and 3,288 spots with 18,942 genes. The human sections contain 4,906 spots with 21,998 genes and 4,968 spots with 22,395 genes. Their source identifiers are `V1_Mouse_Brain_Sagittal_Posterior`, its `Section_2` counterpart, and `V1_Human_Brain_Section_1` and `Section_2`. Table 6 reports the human brain transfer comparison. The Visium spatial-clustering example is mouse brain, a separate dataset within this platform.

**Slide-seqV2.** The mouse hippocampus section contains 29,309 beads and 20,412 genes; the somatosensory cortex section contains 21,941 beads and 19,346 genes (Stickels et al., 2021). The transfer experiment trains on hippocampus and evaluates somatosensory cortex. Its in-section comparison fits and evaluates the cortex section. The gene-expression and clustering examples instead display hippocampus. These tissue identities distinguish the spatial examples from the target in the pairwise quantitative comparison.

**Stereo-seq MOSTA.** The MOSTA collection spans mouse embryonic stages E9.5–E16.5 (Chen et al., 2022). Two training protocols are reported. In the single-source cross-section comparison, E15.5 is the training section and the other evaluated stages are targets. In the multi-section generalization setting, the first section of each stage is reserved for testing, with 47 sections used for training and 3 for validation. These protocols are displayed as separate blocks in Table 4. The ablation is a specific E15.5-to-E13.5 transfer experiment, not the mean across target stages. The *H19* spatial example displays E15.5 as a target for cross-section TissueField and therefore uses a different source–target arrangement from the single-source quantitative benchmark.

**MERFISH human heart.** The three sections have a targeted panel of 238 genes (Farah et al., 2024): `R77_4C4` contains 72,962 cells, `R78_4C12` contains 75,782 cells, and `R78_4C15` contains 79,891 cells, totaling 228,635. Each section serves once as the single training section and the remaining two as test sections. The three resulting splits are summarized for the cross-section benchmark. In-section reconstruction instead fits each section from its available observations and evaluates held-out locations in that section.

**Preprocessing and evaluation support.** All platforms are represented by two-dimensional coordinates and expression vectors. Within a dataset split, methods use the same processed coordinate–expression matrices and the same held-out locations. Expression losses and metrics exclude unavailable genes using a section-specific gene mask. The reported spot and gene counts describe the input sections; platform-specific filtering can lead to different gene availability. Context and query locations are separated so that query responses are not supplied as conditioning observations. MAE and MSE quantify expression error, while cosine similarity and Pearson correlation measure expression-profile agreement. The tables preserve the reported units and aggregation: MOSTA and MERFISH main results are section-level summaries, and the pairwise, ablation, and sensitivity comparisons are displayed separately. No significance test or confidence interval is inferred from these summaries.

# B Model and Optimization Details

The model consists of an expression embedding, context tokenizer, Transformer hypernetwork, modulated SIREN, and expression decoder. Context tokens sum coordinate and latent-expression projections with a spot-type embedding. Parameter tokens are appended to this sequence, with one token for each modulated layer. Their output states are mapped to scale and bias-shift vectors by shallow heads. Equations 4–6 specify the conditional computation. The shared decoder maps each predicted latent embedding back to expression space.

The Transformer hypernetwork uses six pre-layer-normalized encoder layers with model dimension 256, eight attention heads (32 dimensions per head), feed-forward dimension 1024, and dropout 0.1. The SIREN has architecture $2 \to 2048 \to 2048 \to 2048 \to 2048 \to 32$: four sinusoidal hidden layers followed by a linear output layer. The frequency factor is $\omega_0 = 30$ in the first hidden layer

and $\omega_0 = 1$ in the remaining hidden layers. In the main weight-row-wise (`wrow`) configuration, four parameter tokens each produce a 2048-dimensional row-modulation vector and a 2048-dimensional bias-shift vector for one hidden layer. The linear output layer is not modulated.

**Training objective.** Latent warmup uses the query-embedding squared error in Equation 7, keeping the decoder fixed. Expression-space fine-tuning updates the hypernetwork, field, and decoder together. For a gene present in the section, the positive-expression MSE term uses query entries with nonzero measured expression. The MAE term includes both zero and nonzero entries, and the Dice term compares expression support. The gene mask restricts expression-space loss terms to available genes. Padded context tokens are masked in attention, and padded query slots are excluded from both training losses. Equation 8 summarizes the composite expression-space objective. The loss follows the sparse-expression reconstruction approach of SUICA (Zhu et al., 2025).

Table 3: **Reported TissueField training and inference settings.** These settings apply unless varied in a specific sensitivity experiment. The window side length is estimated based on the size of the training sections and kept unchanged during inference on target sections.

| Setting | Value or procedure |
|---|---|
| Coordinate backbone | SIREN with weight-row-wise modulation |
| Conditioning network | Transformer encoder with parameter tokens |
| Optimizer | AdamW |
| Weight decay | 0.01 |
| Learning-rate schedule | Cosine decay |
| Gradient clipping | Maximum norm 5 |
| Batch size | 256 window episodes |
| Latent warmup | 20,000 steps; peak learning rate $5 \times 10^{-4}$ |
| Expression fine-tuning | 2,000 steps; peak learning rate $2 \times 10^{-5}$ |
| Default context budget | $N_c = 512$ |
| Default query budget | $N_q = 2048$ |
| Window sampling | Square windows in source sections |
| Coordinate frame | Window-centered and scaled by window size |
| Modulation initialization | Near identity |
| Inference overlap | 50% |
| Overlap aggregation | Arithmetic mean of expression predictions |
| Target parameter updates | None |

**Inference.** Observed spots within or near each inference window provide its context. The hypernetwork produces modulation once for that window. The field and decoder then predict all covered query coordinates, and overlapping predictions are averaged. Additional query coordinates reuse the window-specific modulation and require only field and decoder evaluations. The procedure requires target observations, even though it does not optimize model parameters on the target.

# C Complete Reconstruction Results

## C.1 MOSTA and MERFISH

Tables 4 and 5 report the full comparisons. In-section reconstruction fits a model using available observations from the section being evaluated. Cross-section reconstruction fixes the source-trained parameters and evaluates another section; TissueField uses its target observations as context. MOSTA additionally reports multi-section training on a separate train/validation/test split. All expression metrics are evaluated at held-out measured locations.

TissueField has the lowest displayed errors and highest cosine similarity in both in-section blocks, allowing for a displayed MSE tie on MOSTA. Pearson correlation is not uniformly highest: in-section MOSTA SUICA-FFN reaches 0.794 versus 0.782 for TissueField, and in-section MERFISH SIREN reaches 0.788 versus 0.764. In the cross-section blocks, TissueField achieves the lowest MAE and MSE and the highest cosine similarity and Pearson correlation on both platforms.

Table 4: **Complete held-out reconstruction results on Stereo-seq MOSTA under three evaluation settings.** In-section: each method is fitted on the test section itself. Cross-section: models are trained on E15.5 and transferred to other developmental stages without re-optimization. Generalization: TissueField is trained on all training sections and evaluated on held-out test sections. Results are reported as mean (standard deviation) over test sections at held-out measured locations; arrows indicate metric direction.

| **Method** | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
|---|---|---|---|---|
| **In-section**: fit each test section independently | | | | |
| FFN | 0.080(0.028) | 0.020(0.011) | 0.720(0.052) | 0.717(0.044) |
| SIREN | 0.087(0.028) | 0.021(0.011) | 0.687(0.051) | 0.734(0.042) |
| STAGE | 0.071(0.016) | 0.019(0.013) | 0.813(0.056) | 0.702(0.063) |
| SUICA-FFN | 0.069(0.021) | 0.014(0.007) | 0.821(0.026) | 0.794(0.038) |
| SUICA-SIREN | 0.070(0.021) | 0.015(0.007) | 0.819(0.026) | 0.791(0.038) |
| TissueField | 0.054(0.016) | 0.014(0.008) | 0.861(0.030) | 0.782(0.064) |
| **Cross-section**: train on E15.5 and transfer to other test sections | | | | |
| FFN | 0.082(0.026) | 0.024(0.016) | 0.586(0.048) | 0.575(0.063) |
| SIREN | 0.088(0.026) | 0.026(0.016) | 0.504(0.049) | 0.521(0.064) |
| STAGE | 0.073(0.002) | 0.023(0.008) | 0.678(0.056) | 0.535(0.021) |
| SUICA-FFN | 0.072(0.022) | 0.021(0.014) | 0.691(0.042) | 0.613(0.059) |
| SUICA-SIREN | 0.074(0.022) | 0.021(0.014) | 0.693(0.041) | 0.611(0.058) |
| TissueField | 0.065(0.017) | 0.019(0.009) | 0.835(0.042) | 0.763(0.067) |
| **Generalization**: train on training sections and test on held-out sections | | | | |
| TissueField | 0.060(0.021) | 0.015(0.009) | 0.844(0.044) | 0.773(0.067) |

Table 5: **Complete held-out reconstruction results on MERFISH human heart sections.** In-section: each method is fitted on the test section itself. Cross-section: models are trained on one heart section and evaluated on the remaining two, averaged over all three train-test splits. Results are reported as mean (standard deviation) over test sections at held-out measured locations; arrows indicate metric direction.

| **Method** | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
|---|---|---|---|---|
| **In-section**: fit each section independently | | | | |
| FFN | 0.656(0.052) | 0.687(0.084) | 0.851(0.007) | 0.728(0.006) |
| SIREN | 0.572(0.051) | 0.539(0.077) | 0.883(0.010) | 0.788(0.003) |
| STAGE | 0.775(0.075) | 0.641(0.017) | 0.873(0.009) | 0.669(0.010) |
| SUICA-FFN | 0.465(0.008) | 0.391(0.004) | 0.914(0.007) | 0.756(0.009) |
| SUICA-SIREN | 0.483(0.015) | 0.421(0.020) | 0.907(0.006) | 0.734(0.008) |
| TissueField | 0.435(0.007) | 0.331(0.014) | 0.930(0.005) | 0.764(0.010) |
| **Cross-section**: train on one section and test on the other two | | | | |
| FFN | 0.776(0.038) | 0.963(0.086) | 0.769(0.005) | 0.568(0.027) |
| SIREN | 0.812(0.050) | 1.065(0.121) | 0.727(0.022) | 0.509(0.055) |
| STAGE | 0.857(0.064) | 1.072(0.018) | 0.830(0.008) | 0.506(0.020) |
| SUICA-FFN | 0.665(0.036) | 0.765(0.074) | 0.815(0.018) | 0.505(0.054) |
| SUICA-SIREN | 0.715(0.069) | 0.879(0.154) | 0.791(0.033) | 0.474(0.082) |
| TissueField | 0.436(0.021) | 0.335(0.027) | 0.929(0.002) | 0.762(0.005) |

### C.2 PAIRWISE VISIUM AND SLIDE-SEQV2 TRANSFER

Table 6 compares fitting Visium human brain Section 2 with transfer from Section 1 to Section 2. Table 7 compares fitting the Slide-seqV2 somatosensory cortex section with transfer from hippocampus to cortex. Both settings preserve the same target identity within each table. Visium errors use units of $10^{-2}$; Slide-seqV2 errors are shown on their reported expression scale. Numerical magnitudes should not be compared across platforms as though preprocessing and gene panels were identical.

TissueField's Visium cosine similarity changes from 0.879 with in-section training to 0.872 under transfer. In Slide-seqV2 it changes from 0.785 to 0.725. The corresponding baseline values decrease more strongly in these displayed comparisons. For in-section cortex reconstruction, SIREN has the highest Pearson correlation, while TissueField has the lowest MAE and MSE and the highest cosine similarity. Under hippocampus-to-cortex transfer, TissueField has the best displayed value on each of the four metrics.

Table 6: **Pairwise cross-section transfer on 10x Visium human brain.** *Section1* and *Section2* are consecutive human brain sections. The Train model is fitted and tested on the same section; the Transfer model is trained on one section and evaluated on the other without re-optimization. MAE and MSE are reported in units of $\times 10^{-2}$. Both fitting settings are displayed separately.

| Method | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
|---|---|---|---|---|
| **Train model** (Train on *Section2* and Test on *Section2*) | | | | |
| FFN | 5.928 | 0.744 | 0.794 | 0.823 |
| SIREN | 6.947 | 0.967 | 0.677 | 0.769 |
| STAGE | 5.561 | 0.824 | 0.733 | 0.790 |
| SUICA-FFN | 5.453 | 0.665 | 0.817 | 0.821 |
| SUICA-SIREN | 5.445 | 0.662 | 0.819 | 0.825 |
| **TissueField** | **4.255** | **0.513** | **0.879** | **0.841** |
| **Transfer model** (Train on *Section1* and Test on *Section2*) | | | | |
| FFN | 6.133 | 0.829 | 0.765 | 0.799 |
| SIREN | 7.161 | 1.074 | 0.642 | 0.730 |
| STAGE | 5.892 | 1.077 | 0.705 | 0.754 |
| SUICA-FFN | 5.494 | 0.966 | 0.707 | 0.630 |
| SUICA-SIREN | 5.510 | 0.978 | 0.703 | 0.623 |
| **TissueField** | **4.355** | **0.558** | **0.872** | **0.835** |

Table 7: **Pairwise cross-section transfer on Slide-seqV2 across brain regions.** The *hippocampus* and *somatosensory cortex* sections are from different brain regions. The Train model is fitted and tested on the same section; the Transfer model is trained on one region and evaluated on the other without re-optimization. Both fitting settings are displayed separately.

| Method | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
|---|---|---|---|---|
| **Train model** (Train on *somatosensory cortex* and Test on *somatosensory cortex*) | | | | |
| FFN | 0.417 | 0.336 | 0.507 | 0.586 |
| SIREN | 0.418 | 0.313 | 0.527 | **0.645** |
| STAGE | 0.406 | 0.289 | 0.583 | 0.634 |
| SUICA-FFN | 0.314 | 0.225 | 0.739 | 0.623 |
| SUICA-SIREN | 0.319 | 0.231 | 0.722 | 0.585 |
| **TissueField** | **0.300** | **0.184** | **0.785** | 0.589 |
| **Transfer model** (Train on *hippocampus* and Test on *somatosensory cortex*) | | | | |
| FFN | 0.442 | 0.464 | 0.242 | 0.172 |
| SIREN | 0.442 | 0.451 | 0.286 | 0.282 |
| STAGE | 0.432 | 0.371 | 0.507 | 0.450 |
| SUICA-FFN | 0.353 | 0.380 | 0.559 | 0.248 |
| SUICA-SIREN | 0.350 | 0.377 | 0.566 | 0.258 |
| **TissueField** | **0.318** | **0.269** | **0.725** | **0.576** |

# D Ablation and Sensitivity Analyses

## D.1 Component Ablation

Table 8 complements Table 2 with in-section results for MOSTA and Visium. The variants distinguish training episodes, context modulation, and inference conditioning. In *w/o local-window training*, whole-section episodes replace local training episodes. In *w/o hypernetwork*, a trainable shared base SIREN INR without context-dependent modulation is followed by the original decoder. In *w/o local-window inference*, training is unchanged: a single window covering the target section produces one modulation shared by all query locations, replacing local-window conditioning and overlap aggregation. This inference variant retains target-context conditioning. The MOSTA in-section evaluation uses E15.5, while cross-section evaluation uses E13.5 after training on E15.5. The Visium experiment trains on Brain 1 and evaluates Brain 1 or Brain 2. The hypernetwork ablation gives the largest deterioration in cross-section cosine similarity and Pearson correlation on both platforms. The other changes have smaller effects, and not every metric improves monotonically in the in-section block.

Table 8: **Complete ablation study with both in-section and cross-section evaluation.** For MOSTA, variants are trained on E15.5 and evaluated on E13.5 (cross-section) or on E15.5 itself (in-section). For Visium, variants are trained on Brain1 and evaluated on Brain2 (cross-section) or on Brain1 itself (in-section). The in-section block isolates component contributions under standard fitting, while the cross-section block reveals their importance for cross-section reconstruction.

| Variant | MOSTA | | | | Visium | | | |
|---|---|---|---|---|---|---|---|---|
| | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
| **In-section** | | | | | | | | |
| TissueField | 0.105 | 0.044 | 0.802 | 0.729 | 0.045 | 0.006 | 0.907 | 0.896 |
| w/o local-window training | 0.106 | 0.042 | 0.751 | 0.656 | 0.047 | 0.007 | 0.895 | 0.885 |
| w/o hypernetwork | 0.119 | 0.054 | 0.692 | 0.628 | 0.050 | 0.009 | 0.858 | 0.848 |
| w/o local-window inference | 0.105 | 0.044 | 0.742 | 0.663 | 0.048 | 0.008 | 0.890 | 0.878 |
| **Cross-section** | | | | | | | | |
| TissueField | 0.041 | 0.008 | 0.854 | 0.812 | 0.043 | 0.005 | 0.875 | 0.837 |
| w/o local-window training | 0.046 | 0.010 | 0.815 | 0.784 | 0.046 | 0.006 | 0.856 | 0.816 |
| w/o hypernetwork | 0.053 | 0.013 | 0.660 | 0.526 | 0.103 | 0.017 | 0.444 | 0.708 |
| w/o local-window inference | 0.042 | 0.009 | 0.851 | 0.807 | 0.044 | 0.006 | 0.862 | 0.817 |

## D.2 Context Budget and Held-out Fraction

Table 9 varies $N_c$ from 64 to 768 on MERFISH. These results show generally modest gains with more context, with differences becoming small around 512 observations. Table 10 varies the held-out fraction from 40% to 80%. In-section and cross-section results remain separate, and the baseline inventory here excludes STAGE because it was not reported in this experiment. Increasing the held-out fraction reduces observed spatial support; it does not mask selected genes within an otherwise observed spot. TissueField achieves the lowest cross-section MAE and MSE and the highest cross-section cosine similarity and Pearson correlation at all three held-out fractions. From 40% to 80%, its cross-section cosine similarity decreases from 0.925 to 0.914 and Pearson correlation from 0.734 to 0.724.

Table 9: **Sensitivity to the number of context spots ($N_c$) on MERFISH.** TissueField is evaluated with context budgets ranging from 64 to 768 spots per window under both in-section and cross-section settings. The two protocols are reported separately; there is little additional benefit beyond $N_c = 512$ on most displayed metrics.

| | **In-section** | | | | **Cross-section** | | | |
|---|---|---|---|---|---|---|---|---|
| $N_c$ | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
| 64 | 0.493 | 0.440 | 0.905 | 0.728 | 0.432 | 0.358 | 0.918 | 0.725 |
| 128 | 0.477 | 0.401 | 0.918 | 0.734 | 0.431 | 0.342 | 0.921 | 0.730 |
| 256 | 0.469 | 0.391 | 0.925 | 0.743 | 0.429 | 0.339 | 0.920 | 0.745 |
| 512 | 0.464 | 0.378 | 0.925 | 0.745 | 0.428 | 0.338 | 0.927 | 0.752 |
| 768 | 0.465 | 0.377 | 0.928 | 0.744 | 0.429 | 0.337 | 0.924 | 0.748 |

Table 10: **Robustness to increasing held-out ratios on MERFISH.** Models are evaluated when 40%, 60%, or 80% of observed spots are held out for testing, progressively reducing the available context. In-section fitting and cross-section transfer are shown in separate blocks; STAGE is not included in this sensitivity experiment.

| **Method** | **Mask ratio** | **In-section** | | | | **Cross-section** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ | MAE↓ | MSE↓ | Cosine↑ | Pearson↑ |
| FFN | 40% | 0.709 | 0.773 | 0.841 | 0.729 | 0.757 | 0.914 | 0.757 | 0.571 |
| FFN | 60% | 0.719 | 0.790 | 0.836 | 0.718 | 0.756 | 0.907 | 0.758 | 0.572 |
| FFN | 80% | 0.788 | 0.943 | 0.796 | 0.655 | 0.770 | 0.928 | 0.752 | 0.575 |
| SIREN | 40% | 0.621 | 0.617 | 0.872 | 0.783 | 0.802 | 1.032 | 0.714 | 0.517 |
| SIREN | 60% | 0.755 | 0.881 | 0.800 | 0.668 | 0.854 | 1.114 | 0.684 | 0.474 |
| SIREN | 80% | 0.796 | 0.956 | 0.780 | 0.634 | 0.850 | 1.098 | 0.688 | 0.477 |
| SUICA-FFN | 40% | 0.479 | 0.404 | 0.919 | 0.763 | 0.617 | 0.669 | 0.829 | 0.540 |
| SUICA-FFN | 60% | 0.531 | 0.490 | 0.898 | 0.714 | 0.629 | 0.691 | 0.824 | 0.541 |
| SUICA-FFN | 80% | 0.581 | 0.574 | 0.880 | 0.670 | 0.657 | 0.737 | 0.810 | 0.517 |
| SUICA-SIREN | 40% | 0.501 | 0.445 | 0.909 | 0.739 | 0.656 | 0.753 | 0.802 | 0.503 |
| SUICA-SIREN | 60% | 0.509 | 0.451 | 0.907 | 0.733 | 0.668 | 0.773 | 0.794 | 0.487 |
| SUICA-SIREN | 80% | 0.566 | 0.549 | 0.883 | 0.699 | 0.700 | 0.842 | 0.775 | 0.469 |
| TissueField | 40% | 0.450 | 0.374 | 0.929 | 0.748 | 0.425 | 0.330 | 0.925 | 0.734 |
| TissueField | 60% | 0.469 | 0.379 | 0.925 | 0.742 | 0.431 | 0.333 | 0.921 | 0.730 |
| TissueField | 80% | 0.475 | 0.388 | 0.916 | 0.738 | 0.438 | 0.344 | 0.914 | 0.724 |

## E COMPUTATIONAL COST

Table 11 reports target-section processing on MERFISH. The fitted baselines optimize a model for each new section in this timing comparison. TissueField instead conditions a source-trained model and reconstructs through forward evaluation, taking approximately 9 seconds per section. These times are not inference-only measurements for all methods and do not correspond to running the fixed-parameter baseline transfer protocol in Table 1. TissueField's initial source-training cost is additional and shared across target sections.

Table 11: **Target-section processing time on MERFISH.** The baseline measurements include section-specific optimization; TissueField uses source-trained parameters and forward computation. The reported times exclude TissueField's source training.

| Method | Test-time optimization | Runtime per section |
|---|---|---|
| FFN | Yes | 57.95 min |
| SIREN | Yes | 58.56 min |
| SUICA-FFN | Yes | 54.40 min |
| SUICA-SIREN | Yes | 53.83 min |
| STAGE | Yes | 41.26 min |
| TissueField | No | $\sim$9 s |

## F SPATIAL EXAMPLES AND BIOLOGICAL ORGANIZATION

The main gene-expression figure uses *Cnp* in mouse hippocampus, *HEY2* in MERFISH heart, and *H19* in E15.5 embryo. TissueField is evaluated across sections, whereas the displayed baselines are fitted within the target section; the SUICA row is SUICA-SIREN. The MOSTA clustering example uses the same comparison of cross-section TissueField with in-section baselines. These spatial comparisons assess the field produced by model reuse against section-specific reconstructions, separately from the all-method cross-section quantitative tables.

Figure 4 reproduces the full clustering panel. For MOSTA and MERFISH, ARI compares expression-derived clusters to tissue-domain and cell-type annotations, respectively. The brain examples compare against reference clusterings. Consequently, the reference labels have different meanings across datasets, and their ARI values should not be pooled into a single biological score. The GT columns show the measured-expression analysis. On MOSTA, its ARI is 0.3110; on MERFISH it is 0.7181. These are empirical reference results under the downstream procedure, not theoretical upper bounds.

The displayed TissueField and SUICA values are 0.3209 and 0.3147 for MOSTA, 0.6676 and 0.3092 for MERFISH, 0.5395 and 0.3172 for hippocampus, and 0.6198 and 0.3809 for Visium mouse brain. The MOSTA difference is small, while the other examples show larger separation. The spatial distributions and ARI values support the organization visible in these specific examples.

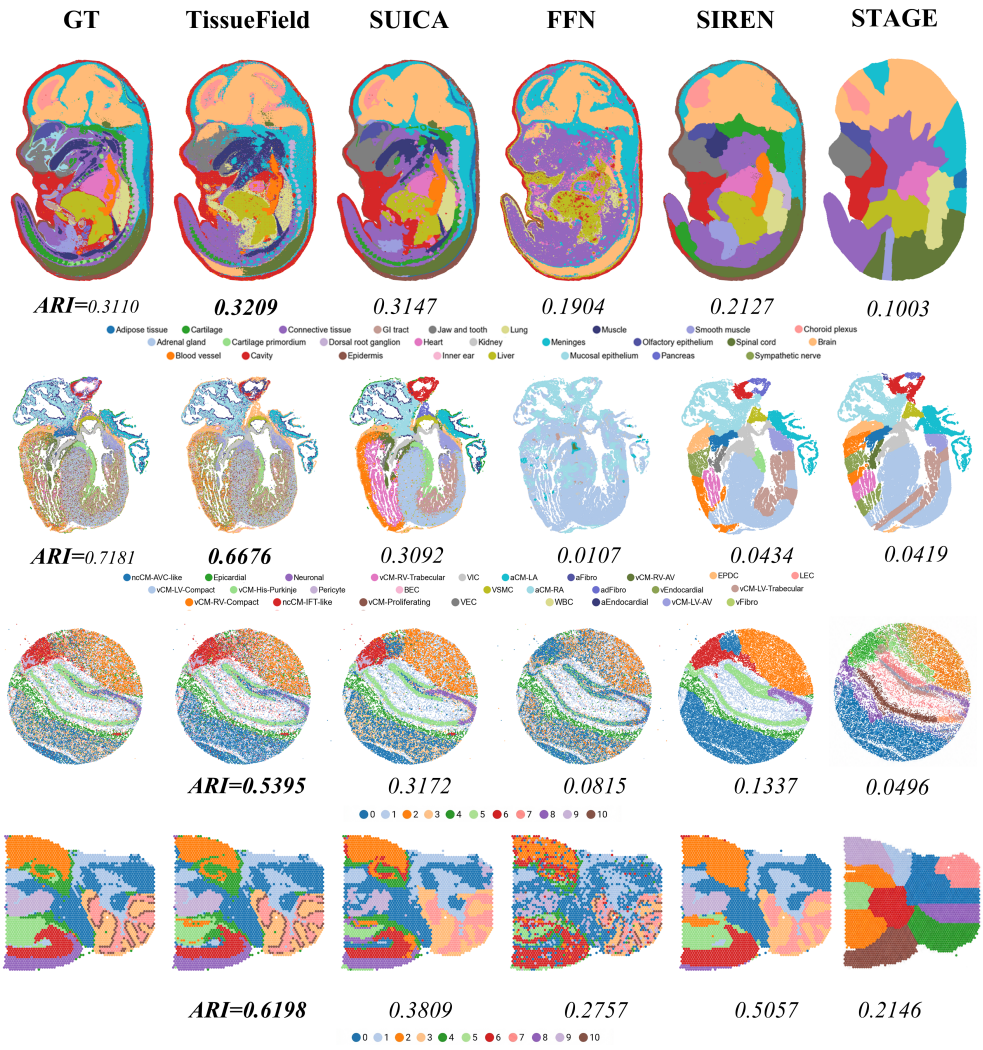


Figure 4: **Complete spatial clustering comparison.** Rows show MOSTA embryo, MERFISH human heart, Slide-seqV2 mouse hippocampus, and Visium mouse brain. Columns show measured-expression analysis (GT), TissueField, SUICA, FFN, SIREN, and STAGE. The first row is the expanded version of Figure 3, with cross-section TissueField and in-section baselines; SUICA is the SIREN variant in that example. ARI uses tissue-domain annotations for MOSTA, cell-type annotations for MERFISH, and reference clusterings for the brain datasets. Numbers below the maps give ARI values.