GENE-AMORTIZED FIELDS FOR HELD-OUT GENE IMPUTATION IN SPATIAL TRANSCRIPTOMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatial transcriptomics profiles gene expression across intact tissue but measures only a bounded gene panel. We study gene-level generalization within a tissue domain, where a gene that is described by domain-level side information but withheld entirely from model supervision must have its spatial expression reconstructed at every location from the remaining genes. A gene held out in full has no measurements of its own to fit, which makes the task out-of-column matrix completion. Completion is feasible only if genes carry side information relating masked columns to observed ones. Existing methods take that side information from an external single-cell reference in which the masked gene is measured, and none of the eight tissue domains studied here has a paired one. We instead learn a single gene-independent coordinate field, the shared field, and decode every gene from it through a map amortized from a fixed gene embedding. The embedding is computed from the panel's own co-expression rather than fitted to the gene it describes, so a gene the model was never supervised on is decoded by the same map with no code fitted to it. It need not appear in a single-cell atlas, though it must have been measured in spatial data of the domain, and a cell-type reference still anchors the count likelihood. The imputation is therefore reference-light rather than reference-free. Expression is structured at multiple non-stationary scales, so the field is multi-resolution, instantiated as a gene-amortized multi-scale wavelet representation. On eight HEST-1K tissue domains spanning two species and five organs under gene-level cross-validation, the model surpasses the strongest of ten reproduced baselines on average (0.300 against 0.265) and on seven of eight domains. The per-domain differences are significant under a paired signed-rank test ($p = 0.016$). A single-scale variant is worse on all eight. Over the two hundred genes whose measured expression is most spatially autocorrelated, the mean correlation is 0.47. A low-rank spectral coupling added to the readout leaves accuracy unchanged but exposes a factorization over genes whose modes align with measured co-expression in all eight domains once the input embedding is controlled for, and recover programs recurring across organs and species.

1 INTRODUCTION

Spatial transcriptomics (ST) resolves gene expression across intact tissue (Ståhl et al., 2016), but each assay profiles only a bounded gene panel, set by platform chemistry, cost, and detection sensitivity (Chen et al., 2015). We study the recovery of the omitted genes in its held-out form. Within a tissue domain whose slides share a panel, a subset of genes is masked and must be reconstructed at every location. The masked genes receive no supervision, but they remain part of the panel whose co-expression structure supplies the side information, so the setting is transductive along the gene axis (Sec. 3.1, Figure 1a). One family of methods transfers such a gene from an external single-cell atlas (Abdelaal et al., 2020; Biancalani et al., 2021; Lopez et al., 2019; Li et al., 2024a;b; 2025; Zeng et al., 2024), recovering it from cells that measured it. Imputation is then available only where an atlas of the same tissue carries every gene of interest, and none of the eight tissue domains we evaluate on has a paired single-cell dataset (Sec. 4.1). A second fits a model to the spatial data alone, as a continuous function of coordinates, as in implicit neural representations (INRs) (Mildenhall et al., 2020; Sitzmann et al., 2020; Luo et al., 2025; Zhu et al., 2024), or over a spatial neighbor graph (Dong & Zhang, 2022), and is thereby fit to a gene's own measurements, which the held-out

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.

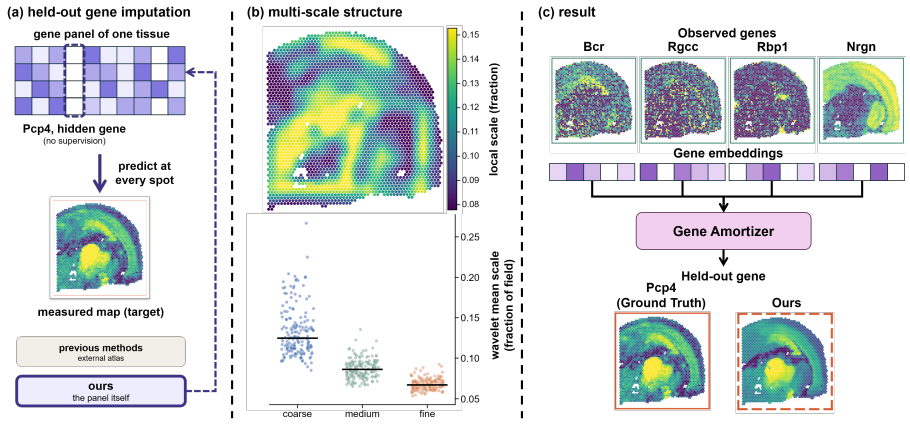

Figure 1: Held-out gene imputation on one section of mouse brain aging (C8). **(a)** The task (Sec. 3.1). **(b)** Expression scale varies across the tissue and across genes (App. A). **(c)** Four observed genes train the shared field, and the held-out gene Pcp4 is decoded from its own fixed embedding through the same amortizer, with no weight fitted to it (PCC = 0.86).

setting withholds. Each family requires exactly what the setting removes, so sharing across genes is a requirement here rather than an efficiency choice. The task is out-of-column matrix completion, in which an entire column is missing and recovery is possible only given side information relating masked columns to observed ones. This is the cold-start case of inductive matrix completion (Jain & Dhillon, 2013; Chiang et al., 2015). We take that side information from the panel itself, through a gene embedding built from co-expression across a domain's slides, and that embedding is all the decoder receives about a gene (Sec. 3.1).

The field that embedding indexes must serve every gene, and its form follows from how expression is organized in tissue. Expression is spatially structured, but not at a single characteristic scale. The dominant scale of one gene is rarely that of another, and a single gene may vary at different scales in different regions of the same section. Figure 1(b) shows both on one tissue. Within a single slide the local wavelet scale of the dominant expression program varies across locations. Across genes, the mean wavelet scale differs by a factor of 1.9 between the least and the most spatially autocorrelated tertile, and the scale is measured independently of the autocorrelation that defines the tertiles. A coordinate network tied to a single frequency band is spectrally biased and does not span this range (Rahaman et al., 2019; Tancik et al., 2020), so the shared field is multi-resolution. We instantiate it as a gene-amortized multi-scale wavelet INR anchored in raw counts by a cell-type deconvolution path (Sec. 3).

We make three contributions. First, we formulate held-out gene imputation as readout from a single gene-amortized multi-resolution coordinate field, in which the code that identifies a gene is derived from side information rather than fitted to the gene (Sec. 3.1). Second, we assemble a multi-domain benchmark of eight HEST-1K tissue domains under gene-level cross-validation, with a unified evaluation protocol and ten reproduced baselines. Earlier benchmarks of this task score against a paired single-cell reference (Li et al., 2022), whereas ours holds out whole genes within a domain. Third, we characterize which genes are recoverable, and attribute the dependence of accuracy on a gene's spatial autocorrelation to the genes rather than to a model (Sec. 4.3). On this benchmark our method reaches an average held-out correlation of 0.300 against 0.265 for the strongest reproduced baseline, surpassing it on seven of the eight domains ($p = 0.016$), and a single-scale variant is worse on all eight. A low-rank coupling over genes, which leaves accuracy unchanged, yields modes that align with measured co-expression and recover programs recurring across organs and species (App. E).

2 Related Work

Reference-based gene imputation. Most work recovers an unmeasured ST gene from an external scRNA-seq reference of the same tissue. The reference is used by aligning the two modalities (SpaGE, Abdelaal et al., 2020), by a soft assignment of reference cells to locations (Tangram, Bian-

calani et al., 2021), by a joint generative model over both (gimVI, Lopez et al., 2019), by conditional diffusion (stDiff, Li et al., 2024a; SpaDiT, Li et al., 2024b), or by style transfer with the gene as the sample (SpaIM, Li et al., 2025). All define a masked gene through its counterpart in the reference and therefore require an atlas containing every gene to be imputed. We take the side information from the domain's own spatial slides instead, so the requirement moves rather than disappears. The gene must have been measured in spatial data of the domain, and it is represented by a 64-dimensional co-expression code rather than by an expression profile over reference cells (Sec. 3.1, App. B). A cell-type signature reference is still used, for the count likelihood rather than to supply the held-out gene (Sec. 3.4).

Gene representations as side information. stImpute (Zeng et al., 2024) is the closest mechanism. It gives each gene a representation, a graph over genes built from protein language model embeddings, and propagates expression along it. It differs from ours on two axes that point in opposite directions. Its representation is external to the assay and needs no measurement of the gene, whereas our code is computed from the panel and needs the gene measured in the domain. Its graph, however, only reweights an average over reference cells that measured the held-out gene. Here the code is the whole of the decoder's input, and nothing carrying the gene's measurements is read at inference.

Spatial fields and reconstruction. A second line fits a model to the spatial data alone. STINR (Luo et al., 2025) represents expression as a single-scale sinusoidal field of tissue coordinates with joint cell-type deconvolution, and STAGATE (Dong & Zhang, 2022) reconstructs and denoises expression over a spatial neighbor graph. Both are fit to a gene's own measurements. STINR carries a per-slide offset and a decoder row for each gene on the output side, so an objective masked to the observed genes sends no gradient to either for a masked gene, and the value it returns is a function of the initialization. STINR is our closest predecessor. We make the field multi-resolution and share it across genes through amortization, so that held-out genes are decoded from their embeddings alone. STAGATE serves as our reconstruction baseline.

Conditioning a field per signal. A field can be shared across many signals and specialized to each by a code, whether a hypernetwork emits the weights (Ha et al., 2017), a latent modulates a shared synthesis network (Mehta et al., 2021), or a modulation vector is attached to a meta-learned base (Dupont et al., 2022; Tancik et al., 2021). Our readout is of this kind, but these methods obtain the code by fitting it to the signal being represented, and a held-out gene provides nothing to fit against. SUICA (Zhu et al., 2024) applies this idea to ST with the spot as the unit that receives a code: a graph auto-encoder compresses each spot's expression vector into a code that a coordinate network predicts from location, so an unmeasured spot can be decoded. Genes are only the input and output dimensions of that auto-encoder, each with its own weights, so a gene that is never observed leaves those weights untrained. All three lines therefore fail on a held-out gene for the same reason. Each needs something that comes from the gene being predicted, whether an entry in an atlas, its own measurements on the slide, or a code fitted to it. This work needs none of the three. Its cost, stated in Sec. 3.1, is that the gene must have been measured somewhere in the domain.

3 METHOD

We treat held-out gene imputation as *out-of-column matrix completion* and solve it with a single gene-independent multi-resolution coordinate field, *the shared field*, that every gene, observed or held out, decodes through a map produced from a fixed gene embedding. The architecture follows from one requirement. A held-out gene has no column in the training objective (Sec. 3.1), so nothing on the path that produces its prediction may be a parameter fitted to that gene. A per-gene weight on the output side, as in STINR and SUICA, would be left at its initialization. What remains is a spatial representation shared by every gene and fit from the observed ones (Sec. 3.2), and gene-specific decoding parameters produced from a fixed side-information embedding rather than learned (Sec. 3.3).

The forward pass is as follows. A coordinate enters a wavelet encoder and comes out as K spatial feature streams that carry no gene index. Each gene's fixed embedding enters an amortizer and comes out as that gene's readout weights. The two meet in one inner product per stream, which

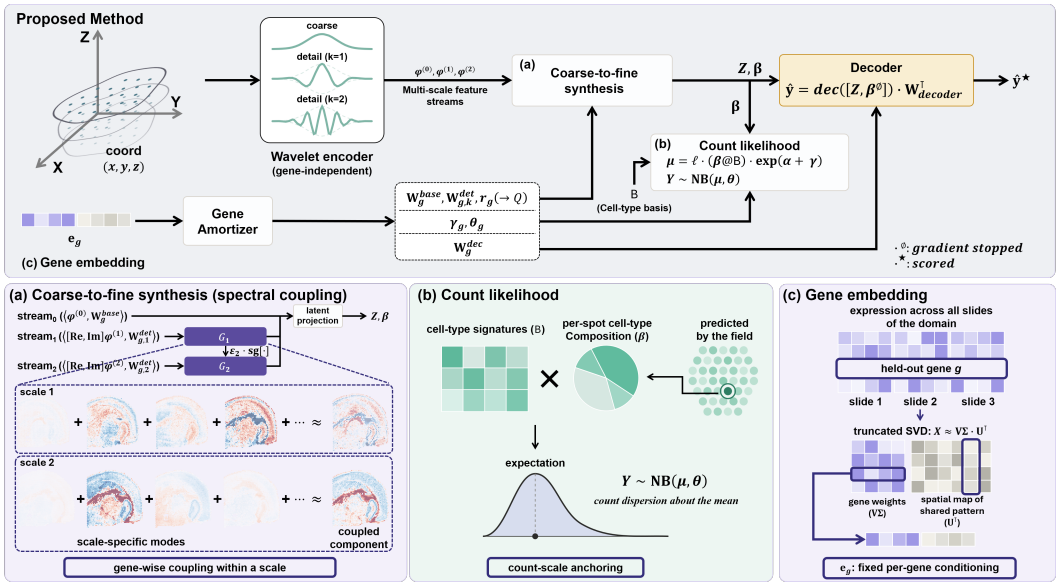

Figure 2: The model. Coordinates enter the wavelet encoder and each gene's fixed embedding enters the gene amortizer (Sec. 3.3). The branches meet in the coarse-to-fine synthesis (a), whose latent projection yields the spot latent and the composition that feed the count likelihood (b) and the decoder (Sec. 3.4). (c) The gene embedding (Sec. 3.1). A star marks the scored quantity and a null sign a stopped gradient.

gives the multi-resolution readout $\hat{x}_{ig}$. The readout of a spot over the whole panel is projected to a spot latent $\mathbf{z}_i$, from which a cell-type composition and the scored prediction $\hat{y}_{ig}$ are decoded, and a negative-binomial likelihood on the raw counts ties the composition to a fixed cell-type basis (Sec. 3.4). A low-rank coupling across genes is added to the readout (Sec. 3.5) and analysed in App. E (Figure 2 draws the two branches).

3.1 Out-of-column matrix completion

A tissue domain provides slides that share a common panel of G genes at spatial coordinates $\mathbf{x}_i \in [0,1]^2$ (with a slide-depth coordinate appended, so inputs are in $\mathbb{R}^3$ across serial slides). Let $\mathbf{Y} \in \mathbb{N}^{N\times G}$ be observed counts with library sizes ℓ_i and slide labels s_i, and let $\mathbf{X} \in \mathbb{R}^{N\times G}$ be the corresponding log-normalized expression. Under gene-level cross-validation the panel is split into an observed set $\mathcal{O}$ and a disjoint held-out set $\mathcal{H}$. The columns $\mathbf{Y}_{:,\mathcal{H}}$ are withheld from supervision and must be predicted at every spot. Because an entire column is missing, completion is possible only if genes carry side information relating held-out to observed columns, which we supply as a *fixed* embedding $\mathbf{e}_g \in \mathbb{R}^{68}$ per gene (defined below). Imputing a gene then reduces to a readout of the shared field indexed by its embedding.

Scope of held-out set. A gene $g \in \mathcal{H}$ is withheld from supervision on every slide of the domain. No term of the objective is evaluated on $\mathbf{Y}_{:,g}$, and no parameter of the model is fitted to those measurements, which enter the model in one place only, the decomposition over the full panel that produces the gene embeddings (Figure 2c). The setting is therefore transductive along the gene axis. A gene measured nowhere in the domain lies outside this formulation, since a code relating a gene to the rest of the panel can only be formed where that relation is measurable (Sec. 5).

Gene embedding. The embedding is fixed rather than learned. It concatenates a 64-dimensional co-expression code, taken from a truncated SVD of the domain's expression matrix over the full panel, with the gene's standardized Moran's I (Moran, 1950) and a one-hot of its spatial scale group, giving $\mathbf{e}_g \in \mathbb{R}^{68}$. The scale group is a tertile of that same Moran's I, so the last four dimensions carry one measurement at two resolutions. App. B gives the construction and the control that would separate what the code carries from the slides on which it is computed.

3.2 A SHARED MULTI-RESOLUTION FIELD

The field is the part of the model that every gene shares. It turns a coordinate into spatial features that no gene owns, so that what a held-out gene needs to know about location is learned from the observed genes. The core object is a gene-independent field $\boldsymbol{\phi} : \mathbb{R}^3 \to (\mathbb{R}^b)^K$ that maps a coordinate to K=3 multi-scale feature streams $\boldsymbol{\phi}(\mathbf{x}) = (\boldsymbol{\phi}^{(0)}(\mathbf{x}), \ldots, \boldsymbol{\phi}^{(K-1)}(\mathbf{x}))$, b=64. Expression over tissue is non-stationary and multi-scale (Fig. 1), a range a single-frequency encoder does not span (Sitzmann et al., 2020; Tancik et al., 2020; Rahaman et al., 2019). We use a stack of multi-scale wavelet layers (Saragadam et al., 2023; Fathony et al., 2021; Lindell et al., 2022). The coarse stream applies a Gaussian envelope, and each finer stream k applies a complex Gabor activation with carrier frequency $\omega_k = 2^{k-1}$, so the streams span coarse to fine. The field carries no gene index. It is fit from the observed genes and is the same field a held-out gene is decoded from, which is what makes out-of-column completion well posed.

3.3 GENE AMORTIZATION

The amortizer is what makes a held-out gene decodable. It replaces the per-gene weights a field would otherwise learn with weights computed from the gene's fixed embedding, so genes with and without supervision are decoded by the same map. Each gene decodes the field through parameters produced from its embedding by a shared amortization network with a trunk and linear heads,

$$\mathbf{r}_g = \mathrm{trunk}(\mathbf{e}_g) \in \mathbb{R}^{64}, \qquad \big(\log \theta_g,\ \boldsymbol{\gamma}_g,\ \mathbf{w}_g^{\mathrm{base}},\ \{\mathbf{w}_{g,k}^{\mathrm{det}}\}_{k\geq 1},\ \mathbf{w}_g^{\mathrm{dec}}\big) = \mathrm{heads}(\mathbf{r}_g),$$

where trunk is a two-layer MLP. Every quantity through which gene g is decoded is a function of $\mathbf{e}_g$, namely the negative-binomial dispersion θ_g, the per-slide bias $\boldsymbol{\gamma}_g$, the per-stream readouts $\mathbf{w}_g^{\mathrm{base}} \in \mathbb{R}^b$ and $\mathbf{w}_{g,k}^{\mathrm{det}} \in \mathbb{R}^{2b}$, and the decoder projection $\mathbf{w}_g^{\mathrm{dec}}$. The one gene-indexed weight that is free, the projection that aggregates the field readout into the per-spot latent of Sec. 3.4, sits on the input side and is shared by every gene. The coordinate-field methods we compare against place their free per-gene parameters on the output side, where a masked gene leaves them at their initialization and no prediction that depends on the gene is defined.

The gene's expression field is synthesized coarse to fine, as a coarse baseline plus finer detail,

$$\hat{x}_{ig} = \big\langle \boldsymbol{\phi}^{(0)}(\mathbf{x}_i),\, \mathbf{w}_g^{\mathrm{base}} \big\rangle \;+\; \sum_{k=1}^{K-1} d_{ig}^{(k),\mathrm{final}}, \qquad d_{ig}^{(k)} = \big\langle [\Re, \Im]\, \boldsymbol{\phi}^{(k)}(\mathbf{x}_i),\, \mathbf{w}_{g,k}^{\mathrm{det}} \big\rangle,$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product and $[\Re, \Im]$ concatenates the real and imaginary parts of the complex detail stream. Each gene therefore reads each stream through one linear functional whose weights come from its embedding. Here $d^{(k)}$ is the raw detail readout, and $d^{(k),\mathrm{final}}$ equals $d^{(k)}$ unless the coupling of Sec. 3.5 is applied. We write $\hat{x}_{ig}^{(\leq k)}$ for the partial sum that stops after scale k, so that $\hat{x}^{(\leq K-1)} = \hat{x}$, and the objective of Sec. 3.4 uses these partial sums. The code that specializes the field to a gene is therefore read rather than fitted, which is the step Sec. 2 identifies as undefined for a held-out gene, and the trunk and heads are fit on observed genes only. A gene the model has never been supervised on is decoded by that same map.

3.4 PREDICTION AND OBJECTIVE

This section turns the readout into the two quantities that are trained, the scored prediction and a count likelihood that anchors a cell-type composition in the raw counts. The readout of spot i over the whole panel, $\hat{\mathbf{x}}_i \in \mathbb{R}^G$, is projected to a per-spot latent $\mathbf{z}_i = \mathbf{P}\hat{\mathbf{x}}_i$ by the one free gene-indexed weight $\mathbf{P}$. From $\mathbf{z}_i$ we form a cell-type composition $\boldsymbol{\beta}_i = \mathrm{softmax}(\mathbf{W}_\beta \mathbf{z}_i) \in \Delta^{C-1}$ and a per-spot scale a_i (written α in Figure 2), a linear function of $\mathbf{z}_i$ and a learned slide embedding. Two quantities are decoded from these. The first is the per-gene prediction we impute and score,

$$\hat{y}_{ig} = \big\langle \mathrm{dec}([\mathbf{z}_i,\ \mathrm{sg}[\boldsymbol{\beta}_i]]),\, \mathbf{w}_g^{\mathrm{dec}} \big\rangle,$$

where dec is a single wavelet layer, $\mathrm{sg}[\cdot]$ stops the gradient, and the projection $\mathbf{w}_g^{\mathrm{dec}}$ is a function of the gene embedding and so applies unchanged to held-out genes. The composition is detached where it enters the decoder, so it is estimated by the count likelihood below and read here as a

fixed per-spot feature. The second quantity grounds the model in raw counts. With a fixed cell-type signature matrix $\mathbf{B} \in \mathbb{R}^{C\times G}$ the negative-binomial mean is

$$\mu_{ig} = \ell_i \, (\boldsymbol{\beta}_i \mathbf{B})_g \, \exp\big(a_i + \gamma_{g,s_i}\big), \qquad Y_{ig} \sim \mathrm{NB}(\mu_{ig}, \theta_g),$$

following the INR-with-deconvolution lineage (Luo et al., 2025). $\mathbf{B}$ holds the mean log-normalized expression of each labelled cell type in the reference (App. A). Its rows are not normalized to proportions, and the per-spot scale a_i and the per-gene, per-slide bias $\gamma_{g,s}$ absorb the resulting scale. The per-spot $\boldsymbol{\beta}$ is what App. F reports. The scored prediction $\hat{y}$ is never multiplied by $\mathbf{B}$, so a held-out gene is imputed without appearing in any single-cell atlas. The prediction is reference-light rather than reference-free.[1]

Objective. The model is a single network fit jointly on all slides of a domain. Its objective has three data terms, each evaluated on the observed genes $\mathcal{O}$ only (and, under spot-level hold-out, on observed spots):

$$\mathcal{L}_{\mathrm{NB}} = -\frac{1}{N|\mathcal{O}|} \sum_i \sum_{g\in\mathcal{O}} \log \mathrm{NB}(Y_{ig}; \mu_{ig}, \theta_g), \qquad \mathcal{L}_{\mathrm{dec}} = \frac{1}{N|\mathcal{O}|} \sum_i \sum_{g\in\mathcal{O}} \big(\hat{y}_{ig} - X_{ig}\big)^2,$$

$$\mathcal{L}_{\mathrm{mr}} = \frac{1}{K} \sum_{k=0}^{K-1} \frac{1}{N|\mathcal{O}|} \sum_i \sum_{g\in\mathcal{O}} \big(\hat{x}_{ig}^{(\le k)} - (\mathcal{S}_k \mathbf{X})_{ig}\big)^2,$$

where $\hat{x}^{(\le k)}$ are the partial sums of the coarse-to-fine synthesis and $\mathcal{S}_k$ is a spatial smoothing operator applied $2(K-1-k)$ times. The coarse partial sum is therefore matched to a heavily smoothed target and the full synthesis to the measurement itself (App. A). The count term and the reconstruction term $\mathcal{L}_{\mathrm{dec}} + \mathcal{L}_{\mathrm{mr}}$ are balanced by learned uncertainty weights (Kendall et al., 2017), and the two coupling regularizers of Sec. 3.5 are added with fixed weights. The forward pass always computes all G columns, so nothing about a held-out gene has to be added back at inference. Held-out genes $g \in \mathcal{H}$ are decoded from their embeddings through the same amortizer and shared field, with no retraining, and are scored exactly as the baselines are, by comparing $\hat{y}$ to the measured expression on the held-out columns only under a shared metric. The same configuration is used for every domain, and App. A gives it.

3.5 Scale-specialized low-rank spectral coupling

The coupling is an analysis module rather than a source of accuracy. The readout of Sec. 3.3 treats each gene independently given the field, and the coupling adds on each detail scale a low-rank term that rebuilds a gene's detail from modes shared by the panel, so that the model exposes a gene-axis factorization that can be inspected (App. E) at a cost of about two thousand parameters and no measurable change in accuracy (Sec. 4.5). Writing $\mathbf{r} \in \mathbb{R}^{G\times 64}$ for the stacked representations $\mathbf{r}_g$, the coupling operator at scale k is

$$\mathbf{G}_k = \mathbf{Q}^{(k)} \, \mathrm{diag}(\mathbf{g}^{(k)}) \, (\mathbf{Q}^{(k)})^\top, \qquad \mathbf{Q}^{(k)} = \mathrm{head}_Q^{(k)}(\mathbf{r}) \in \mathbb{R}^{G\times m}, \; m = 16,$$

with per-scale gains $\mathbf{g}^{(k)} = \mathrm{softplus}(\cdot)$. In form, $\mathbf{G}_k$ is a spectral graph filter on the gene axis (Bruna et al., 2014; Defferrard et al., 2016; Kipf & Welling, 2017) whose basis is produced from the gene representation by the amortizer rather than taken from a given graph (Ha et al., 2017; Franceschi et al., 2019; Kazi et al., 2023), which is where the module departs from stImpute (App. A). The gated detail is combined with the raw detail through a single learned scalar $\lambda = \sigma(\cdot)$, and carries a small stopped-gradient residual from the previous scale:

$$\mathbf{d}_{\mathrm{gated}}^{(k)} = \mathbf{d}^{(k)} \mathbf{G}_k + \varepsilon_k \, \mathrm{sg}\Big[\mathbf{d}_{\mathrm{gated}}^{(k-1)}\Big] \;\; (k \ge 2), \qquad d^{(k),\mathrm{final}} = (1-\lambda)\, d^{(k)} + \lambda \, \mathbf{d}_{\mathrm{gated}}^{(k)}.$$

The default specializes $\mathbf{Q}^{(k)}$ per scale and adds two soft regularizers that keep the scales in distinct gene subspaces, and a "shared" routing that ties $\mathbf{Q}^{(k)}$ across scales recovers a single-operator variant (App. A). The mixing weight settles between 0.12 and 0.26 across the eight domains, so the coupled detail stays a minority of the readout. The module's effect on held-out accuracy is within run-to-run noise (Sec. 4.5). Its role in this paper is as an inspectable factorization over genes, read as co-expression programs in App. E.

[1] $\boldsymbol{\beta}$ is estimated through the $\mathbf{B}$-based likelihood and enters the scored prediction, so the reference influences $\hat{y}$ even though $\mathbf{B}$ is never multiplied into it. The detachment fixes the direction of that influence rather than removing it.

4 EXPERIMENTS

4.1 SETUP

We evaluate on eight tissue domains from HEST-1K (Jaume et al., 2024) spanning two species and five organs. The human domains are DLPFC (C1), kidney ccRCC (C2), liver (C3), bowel CRC (C4) and heart (C5), and the mouse domains are liver regeneration (C6), liver hepatoblastoma (C7) and brain aging (C8). The domains comprise 90 Visium sections from HEST-1K in total (10, 11, 8, 19, 15, 11, 10 and 6 for C1 to C8, with sample identifiers in App. A). One model with one set of hyperparameters is fit per domain, and nothing is tuned per tissue. Each domain's panel of 1000 genes is drawn from the domain-wide highly variable genes by stratified sampling across the tertiles of Moran's I, about 333 genes per tertile, so it spans the full autocorrelation range rather than only spatially variable genes (App. A). The primary protocol is gene-level cross-validation. Within a domain a fold of the panel is masked, the model is fit on the remaining genes only, and the masked columns are predicted at every spot. We report the mean Pearson correlation between predicted and measured expression over held-out genes, averaged over slides. All baselines are reproduced from the authors' repositories and run through one driver, so every method sees the same panel, gene split and evaluator. The reference-based baselines require a reference in which the masked gene is measured. None of these domains has a paired single-cell dataset, so for all of them alike the reference is the domain's remaining slides. Every method therefore receives some information about a held-out gene from measurements of it. The reference-based baselines receive its cell-level expression on the other slides of the domain, whereas our model receives only its row of the co-expression code, sixty-four coordinates computed over all slides that carry no spot index (Sec. 3.1). The side information our model uses is thus no more informative about the held-out gene than what the baselines are given, and the transductive element it introduces is the same one the reference-based setting already contains. STINR and SUICA define no prediction for a held-out gene (Sec. 2), so we compare against each on the protocols where it is defined (Sec. 4.4).

4.2 HELD-OUT GENE IMPUTATION

Table 1: Held-out gene imputation under gene-level cross-validation: mean Pearson correlation over held-out genes, per tissue domain and averaged, best per column in bold. Cluster legend, used by every later table: C1 DLPFC, C2 kidney (ccRCC), C3 human liver, C4 bowel (CRC), C5 heart, C6 mouse liver (regen.), C7 mouse liver (hepatoblastoma), C8 mouse brain (aging).

Method	C1	C2	C3	C4	C5	C6	C7	C8	Avg
SpaGE	0.2258	0.1841	0.1907	0.2459	0.2579	0.2321	0.2448	**0.5386**	0.2650
Tangram	0.1862	0.1753	0.1858	0.2422	0.2283	0.2015	0.2095	0.4980	0.2409
SpaIM	0.2127	0.1769	0.1709	0.2414	0.2102	0.1741	0.2113	0.4183	0.2270
gimVI	0.1726	0.1322	0.1594	0.1921	0.1872	0.2322	0.2234	0.4996	0.2248
stImpute	0.1294	0.0937	0.1807	0.1771	0.1727	0.2066	0.2023	0.4684	0.2039
stDiff	0.1481	0.1257	0.1364	0.1667	0.1850	0.1649	0.1611	0.4261	0.1893
SpaDiT	-0.1047	-0.0516	-0.0143	-0.0911	-0.0488	-0.0217	-0.0573	-0.1652	-0.0693
Ours	**0.2384**	**0.2598**	**0.2655**	**0.2815**	**0.2839**	**0.2758**	**0.2607**	0.5324	**0.2997**

The shared field attains the highest average held-out correlation on the benchmark, 0.2997 against 0.2650 for SpaGE, the strongest baseline on this protocol, and leads on seven of the eight domains. With the domain as the unit of analysis the per-domain differences are significant ($p = 0.016$, Table 4). The margins are largest where the reference-based baselines do worst, kidney ($+0.076$) and human liver ($+0.075$).

Against the two most recent methods of the reference-based class the result is unanimous, $+0.073$ over SpaIM and $+0.096$ over stImpute, in all eight domains each ($p = 0.008$). Neither surpasses SpaGE, which leads both on all eight domains, so the comparison above is against the strongest member of that class. Figure 3 shows the margin on tissue for one gene per quartile of spatial autocorrelation. The baselines lead on one domain, C8, where SpaGE is ahead by 0.006. C8 is also the domain that separates methods least, since every method scores near 0.5 there, and excluding it our margin over SpaGE widens from $+0.035$ to $+0.041$, but we report it in every table. SpaDiT is negative on all eight domains and is reported as run.

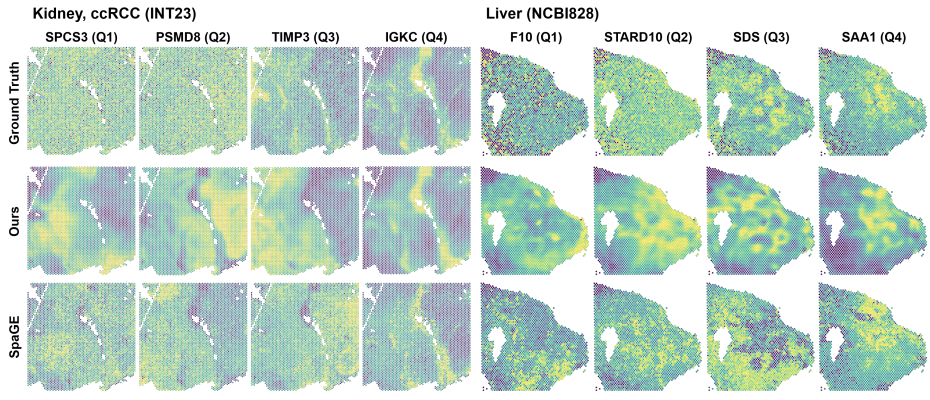

Figure 3: Held-out gene imputation on one section of two tissues, one gene per Moran's I quartile of the measured expression. Columns are grouped by tissue and run Q1 to Q4 within each group, and rows are the measurement, our method and SpaGE. Our correlations on these sections run 0.16, 0.26, 0.47, 0.89 in kidney and 0.16, 0.29, 0.49, 0.52 in liver, against 0.10, 0.04, -0.11, 0.63 and 0.07, 0.20, -0.08, 0.38 for SpaGE. Colour is scaled per panel between the 2nd and 98th percentile of that panel's own values, so intensity is comparable within a panel and not across panels. Sections are the ones nearest the median of our per-slide correlation for these genes, so the panels are typical rather than best.

4.3 Determinants of gene-level recoverability

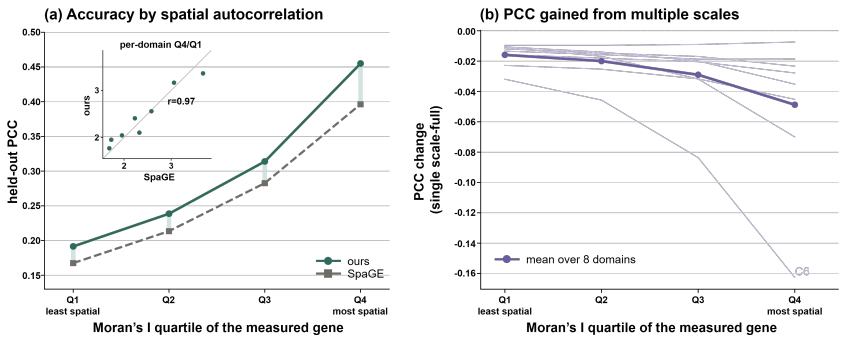

Figure 4: **(a)** Held-out accuracy against the Moran's I quartile of the gene's measured expression, averaged over the eight domains, and the inset compares the two methods' per-domain Q4/Q1 ratios. **(b)** The accuracy a single-scale variant gives up, by the same quartiles, plotted as single scale minus full so that a more negative value is a larger loss. Thin lines are the eight domains and C6 is the outlier at Q4.

Accuracy on this task is predictable from a property of the gene that is fixed before any model is fit. Binned by the Moran's I of their measured expression (Table 3, Fig. 4a), held-out genes rise from 0.192 in the least autocorrelated quartile to 0.455 in the most, monotonically in every domain, and reach 0.522 on the fifty genes with the highest Moran's I. SpaGE, a method of an unrelated lineage evaluated identically, returns the same gradient, $2.37\times$ against our $2.38\times$, monotone in all eight domains and with per-domain ratios correlating at 0.97, which places the gradient in the genes rather than in a model. It is not an artefact of holding out columns either, since the stratification is $2.13\times$ on reconstruction, where every gene is observed at training, and $2.5\times$ under within-slide spot masking, monotone in all eight domains throughout.

The panel average is therefore a summary of a wide distribution rather than a level (Fig. 5). Over the k genes with the highest Moran's I the mean correlation is 0.505 at $k = 10$ and still 0.468 at $k = 200$, and 39% of held-out genes clear a correlation of 0.3 against 32% for SpaGE. The Pcp4 example of Figure 1(c) belongs to this top stratum, whose mean in that domain is 0.83 against 0.53

for the panel. Our lead holds in every quartile on the same seven domains, each significant on its own (Table 4), and is flat at 11% to 15% in relative terms, so it is neither confined to the spatially smooth genes nor produced by them.

4.4 Other protocols

Two further protocols place the method against the two spatial fields that define no held-out gene prediction (App. D). Under STINR's own protocol, which hides 60 to 80% of the spots in each slide and scores the hidden ones on identical spots, we lead STINR by 0.061, 0.060 and 0.058 at the three ratios and in seven of eight domains at each. SUICA accepts no external spot mask and is reported twice, 0.204 on the full slide, where 80% of the spots were seen in its own training, and 0.022 against our 0.261 on the spots both procedures withheld, a difference of 0.239 in all eight domains. Scored on the genes observed at training, we lead STINR on all eight domains and by 0.080 on average, so sharing the field across genes does not reduce accuracy on the observed genes (STAGATE, a reconstruction method, leads that table). Holding out an entire slide instead yields 0.046, which lies on a different generalization axis from the one this paper addresses (App. G, Table 11).

4.5 Ablation

Table 2 (App. C) separates the three components under gene-CV. The multi-resolution representation carries the design. A single-scale variant is worse in all eight domains, by 0.028 on average ($p = 0.008$) and by as much as 0.081 on C6, an effect consistent across tissues rather than driven by one. Slicing that loss by the bins of Sec. 4.3 localizes it (Fig. 4b). Dropping to a single scale costs 0.049 in the most autocorrelated quartile against 0.016 in the least, a larger loss in Q4 in seven of eight domains and reaching 0.163 on C6. The ordering survives normalizing by each bin's own level (-10.7% against -8.3%). The scales contribute where there is spatial structure to represent.

The spectral coupling is measured on the accuracy axis here. Removing it changes the average by 0.001 and tying its modes across scales by 0.0003. Neither reaches significance ($p = 0.195$ and $p = 0.445$, Table 4), and each variant exceeds the full model on three of the eight domains, so the direction is not determined rather than small and fixed. What the modes recover is a separate question, answered in App. E. The sixteen modes of every tissue reach enrichment significance, the programs shared between tissues follow shared biology (the two brains, the liver pairs and the two carcinomas), and the coupling still tracks measured co-expression in all eight domains after the input embedding is partialled out. The deconvolution path also yields a per-spot cell-type composition, which is spatially organised in five of the eight domains and is evaluated qualitatively in App. F.

5 Conclusion and Limitations

We learn one gene-independent multi-resolution coordinate field and decode every gene from it through a map amortized from a fixed embedding. The embedding is computed from the panel's own co-expression rather than fitted to the gene, so a gene the model was never supervised on is decoded by the same map as an observed one. Across eight HEST-1K domains spanning two species and five organs, one model with one set of hyperparameters attains the highest average held-out correlation and leads on seven of the eight, with a single-scale variant worse on all eight. The consistency, more than the margin, is the evidence that the field captures a general property of tissue expression rather than the particulars of one tissue.

Correlations on this task are low for every method, 0.189 to 0.265 for the reproduced baselines against our 0.300. Part of that level is irreducible (Sec. 4.3), but we have not established where the ceiling lies for the recoverable genes, and we do not claim that the task is solved. Imputation here is reference-light rather than reference-free. A held-out gene need not appear in a single-cell atlas, but its co-expression code exists only if the gene is measured in spatial data of the domain. What changes is therefore which reference is required rather than whether one is (Sec. 3.1, App. B), and a cell-type signature matrix still anchors the count likelihood (Sec. 3.4). App. G states four further limitations, among them that the shared field does not transfer across slides.

AI USE STATEMENT

We used a large language model as a writing aid, to tighten prose, shorten captions, and check internal consistency between the reported numbers and the tables. All research ideas, experimental design, implementation, analysis, and claims are the authors' own, and the authors verified every number and statement in the paper against the source data.

REFERENCES

Tamim Abdelaal, Soufiane Mourragui, Ahmed Mahfouz, and Marcel J. T. Reinders. SpaGE: Spatial gene enhancement using scRNA-seq. *Nucleic Acids Research*, 48(18):e107, 2020. doi: 10.1093/nar/gkaa740.

Michael Ashburner, Catherine A. Ball, Judith A. Blake, et al. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, 2000. doi: 10.1038/75556.

Tommaso Biancalani, Gabriele Scalia, Lorenzo Buffoni, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nature Methods*, 18(11):1352–1362, 2021. doi: 10.1038/s41592-021-01264-7.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR)*, 2014.

Edward Y. Chen, Christopher M. Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela V. Meirelles, Neil R. Clark, and Avi Ma'ayan. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14(1):128, 2013. doi: 10.1186/1471-2105-14-128.

Kok Hao Chen, Alistair N. Boettiger, Jeffrey R. Moffitt, Siyuan Wang, and Xiaowei Zhuang. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 348(6233):aaa6090, 2015. doi: 10.1126/science.aaa6090.

Kai-Yang Chiang, Cho-Jui Hsieh, and Inderjit S Dhillon. Matrix completion with noisy side information. *Advances in neural information processing systems*, 28, 2015.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, 2016.

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. doi: 10.1038/s41467-022-29439-6.

Emilien Dupont, Hyunjik Kim, S. M. Ali Eslami, Danilo Jimenez Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one. In *International Conference on Machine Learning (ICML)*, volume 162 of *PMLR*, pp. 5694–5725, 2022.

Rizal Fathony, Anit Kumar Sahu, Devin Willmott, and J. Zico Kolter. Multiplicative filter networks. In *International Conference on Learning Representations (ICLR)*, 2021.

Luca Franceschi, Mathias Niepert, Massimiliano Pontil, and Xiao He. Learning discrete structures for graph neural networks. In *International Conference on Machine Learning (ICML)*, volume 97 of *PMLR*, 2019.

David Ha, Andrew M. Dai, and Quoc V. Le. Hypernetworks. In *International Conference on Learning Representations (ICLR)*, 2017.

Prateek Jain and Inderjit S Dhillon. Provable inductive matrix completion. *arXiv preprint arXiv:1306.0626*, 2013.

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. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 2024.

Anees Kazi, Luca Cosmo, Seyed-Ahmad Ahmadi, Nassir Navab, and Michael M. Bronstein. Differentiable graph module (DGM) for graph convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1606–1617, 2023.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics (2017). *arXiv preprint arXiv:1705.07115*, 2017.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Maxim V. Kuleshov, Matthew R. Jones, Andrew D. Rouillard, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1):W90–W97, 2016. doi: 10.1093/nar/gkw377.

Bin Li, Wen Zhang, Chuang Guo, Hao Xu, Longfei Li, Minghao Fang, Yinlei Hu, Xinye Zhang, Xinfeng Yao, Meifang Tang, et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nature methods*, 19(6):662–670, 2022.

Bo Li, Ziyang Tang, Aishwarya Budhkar, Xiang Liu, Tonglin Zhang, Baijian Yang, Jing Su, and Qianqian Song. Spaim: single-cell spatial transcriptomics imputation via style transfer. *Nature Communications*, 16(1):7861, 2025.

Kongming Li, Jiahao Li, Yuhao Tao, and Fei Wang. stDiff: a diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Briefings in Bioinformatics*, 25(3):bbae171, 2024a. doi: 10.1093/bib/bbae171.

Xiaoyu Li, Fangfang Zhu, and Wenwen Min. SpaDiT: diffusion transformer for spatial gene expression prediction using scRNA-seq. *Briefings in Bioinformatics*, 25(6):bbae571, 2024b. doi: 10.1093/bib/bbae571.

David B. Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. BACON: Band-limited coordinate networks for multiscale scene representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16252–16262, 2022.

Romain Lopez, Achille Nazaret, Maxime Langevin, Jules Samaran, Jeffrey Regier, Michael I. Jordan, and Nir Yosef. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *arXiv preprint arXiv:1905.02269*, 2019. ICML Workshop on Computational Biology.

Yisi Luo, Xile Zhao, Kai Ye, and Deyu Meng. STINR: Deciphering spatial transcriptomics via implicit neural representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25930–25939, 2025.

Ishit Mehta, Michaël Gharbi, Connelly Barnes, Eli Shechtman, Ravi Ramamoorthi, and Manmohan Chandraker. Modulated periodic activations for generalizable local functional representations. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14214–14223, 2021.

Marija Milacic, Deidre Beavers, Patrick Conley, et al. The Reactome pathway knowledgebase 2024. *Nucleic Acids Research*, 52(D1):D672–D678, 2024. doi: 10.1093/nar/gkad1025.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, pp. 405–421, 2020.

P. A. P. Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950. doi: 10.2307/2332142.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning (ICML)*, volume 97 of *PMLR*, pp. 5301–5310, 2019.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G. Baraniuk. WIRE: Wavelet implicit neural representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18507–18516, 2023.

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

Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016. doi: 10.1126/science.aaf2403.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, et al. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:7537–7547, 2020.

Matthew Tancik, Ben Mildenhall, Terrance Wang, Divi Schmidt, Pratul P. Srinivasan, Jonathan T. Barron, and Ren Ng. Learned initializations for optimizing coordinate-based neural representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2846–2855, 2021.

Yuansong Zeng, Yujie Song, Chengyang Zhang, Haoxuan Li, Yongkang Zhao, Weijiang Yu, Shiqi Zhang, Hongyu Zhang, Zhiming Dai, and Yuedong Yang. Imputing spatial transcriptomics through gene network constructed from protein language model. *Communications Biology*, 7 (1):1271, 2024.

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. *arXiv preprint arXiv:2412.01124*, 2024.

A Implementation

One configuration is used for every domain and nothing is tuned per tissue. The wavelet encoder has three layers of width 256 with Morlet activations, mapping to $K = 3$ streams of width $b = 64$. The coarse stream carries a Gaussian envelope and the two detail streams complex Gabor carriers at $\omega_k = 2^{k-1}$. The gene embedding is $\mathbf{e}_g \in \mathbb{R}^{68}$ and the amortizer trunk is a two-layer MLP of width 128 with SiLU activation onto $\mathbf{r}_g \in \mathbb{R}^{64}$. Every head is linear, and the coupling heads carry no bias. The spectral coupling has rank $m = 16$ with per-scale routing, gains initialized at unit softplus, cross-scale residual weights initialized at 0.05, and the mixing weight λ initialized at $\sigma(-2.2) = 0.10$.

In the language of spectral graph convolution, $\mathbf{G}_k$ is the $\mathbf{U}\, h(\mathbf{\Lambda})\, \mathbf{U}^\top$ form with the m modes $\mathbf{Q}^{(k)}$ in place of the Laplacian eigenbasis and the gains in place of the spectral response, and, as in graph-structure learning, the basis is learned rather than given. The two coupling regularizers of Sec. 3.5 are cross-scale subspace orthogonality $\sum_{k<k'} \|(\mathbf{Q}^{(k)})^\top \mathbf{Q}^{(k')}\|_F^2$ at weight 0.01 and gain smoothness $\sum_k \|\mathbf{g}^{(k)} - \mathbf{g}^{(k-1)}\|^2$ at weight 0.001. Both penalties act between scales and neither acts within one, so the m modes of a single scale are not constrained against each other and are not m independent programs, and under shared routing the orthogonality term is identically zero. With the default routing the normalized cross term between the two detail scales falls between 5×10^{-6} and 4×10^{-4} over the eight domains. The mixing weight is initialized at 0.10. The operator $\mathbf{G}_k$ is unchanged when a column of $\mathbf{Q}^{(k)}$ flips sign, so which end of a mode counts as positive is not fixed by the model. The direction reported in Table 8 separates the two ends of a mode and no more. The per-spot latent is 128-dimensional and the slide embedding 16-dimensional. The model also accepts, in place of a labelled signature $\mathbf{B}$, non-negative archetypes initialized by NMF on the panel and fine-tuned under a softplus parameterization, but every reported run uses the labelled signature.

Training uses Adamax at a learning rate of 10^{-3} for 30k steps, with the negative-binomial and reconstruction terms balanced by uncertainty weighting, one learned log-variance per term (Kendall et al., 2017). The smoothing operator $\mathcal{S}$ of the multi-resolution term averages each spot with its 15 nearest neighbours under Gaussian weights whose bandwidth is the median neighbour distance, mixing one half of the current value with one half of that average per application, and is applied four, two and zero times to the targets of the three partial sums. The gene-level split is a five-fold shuffled partition of the panel with a fixed seed, so each reported per-domain value aggregates five models, each scoring the fold it did not see. Spot-level masking draws its held-out spots from a fixed seed shared with the STINR runs, so both methods are scored on identical spots. Each domain is trained once, so no reported number carries a seed variance. Every linear layer of the field, the latent projection and the decoder pre-layer is reparameterized as $\mathbf{\Lambda}\mathbf{B}_{\mathrm{w}}$ with a fixed orthonormal wavelet basis $\mathbf{B}_{\mathrm{w}}$ and a learned coefficient matrix $\mathbf{\Lambda}$ (run configuration `reparam=wavelet`). Six domains rerun with an identical configuration differ from the reported values by 0.0004 on average and 0.0009 at most, which is the run-to-run noise quoted in App. C.

Panel and preprocessing. Within a domain the slides are intersected to a common gene set, filtered to genes detected in a minimum number of spots, and reduced to the domain-wide highly variable genes (`seurat_v3`, up to 3,000, fewer where the shared set is smaller). Counts are then total-count normalized to 10^4 per spot and log-transformed, which is the expression $\mathbf{X}$ of Sec. 3.1. Moran's I is computed per gene on the pooled spots of the domain, the spatial scale group is its tertile, and the 1000-gene panel is a stratified sample of about 333 genes per tertile from the highly variable set.

Slides. The HEST-1K sample identifiers are as follows. C1: MISC2 to MISC6 and MISC8 to MISC12. C2: INT13 to INT15 and INT17 to INT24. C3: NCBI826 to NCBI833. C4: MISC34 to MISC37, MISC39 to MISC41, MISC45 to MISC50, MISC62, MISC66, MISC69, MISC70, MISC72 and MISC73. C5: MISC118 to MISC127, MISC131 and MISC133 to MISC136. C6: NCBI834 to NCBI844. C7: NCBI618 to NCBI627. C8: NCBI715 to NCBI720.

Cell-type signatures. $\mathbf{B}$ is the mean log-normalized expression of each annotated cell type in a single-cell reference of the same tissue, restricted to the panel, with the number of types per domain given in Table 10.

Scale measurement of Figure 1(b). Expression is rasterized onto a 96×96 grid over the section and decomposed with a two-dimensional undecimated (stationary) wavelet transform (`db2`, four levels). A field's local scale at a grid cell is the energy-weighted mean of the level scales, each level's scale being 2^{ℓ} cells as a fraction of the grid, with the detail energies of that cell as weights. The bottom panel applies this to each gene's own expression map and reports the mean over the tissue, for 170 genes drawn from each Moran's I tertile, which is the spatial scale group of Sec. 3.1. The vertical axis is therefore measured independently of the variable that defines the groups, and the coarse and fine group medians differ by a factor of 1.9. The top panel applies it to the dominant expression program of the section, the leading component of a non-negative matrix factorization of the panel, and draws the local scale after Gaussian smoothing within the tissue mask ($\sigma = 2.5$ cells), read back at each spot.

B GENE EMBEDDING: CONSTRUCTION AND CONTROL

Construction. For a tissue domain whose slides share a panel of G genes, let $\mathbf{X} \in \mathbb{R}^{N \times G}$ stack the log-normalized expression of every spot of every slide. We center each gene, take a truncated SVD $\mathbf{X} - \bar{\mathbf{X}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$ retaining $d = 64$ components, and set the co-expression code of gene g to the g-th row of $\mathbf{V}\mathbf{\Sigma}$, normalized to unit ℓ_2 norm. Concatenating the gene's standardized Moran's I and a one-hot of its spatial scale group, itself a tertile of that Moran's I, gives $\mathbf{e}_g \in \mathbb{R}^{68}$. The result is a fixed buffer: no gradient reaches it and it is identical for every spot and every training step.

What the code carries. The code has no spot index. It cannot express where in the tissue a gene is expressed, only how the gene covaries with the rest of the panel, so every spatial structure in a prediction is produced by the shared field, which is fit to the observed genes. What it does

carry is the held-out gene's loading on the panel's dominant co-expression modes, computed from a decomposition in which that gene's own column participates. This is the sense in which the setting is transductive along the gene axis (Sec. 3.1): the panel is known and a subset of it is withheld from supervision. Any co-expression code requires the gene to have been measured somewhere, co-expression being a relation between it and the rest of the panel, and the reference-based methods meet that requirement in an external atlas and less compressively, taking a masked gene's expression from matched reference cells where our code retains 64 numbers. The setting this addresses is a panel profiled on some sections and extended to others, not a gene absent from every assay.

The control this raises. The construction leaves open whether accuracy depends on the code being computed from the evaluation slides themselves, or only on the co-expression structure the code summarizes, which any measurement of the gene would recover. The two are separable without changing the model, and without holding the field out of the scored slide: the code is rebuilt from the domain's other slides, the field is still trained on the observed genes of every slide including the scored one, and the held-out genes of the scored slide are scored as before. The gene's expression on the scored slide then enters neither the training terms nor the code, which is the setting of a panel profiled on some sections and extended to others. We state the control rather than report it. The claim we make is accordingly the narrower one, that the side information is a 64-dimensional summary of co-expression rather than the gene's expression profile, and we do not claim that the set of slides over which that summary is computed is immaterial.

C Additional held-out gene results

Table 3 slices the held-out gene result of Table 1 by the Moran's I of the measured expression, Figure 5 gives the top-k curves and the fraction of genes above 0.3 that Sec. 4.3 reports, and Table 4 collects the paired tests behind every comparison in the main text.

stImpute coverage. stImpute's protein-language-model graph resolved 949 to 997 of the 1000 panel genes per domain, so it ran on its intended input.

Paired tests. Every row of Table 4 compares two methods on the same eight domains with a Wilcoxon signed-rank test, the domain being the unit of analysis because domains are separate datasets while the two methods are paired within a domain on the same slides and folds. At $n = 8$ the smallest attainable two-sided p-value is 0.008, so a row reading 0.008 means every domain favours our method and the test has no resolution left to distinguish such rows from one another. The first row is the primary comparison, the second block repeats it against the two most recent baselines, and the remaining held-out gene blocks slice that same result along different axes, so these are not independent hypotheses and no correction for multiplicity is applied. The two ablations that do not reach significance are included because they carry the claim of Sec. 4.5 that the spectral coupling and the routing are within run-to-run noise, which reruns of the same configuration on six domains put at 0.0004 on average and 0.0009 at most. On that scale the single-scale effect (0.028) is about seventy times the mean run-to-run difference and thirty times the largest. The final block repeats the two spot-level comparisons with the slide as the unit ($n = 90$), which addresses the objection that the spread across slides within a domain (0.051 on average) is larger than the difference being tested. Slides within a domain are not independent of one another, so that block supports the domain-level rows rather than replacing them.

D Held-out spots and reconstruction

Tables 5 and 6 give the two protocols of Sec. 4.4. Both withhold something other than whole genes, so neither is comparable to the primary protocol or to the other.

The second protocol is STINR's own, which hides a fraction of the spots in each slide and scores the hidden ones. We run STINR itself at the three ratios rather than only borrowing its evaluation, reading the masking seed from each of our configurations so that both methods are scored on identical spots (Table 5). We lead on average at every ratio, by 0.061, 0.060 and 0.058, and in seven of eight domains at each. Excluding C7, where STINR degrades most, the margin is still $+0.045$. Accuracy falls as the masking ratio rises (0.2606, 0.2532, 0.2391), the field remaining constrained by

Table 2: Ablation of our method under held-out gene-CV, mean Pearson correlation over held-out genes. The single-scale variant keeps every layer width and sets $K = 1$ with the coupling off, so it is the amortized counterpart of a single-scale field. Tying the routing across scales also removes the cross-scale orthogonality penalty, which is identically zero under that setting (Sec. 3.5), so that row bounds the pair rather than isolating the tying. Paired tests are in Table 4, and clusters are as in Table 1.

Variant	C1	C2	C3	C4	C5	C6	C7	C8	Avg
Full (multi-scale)	0.2384	0.2598	0.2655	0.2815	0.2839	0.2758	0.2607	0.5324	0.2997
Shared routing	0.2392	0.2591	0.2651	0.2796	0.2822	0.2761	0.2603	0.5336	0.2994
No coupling	0.2361	0.2599	0.2669	0.2788	0.2830	0.2728	0.2600	0.5325	0.2987
Single scale	0.2218	0.2431	0.2456	0.2503	0.2634	0.1949	0.2283	0.5235	0.2713

Table 3: Held-out gene imputation sliced by how spatially structured the gene is. Held-out genes are binned by the Moran's I of their *measured* expression, where Q1 is the least and Q4 the most spatially autocorrelated quartile, and top-50 is the fifty genes with the highest Moran's I. Entries are mean Pearson correlation within the bin, under the same protocol and the same runs as Table 1, whose overall values these rows average to. The Q4/Q1 ratios quoted in Sec. 4.3 are ratios of the domain means. Best per cell pair in bold, and clusters as in Table 1.

	Q1 (least spatial)		Q4 (most spatial)		top-50 by Moran's I	
Domain	Ours	SpaGE	Ours	SpaGE	Ours	SpaGE
C1	**0.1618**	0.1427	**0.3402**	0.3318	**0.4348**	0.4069
C2	**0.1880**	0.1413	**0.3670**	0.2442	**0.3862**	0.2367
C3	**0.1731**	0.1277	**0.4168**	0.2848	**0.5378**	0.3714
C4	**0.2119**	0.1867	**0.3752**	0.3149	**0.3884**	0.3095
C5	**0.1970**	0.1799	**0.4029**	0.3519	**0.5376**	0.4385
C6	**0.1515**	0.1279	**0.4794**	0.3915	**0.5729**	0.4193
C7	**0.1373**	0.1206	**0.4617**	0.4435	**0.4866**	0.4604
C8	0.3115	**0.3131**	0.7968	**0.8081**	0.8278	**0.8431**
Mean	**0.1915**	0.1675	**0.4550**	0.3963	**0.5215**	0.4357

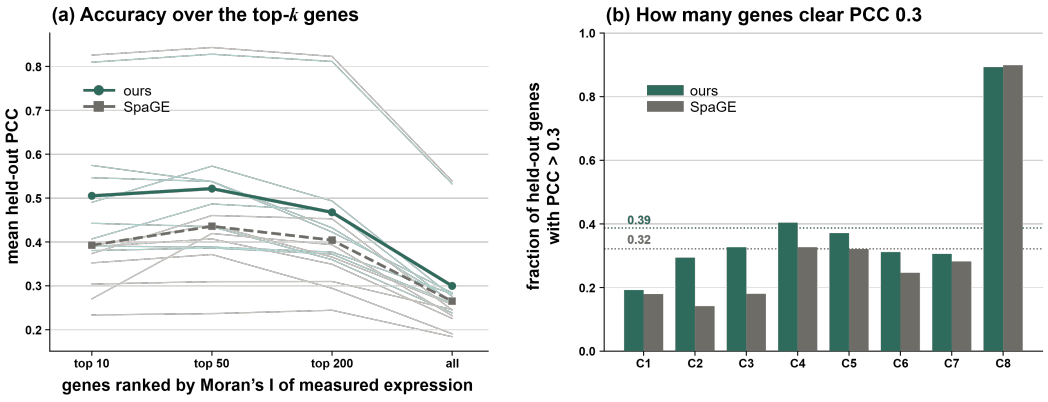

Figure 5: **(a)** Mean held-out correlation over the k genes with the highest Moran's I, for $k = 10$, 50, 200 and the whole panel. Thin lines are the eight domains, for both methods. **(b)** The fraction of held-out genes reaching a correlation above 0.3, per domain, with dotted lines at the two means.

the spots that were not hidden. SUICA accepts no external spot mask and so cannot be placed on the same held-out spots by construction. We therefore report it twice rather than choosing between two figures that order it differently. It reaches 0.204 on the full slide, a setting in which 80% of the spots were observed during its own training, and 0.022 against our 0.261 on the spots both procedures withheld, a difference of 0.239 in all eight domains ($p = 0.008$). The distance between those two rows locates SUICA's limit in the axis it amortizes over, since its graph autoencoder forms a spot's

Table 4: Paired tests over the eight tissue domains: mean difference, number of domains favouring the first method, and Wilcoxon signed-rank p-value with the domain as the unit of analysis ($n = 8$, so the smallest attainable two-sided p-value is 0.008). The final block repeats the two spot-level comparisons with the slide as the unit ($n = 90$). The paragraph above states how the rows relate.

Comparison	Δ	Domains	p
Primary			
Held-out gene imputation vs. SpaGE	$+0.035$	7/8	0.016
The same protocol against the two most recent baselines			
vs. SpaIM	$+0.073$	8/8	0.008
vs. stImpute	$+0.096$	8/8	0.008
The same comparison, by how spatially structured the gene is			
Q1, least autocorrelated quartile	$+0.024$	7/8	0.016
Q2	$+0.025$	7/8	0.016
Q3	$+0.031$	7/8	0.016
Q4, most autocorrelated quartile	$+0.059$	7/8	0.023
The same comparison, restricted to the top-k genes			
top-10	$+0.113$	7/8	0.016
top-50	$+0.086$	7/8	0.016
top-200	$+0.063$	7/8	0.023
fraction of genes above 0.3	$+0.065$	7/8	0.016
Other protocols			
Reconstruction vs. STINR	$+0.080$	8/8	0.008
Held-out spot vs. STINR (sr06)	$+0.061$	7/8	0.016
Held-out spot vs. SUICA, spots held out by both	$+0.239$	8/8	0.008
Ablations, against the full model			
Single scale	-0.028	8/8	0.008
No spectral coupling	-0.001	5/8	0.195
Routing tied across scales	-0.000	5/8	0.445
Supporting, with the slide as the unit ($n = 90$)			
Held-out spot vs. STINR (sr06)	$+0.061$	83/90	1×10^{-15}
Held-out spot vs. SUICA, spots held out by both	$+0.230$	90/90	2×10^{-16}

embedding from that spot's own expression and a spot it never observed supplies no such input. The LOSO row holds out an entire slide and yields a much lower average (0.046). It lies on a different generalization axis from the one this paper addresses. The primary task is a held-out gene on an observed slide, and the spot protocol is a held-out spot on an observed slide. A held-out gene on an unseen slide would additionally require the field to transfer across slides, which the shared field as trained here does not do (App. G). We report the row so that this boundary is explicit rather than implied.

Scored instead on the genes observed at training, our method leads STINR, the single-scale non-amortized field this work builds on, on all eight domains and by 0.080 on average ($p = 0.008$), so what Sec. 3.3 argues is necessary is not paid for on the observed genes. A reconstruction-specialized method, STAGATE, leads that table on average by denoising each gene from its own measurements over a spatial neighbour graph, which is exactly the information a held-out gene does not have. Table 6 gives it. On the spots it did not observe, SUICA's decoder returns a spatially constant column for 182 of the 1000 genes on average, which is the mechanism behind its lower-block values in Table 5.

E Enrichment

E.1 Programs recovered by the spectral coupling

The coupling is the one place where the model's gene axis is explicit rather than implicit, since its modes are a factorization over genes. Two analyses characterize what they capture.

Table 5: Held-out spot imputation with within-slide masking. sr06, sr07 and sr08 hide 60%, 70% and 80% of the spots in each slide. STINR and our method are scored on identical held-out spots, the masking seed being read from each of our configurations. SUICA accepts no external spot mask and so appears twice. The upper block reports each method on the spots its own protocol held out, which for SUICA is a full-slide reconstruction in which 80% of the spots were observed during its training, and the lower block restricts all three to the spots both procedures withheld, 12.1% of each slide and 189 to 598 spots per slide. Best per column in bold within each block, excluding the full-slide row, which is not a held-out setting. The LOSO row is cross-slide transfer, a different protocol comparable to neither block. Clusters as in Table 1.

Method	C1	C2	C3	C4	C5	C6	C7	C8	Avg
Each method on the spots its own protocol held out									
SUICA (full slide)	0.1156	0.2036	0.1606	0.2341	0.1753	0.1758	0.1682	0.3963	0.2037
STINR (sr06)	**0.1826**	0.2101	0.1514	0.2264	0.1946	0.1215	0.0519	0.4562	0.1993
STINR (sr07)	0.1804	0.2015	0.1612	0.2101	0.1880	0.1163	0.0427	0.4424	0.1928
STINR (sr08)	0.1633	0.1981	0.1397	0.2074	0.1764	0.1021	0.0470	0.4157	0.1812
Ours (sr06)	0.1803	**0.2440**	**0.2321**	0.2743	**0.2523**	**0.1916**	**0.2275**	**0.4824**	**0.2606**
Ours (sr07)	0.1755	0.2389	0.2177	**0.2761**	0.2445	0.1734	0.2240	0.4759	0.2532
Ours (sr08)	0.1632	0.2298	0.2038	0.2708	0.2317	0.1465	0.2100	0.4568	0.2391
All three methods on the spots held out by both procedures (sr06)									
SUICA	0.0139	0.0252	0.0147	0.0289	0.0243	0.0089	0.0207	0.0401	0.0221
STINR	**0.1804**	0.2117	0.1528	0.2271	0.1943	0.1194	0.0508	0.4577	0.1993
Ours	0.1792	**0.2458**	**0.2277**	**0.2765**	**0.2524**	**0.1951**	**0.2307**	**0.4813**	**0.2611**
Ours (LOSO)	0.0602	0.0090	0.0010	0.0009	0.0041	0.0214	0.0311	0.2435	0.0464

Table 6: Reconstruction (denoising): mean Pearson correlation over all genes, which are observed at training and not held out. STAGATE denoises each gene from its own measurements over a spatial neighbour graph, the information a held-out gene does not have, so it is expected to lead here (Sec. 4.4). Best per column in bold, and clusters as in Table 1.

Method	C1	C2	C3	C4	C5	C6	C7	C8	Avg
STAGATE	**0.2731**	**0.3186**	**0.3098**	**0.3560**	**0.3392**	**0.3090**	**0.3047**	0.5425	**0.3441**
STINR	0.2178	0.2261	0.1848	0.2386	0.2245	0.1438	0.0577	0.5020	0.2244
Ours	0.2445	0.2635	0.2677	0.2887	0.2876	0.2760	0.2654	**0.5433**	0.3046

The first is which tissues recover the same biology. Every domain is fit independently, one model per domain, so a program that appears in two of them is not a shared parameter. Folding GO and Reactome names of the same program together and setting aside the ubiquitous translation and ribosomal terms, half of the twenty-four pairs of tissues drawn from different organs share no program at all, and the mean over those pairs is one. The pairs that do share are the ones with biology in common. The two brains return the same synaptic-release and substantia-nigra programs across species. The human and the mouse hepatoblastoma liver share three, among them platelet degranulation and biological oxidations. The two carcinomas, in kidney and in bowel, share six, among them extracellular matrix organization, metallothionein binding and neutrophil degranulation. Extracellular matrix organization appears in four of the eight tissues rather than in the carcinomas alone. Within a tissue the programs are appropriate to it, with neurotransmitter release in DLPFC, platelet and calcium response in human liver, muscle contraction in heart and glutathione conjugation in hepatoblastoma. All sixteen modes reach significance in every tissue, and Table 8 lists the most significant tissue-specific program of each.

The second is quantitative. The modes are produced from the gene embedding by the learned amortizer, so they are a deterministic function of it and independence from it is not available by construction. What is available is whether the learned transform tracks co-expression more closely than the code it was given. Partialling out both the gene embedding and expression strength linearly, the correlation between the coupling and measured co-expression remains positive in all eight domains (0.12 to 0.51, Table 9), while the embedding on its own agrees with measured co-expression at 0.24 to 0.30. Holding the learned embedding fixed and redrawing the amortizer trunk and the coupling

heads twenty times leaves a residual whose per-domain mean is positive in four of the eight, and the trained value exceeds it in every domain ($p = 0.008$). The agreement across domains is what carries this rather than separation within any one, since individual redraws reach ± 0.47 and about one in seven exceeds its own domain's trained value.

The sixteen modes of a tissue are not sixteen independent programs, and we do not read them as such. Nothing orthogonalizes them within a scale (Sec. 3.5), and a program that dominates a panel loads on many of them, as muscle contraction does on nine of the sixteen heart modes. Ubiquitous translation and ribosomal programs appear everywhere and we treat them as a shared baseline signal rather than as tissue-specific findings. Table 8 lists the top tissue-specific program of each domain and Table 7 the programs shared between domains.

E.2 Procedure and tables

Procedure. For each tissue and each of the 16 coupling modes we take the genes at each end of the mode, defined by the sign of their loading on $\mathbf{Q}$, and test them against GO Biological Process 2021 (Ashburner et al., 2000) and Reactome 2022 (Milacic et al., 2024) with Enrichr (Chen et al., 2013; Kuleshov et al., 2016), using the domain's own 1000-gene panel as the background. Terms are reported at Benjamini–Hochberg $p_{\text{adj}} < 0.05$, which yields 248 hits over the eight domains, 30 to 32 per domain, with every one of the 16 modes reaching significance in every tissue.

Which tissues share programs. Table 7 gives the overlap for every pair of tissues that shares more than one program, after folding GO and Reactome names of the same program together and setting aside the ubiquitous translation and ribosomal terms together with five Reactome root labels that are categories rather than programs. Twelve of the twenty-eight pairs share nothing and six share one. The pairs at the top of the table are the ones with biology in common: the two carcinomas, the three liver pairs and the two brains. Each domain appears in seven pairs, so the twenty-eight are not independent and we report the distribution rather than a test over it.

Table 7: Programs shared between tissue domains, for the ten pairs of the twenty-eight that share more than one. Same-organ pairs are marked. The remaining eighteen pairs share one program (six pairs) or none (twelve). Clusters as in Table 1.

Pair	Same organ	n	Programs
C2–C4		6	extracellular matrix organization; extracellular structure organization; cytokine-mediated signalling; metallothioneins bind metals; negative regulation of programmed cell death; neutrophil degranulation
C3–C7	yes	3	biological oxidations; peptidyl-cysteine S-nitrosylation; platelet degranulation
C4–C5		3	extracellular matrix organization; metallothioneins bind metals; smooth muscle contraction
C1–C8	yes	2	neurotransmitter release cycle; substantia nigra development
C3–C6	yes	2	binding and uptake of ligands by scavenger receptors; regulation of TLR by endogenous ligand
C6–C7	yes	2	metabolism of amino acids and derivatives; neutrophil mediated immunity
C1–C5		2	aerobic electron transport chain; axon guidance
C2–C5		2	extracellular matrix organization; metallothioneins bind metals
C2–C6		2	collagen formation; neutrophil degranulation
C3–C5		2	binding and uptake of ligands by scavenger receptors; extracellular matrix organization

Table 8: Biological programs recovered by the spectral coupling. For each tissue, the most significant *tissue-specific* Enrichr program (GO BP / Reactome) among the 16 coupling modes, with the mode index, the end of the mode its genes load on, and the Benjamini–Hochberg adjusted p-value. All 16 modes reach significance ($p_{\text{adj}} < 0.05$) in every tissue. The sixteen are not sixteen independent programs, since nothing orthogonalizes modes within a scale (Sec. 3.5), and the direction column separates the two ends of a mode and no more, the coupling being unchanged when a column of $\mathbf{Q}^{(k)}$ flips sign. Ubiquitous translation and ribosomal programs are omitted as a shared baseline signal, and the recurrence across tissues is reported in App. E.1. Program names are as they appear in the source databases, with accession numbers stripped. Clusters as in Table 1.

Tissue	Top tissue-specific program	Mode	End	p_{adj}
Brain, DLPFC (human)	Neurotransmitter Release Cycle	15	neg	$6e{-7}$
Kidney, ccRCC (human)	Extracellular Matrix Organization	12	pos	$4e{-14}$
Liver (human)	Response To Elevated Platelet Cytosolic Ca2+	9	pos	$9e{-18}$
Bowel, CRC (human)	Extracellular Structure Organization	9	pos	$4e{-19}$
Heart (human)	Muscle Contraction	0	pos	$7e{-25}$
Mouse liver (regen.)	Binding And Uptake Of Ligands By Scavenger Receptors	4	pos	$1e{-13}$
Mouse liver (HB)	Glutathione Conjugation	14	pos	$3e{-6}$
Mouse brain (aging)	Neurotransmitter Release Cycle	14	pos	$4e{-15}$

Table 9: What the spectral coupling tracks, by tissue domain. The first column is the partial correlation between the coupling and measured co-expression after linearly removing the gene embedding and expression strength. It is positive in all eight domains, against 0.2372 to 0.3010 for the agreement between the embedding alone and measured co-expression, which is the part the partial removes. The remaining columns repeat that quantity for twenty redraws of the amortizer trunk and the coupling heads with the learned gene embedding held fixed, each layer drawn at its learned standard deviation. Their per-domain means are positive in four of the eight and the trained value exceeds them in all eight (paired Wilcoxon $p = 0.008$, and the default initialization gives the same test with means positive in six of eight). The last column bounds the claim, since individual redraws are not small and in the weakest domains a fifth to a third of them exceed the trained value: the evidence is the agreement across domains rather than separation within any one. Clusters as in Table 1.

		Untrained, 20 redraws		
Domain	Coupling, partial	Mean	SD	Above trained
C1	0.1186	$+0.019$	0.197	5/20
C2	0.3344	-0.026	0.202	0/20
C3	0.2422	$+0.138$	0.220	6/20
C4	0.5118	-0.011	0.208	0/20
C5	0.3669	$+0.007$	0.176	0/20
C6	0.3731	$+0.046$	0.189	1/20
C7	0.1719	-0.025	0.279	6/20
C8	0.2079	-0.034	0.204	4/20
Mean	0.2908	$+0.014$	0.209	22/160

F CELL-TYPE COMPOSITION

F.1 COMPOSITION ACROSS DOMAINS

The deconvolution path produces a per-spot cell-type composition as a by-product. In five of the eight domains it spreads over several cell types and is spatially organised, with tumour and immune compartments occupying mutually exclusive regions in the kidney and pacemaker, fibroblast and neural compartments laid out anatomically in the heart. It concentrates on a single type where one compartment dominates the tissue (C7 and C4, entropy ratios 0.14 and 0.26), and in C6 it stays diffuse while losing spatial coherence. Coarse compartments are recovered and fine gradients are

not, and there is no per-spot ground truth for these slides. Table 10 summarizes all eight domains and Figure 6 maps four.

Figure 6 shows the composition on one slide from each of four domains and Table 10 summarizes all eight, including the three where it fails. Where it works the structure is interpretable, and the failures are of two kinds: concentration on a single type where one compartment dominates the tissue (C7 and C4), and a diffuse but spatially incoherent composition in C6. Coarse compartments are recovered and fine gradients are not, and cortical layers in DLPFC do not emerge.

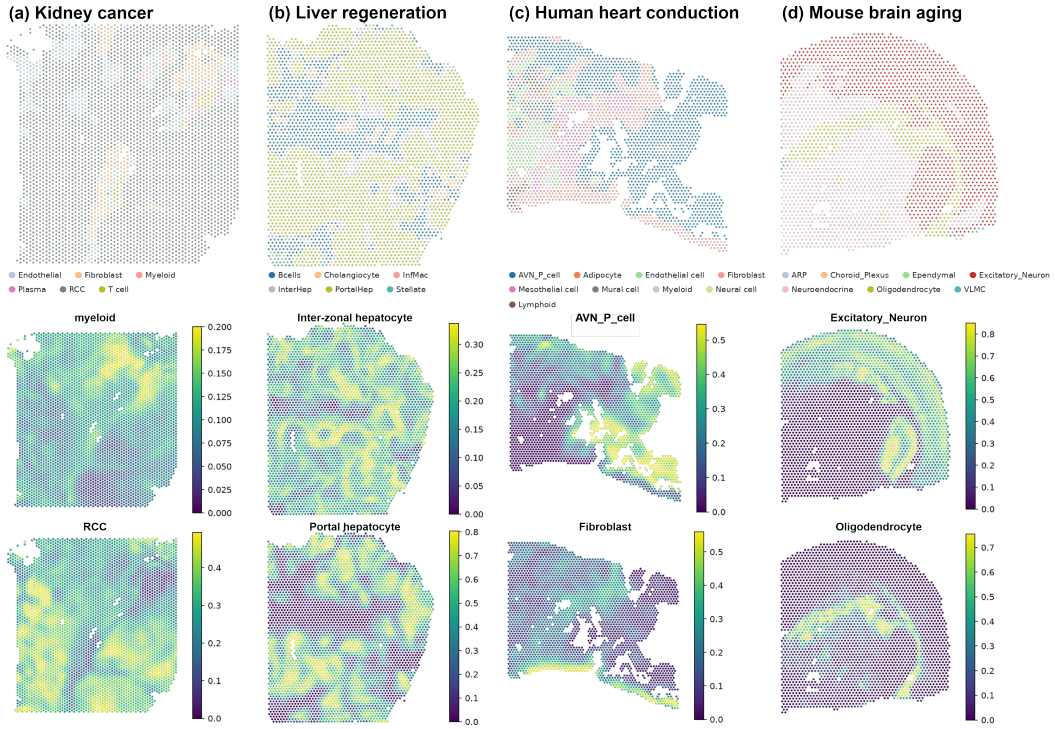

Figure 6: Cell-type composition on one slide from each of four domains: (a) kidney ccRCC (C2), (b) human liver (C3), (c) heart (C5), (d) mouse brain aging (C8). Top row: the dominant cell type at each spot, with the legend lists only the types that win somewhere on that slide. Lower rows: proportion maps of the two types that carry the spatial structure in that domain (carcinoma and myeloid in the kidney, portal and inter-zonal hepatocytes in the liver, pacemaker and fibroblast in the heart, excitatory neurons and oligodendrocytes in the brain), which show the separation that an argmax hides. Colour scales are per panel and the cell-type sets differ between domains. There is no per-spot ground truth, so no accuracy is claimed.

G Further limitations

Sec. 5 states the two limitations that bear on the central claim. This appendix gives the remainder with their measurements.

Cross-slide transfer. The shared field does not transfer across slides. Holding out an entire slide and scoring its genes yields an average of 0.046 (Table 5), against 0.261 when the withheld spots are drawn from within each slide. The two hold-out axes are distinct, and Table 11 places each protocol of this paper on them.

The coordinate field is fit to the slides it is trained on and the embedding is built from those slides' co-expression, so nothing in the current design aligns one section's coordinates or batch to another's. Addressing this means adding what is missing rather than tuning: a slide-level alignment or registration step so that coordinates are comparable across sections, and a term separating slide-specific variation from the shared field. Whether the shared field can then transfer is an open question.

Table 10: Cell-type composition recovered by the deconvolution path, one representative slide per domain. Median max-prop is the largest proportion at a spot, median over spots, with the value expected under a uniform composition in brackets, and the entropy ratio is the entropy of the composition divided by that of a uniform one, so 1 is flat and 0 is a single cell type. The last column is Moran's I of the six most abundant types, averaged. There is no per-spot ground-truth composition, so no accuracy is claimed.

	Tissue	Types	Dominant type	Mean	Max-prop	Entropy	Moran's I
C1	Brain, DLPFC	33	Ex_2_L5	0.193	0.241 (0.030)	0.687	0.94
C2	Kidney, ccRCC	12	RCC	0.313	0.328 (0.083)	0.743	0.94
C3	Liver	12	PortalHep	0.380	0.426 (0.083)	0.653	0.88
C4	Bowel, CRC	19	Epi	0.773	0.827 (0.053)	0.262	0.97
C5	Heart	15	AVN pacemaker	0.224	0.307 (0.067)	0.704	0.95
C6	Mouse liver, regen.	11	MPs	0.273	0.408 (0.091)	0.701	0.58
C7	Mouse liver, HB	11	Hepatocyte	0.870	0.912 (0.091)	0.141	0.92
C8	Mouse brain, aging	20	Neuroendocrine	0.398	0.628 (0.050)	0.386	0.88

Table 11: Generalization axes. Rows hold out along the slide axis and columns along the gene axis, and each cell names the protocol and where it is reported. This paper addresses the upper-right cell.

	Observed gene	Held-out gene
Observed slide	reconstruction (Table 6) and held-out spots (Table 5)	**this work** (Table 1)
Unseen slide	cross-slide transfer (LOSO, Table 5)	not addressed

Smoothness of the predictions. The predictions are smoother than the measurements by a wide margin, the Moran's I of a predicted gene averaging 0.88 against 0.13 for the same gene measured. A continuous field does not reproduce counting noise and Pearson correlation is invariant to that, so this does not affect the reported accuracy and is arguably what a denoising imputation should do. It has one consequence that was not intended. Because the predictions are uniformly smooth they differentiate less between genes that are and are not spatially structured: the per-gene agreement between predicted and measured Moran's I is 0.24 for our method against 0.37 for SpaGE, the one axis on which the baseline leads, though the difference does not reach significance ($p = 0.31$). Our accuracy is nonetheless strongly gene-dependent, so what is similar between genes is the noise floor of the prediction rather than its signal. A downstream analysis that reads spatial statistics off imputed values rather than the values themselves should account for this.

Cell-type composition. The compositions are evaluated qualitatively, since no per-spot ground truth exists for these slides, and they are not uniform in quality (App. F.1, App. F). The analysis therefore supports a correspondingly limited claim. A quantitative claim needs ground truth the assay does not provide, and the closest available test would be a domain with paired annotations, such as expert layer labels in DLPFC, scored as agreement with a known regional structure rather than as per-spot accuracy.

Interpretability analysis. The co-expression analysis is internal to our model and is not a comparison, since the reference-based baselines expose no analogous gene-gene modes to enrich, so we claim only that this module is interpretable, not that the method is more interpretable than theirs. The claim is narrow in a second way. The modes are a function of the gene embedding, which is itself built from the panel's co-expression, so the analysis cannot establish independence from co-expression and does not try to (App. E.1). We do control for the possibility that any map of this shape would align with co-expression equally well. Holding the learned gene embedding fixed and redrawing the amortizer trunk and the coupling heads, twenty draws per tissue leave a residual whose sign is not consistent across tissues: the per-tissue mean is positive in four of eight when each layer is drawn at its learned scale and in six of eight under the default initialization, spanning -0.04 to $+0.14$, against $+0.291$ for the trained coupling, which is positive in all eight (paired Wilcoxon $p = 0.008$ under either initialization). That control establishes consistency across tissues rather than separation within one. Individual draws reach ± 0.47, about one in seven exceeds its own tissue's trained value, and in the weakest tissue the trained value sits half a standard deviation above the

untrained mean under the matched initialization and a quarter of one under the default. A particular random map can therefore align with co-expression by chance, and the claim rests on the agreement across eight tissues rather than on any one comparison. The direction in which a gene loads on a mode is not interpretable either, since the coupling is unchanged when a column of the mode matrix flips sign.

Seed and fold variance. Each domain is trained once on one gene split, so none of the reported numbers carries a seed or fold variance.