

# The covariance environment defines cellular niches for spatial inference

Received: 18 April 2023

Accepted: 28 February 2024

Published online: 02 April 2024

 Check for updates

Doron Haviv<sup>1,2</sup>, Ján Remšík<sup>3</sup>, Mohamed Gatie<sup>4</sup>, Catherine Snopkowski<sup>1</sup>, Meril Takizawa<sup>1</sup>, Nathan Pereira<sup>5</sup>, John Bashkin<sup>5</sup>, Stevan Jovanovich<sup>6</sup>, Tal Nawy<sup>1</sup>, Ronan Chaligne<sup>1</sup>, Adrienne Boire<sup>3,6,7</sup>, Anna-Katerina Hadjantonakis<sup>4</sup> & Dana Pe'er<sup>1,8</sup> ✉

A key challenge of analyzing data from high-resolution spatial profiling technologies is to suitably represent the features of cellular neighborhoods or niches. Here we introduce the covariance environment (COVET), a representation that leverages the gene–gene covariate structure across cells in the niche to capture the multivariate nature of cellular interactions within it. We define a principled optimal transport-based distance metric between COVET niches that scales to millions of cells. Using COVET to encode spatial context, we developed environmental variational inference (ENVI), a conditional variational autoencoder that jointly embeds spatial and single-cell RNA sequencing data into a latent space. ENVI includes two decoders: one to impute gene expression across the spatial modality and a second to project spatial information onto single-cell data. ENVI can confer spatial context to genomics data from single dissociated cells and outperforms alternatives for imputing gene expression on diverse spatial datasets.

Intense interest in cellular interactions and tissue context has spurred the growth of multiplexed spatial transcriptomics and antibody-based technologies, sparking the need for computational approaches to identify biological patterns within tissues<sup>1–5</sup>. The local neighborhood, or niche, of a cell is a useful resolution for defining cell interactions; it may represent functional anatomical subunits (such as stem cell niches) and is a basis for identifying larger spatial patterns. However, efficient representations of the cellular microenvironment that retain the full richness of the data and can be used to effectively compare niches are lacking<sup>6</sup>. At the same time, there is a need to address the limited molecular plexity of high-resolution spatial profiling technologies<sup>7</sup>.

Most methods for analyzing spatial data characterize each niche by tabulating discrete cell types within a given region<sup>8–11</sup>. Although these have generated important discoveries<sup>2,8</sup>, they were developed for

low-plex antibody-based imaging methods that devote most markers to cell typing. Spatial transcriptomics methods, including commercial platforms, can now profile hundreds of genes<sup>12–18</sup>, meaning that analysis at the cell type level leads to substantial information loss. In single-cell genomics, the switch from discrete cell typing to continuous approaches, such as diffusion maps<sup>19</sup> and pseudotime<sup>20,21</sup>, has driven remarkable discovery. Moreover, setting thresholds for continuous cellular phenotypes is subjective and invokes problems of instability and bias. Even within highly discrete cell types, vast and meaningful variation often exists, such as the spectrum of activated and metabolic states within immune cell types<sup>22–24</sup>.

Thus, a niche representation is needed that considers the full measured expression and its continuous nature and that enables robust, efficient comparisons. We propose a representation that goes

<sup>1</sup>Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>2</sup>Tri-Institutional Training Program in Computational Biology and Medicine, Weill Cornell Medicine, New York, NY, USA. <sup>3</sup>Human Oncology & Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>4</sup>Developmental Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>5</sup>S2 Genomics, Livermore, CA, USA. <sup>6</sup>Department of Neurology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>7</sup>Brain Tumor Center, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>8</sup>Howard Hughes Medical Institute, New York, NY, USA.

✉ e-mail: [peerd@mskcc.org](mailto:peerd@mskcc.org)

beyond cell typing and preserves complex patterns of gene expression, including covariation in genes across cell states. Specifically, we developed the covariance environment (COVET), a compact representation of a cell's niche that assumes that interactions between the cell and its environment create biologically meaningful covariate structure in gene expression between cells of the niche. We developed a corresponding distance metric that unlocks the ability to compare and analyze niches using the full toolkit of approaches currently employed for cellular phenotypes, including dimensionality reduction, spatial gradient analysis and clustering.

Imaging-based spatial transcriptomics technologies face issues that practically limit quantification to hundreds of genes. Some methods can impute spatial information for genes not measured in the spatial modality, by integrating matched single-cell RNA sequencing (scRNA-seq) data<sup>9,25,26</sup>. However, integration methods do not explicitly model cellular microenvironment context from the spatial data, thereby limiting inference power.

To achieve transcriptome-wide spatial inference, we developed environmental variational inference (ENVI), a conditional variational autoencoder (CVAE)<sup>27,28</sup>, that simultaneously incorporates scRNA-seq and spatial data into a single embedding. ENVI leverages the covariate structure of COVET as a representation of cell microenvironment and achieves total integration by encoding both genome-wide expression and spatial context (the ability to reconstruct COVET matrices) into its latent embedding. Our approach is effective on data from a variety of multiplexed spatial technologies and outperforms other methods in accurately imputing the expression of genes in diverse developmental contexts. ENVI can also be used to project valuable spatial information onto dissociated scRNA-seq data and can capture continuous variation along spatial axes across large complex tissue regions.

## Results

### COVET defines spatial neighborhoods

To move beyond cell type fraction and to characterize niches in a manner that leverages measured genes and enables quantitative comparison, we developed the COVET framework. Our core assumption is that a cell affects—and is also affected by—cells in its vicinity, generating covarying patterns of expression among the interacting cells. Our framework includes three components: (1) COVET, a robust per-cell representation of neighborhood information based on a modified formulation of gene–gene covariance among niche cells; (2) a distance metric that is essential for comparing and interpreting niches; and (3) an algorithm to efficiently compute this distance metric. Unlike mean expression, gene–gene covariance captures the relationships among genes and cell states that are shaped by cellular interplay within the niche. These relationships are rich, stable and enriched for biological signal; moreover, they contain substantial hidden information from unmeasured genes, providing an advantage for imputation tasks.

To calculate COVET, we first define the niche of each cell in a dataset by the  $k$  spatial nearest neighbors of that cell and then compute each niche's gene–gene shifted covariance matrix (Fig. 1a and Methods). Shifted covariance modifies the classic covariance formulation by using mean expression across the entire dataset rather than local mean expression as a reference. This constructs each cell's covariance matrix relative to the entire population and critically enables direct comparison between niches, highlighting their shared and unique features. Gene–gene covariance provides the additional benefit of being more robust to technical artifacts<sup>23</sup>, facilitating integration across technologies.

Despite being a compact and powerful representation of the niche, COVET requires a metric for comparison. Niche similarity cannot be determined by simply subtracting the cell-by-gene expression of matrices of two niches, because the result depends on cell order, which is set arbitrarily (it will change if an image is rotated, for example). We, thus, seek to quantify niche similarity in a permutation-invariant manner,

for which the Fréchet distance provides a closed-form solution<sup>29</sup>. However, calculating Fréchet distance is computationally intractable, so we developed an approximation (approximate optimal transport (AOT)) that reduces runtime by over an order of magnitude and is substantially faster than another common metric, the Bhattacharyya distance<sup>30</sup> (Extended Data Fig. 1a). AOT yields similar results to true optimal transport, and a GPU implementation takes under 1 min to compute the cell–cell AOT distance matrix of 100,000 cells (Extended Data Fig. 1b–d).

As AOT can be computed via Euclidean distance, which underlies many standard single-cell analyses, such as clustering<sup>31</sup>, diffusion components<sup>32</sup> and uniform manifold approximation and projection (UMAP)<sup>33</sup>, niches can now be analyzed with the same algorithms designed to analyze phenotypes. Clustering niches can characterize canonical environments; visualization can be used to observe their relationships; and trajectory analysis can capture continuous trends, enabling facile interpretation. COVET thus provides a rich, robust and computationally efficient representation of cellular niches, derived from a mathematically principled formulation based on optimal transport.

### The ENVI algorithm

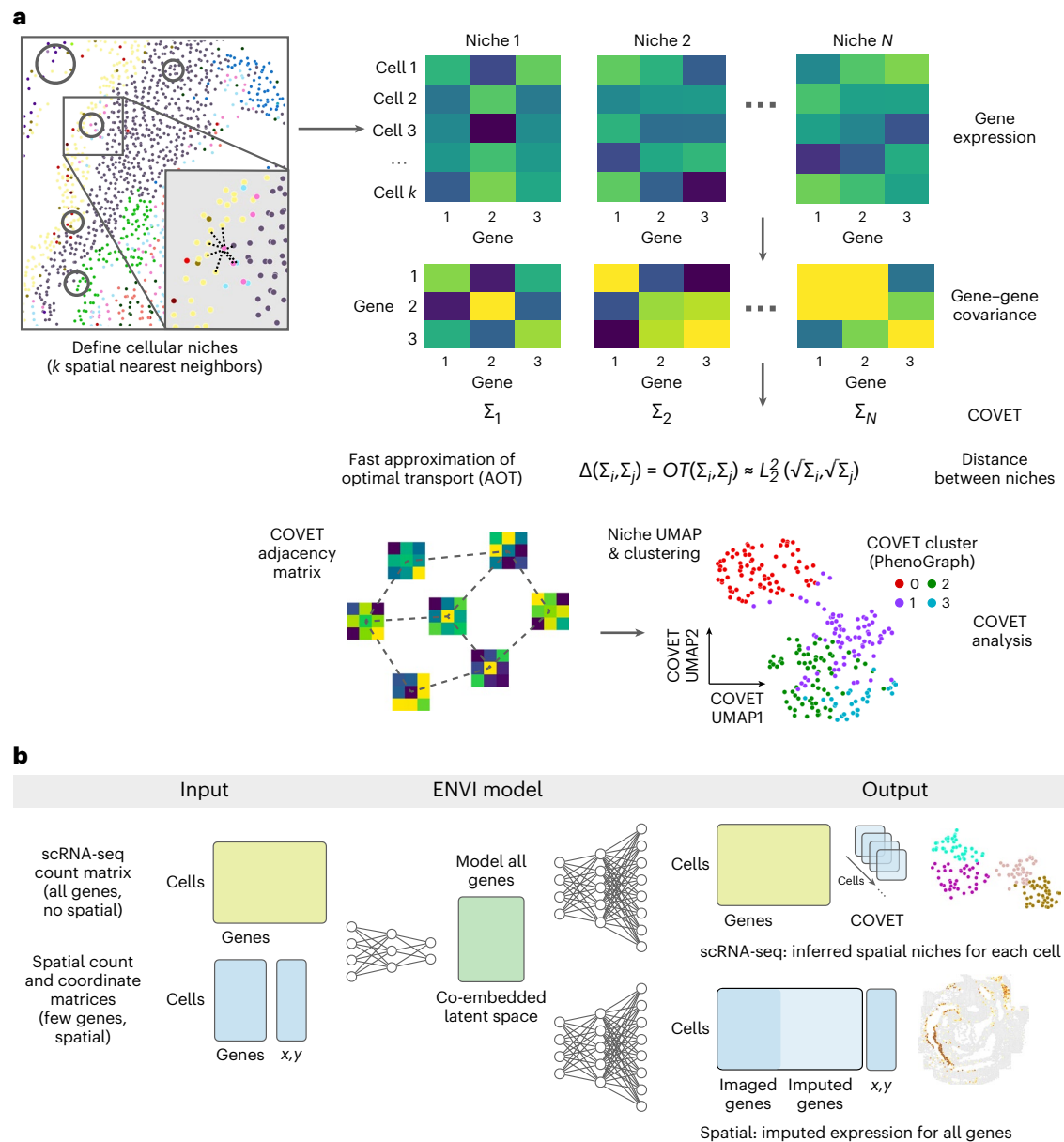
ENVI employs a conditional variational autoencoder to infer spatial context in scRNA-seq data and impute missing genes in spatial data, by mapping both modalities to a common embedding (Fig. 1b and Methods). Unlike other CVAEs used for spatial inference<sup>25,34,35</sup>, which only model genes measured in both modalities, ENVI explicitly models spatial information and gene expression genome wide. More importantly, it uses the COVET matrix to represent spatial information and simultaneously trains on samples from both spatial and single-cell datasets, optimizing a single latent space to decode the full transcriptome and spatial context for both modalities.

ENVI architecture includes a single encoder for both spatial and single-cell genomics data and two decoder networks: one for the full transcriptome and the second for the COVET matrix, providing spatial context. The requisite for decoding the spatial niche (and the use of a second decoder) is a unique aspect of ENVI. Intuitively, ENVI uses gene expression in the cell paired with its niche information (COVET) to learn an 'environment' regression model, which infers spatial context from gene expression input, and, simultaneously, an 'imputation' regression model trained to reproduce the full scRNA-seq dataset from the gene subset profiled by spatial transcriptomics. The nonlinear network architecture can capture complex dependencies between the variables.

Sequencing and spatial technologies measure different parameters and produce different data distributions and dynamic ranges (Extended Data Fig. 2a). ENVI takes this into account by marginalizing technology-specific effects on expression, augmenting the standard variational autoencoder (VAE) by adding an auxiliary binary neuron to the input layers of encoding and decoding networks for each modality. Moreover, ENVI parameterizes each modality with different probabilistic distributions, modeling single-cell data with a negative binomial by default to account for dropout<sup>36</sup>, and spatial data with a Poisson by default to reflect the high capture rate of fluorescence in situ hybridization (FISH)-based technologies<sup>3</sup>. ENVI thus integrates, imputes and reconstructs spatial context with a single end-to-end model, using deep learning for high-dimensional regression and variational inference for optimal integration of scRNA-seq and spatial data. The method scales to atlas-size datasets including millions of cells with constant time computational complexity (Extended Data Fig. 2b) while being robust to technology-specific artifacts, such as data sparsity (Extended Data Fig. 2c,d and Methods).

### ENVI imputes spatial patterns underlying gastrulation

We used ENVI to analyze a 350-gene sequential FISH (seqFISH)<sup>37</sup> of mouse organogenesis at embryonic day 8.75 (E8.75) and matched scRNA-seq



**Fig. 1 | A covariance-based framework characterizes spatial niches and powers single-cell and spatial data integration for robust transcriptome-wide spatial inference. a**, Schematics indicate steps in spatial covariance computation and ENVI operation. Each COVET matrix characterizes a cell's niche, comprising  $k$  nearest spatial neighbors, based on the shifted covariance of gene expression within the niche. Shifted covariance is calculated relative to mean expression in the sample, enabling meaningful comparison of niches. Distance between niches is determined by an efficient approximation of optimal transport. The COVET adjacency matrix, based on AOT as the spatial similarity metric, can be used directly for other downstream spatial analyses, such as

dimensionality reduction and clustering, where cells are grouped together by similar environment rather than by expression. **b**, ENVI is a conditional autoencoder that simultaneously embeds scRNA-seq and multiplexed spatial transcriptomic data into a unified latent embedding. ENVI models all genes (including those not imaged with spatial transcriptomics) and uses the COVET framework to represent information about cellular environment. An environment decoder allows ENVI to project spatial information onto single-cell data, and an expression decoder, which also includes genes captured only in the single-cell data, enables imputation of spatial expression transcriptome wide.

dataset<sup>38</sup> at E8.5 (Fig. 2a). Unlike the discrete layers of adult brain tissue<sup>3,39,40</sup> that dominate current spatial transcriptomic datasets, cells in developing embryos undergo rapid proliferation, differentiation and movement to create complex patterns and spatial gradients, presenting a challenging context for performance assessment. The most basic evaluation of any embedding-based data integration method is how well data across technologies co-embed, because this is critical for successful information transfer between modalities. The embedding learned by ENVI correctly maps major cell types to the combined latent space (Fig. 2b), as measured by average batch silhouette score<sup>41</sup> (Methods).

Current FISH-based technologies only quantify the expression of hundreds of genes<sup>12,37,40</sup>, prompting the development of algorithms to impute the spatial patterns of unmeasured genes<sup>9,25,26,42,43</sup>. Previous studies<sup>9,25,44</sup> used Pearson correlation and mean squared error between imputed and ground truth expression to evaluate the quality of imputation. However, both metrics are computed on a per-cell basis and ignore spatial context. To evaluate concordance between spatial patterns, we developed the multiscale spectral similarity index (MSSI), a metric that can capture similarity between spatial patterns by taking cell-cell proximity into account (Fig. 2c and Methods). MSSI borrows

from the multiscale structure similarity index measure (MS-SSIM)<sup>45</sup>, a spatial pattern similarity metric widely used in computer vision that iteratively subsamples an image and assesses similarity at multiple resolutions. Our MSSI metric uses a cell–cell neighbor graph based on spatial proximity to generate a series of images at progressively lower resolutions by aggregating proximal cells and then applies SSIM to compare similarity at each resolution. MSSI is, thus, a spatially aware similarity metric that uses full count matrices and incorporates patterning at the cellular rather than pixel level and has multiple use cases, such as comparing the similarity of different gene expression patterns<sup>46</sup>.

We used five-fold cross-validation (Methods) to compare ENVI imputation with measurements of held-out genes using both MSSI and Pearson correlation. The imputed expression of representative genes with clear spatial expression in endoderm (*Krt18*), neural stem (*Sox2*) and posterior section (*Hoxb9*) was visually similar to ground truth (Fig. 2d) and expressed in the correct organ. We found that some genes with correctly predicted organ-specific expression have high MSSI score but low Pearson correlation, supporting the importance of a spatially aware metric (Extended Data Fig. 3a).

We compared ENVI against Tangram<sup>9</sup>, gimVI<sup>25</sup> and uniPort<sup>47</sup>, which were recently shown to outperform other integration methods<sup>44</sup>; NovoSpaRc<sup>26</sup>, because it uses fused optimal transport to explicitly model spatial context; deepCOLOR<sup>48</sup>, because it uses a deep generative model; and Harmony<sup>49</sup>, for its widespread use as a batch correction method<sup>50</sup>. ENVI significantly outperforms all other methods based on both MSSI and Pearson correlation (Fig. 2e).

Finally, we evaluated ENVI's ability to impute genes beyond the 350-gene panel by assessing canonical markers of the developing lung (*Ripply3*)<sup>51</sup>, heart (*Nkx2-5*)<sup>52</sup> and intestine (*Tlx2*)<sup>53</sup>. The expression of all three genes was validated as organ specific at E8.75 (before organ formation) by *in situ* hybridization chain reaction (HCR) imaging<sup>54</sup> and was correctly imputed by ENVI (Fig. 2f). By contrast, Tangram and gimVI predicted weaker expression in the relevant region and anomalous expression beyond the organ (Extended Data Fig. 3b).

### ENVI ascribes spatial patterns to single-cell genomics data

In addition to gene imputation, ENVI can uniquely project spatial information onto dissociated cells profiled by scRNA-seq, by using its second decoder to reconstruct COVET matrices from the latent space. This approach can use limited spatial profiling data to confer spatial context onto the millions of cells in single-cell atlases. COVET represents gene–gene covariation between neighboring cells; thus, beyond deducing the cell type of neighbors, it can also infer their gene expression.

To demonstrate this ability, we used the mouse embryo dataset and focused on the gut tube, which generates the thymus, thyroid, lung, liver, pancreas, small intestine and colon in a stereotypical anterior-to-posterior sequence. Although E8.75 gut tube cells are anatomically indistinguishable, spatially delimited expression

reveals that precursors are poised for their organ fates<sup>55</sup>. We computed COVET matrices for (measured) seqFISH data and used ENVI to infer COVET matrices for scRNA-seq data and then applied the AOT metric to co-embed matrices from both modalities (Fig. 3a). The rich transcriptional information in scRNA-seq data facilitated the assignment of endodermal organ identity to these cells<sup>55</sup>, and ENVI's highly concordant co-embedding allowed for label transfer to those cells measured with seqFISH, as confirmed by anatomical localization; thymus and thyroid cells fall into the most anterior ventral gut tube, followed by dorsal and ventral lung clusters and, finally, intestine (Fig. 3b).

Using only endodermal scRNA-seq data, we plotted the average COVET matrices of dorsal and ventral lung and observed that these closely match empirical matrices computed from the seqFISH data (Fig. 3c). These COVET matrices infer modules of covarying genes in the niche environment and notably include genes expressed by adjacent mesodermal cells, which are known to provide spatial patterning cues to the endoderm<sup>56</sup>. To validate these inferred gene modules, we used the seqFISH data from mesodermal cells proximal to the gut tube (ignoring endodermal cells) and found that average ventral COVET gene expression is enriched in the ventral pharyngeal mesoderm, whereas average dorsal COVET gene expression is enriched in the dorsal brain and paraxial mesoderm (Fig. 3d). Our observations validate the predicted dorsal and ventral subdomains within the gut tube and demonstrate that ENVI can identify biologically important signaling originating from cells that were not sampled directly.

Using spatial covariance can also markedly improve the simpler task of labeling organ identity for cells from the spatial modality, as seqFISH measures fewer genes and cells than scRNA-seq and is, thus, more difficult to label. Labeling scRNA-seq cells from the gut tube with organ-specific gene sets<sup>55</sup> (Methods) revealed an almost one-to-one matching between organ precursors and COVET clusters, whereas ENVI without COVET failed to generate accurate labels, and alternative approaches were even less accurate (Extended Data Fig. 4). ENVI-based label transfer is also robust to variation in neighborhood size when computing COVET (Extended Data Fig. 5).

### ENVI learns spatial gradients from single-cell data

Although the gut tube is defined by relatively discrete primordial organs, many processes—such as the specification of spinal cord cells and their precursors, the neuromesodermal progenitors (NMPs), along the anteroposterior (AP) axis—are organized by continuous spatial gradients<sup>57</sup>. To highlight ENVI's ability to model gradients, we co-embedded empirical seqFISH COVET matrices with ENVI-inferred scRNA-seq COVET matrices for NMPs and spine cells using a force-directed layout (FDL)<sup>58</sup> and calculated their diffusion components (DCs) (Fig. 4a). The first DC is highly congruent with the AP axis (Pearson correlation = 0.86), demonstrating that COVET can capture gradual

**Fig. 2 | ENVI accurately recovers the expression of embryonic genes not imaged by multiplexed FISH. a**, seqFISH<sup>37</sup> image of an E8.75 mouse embryo sagittal section (left) and UMAP embedding of matched scRNA-seq data<sup>38</sup> at E8.5 (right), both colored by major cell type compartment. **b**, UMAPs of ENVI latent embedding learned from mouse embryo data. Cells from seqFISH (left) and scRNA-seq (right) data are colored as in **a**. Average batch silhouette score  $bASW = 0.86$  (Methods). **c**, Schematic of MSSI computation (Methods) for comparing two spatial expression profiles. Each profile is iteratively downsampled using spectral pooling on the cell proximity graph, and the SSIM is computed at each scale. MSSI is a weighted geometric mean of the SSIM computed at five scales, providing a spatially aware similarity metric on a scale from 0 to 1. **d**, seqFISH measurement in log counts (ground truth) and ENVI imputation for three withheld genes, marking endoderm (*Krt18*), neural stem (*Sox2*) and posterior section (*Hoxb9*) cells. MSSI values for each gene appear in parentheses. **e**, Pearson correlation and MSSI scores between log of seqFISH and

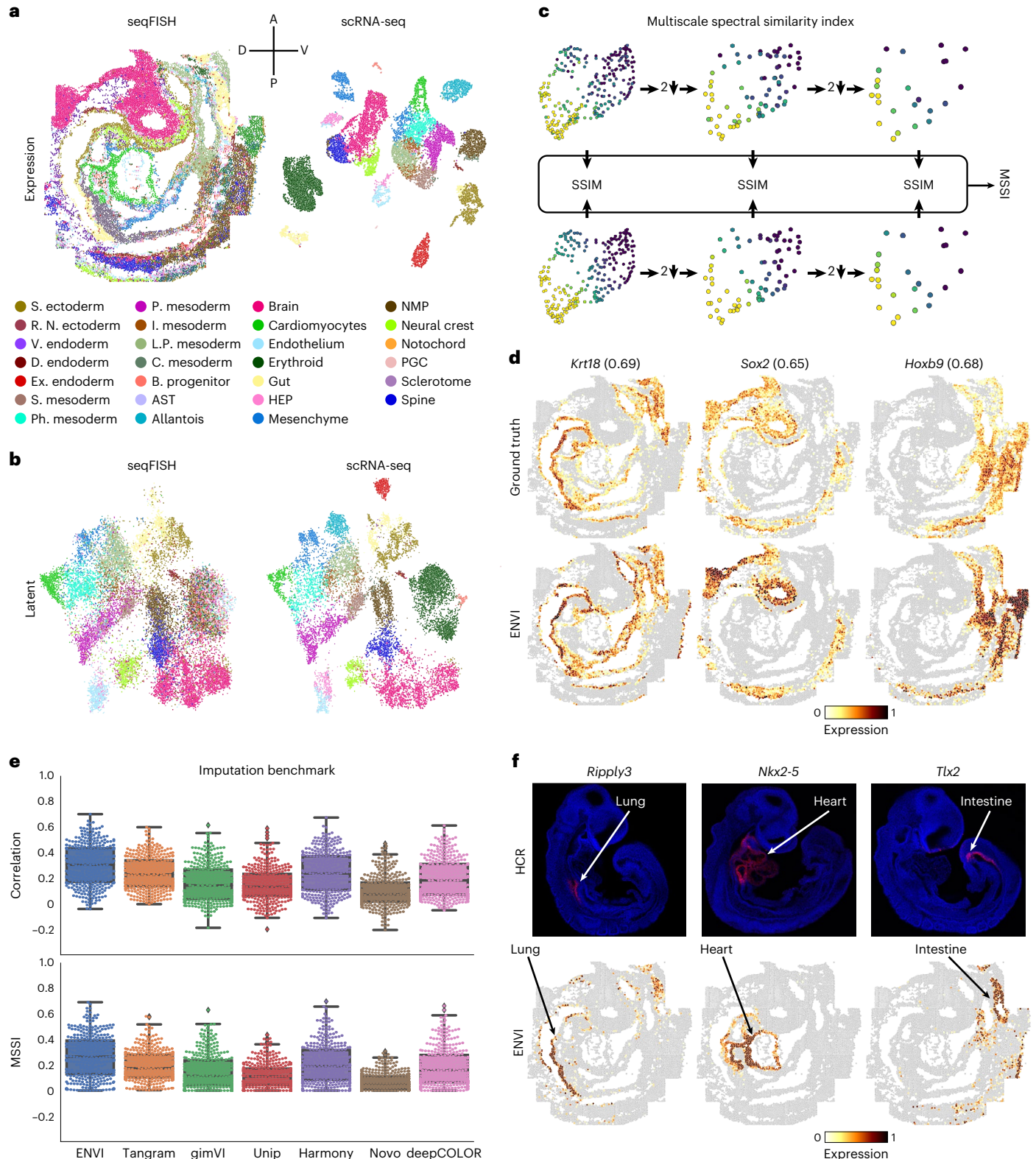
imputed expression across all genes predicted from five-fold cross-validation, comparing four algorithms run with default parameters (Methods). Novo, NovoSpaRc; Tg, Tangram; Unip, uniPort. Boxes and lines represent interquartile range (IQR) and median, respectively; whiskers represent  $\pm 1.5 \times$  IQR. In order, MSSI/Pearson correlation  $P$  values (one-sided  $t$ -test,  $n = 351$ ) are:  $4.45 \cdot 10^{-11}$ ,  $4.75 \cdot 10^{-9}$ ,  $3.45 \cdot 10^{-58}$ ,  $3.02 \cdot 10^{-94}$ ,  $1.44 \cdot 10^{-76}$ ,  $6.68 \cdot 10^{-93}$ ,  $4.83 \cdot 10^{-18}$ ,  $3.44 \cdot 10^{-50}$ ,  $4.12 \cdot 10^{-79}$ ,  $15.47 \cdot 10^{-85}$  and  $6.48 \cdot 10^{-32}$ ,  $4.51 \cdot 10^{-69}$ . **f**, ENVI imputation (bottom) of organ marker genes not profiled by seqFISH: *Ripply3* (lung), *Nkx2-5* (heart) and *Tlx2* (intestine). Imputed expression matches associated organs and is validated by whole-mount HCR *in situ* (top). AST, anterior somitic tissue; B., blood; C., caudal; D., definitive; Ex., extraembryonic; HEP, hematoendothelial progenitor; I., intermediate; L.P., lateral plate; Novo, NovoSpaRc; P., paraxial; PGC, primordial germ cell; Ph., pharyngeal; R.N., rostral neuro; S. Ectoderm, surface ectoderm; S. Mesoderm, somitic mesoderm; Unip, uniPort; V., visceral.

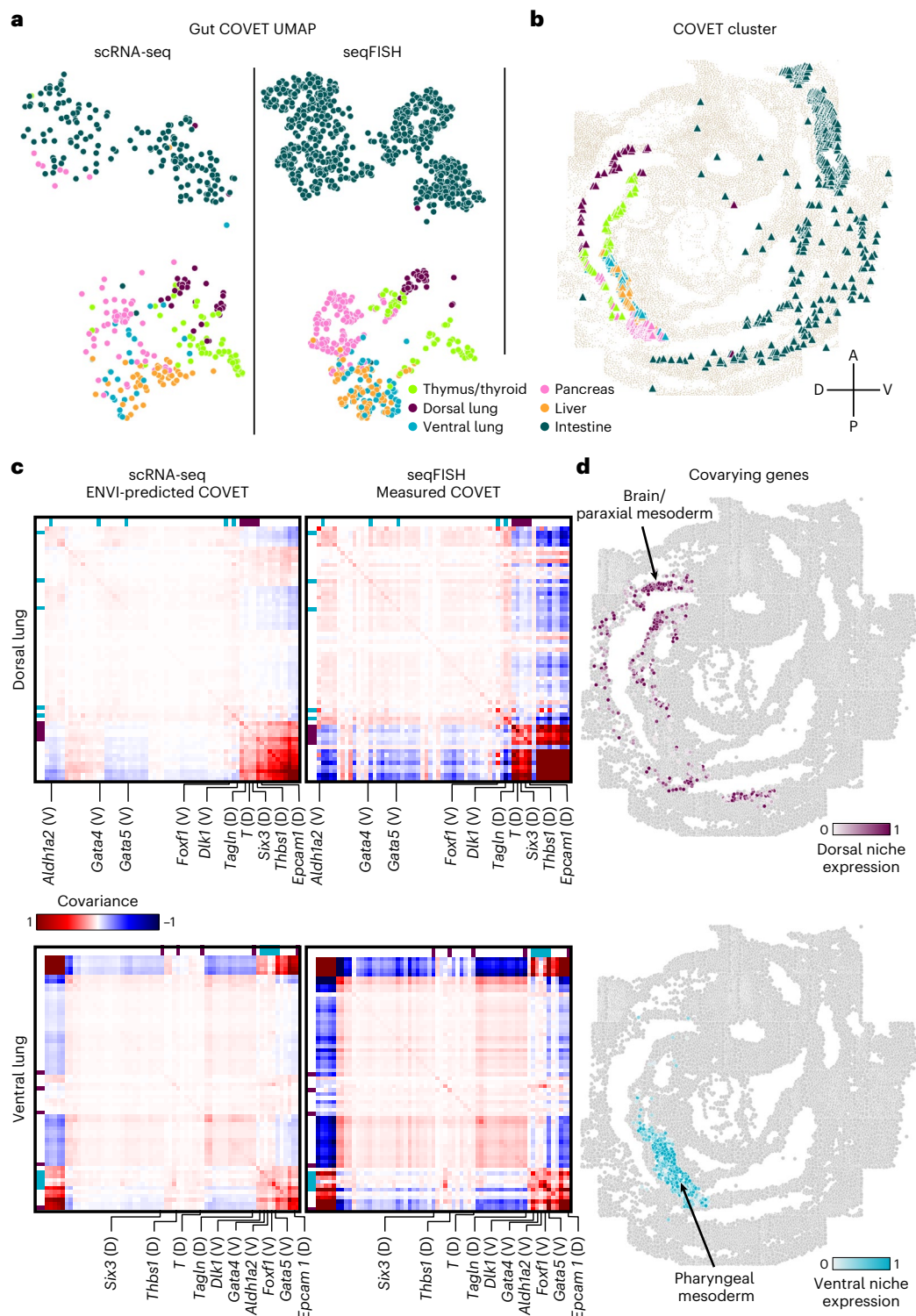
spatial trends (Fig. 4a,b and Extended Data Fig. 6a). As the COVET DC is calculated from both seqFISH and scRNA-seq datasets, we can use it to assign AP pseudo-coordinates to NMPs and spine cells from scRNA-seq data.

The first COVET DC correctly reveals that scRNA-seq cells are enriched for *Hoxd4* (refs. 55,59) (anterior) and *Hoxb9* (ref. 60) (posterior) markers in their respective domains, consistent with seqFISH expression in NMPs and spine cells (Fig. 4c). Furthermore, ENVI

correctly mapped high expression of *Hoxd3* (ref. 59) (anterior) and *Hoxb5os<sup>sl</sup>* (posterior) markers to scRNA-seq cells in their corresponding AP domains, demonstrating that ENVI spatial modeling extends to genes that are not imaged (Fig. 4d). Conversely, ENVI-imputed *Hoxb5os* and *Hoxd3* expression for the seqFISH data mirrors the predicted spatial context of the scRNA-seq data.

We found that the major axis of variation (first DC) between the COVET matrices that model the niche reflects the spatial organization



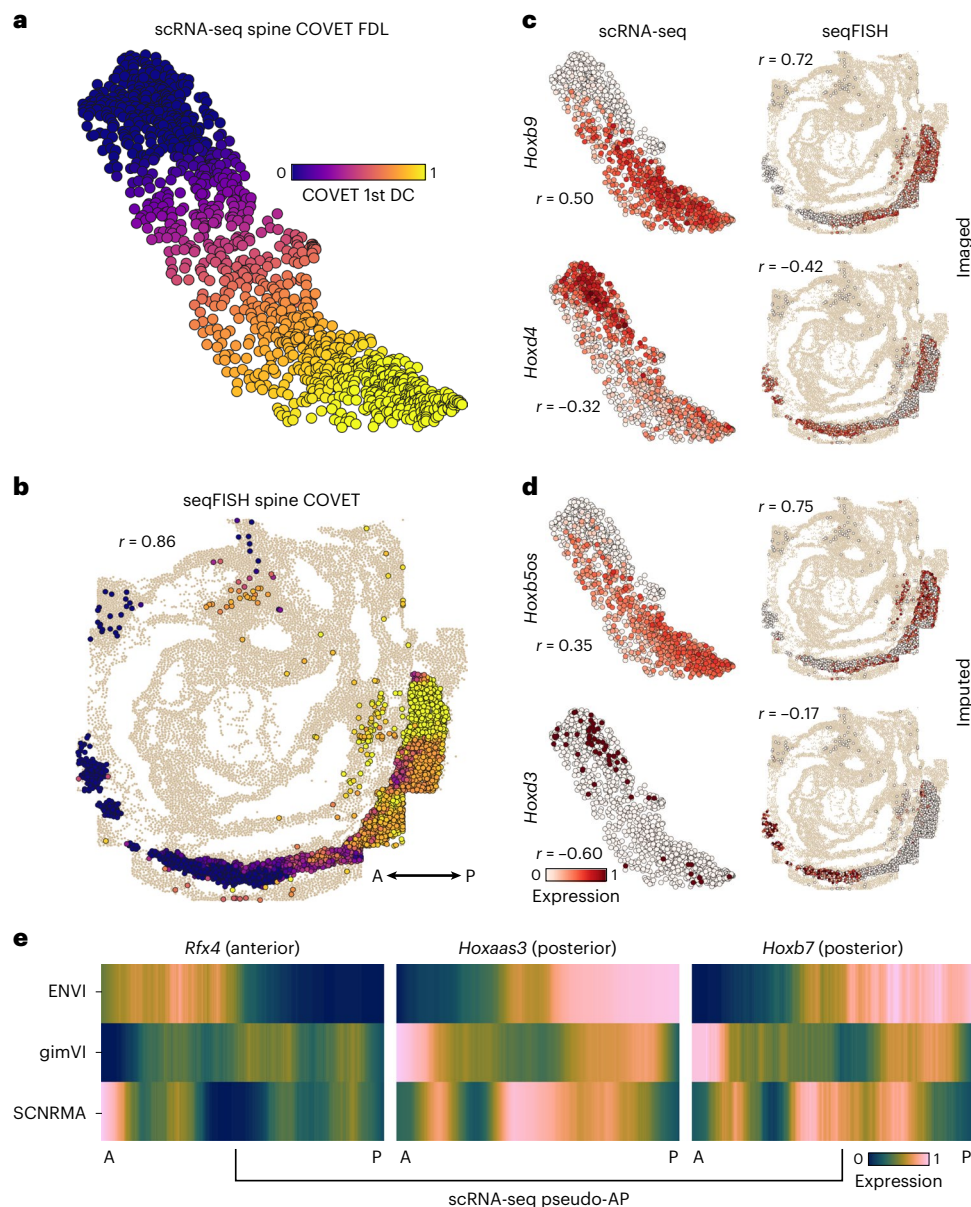


**Fig. 3 | ENVI confers spatial context to single-cell samples from mouse gut organogenesis.** **a**, COVET UMAP of gut tube cells. Each cell is represented by its directly measured (seqFISH) or ENVI-inferred (scRNA-seq) COVET, embedded in 2D coordinates by UMAP using the AOT metric. Cells are colored by emerging organ. **b**, seqFISH endoderm cells, colored by organ (based on label transfer from scRNA-seq), delineating spatial regions of the gut tube and showing coherence of organ assignments in distinct spatial niches. **c**, Average ENVI-imputed (scRNA-

seq, left) or empirically measured (seqFISH, right) COVET matrices of dorsal (top) and ventral (bottom) lung cells from scRNA-seq (left) and seqFISH (right) datasets. Labeled gene modules (top row of each matrix) are uniquely present in either dorsal (purple) or ventral (cyan) lung COVET representations. **d**, Mean expression of the uniquely covarying genes in either dorsal (top) or ventral (bottom) lung COVET matrices in **c**. Only mesoderm cells surrounding the endoderm gut tube (but not part of it) are colored.

of the tissue; ordering NMPs and spine cells along DC 1 recovers a pseudo-AP axis that can be used to visualize predicted expression trends<sup>57</sup> (Fig. 4e). Similar analysis using the gimVI latent space and

Scanorama<sup>62</sup> integration (Methods) led to inferior alignment with the true AP axis (Extended Data Fig. 6b), despite selecting the gimVI and Scanorama DCs most correlated ( $r = 0.76$  and  $r = 0.7070$ , respectively)



**Fig. 4 | ENVI maps continuous spatial gradients in spine development from single-cell and spatial data. a**, FDL based on ENVI-predicted COVET matrices of scRNA-seq spinal cord cells and their NMP precursors, colored by the first COVET diffusion component (pseudo-AP). FDL and DCs were calculated using both the seqFISH and scRNA-seq COVET matrices. **b**, seqFISH NMPs and spine cells colored by the first COVET DC, which traverses the AP axis. Pearson correlation ( $r$ ) is calculated between pseudo-AP and true AP. **c**, Expression of *Hoxd4*

(anterior) and *Hoxb9* (posterior) markers in NMPs and spine cells from scRNA-seq and seqFISH data, recapitulating expected AP localization. **d**, Same as **c**, colored by ENVI-imputed *Hoxb5os* (posterior) and *Hoxd3* (anterior) expression. These markers were not imaged. **e**, scRNA-seq expression of canonical AP axis markers in NMPs and spine cells ordered along pseudo-AP axis according to different integration methods. SCNRMA, Scanorama.

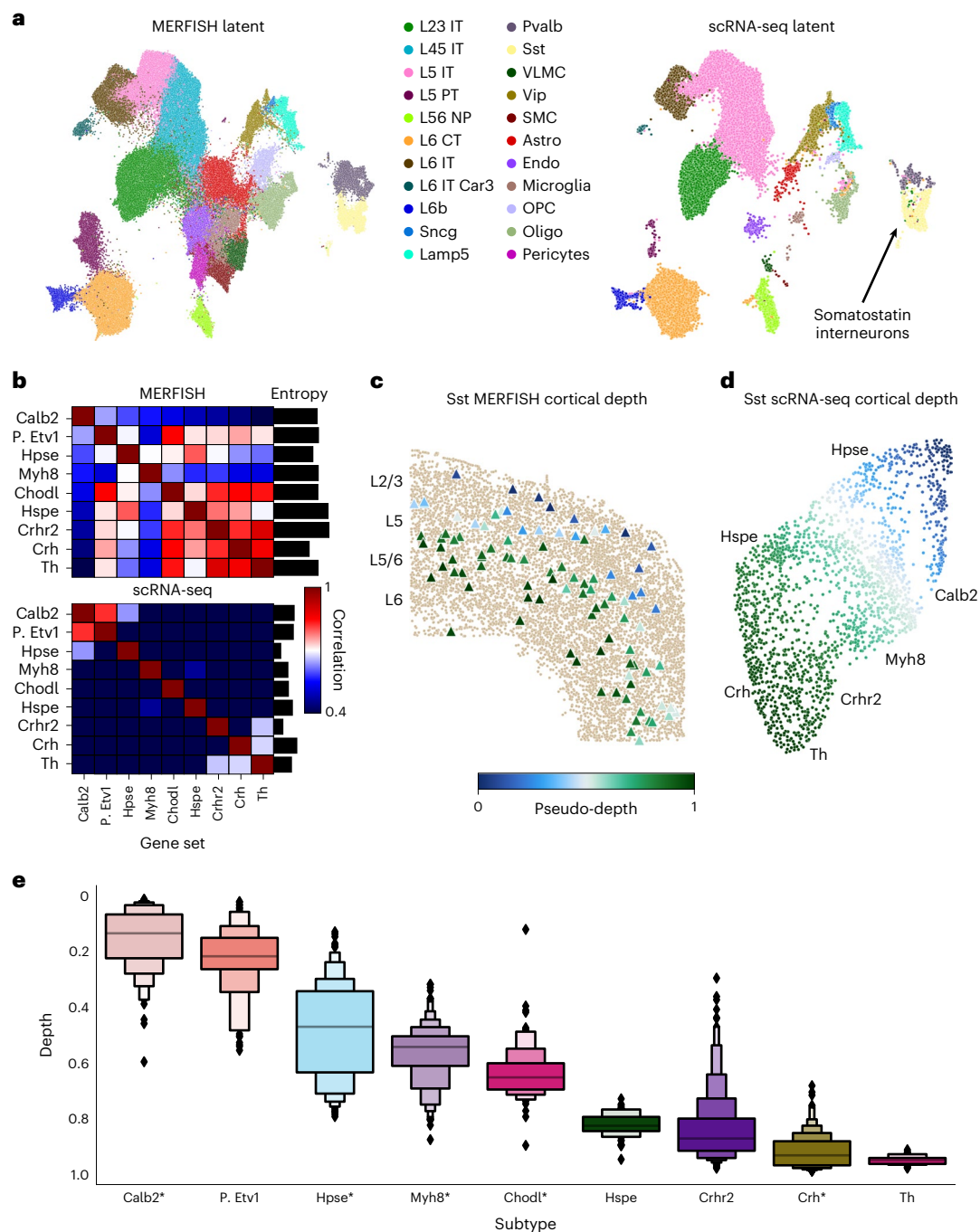
with true AP polarity. This slightly lower correlation with the AP axis propagates into more pronounced inaccuracies in expression patterns; only ENVI correctly derived expected AP trends for *Rfx4* (ref. 63), *Hoxaas3* (ref. 61) and *Hoxb7* (ref. 64) (Fig. 4e). More generally, both anterior and posterior canonical markers are more correlated (or anti-correlated) with ENVI COVET pseudo-AP than with axes defined by gimVI and Scanorama (Extended Data Fig. 6c). ENVI can, thus, correctly uncover AP polarity within single-cell NMPs and spine cells and correctly place them along this spatial axis.

#### ENVI delineates tissue-scale patterning in the motor cortex

Although data integration is typically evaluated on abundant neural cell types that dominate spatial regions, we challenged ENVI to recover

rare cell types. Somatostatin (*Sst*)-expressing interneurons are a cardinal class of inhibitory neurons in the cortex<sup>65</sup> that are implicated in Alzheimer's disease and depression<sup>66</sup> and encompass substantial diversity<sup>67–69</sup>. Although we know that *Sst* interneurons influence their environment, their localization and its relationship to function and transcriptional states have not been fully explored.

To localize *Sst* interneurons, we analyzed the scRNA-seq (71,183-cell) and 252-gene MERFISH (276,556-cell) atlases of the motor cortex of the Brain Initiative Cell Census Network (BICCN)<sup>40,70</sup>. ENVI outperformed all other tested methods in both speed (training on this large atlas in minutes) and imputation (Extended Data Figs. 2b and 7), and it successfully co-embedded the 22 BICCN-annotated coarse cell types (Fig. 5a). Notably, only scRNA-seq data can distinguish the nine



**Fig. 5 | ENVI predicts the cortical localization of *Sst* interneuron subtypes.**

**a**, ENVI latent embedding of motor cortex MERFISH (left) and scRNA-seq (right) data, labeled by cell types from Zhang et al.<sup>40</sup> and Yao et al.<sup>70</sup>, and highlighting *Sst* interneurons. Latent integration bASW = 0.78. OPC, oligodendrocyte precursor cell. **b**, Correlation between subtype-specific gene sets (Supplementary Table 3) in *Sst* interneurons from the MERFISH and scRNA-seq datasets. Bar plots mark the entropy of overlap for each subgroup (higher values denote poorer delineation into distinct subtypes). **c**, *Sst* interneurons from MERFISH data, colored by first COVET DC representing pseudodepth. L2/3 to L6 indicate approximate location

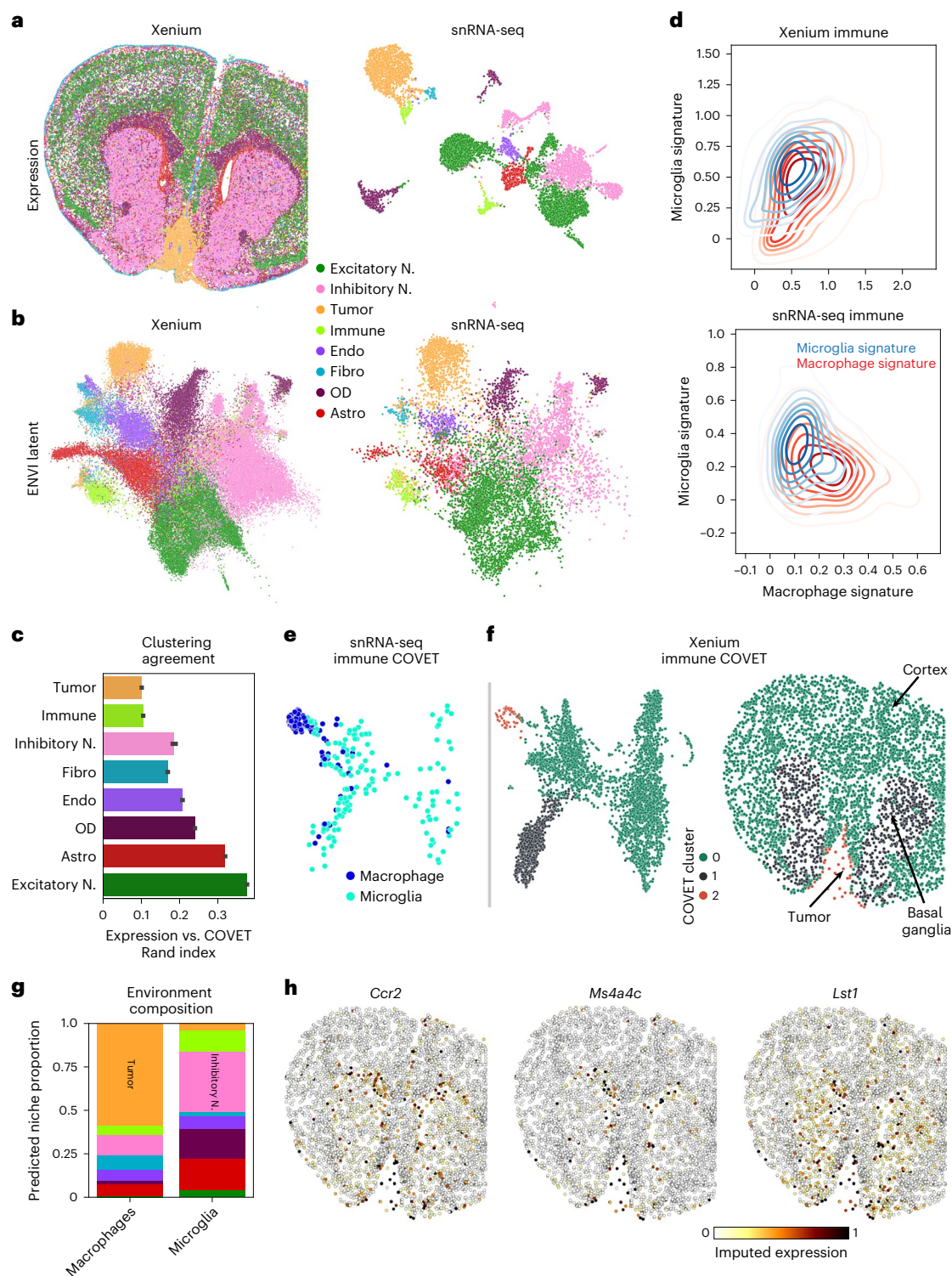
of cortical layers. **d**, FDL of scRNA-seq COVET for all *Sst* interneurons, colored by first COVET DC. COVET FDL and DCs are calculated based on both scRNA-seq and MERFISH datasets. Specific subtypes, labeled by Yao et al.<sup>70</sup>, are marked. **e**, COVET DC-predicted cortical depth of scRNA-seq *Sst* interneurons, grouped by labeled subtype. \*, cell types imaged in Wu et al.<sup>71</sup>; all empirical depths match ENVI predictions. In the box plot, line is median; first boxes include the 50th percentile; and each successive pair of boxes outward contain half of the remaining data.

distinct *Sst* subpopulations, as the MERFISH panel lacks requisite marker genes (Fig. 5b and Methods).

Using ENVI-imputed COVET matrices, we mapped *Sst* interneurons labeled in the scRNA-seq dataset to their location within the cortex. We found that—despite being interspersed throughout the cortex, where cell types, such as excitatory neurons, dominate—the first DC of

COVET matrices is highly correlated with cortical depth, thus defining a ‘pseudodepth’ axis (Fig. 5c,d), and that *Sst* subtypes are predicted to stratify by depth (Fig. 5e). Molecular imaging by genetic strategies targeting *Sst* subtypes<sup>71</sup> validates a number of our predictions, including the localization of *Calb2* interneurons to the L2/3 layers and *Crh* interneurons to L6. Beyond these, ENVI predicted the cortical depth of





**Fig. 6 | ENVI integrates Xenium and snRNA-seq data to localize neuroimmune cell types during metastasis. a**, Xenium image and UMAP embedding of snRNA-seq data from mouse brain bearing a melanoma metastasis, colored by major cell type. **b**, UMAP embeddings of ENVI latent space showing cells from the spatial (left) and snRNA-seq (right) datasets. Similar cell types, including malignant cells, co-embed across modalities. bASW = 0.87. **c**, Average concordance of technical replicates between the expression of each cell type and its environment in the Xenium data, measured as ARI between  $k$ -means clusters of gene expression and COVET representation. Error bars, 95% confidence interval. **d**, Density plots of microglia and macrophage cell signature expression in immune-labeled cells from Xenium (top) and snRNA-seq (bottom) datasets. Only snRNA-seq data measure enough genes to separate cell types. **e**, UMAP

embedding of the ENVI-predicted COVET representation of snRNA-seq immune cells, colored by subtype. **f**, COVET UMAP (left) and spatial coordinates (right) of Xenium immune cells, colored by COVET clusters representing major immune cell microenvironments: cortex (C0), basal ganglia (C1) and tumor (C2). COVET UMAP and clusters are calculated from both snRNA-seq and Xenium datasets. The vast majority (85%) of snRNA-seq macrophages were assigned to C2, predicting their localization to the tumor. **g**, Cell type fractions of ENVI-predicted environments for snRNA-seq macrophages and microglia, colored as in **a**, highlighting their predominant tumor and inhibitory neuron environments, respectively. **h**, ENVI-imputed macrophage infiltration marker genes on Xenium immune cell types are distinctly enriched within the metastasized tumor. Astro, astrocyte; Endo, endoderm; Fibro, fibroblast; OD, oligodendrocyte; N., neuron.

many subtypes identified in the scRNA-seq atlas with unknown localization. For example, it placed *Sst* interneurons expressing high levels of the neurotransmitter metabolism gene *Tyrosine hydroxylase* in the deep L6 layer, as might be expected, suggesting that ENVI can articulate the interplay between transcriptional state and microenvironment.

ENVI can also capture spatial patterns within the cortex from datasets that include only a few imaged genes. Applied to a 33-gene osmFISH and matched scRNA-seq dataset of the somatosensory cortex<sup>3</sup>, ENVI successfully integrated the small datasets (fewer than 10,000 cells combined) into a unified embedding (Extended Data Fig. 8a,b) and outperformed alternative methods in cell type resolution and spatial gene imputation (Extended Data Fig. 8c–e). To determine whether ENVI can impute unimaged genes, we leveraged Allen Brain Atlas ground truth data for mouse brain cortex (<https://mouse.brain-map.org/>) and confirmed that ENVI correctly imputes layer-specific spatial expression for *Dti4l*, *Rprm* and *Ntst4* in the L2/3, L5/6 and CA1 regions, respectively (Extended Data Fig. 8f).

### ENVI integrates Xenium data on brain metastasis

Leptomeningeal metastasis (LM) is a lethal condition in which distant tumor cells spread into the fluid-filled space surrounding the central nervous system<sup>72,73</sup>. The poor understanding of interactions among tumor, immune and underlying brain parenchyma cells limits the discovery of therapeutics. We used the Xenium platform (10x Genomics)<sup>16</sup> to perform in situ hybridization (ISH) of 243 genes in a mouse model of melanoma LM<sup>74</sup> and also sequenced cells from an adjacent section using a custom single-nucleus RNA sequencing (snRNA-seq) protocol (Methods) that we developed by optimizing RNA extraction from formalin-fixed paraffin-embedded (FFPE) samples, followed by 10x Genomics Flex probe-based library preparation. We separately clustered and annotated the spatial and single-cell samples into major cell types based on marker genes (Fig. 6a and Extended Data Fig. 9a). Even in this pathological context, ENVI performance with default parameters matches or exceeds competing methods on gene imputation (Extended Data Fig. 9b) and harmonizes the two datasets into a unified latent space (Fig. 6b and Extended Data Fig. 9c).

Our approach provides two representations of the Xenium data; we can visualize and cluster each cell based either on its gene expression or on its COVET matrix (representing the local niche). Measuring the agreement between clustering of the two representations reveals that, as expected, excitatory neuron expression depends strongly on spatial context, due to the association between distinct cortical layers and molecular markers<sup>70</sup>, whereas tumor and immune cell types show little concordance between expression and environmental context (Fig. 6c).

Melanoma LM interacts with two key immune populations: tissue-resident brain macrophages, known as microglia, and monocyte-derived macrophages that are recruited to the tumor lesion and colonize it from the periphery<sup>74</sup>. The snRNA-seq data clearly distinguish these myeloid subtypes based on curated gene sets<sup>75</sup>, whereas the Xenium brain panel lacks the markers to distinguish macrophages and microglia (Fig. 6d). To resolve where the subtypes localize, we co-embedded ENVI-imputed snRNA-seq COVET matrices with observed Xenium COVET matrices and clustered the data, revealing three distinct immune microenvironments consisting of cortical, basal ganglia or tumor cells (Fig. 6e,f). The snRNA-seq data enabled the labeling of cluster 2 as non-resident macrophage, and the Xenium data allowed us to visualize the localization of cells from this COVET cluster. Confirming known patterns of neuroimmune cell types<sup>76–78</sup>, most microglia were assigned to the basal ganglia and cortex, whereas most macrophages were localized to the tumor and its boundary. COVET allowed us to infer the niche composition for each immune cell in the snRNA-seq data, which corroborates that macrophages are found mainly near tumor cells, whereas microglia are found mainly near neurons and other glial cells (Fig. 6g).

Beyond localizing macrophages and microglia, ENVI can distinguish the transcriptional patterns of tumor-infiltrating macrophages from those on the boundary by imputing gene expression in the Xenium data (Fig. 6h). For instance, imputation of *Ccr2*, a chemokine receptor that recruits monocytes to the tumor and promotes their differentiation into tumor-associated macrophages<sup>78</sup>, was enriched in immune cells within the tumor and its vicinity. In contrast, clustering-based analysis of the gimVI latent on the immune cells does not clearly assign macrophages to a malignant microenvironment, and its gene imputation is also inaccurate, predicting that tumor infiltration genes are broadly expressed across the brain (Extended Data Fig. 9d,e). Harmony and gimVI also fail to localize infiltration marker expression to immune cells within the tumor (Extended Data Fig. 10).

## Discussion

ENVI robustly integrates scRNA-seq and spatial transcriptomics data, overcoming technical biases while retaining biological information. The algorithm provides superior performance for imputing missing gene expression in spatial modalities; it scales to millions of cells; and it has the distinctive ability to infer the spatial context of dissociated cells, even across multiple cell types in complex tissues.

ENVI's capabilities rely on COVET as a representation of spatial niches. Although most spatial representations are based on discrete cell typing, COVET takes full advantage of the quantitative nature of gene expression data. The COVET matrix captures covariation between markers in a cell's niche and uses optimal transport to derive a principled and quantitative model of cellular neighborhoods. COVET powers a shift from discrete cell type to continuous cell state paradigms and the discovery of continuous trends in spatial microenvironments.

ENVI performance is primarily driven by three factors: (1) deep Bayesian inference to regress out modality-related confounders while learning nonlinear relationships between genes and niches; (2) explicit modeling of the entire transcriptome from scRNA-seq data; and (3) direct incorporation of spatial context via COVET. Whereas current methods only learn the genes that overlap between scRNA-seq and spatial datasets, ENVI models all available information and does not rely on post hoc inference. This proves invaluable, as the ENVI model is imbued with both spatial context and full transcriptome information, allowing for reliable transfer of information between modalities.

The ENVI COVET space can correctly predict primordial organ niches from seqFISH and scRNA-seq data of mouse gastrulation, and COVET-based DC analysis can highlight continuous AP trends of both expression and environment in the developing spine. ENVI's critical ability to confer spatial context onto dissociated single cells drives the inference of circuits of *Sst* interneuron subtypes in the motor cortex. Moreover, it provides an accurate representation of both discrete and diffuse signals in healthy and pathological tissue contexts, enabling spatial reasoning along the full transcriptome, including the spatial distinction of subtly different tumor and non-tumor-associated macrophage cell states in metastatic tissue.

One caveat is that the range of spatial factors can vary, whereas COVET is currently defined at a single scale set by the neighborhood size,  $k$ . Although COVET is relatively robust to small changes of  $k$ , larger differences may lead to different outcomes, and its value should be tuned to the spatial questions of interest.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-024-02193-4>.

## References

- Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet.* **23**, 741–759 (2022).
- Keren, L. et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell* **174**, 1373–1387 (2018).
- Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
- Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl Acad. Sci. USA* **113**, 11046–11051 (2016).
- Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376 (2018).
- Atta, L. & Fan, J. Computational challenges and opportunities in spatially resolved transcriptomic data analysis. *Nat. Commun.* **12**, 5283 (2021).
- Fang, S. et al. Computational approaches and challenges in spatial transcriptomics. *Genomics Proteomics Bioinformatics* **21**, 24–47 (2023).
- Goltsev, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174**, 968–981 (2018).
- Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
- Wu, Z. et al. Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nat. Biomed. Eng.* **6**, 1435–1448 (2022).
- Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* **41**, 332–336 (2023).
- Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
- Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
- Merritt, C. R. et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).
- Liu, J. et al. Concordance of MERFISH spatial transcriptomics with bulk and single-cell RNA sequencing. *Life Sci. Alliance* **6**, e202201701 (2023).
- Janesick, A. et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).
- Lee, J. H. et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
- Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
- Haghverdi, L., Büttner, M., Wolf, F. A., Büttner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
- Bendall, S. C. et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014).
- Setty, M. et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
- Gaublomme, J. T. et al. Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* **163**, 1400–1412 (2015).
- Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 (2018).
- Geltink, R. I. K., Kyle, R. L. & Pearce, E. L. Unraveling the complex interplay between T cell metabolism and function. *Annu. Rev. Immunol.* **36**, 461–488 (2018).
- Lopez, R. et al. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.1905.02269> (2019).
- Nitzan, M., Karaiskos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).
- Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at *arXiv* <https://doi.org/10.48550/arXiv.1312.6114> (2013).
- Sohn, K., Yan, X. & Lee, H. Learning structured output representation using deep conditional generative models. In *Proc. of the 28th International Conference on Neural Information Processing Systems* 3483–3491 (MIT Press, 2015).
- Dowson, D. C. & Landau, B. V. The Fréchet distance between multivariate normal distributions. *J. Multivar. Anal.* **12**, 450–455 (1982).
- Choi, E. & Lee, C. Feature extraction based on the Bhattacharyya distance. *Pattern Recognit.* **36**, 1703–1709 (2003).
- Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
- Coifman, R. R. Special issue on diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 3 (2006).
- McInnes, L., Healy, L., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Soft.* **3**, 861 (2018).
- Lotfollahi, M. et al. Mapping single-cell data to reference atlases by transfer learning. *Nat. Biotechnol.* **40**, 121–130 (2022).
- Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
- Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
- Lohoff, T. et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* **40**, 74–85 (2022).
- Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
- Alon, S. et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656 (2021).
- Zhang, M. et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).
- Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
- Abdelaal, T., Mourragui, S., Mahfouz, A. & Reinders, M. J. T. SpaGE: spatial gene enhancement using scRNA-seq. *Nucleic Acids Res.* **48**, e107 (2020).
- Liu, J. et al. Jointly defining cell types from multiple single-cell datasets using LIGER. *Nat. Protoc.* **15**, 3632–3662 (2020).
- Li, B. et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).
- Wang, Z., Simoncelli, E. P. & Bovik, A. C. Multiscale structural similarity for image quality assessment. In *Thirty-Seventh Asilomar Conference on Signals, Systems & Computers* 1398–1402 (IEEE, 2003).
- DeTomaso, D. & Yosef, N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446–456 (2021).
- Cao, K., Gong, Q., Hong, Y. & Wan, L. A unified computational framework for single-cell data integration with optimal transport. *Nat. Commun.* **13**, 7419 (2022).

48. Kojima, Y. et al. Single-cell colocalization analysis using a deep generative model. *Cell Syst.* **15**, 180–192.e7 (2024).
49. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
50. Tran, H. T. N. et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**, 12 (2020).
51. Okubo, T. et al. Ripply3, a Tbx1 repressor, is required for development of the pharyngeal apparatus and its derivatives in mice. *Development* **138**, 339–348 (2011).
52. Lyons, I. et al. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene *Nkx2-5*. *Genes Dev.* **9**, 1654–1666 (1995).
53. Shirasawa, S. et al. *Enx* (*Hox11L1*)-deficient mice develop myenteric neuronal hyperplasia and megacolon. *Nat. Med.* **3**, 646–650 (1997).
54. Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).
55. Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
56. Han, L. et al. Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis. *Nat. Commun.* **11**, 4158 (2020).
57. Carpenter, E. M. *Hox* genes and spinal cord development. *Dev. Neurosci.* **24**, 24–34 (2002).
58. Fruchterman, T. M. J. & Reingold, E. M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21**, 1129–1164 (1991).
59. López-Delgado, A. C., Delgado, I., Cadenas, V., Sánchez-Cabo, F. & Torres, M. Axial skeleton anterior-posterior patterning is regulated through feedback regulation between Meis transcription factors and retinoic acid. *Development* **148**, dev193813 (2021).
60. Chen, F. & Capecchi, M. R. Targeted mutations in *Hoxa-9* and *Hoxb-9* reveal synergistic interactions. *Dev. Biol.* **181**, 186–196 (1997).
61. Degani, N., Lubelsky, Y., Perry, R. B.-T., Ainbinder, E. & Ulitsky, I. Highly conserved and *cis*-acting lncRNAs produced from paralogous regions in the center of HOXA and HOXB clusters in the endoderm lineage. *PLoS Genet.* **17**, e1009681 (2021).
62. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
63. Ashique, A. M. et al. The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. *Sci. Signal.* **2**, ra70 (2009).
64. Chen, F., Greer, J. & Capecchi, M. R. Analysis of *Hoxa7/Hoxb7* mutants suggests periodicity in the generation of the different sets of vertebrae. *Mech. Dev.* **77**, 49–57 (1998).
65. Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241 (2021).
66. Song, Y.-H., Yoon, J. & Lee, S.-H. The role of neuropeptide somatostatin in the brain and its application in treating neurological disorders. *Exp. Mol. Med.* **53**, 328–338 (2021).
67. Muñoz, W., Tremblay, R., Levenstein, D. & Rudy, B. Layer-specific modulation of neocortical dendritic inhibition during active wakefulness. *Science* **355**, 954–959 (2017).
68. Ma, Y., Hu, H., Berrebi, A. S., Mathers, P. H. & Agmon, A. Distinct subtypes of somatostatin-containing neocortical interneurons revealed in transgenic mice. *J. Neurosci.* **26**, 5069–5082 (2006).
69. Nigro, M. J., Hashikawa-Yamasaki, Y. & Rudy, B. Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex. *J. Neurosci.* **38**, 1622–1633 (2018).
70. Yao, Z. et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021).
71. Wu, S. J. et al. Cortical somatostatin interneuron subtypes form cell-type-specific circuits. *Neuron* **111**, 2675–2692 (2023).
72. Chamberlain, M. C. Leptomeningeal metastasis. *Curr. Opin. Oncol.* **22**, 627–635 (2010).
73. Wilcox, J. A., Li, M. J. & Boire, A. A. Leptomeningeal metastases: new opportunities in the modern era. *Neurotherapeutics* **19**, 1782–1798 (2022).
74. Remsik, J. et al. Leptomeningeal anti-tumor immunity follows unique signaling principles. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.03.17.533041> (2023).
75. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**, baz046 (2019).
76. Guadagno, E. et al. Role of macrophages in brain tumor growth and progression. *Int. J. Mol. Sci.* **19**, 1005 (2018).
77. Goswami, S. et al. Immune profiling of human tumors identifies CD73 as a combinatorial target in glioblastoma. *Nat. Med.* **26**, 39–46 (2020).
78. Noy, R. & Pollard, J. W. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* **41**, 49–61 (2014).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

## Methods

### Computational methods

**MSSI.** When comparing the spatial distribution of genes or markers across a tissue, it is imperative to have a robust metric that takes spatial structure into account. Although ubiquitous metrics, such as Pearson correlation, SSIM and root mean square error, can provide some insight, they lack spatial context (for example, cell–cell proximity or spatial patterns) and measure only per-cell discrepancy.

To devise a metric for spatial data, we borrowed the MS-SSIM<sup>45</sup>, a ubiquitous metric for the quality of image reconstruction, from computer vision. Given two images, MS-SSIM iteratively downsamples each image, creating an image pyramid<sup>79</sup>—a multiscale signal representation consisting of the same image at multiple resolutions. MS-SSIM returns a weighted geometric average of the standard SSIM scores between the two images at each scale of the pyramid. Standard SSIM for two images,  $x$  and  $y$ , is

$$\text{SSIM}(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y),$$

where

$$l(x, y) = \frac{2\mu_x\mu_y + (0.01 \cdot M)^2}{\mu_x^2 + \mu_y^2 + \frac{0.01}{M}}, c(x, y) = \frac{2\sigma_x\sigma_y + (0.03 \cdot M)^2}{\sigma_x^2 + \sigma_y^2 + (0.03 \cdot M)^2},$$

$$s(x, y) = \frac{\sigma_{xy} + \frac{(0.03 \cdot M)^2}{2}}{\sigma_x\sigma_y + \frac{(0.03 \cdot M)^2}{2}}$$

$M$  represents the maximum values between  $x$  and  $y$ ;  $\mu_x$  and  $\mu_y$  are their average values;  $\sigma_x$  and  $\sigma_y$  measure how each varies; and  $\sigma_{xy}$  represents how much they covary.  $l(x, y)$ ,  $c(x, y)$  and  $s(x, y)$  are measures of ‘luminance’ (signal brightness), contrast and structure, respectively. Although SSIM is meant for images, it can also be calculated between any two vectors of similar sizes.

We introduce the MSSI as an adaptation of MS-SSIM to spatial transcriptomics that compares count matrices from segmented cells, rather than pixels, using a neighbor graph of spatially neighboring cells to capture structure. Intuitively, MSSI is a spectral analog of MS-SSIM; by rephrasing image coarsening to its graph-based counterpart, we can apply it to segmented cells and produce a multiscale, spatially driven score of expression reconstruction quality.

MSSI compares the expression profiles of two genes from a multiplexed image:  $X = \{x_i\}_{i=1}^N$  and  $Y = \{y_i\}_{i=1}^N$ , where the index  $i$  enumerates segmented cells, and  $x$  and  $y$  can be either (1) two different genes or (2) a ground truth gene and its imputed value. In addition, the spatial coordinate of each cell is  $D = \{D_i\}_{i=1}^N$ . We first compute the  $k$  nearest neighbor (kNN) graph  $G^1$  of segmented cells from  $\{D_i\}_{i=1}^N$ . To generate a subsampled version of the kNN graph, we use a graph coarsening algorithm<sup>80</sup>, which pools nodes together based on their connectivity pattern, similarly to how image downsampling groups pixels together (Fig. 2c). We iteratively coarsen and blur the graph four times by a factor of 2 and produce the expression of every gene at each scale.

Mathematically, each coarsening step produces a pooled version of the graph  $\{G^s\}_{s=2}^5$  and a coarsening operator  $\{C^s\}_{s=2}^5$ , which is the mapping between nodes at one scale to nodes at the next and allows us to generate pooled versions of the gene expression signals:

$$\chi^{s+1}, \gamma^{s+1} = C^s \chi^s, C^s \gamma^s$$

After MS-SSIM, we compute the MSSI between the expression profiles at each scale and return their weighted geometric mean. In detail, we compute the  $l$ ,  $c$  and  $s$  SSIM-related values at each scale and derive MSSI based on their weighted product, as for the MS-SSIM:

$$\text{MSSI}(X, Y, D) = l_s(X^s, Y^s)^{\alpha_s} \prod_{s=1}^4 c(X^s, Y^s)^{\alpha_s} s(X^s, Y^s)^{\alpha_s},$$

where the weights are equivalent to those in MS-SSIM<sup>45</sup>:

$$\alpha = (0.0448, 0.2856, 0.3001, 0.2363, 0.1333)$$

When  $X^i, Y^i$  are anti-correlated ( $\sigma_{xy} < 0$ ),  $s$  is negative, which prevents computing the weighted geometric mean; we, thus, clip negative values to 0. This implies that if, at any scale,  $X^s, Y^s$  are anti-correlated, the MSSI will be 0, its lowest possible value. We also normalize the original-scale gene expression to be between 0 and 1 but do not re-normalize at each coarsening scale.

### Spatial covariance representation

Our spatial covariance framework includes three components: the COVET statistic, a similarity metric and an algorithm to robustly and efficiently compute the COVET metric. The COVET framework assumes that the interplay between the cell and its environment creates covarying patterns of expression between the cell and its niche, which can be formulated via the gene–gene covariance matrix of niche cells. The COVET statistic constructs a shifted covariance matrix (which preserves algebraic properties of the covariance matrix) and, thus, enables the use of any measure of statistical divergence between covariances to define a principled quantitative similarity metric to compare niches. The key is to build the COVET statistic in such a manner that two COVET matrices are comparable and to design a computationally efficient algorithm to quantify the statistical divergence between them.

**COVET.** The inputs to COVET are (1) the gene expression matrices ( $X \in R^{n \times g}$ ), where  $n$  is the number of cells and  $g$  is the number of genes profiled; (2) the location of each cell in situ; and (3) a parameter ( $k$ ) that defines the number of nearest cells to be included in the niche. For each cell, we identify its  $k$  nearest cells (excluding the cell itself) based on their spatial proximity and construct a niche matrix  $E_i = \{Y_{ij} \in R^g | j \in kNN(i)\}$ , which represents the gene expression vector for each of those nearest neighbors. This produces an  $n \times k \times g$  tensor  $\Omega = \{E_i \in R^{k \times g} | i = 1, \dots, n\}$ , which combines the niche matrices of every cell.

The fundamental goal of COVET is to transform those niche matrices into effective representations of a cell’s niche. To this end, we calculate the ‘shifted’ gene–gene covariance matrix between cells in each niche matrix, where, instead of using the classical formulation

$$\Sigma_i^{\text{classic}} = \text{Cov}(E_i) = \frac{1}{k} (E_i - \bar{E}_i)^T (E_i - \bar{E}_i)$$

we swap the niche mean expression  $\bar{E}_i$  with the total expression average  $\bar{X}$  (computing the mean over the entire dataset). This enables direct comparison between covariance matrices, as they are constructed relative to the same reference:

$$\Sigma_i = \text{ShiftCov}(E_i) = \frac{1}{k} (E_i - \bar{X})^T (E_i - \bar{X}).$$

This creates a representation relative to the entire population, which can better highlight the features that are unique to each niche while also holding the same algebraic properties that the standard covariance matrix holds, namely being positive semi-definite (PSD). Therefore, we can harness measures of statistical divergence to derive a metric on the COVET matrices and quantify differences and similarities between niches. Although we can conceptually use any statistical divergence measure, metrics such as Kullback–Leibler (KL) divergence and Bhattacharyya<sup>30</sup> distance are too computationally intensive and lack interpretability.

**Distance between COVET matrices.** To meaningfully compare between niches, we cannot simply use the sum difference between two niche matrices  $E_i$  and  $E_j$ , as changing the cells' order would change the result (whereas there is no meaning to any given order). An intuitive way to quantify niche similarity is by finding the best matching of cells between niche matrices by solving the assignment problem<sup>81</sup>. Optimal transport (OT)<sup>82</sup> is a relaxed version of the assignment problem, where, instead of matching cells one to one, OT finds the best 'soft assignment' between cells. However, because this approach has no closed-form solution and does not scale to large datasets, we can use the closed-form solution of OT between covariance matrices, known as the Fréchet distance<sup>29</sup>, instead:

$$\Delta_{\text{Fréchet}}(E_i, E_j) = \text{Tr}(\Sigma_i) + \text{Tr}(\Sigma_j) - 2 \cdot \text{Tr}(\sqrt{\Sigma_i \Sigma_j}).$$

The Fréchet distance has time complexity of  $O(k^3)$  and is, thus, computationally intractable for large-scale datasets, which would require billions of pairwise computations between all niches. To speed up computation, we swap the matrix square root (MSQR) and product operation in the last term of the Fréchet distance and define the AOT distance as:

$$\Delta_{\text{AOT}}(E_i, E_j) = \text{Tr}(\Sigma_i) + \text{Tr}(\Sigma_j) - 2 \cdot \text{Tr}(\sqrt{\Sigma_i} \sqrt{\Sigma_j})$$

If  $\Sigma_i$  and  $\Sigma_j$  are commutative, this is no longer an approximation and  $\Delta_{\text{AOT}} = \Delta_{\text{Fréchet}}$ . Both the approximate and true Fréchet distance require  $O(k^3)$  operations between each pair of niches and  $O(n^2 k^3)$  to compute the full distance matrix; however, using the identity that for symmetric matrices,  $\text{Tr}(AB) = \sum_{\gamma, \delta} A_{\gamma\delta} \cdot B_{\gamma\delta}$ , we arrive at:

$$\begin{aligned} \Delta_{\text{AOT}}(E_i, E_j) &= \sum_{\gamma, \delta} (\sqrt{\Sigma_{i\gamma\delta}} \cdot \sqrt{\Sigma_{i\gamma\delta}} + \sqrt{\Sigma_{j\gamma\delta}} \cdot \sqrt{\Sigma_{j\gamma\delta}} - 2 \cdot \sqrt{\Sigma_{i\gamma\delta}} \cdot \sqrt{\Sigma_{j\gamma\delta}}) \\ &= \sum_{\gamma, \delta} (\sqrt{\Sigma_{i\gamma\delta}} - \sqrt{\Sigma_{j\gamma\delta}})^2 = \|\sqrt{\Sigma_i} - \sqrt{\Sigma_j}\|_2^2 \end{aligned}$$

Therefore, when working in square root space, we do not require any computationally extraneous matrix multiplication and many calculations of MSQR. Instead, we first calculate the MSQR of each COVET matrix, which is  $O(nk^3)$ , and then simply calculate pairwise (squared) Euclidean distance for a total time complexity of  $O(nk^3 + n^2 k^2)$ , which is substantially more efficient than  $O(n^2 k^3)$  for large  $n$ . For a given PSD matrix  $A$ , there could be many possible solutions  $B$  that fulfill the equation  $B^2 = A$ . Although this underdetermination is problematic, there is a unique symmetric PSD solution for the MSQR<sup>83</sup>. This solution can be found via spectral decomposition and reconstructing with standard square root of the matrix eigenvalues:

$$\sqrt{A} = \sum_i \sqrt{\lambda_i} v_i v_i^T,$$

where  $\lambda_i, v_i$  are the eigenvalues/vector of  $A$ .

Because AOT can be formalized as the  $L_2^2$  between MSQR of COVET matrices, it allows for direct use of any algorithm that is based on the squared Euclidean distance, such as UMAP, tSNE<sup>84</sup> and FDL<sup>58</sup>, clustering<sup>31</sup> and DC<sup>32</sup> analysis. We can simply compute MSQR of the COVET matrices, flatten the resulting matrices into one-dimensional (1D) vectors and apply the default implementations of all the mentioned algorithms. We can further leverage the squared Euclidean distance representation of the AOT metric and use computational accelerators designed to compute classical pairwise distances for additional speed gains.

We demonstrate that AOT is a good approximation by benchmarking against the true Fréchet distance and the Bhattacharyya distance, another common metric for distances between covariance matrices. Across various sizes of random sets of  $64 \times 64$  covariance matrices, we test the runtime to compute the 10 nearest neighbors matrix in

covariance space. As covariance matrices are PSD, to randomly generate  $n$  covariance matrices of  $64 \times 64$  elements, we first sample  $n$  random  $64 \times 64$  matrices (using the standard normal) and multiply each by its transpose, as a matrix Gramian is always PSD:

$$\Sigma_{i=1}^n = \{X_i \cdot X_i^T\}_{i=1}^n; X_i \sim N(0, I_{64 \times 64}).$$

We find that, whereas AOT produces accurate similarities, its runtime is at least an order of magnitude smaller than that of other metrics, with Fréchet and Bhattacharyya failing on sample sizes larger than 3,000 matrices due to out-of-memory error. Using a GPU implementation of kNN distance built for the Euclidean metric, which can be easily adapted for AOT, the spatial covariance metric is indeed scalable to massive datasets, taking less than 1 min to compute the kNN matrix between 100,000 samples (Extended Data Fig. 1a).

We observe accurate approximation on real COVET matrices, calculated from the eight nearest neighbors of the pharyngeal mesoderm cells from the seqFISH assay<sup>37</sup>, using the 64 most variable genes among the 350 imaged. Despite its efficiency, AOT does not sacrifice accuracy and concurs highly with Fréchet. We calculate the pairwise distance between the pharyngeal mesoderm COVET matrices according to Fréchet, AOT, Bhattacharyya and naive  $L_2$  between matrices. For each pharyngeal mesoderm cell, we find its  $k$  nearest neighbors for every metric and compute their Jaccard index with the Fréchet nearest neighbors. Across a wide range of  $k$ , AOT-based kNN is highly congruent with Fréchet kNN, whereas Bhattacharyya and naive  $L_2$  distances are not (Extended Data Fig. 1b). Qualitatively, using Fréchet, AOT and Bhattacharyya pairwise distances to compute two-dimensional (2D) embeddings and PhenoGraph clusters for the COVET matrices returned similar results (Extended Data Fig. 1c,d).

**Choice of  $k$ .** By default, we select  $k = 8$  neighbors to construct COVET, which usually captures the immediate niche of a cell, but the exact choice of  $k$  should reflect the data. For all datasets analyzed in this study, we kept the value of  $k$  at the default, demonstrating how finding the optimal  $k$  is not required to gain insights from ENVI and COVET. Still, given the computational efficiency of both algorithms, we recommend that users attempt a range of  $k$  values at different scales, such as 8, 20 and 50. Users can visualize the ENVI-learned COVET representations with AOT and choose the most appropriate scale for their biological question. We also implemented an option for COVET to be computed on all cells within a given radius, rather than constant number of neighbors, to account for differences in cell density within a tissue.

### ENVI algorithm

The ENVI algorithm integrates scRNA-seq and spatial transcriptomics data into a common latent embedding, in a manner that can infer spatial context for scRNA-seq and missing genes for spatial data. The core assumption of ENVI is that the interplay between a cell's phenotype and its microenvironment, as captured by the COVET matrix, empowers better data integration.

ENVI is grounded on autoencoder variational inference but diverges from previous work<sup>9,25,47</sup>. Although current methods only model the expression of genes included in both single-cell and spatial datasets, ENVI explicitly incorporates both microenvironment context for spatial data and expression of the full transcriptome for scRNA-seq data. In addition, ENVI contains two decoders: one for expression, which includes additional neurons that learn gene expression only from scRNA-seq data, and one to predict spatial context. Using these decoders, ENVI trains the VAE<sup>27</sup> to reconstruct both full transcriptome expression and spatial context from partial transcriptome samples.

To integrate scRNA-seq and spatial data, ENVI learns a common latent space for both data modalities by marginalizing the technology-specific effect on expression via a CVAE<sup>28</sup>. It achieves this by augmenting the standard VAE with an auxiliary binary neuron in

the input layers to the encoding and decoding networks representing each data modality. Integration is crucial, as each modality harbors technology-specific artifacts (Extended Data Fig. 2a). ENVI takes as input the scRNA-seq count matrix  $X_{sc}$  with  $n_{sc}$  cells and their full transcriptome of  $g_{sc}$  genes as well as counts of segmented cells from spatial transcriptomics matrix  $X_{st}$  from  $n_{st}$  cells and  $g_{st}$  imaged genes. The algorithm is agnostic to the method used to segment cells before input. It uses the spatial data to compute the COVET matrix for each cell and their MSQR to align with the AOT distance formulation.

Next, ENVI's conditional autoencoder builds a shared latent space for both data modalities. As the combined embedding must incorporate spatial context and full transcriptome information and must remove confounders relating to modality, we set the latent dimension to 512, substantially larger than standard VAEs in single-cell genomics, which usually contain around 10 neurons<sup>25,34,35</sup>. As input to the encoder, ENVI takes either spatial or scRNA-seq samples (the latter reduced to the subset of genes that have been imaged), along with the auxiliary neuron  $c$  having value 0 for the spatial data and 1 for scRNA-seq. The expression profile along with the auxiliary neuron are transformed into the latent variable  $l$  using the same encoding neural network, regardless of data modality:

$$l = \begin{cases} \text{Enc}(x_{st}, c = 0) & x_{st} \in X_{st} \\ \text{Enc}(\bar{x}_{sc}, c = 1) & \bar{x}_{sc} \in X_{sc}[:, g_{st}] \end{cases},$$

where the encoder returns two vectors,  $\mu_l$  and  $\sigma_l$ , which parameterize a Gaussian with diagonal covariance describing the posterior distribution of the latent. To calculate gradients through random samples, we use the reparameterization trick, which involves generating a sample from the standard normal  $\varepsilon \sim N(0,1)$  and describing the latent through a function of  $\varepsilon, \mu_l$  and  $\sigma_l$  and treating  $\varepsilon$  as a constant:

$$l \sim N(\mu_l, \sigma_l) \Rightarrow l = \mu_l + \varepsilon \cdot \sigma_l, \varepsilon \sim N(0,1).$$

Through the training process, our goal is to have the latent encode not only gene expression but also information about the spatial context of a given cell while removing confounding effects to allow transfer learning between modalities. This is achieved by optimizing a single latent space to accurately decode both the full transcriptome and COVET matrix for both data modalities, each missing one of these components. The requisite that the latent space be capable of decoding the spatial niche imbues sufficient spatial information into the latent space during training.

The latent of either modality, along with the appropriate auxiliary neurons, is fed into the 'expression' decoder network  $\text{Dec}_{Exp}$ . The loss function, calculated by comparing the activations in the output layer to the true expression profiles, needs to reflect the underlying distribution of each data modality. We use the negative binomial distribution to model scRNA-seq data, similarly to previous work<sup>25,36</sup>, as it suffers from overdispersion and dropout. During training, the scRNA-seq data provide transcriptome-wide expression; therefore, we can include genes whose expression was not provided to the encoder in the loss function, allowing our encoder to model genome-wide expression.

The negative binomial has two parameters per gene: the number of failures,  $r$ , and success probability,  $p$ . Thus, the output layer of the decoder consists of  $2 \cdot g_{sc}$  neurons, where the first  $g_{sc}$  neurons are the parameter  $r$  and the latter  $g_{sc}$  neurons are  $p$ , using the 'softplus' nonlinearity for  $r$  and the sigmoid function for  $p$  to keep it a valid probability:

$$p(\hat{x}_{sc} = k | NB(r, p)) = \frac{k+r-1}{r-1} (1-p)^k p^r$$

where

$$r, p = \text{Dec}_{Exp}(l, c = 1)[:, : g_{sc}], \text{Dec}_{Exp}(l, c = 1)[:, : 2g_{sc}]$$

We use the Poisson distribution to model FISH-based multiplexed imaging data due to its high molecular capture rate<sup>3</sup> and have the first  $g_{st}$  neurons in the output layer parameterize the per-gene rate parameter  $\lambda$  using 'softplus' nonlinearity to ensure that it is a valid rate value:

$$P(\hat{x}_{st} = k | \text{Pois}(\lambda)) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where

$$\lambda = \text{Dec}_{Exp}(l, c = 0)[:, : g_{st}]$$

A standard CVAE, in which all neural parameters are shared aside from the auxiliary neurons, is sufficient to simply integrate between scRNA-seq batches, as demonstrated in scArches<sup>34</sup>. However, to successfully integrate scRNA-seq and multiplexed FISH-based technologies, a single auxiliary neuron is not sufficient to regress out all biases. In ENVI, only the first  $g_{st}$  neurons of the output layer are shared by the two data modalities, whereas the rest are solely trained on the scRNA-seq data. These additional technology-specific parameters improve the ability of ENVI to regress out confounders from the latent embedding, beyond the auxiliary neuron.

Finally, ENVI includes an additional 'environment' decoder network  $\text{Dec}_{Env}$  whose role is to reconstruct the COVET from the latent, which can be trained from the spatial data. The output layer of the environment decoder has  $\frac{g_{spatial} \cdot (g_{spatial} + 1)}{2}$  output neurons parameterizing the lower triangular Cholesky factor. The Gramian matrix of the output layer is the mean parameter of a standard normal, reflecting our AOT distance, as the log likelihood of the standard normal is the  $L_2^2$  distance.

$$P\left(\hat{\Sigma}^{\frac{1}{2}} = \Sigma^{\frac{1}{2}} | N(L \cdot L^T, I)\right) = \left(2\pi^{-\frac{g_{spatial}^2}{2}}\right) \cdot e^{-\frac{1}{2} \|\Sigma^{\frac{1}{2}} - L \cdot L^T\|_F^2},$$

where  $L = \text{Dec}_{Env}(l)$ .

The output of the environment decoder is the MSQR of the COVET matrix, which is trained to minimize the  $L_2^2$  error with the MSQR of the true COVET matrix. Using the AOT metric in this manner involves computing the MSQR of the COVET samples during training, which can be computationally prohibitive. Instead, we first calculate the MSQR of all COVET matrices, which ENVI is directly trained to reconstruct.

We train ENVI simultaneously on samples from both spatial and single-cell datasets, using mini-batch gradient descent on the variational inference loss. With the learned ENVI model, we impute missing genes for the spatial data by treating the latent embedding of the spatial data as if it were from the single-cell data, using the single-cell auxiliary variable and parameterizing as a negative binomial instead of a Poisson. Conversely, we reconstruct spatial context for the single-cell data by applying the 'environment' decoder on its latent, as if it was the latent of the spatial data.

In more detail, we train ENVI to optimize the evidence lower bound (ELBO) with a standard normal prior on the latent, with the goal of increasing the likelihood of the observed data  $\{X_{sc}, X_{st}, \Sigma_{st}\}$  for the parameterization of their decoded distributions  $\{\hat{X}_{sc}, \hat{X}_{st}, \hat{\Sigma}_{st}\}$  while minimizing the KL divergence between the latent distribution and  $N(0, 1)$ :

$$L = \ln NB(X_{sc} | r, p) + \ln \text{Pois}(X_{st} | \lambda) + \ln N(\Sigma_{st} | \mu_{Env}, I) - \beta D_{KL}(N(\mu_l, \sigma_l) | N(0, 1))$$

To train ENVI to impute missing genes for the spatial data, we generate the latent embedding  $l_{st}$  by passing  $X_{st}$  through the encoder, and run the latent layer through the 'expression' decoder, but with the

inverse auxiliary neuron, as if the embedding came from scRNA-seq data:

$$\chi_{st}^{\text{imp}} = E[NB(r_{st}, p_{st})],$$

where

$$r, p = \text{Dec}_{\text{Exp}}(l_{st}, c = 1)[\cdot : g_{sc}], \text{Dec}_{\text{Exp}}(l_{st}, c = 1)[\cdot : 2g_{sc}].$$

Similarly, we reconstruct the spatial context for dissociated scRNA-seq samples by passing the scRNA-seq latent embedding  $l_{sc}$  through the ‘environment’ decoder:

$$\chi_{sc}^{\text{Env}} = E[N(\mu_{sc}^{\text{Env}}, 1)] \text{ where } \mu_{sc}^{\text{Env}} = L_{sc} \cdot L_{sc}^T, L_{sc} = \text{Dec}_{\text{Env}}(l_{sc}).$$

To allow flexibility in modeling technologies with different count distributions and molecular capture rates, we implemented the normal, Poisson, negative binomial and zero-inflated negative binomial (ZINB)<sup>85</sup> distributions, which can be chosen for either modality to reflect pre-processing steps or varying levels of noise or dropout. The rate or mean parameters ( $\lambda$  for Poisson,  $r$  for NB and ZINB and  $\mu$  for normal) must all be defined per cell and per gene and shared across the single-cell and spatial data. However, all other parameters can be chosen to be either per cell and per gene or simply per gene and can be either shared between technologies or made distinct.

By default, the encoder and two decoder networks consist of three hidden layers, each with 1,024 neurons. The latent embedding consists of 512 neurons, and the prior coefficient is set to  $\beta = 0.3$ . For small datasets whose total samples size is fewer than 10,000 cells, we recommend increasing the reliance on the prior and set  $\beta = 1.0$ . We train ENVI for two<sup>14</sup> gradient descent steps with the ADAM optimizer<sup>86</sup> with learning rate  $10^{-3}$  (lowered to  $10^{-4}$  during the last quarter of training steps) and a batch consisting of 1,024 samples, half taken from scRNA-seq and the other half taken from spatial data. To reduce computational complexity, we subset the scRNA-seq dataset to the union of the 2,048 highly variable genes and all genes included in the spatial dataset rather than the full transcriptome.

ENVI training is constant in both time and memory, whereas methods such as Tangram and NovoSpaRc scale quadratically with dataset size and cannot be GPU accelerated on datasets above a few thousand cells. We benchmarked the run times of ENVI, Tangram<sup>9</sup>, NovoSpaRc<sup>26</sup>, gimVI<sup>25</sup>, uniPort<sup>47</sup>, deepCOLOR<sup>48</sup> and Harmony<sup>49</sup> on scRNA-seq datasets of various sizes and on osmFISH, seqFISH, Xenium and MERFISH datasets. All models were trained with their default parameters using a single 12 GB GeForce RTX 2080 GPU, except Tangram, which produced an out-of-memory error above 10,000 cells and was trained with a CPU instead. Model training was stopped prematurely if it exceeded 5 h.

As expected, ENVI’s training time was consistently around 10 min regardless of dataset size (Extended Data Fig. 2b), and Harmony was also constant in time. gimVI runtime grew linearly with dataset size (the model trained for a predefined number of epochs over the datasets), and NovoSpaRc and Tangram were prohibitively slow on larger spatial and scRNA-seq datasets (they learn a cell-to-cell mapping between the spatial and single-cell datasets). We found that GPU acceleration is not possible for Tangram. deepCOLOR and uniPort were also substantially slower than ENVI at larger cell numbers.

### Evaluation of integration quality

Batch average silhouette width (bASW), introduced in a recent benchmarking of batch integration methods for scRNA-seq atlases<sup>41</sup>, evaluates latent integration based on mixing between batches and co-localization of similar cell types within the latent. In brief, bASW computes, for each cell type, how well-mixed batch labels are using the silhouette coefficient and returns the average across all cell types. By treating each modality as a different batch, we could use the bASW score to measure

the quality of ENVI’s learned latent. The latent of ENVI is large, consisting of 512 neurons; because silhouette coefficient is affected by the curse of dimensionality, we first compressed the ENVI latent to the top 10 principal components and computed bASW on them.

### Benchmarking imputation

We benchmarked ENVI gene imputation following previous approaches<sup>25,44</sup> that generate a test set of held-out genes using cross-validation and compared imputed and true expression using Pearson correlation and our spatially aware MSSl metric. We evaluated log expression and imputation profiles, with pseudocount 0.1.

Many algorithms use scRNA-seq data to impute missing genes in spatial transcriptomics data<sup>42,43,87–89</sup>. We compared ENVI to gimVI, Tangram and uniPort for their competitive performance<sup>44,47</sup>, NovoSpaRc for its use of spatial context and optimal transport for data integration, deepCOLOR<sup>48</sup> for its use of a deep generative model and Harmony<sup>49</sup> for its prevalence as a batch correction method<sup>50</sup>.

On the osmFISH dataset, which includes only 33 genes, we performed a full leave-one-out cross-validation by hiding every gene in the imaging panel individually and predicting its expression. On the seqFISH, MERFISH and Xenium datasets, which assay hundreds of genes, we used five-fold cross-validation, whereby the imaged gene set was divided into five random groups, and each model was tested on one withheld group after training on four. To appraise performance, we used a ‘relative’ one-sided *t*-test, as scores are paired across genes.

We benchmarked all models using their default parameters and instructions on all datasets:

- gimVI: We trained for 200 epochs with a batch size of 128 and latent dimension, per author recommendations (<https://docs.scvi-tools.org/en/stable/>), and parameterized spatial and scRNA-seq datasets with NB and ZINB distributions, respectively. To impute genes with the trained model, we followed manuscript instructions and trained a kNN regression model on the scRNA-seq latent and full transcriptome expression, setting  $k$  as 5% of cells in the single-cell dataset. We then applied the regression model on the spatial data latent to predict the expression of unimaged genes.
- Tangram: We trained for 1,000 epochs using default parameters (<https://github.com/broadinstitute/Tangram>). For osmFISH, seqFISH and Xenium datasets, we used the default ‘cells’ mode, and, for the much larger MERFISH atlas, we used the ‘cell-type’ mode, per the tutorial. We set the density prior to be uniform, as our spatial benchmark datasets are single-cell resolution. With the learned mapping, we used the ‘project\_genes’ function to impute genes from scRNA-seq onto the spatial dataset.
- NovoSpaRc: We followed the repository instructions (<https://github.com/rajewsky-lab/novosparc>), using an ‘alpha’ coefficient on a spatial location prior of 0.25 and smoothness parameter ‘epsilon’ of 0.005. To compute the scRNA-seq pairwise distance matrix, we used the union of the 2,048 most variable genes and all genes in the spatial dataset. For spatial datasets consisting of multiple samples, we trained a different model on each sample. Because NovoSpaRc does not scale well to large datasets, we reduced the MERFISH-related scRNA-seq dataset to a tenth of its size, sampling uniformly across each cell type. We applied the learned mapping to impute missing genes using the ‘tissue.sdge’ function.
- uniPort: We replicated tutorial instructions for integrating spatial and single-cell datasets (<https://uniport.readthedocs.io/>) by normalizing each dataset according to library size, log transforming counts, executing the ‘batch\_scale’ function, training the model for 30,000 iterations with a ‘lambda\_kl’ value of 5.0 and, finally, predicting the expression of hidden genes using the ‘predict’ function.



- deepCOLOR: We trained for 500 epochs using default parameters from the tutorial (<https://github.com/kojikoji/deepcolor>). deepCOLOR does not directly impute unimaged genes, so we multiplied the resulting mapping matrix with the scRNA-seq expression of the hidden genes to predict their expression.
- Harmony: We treated spatial and single-cell datasets as separate batches and integrated them using the default Harmony implementation in scanpy<sup>90</sup> (<https://scanpy.readthedocs.io/>). We only included genes from the scRNA-seq data that were also in the spatial data (and removed test genes) to produce Harmony embeddings from the principal components of the concatenated dataset. Mirroring gimVI's imputation procedure, we performed kNN regression on the Harmony embeddings to reconstruct expression of the manually hidden genes.

### Impact of data sparsity on ENVI

To validate ENVI's robustness to single-cell or count data sparsity, which can affect integration<sup>9</sup>, we benchmarked the full embryogenesis seqFISH and scRNA-seq data against random subsampling to 90% or 80% of counts according to the binomial distribution. For all three datasets, we performed five-fold cross-validation (see the 'Benchmarking imputation' subsection for details), finding that removing even 20% of counts does not greatly impact ENVI performance, which still surpasses Tangram on the full dataset (Extended Data Fig. 2c). Using a kNN ( $k = 5$ ) classifier trained to predict cell type from the scRNA-seq latent space, we assigned labels to the seqFISH data and measured balanced accuracy compared to the original assignment, finding that the ENVI latent space remains reliable upon downsampling; datasets with higher sparsity are only slightly less accurate (Extended Data Fig. 2d).

### FDL and DCs

FDL<sup>58</sup> and DCs<sup>32</sup> capture and visualize continuous trends in single-cell data<sup>19</sup>. We calculated FDL following the implementation in Van Dijk et al.<sup>91</sup>, by computing a kNN matrix (using default  $k = 30$ ), converting to an affinity matrix using an adaptive Gaussian kernel with width = 30 and  $k = 10$ , symmetrizing the matrix and using the ForceAtlas<sup>92</sup> function 'force\_directed\_layout' for visualization. DC computation followed a similar process to compute a data affinity matrix, except that we multiplied the affinity matrix by the inverse of its degree matrix to compute the normalized Laplacian. The eigenvectors of the Laplacian matrix, in order of eigenvalue magnitude, are the DCs.

### Applying ENVI to seqFISH embryogenesis data

We started with pre-processed data from the E8.75 mouse gastrulation study<sup>37</sup>, which included 351 genes measured with seqFISH (57,536 imaged cells), and paired it with E8.5 scRNA-seq data (12,995 cells) from a second study<sup>38</sup>. We further processed the scRNA-seq data by removing mitochondrial genes, genes expressed in less than 1% of cells, cells with library size greater than 33,000 (set manually to match the knee point) and cells annotated as 'nan' or representing doublets. To avoid confounding batch effects<sup>50</sup>, we used only the largest scRNA-seq batch (labeled '3'). For the seqFISH dataset, we used only the first of three imaged embryos ('embryo1'), removed cells with abnormally high total expression (threshold set manually to 600) and removed the gene *Cavin3*, which did not appear in the scRNA-seq dataset. For both datasets, we used cell type annotations provided by the authors and visualized the seqFISH data using spatial coordinates and the scRNA-seq data using a UMAP embedding (Fig. 2a). We also renamed several cell types to resolve nomenclature differences, including changing presomitic mesoderm to somitic mesoderm and splanchnic mesoderm to pharyngeal mesoderm.

We trained ENVI on the union of the 2,048 most variable genes in the scRNA-seq data, all seqFISH-measured genes, all *HOX* genes and several organ markers (Supplementary Table 1) using default parameters. We visualized the learned latent posterior of the seqFISH

and scRNA-seq datasets using UMAP and found that cell types tend to co-embed regardless of modality (Fig. 2b).

To test the imputation of unimaged canonical organ markers *Ripply3* (ref. 51) (lung), *Nkx2-5* (ref. 52) (heart) and intestine *Tlx2* (ref. 53) (intestine), we visualized their imputed z-scored, logged expression and thresholded values lower than 2, finding almost exclusive expression in the correct organ (Fig. 2f). To confirm that expression in the correct location at E8.75, before organ formation, we imaged each marker gene using whole-mount HCR (Fig. 2f; see the 'Whole-mount HCR' subsection). HCR produces per-gene three-dimensional (3D) images, which we oriented coronally to match the seqFISH data. We similarly trained gimVI and Tangram on the complete scRNA-seq and seqFISH datasets to impute *Ripply3*, *Nkx2-5* and *Tlx2* and visualized as for ENVI imputation, finding that ENVI imputation more closely matches the experimental data (Extended Data Fig. 3b).

### Spatial organization of emerging organs

At E8.5, scRNA-seq cell clusters correspond to primordial endodermal organs, ordered by where they will later emerge along the gut tube<sup>55</sup>. We identified organ-specific gene sets (Supplementary Table 2) by using the 'rank\_genes' function in scanpy<sup>91</sup> to apply a Wilcoxon test for differentially expressed genes in each organ in a reference scRNA-seq endodermal atlas<sup>55</sup>. Thymus and thyroid are not well delineated at this stage, so we collapsed them into a single thymus/thyroid label, and we assigned small intestine and colon cells to a single 'intestine' label to avoid inconsistencies, as the seqFISH tissue section does not include colon<sup>37</sup>.

We used PhenoGraph to cluster the scRNA-seq gut tube cells into 12 clusters and labeled clusters by best matching organ based on z-scored and logged expression of each gene set, averaged across all cells in that cluster. Most clusters are highly distinct, whereas some co-express several programs. We labeled clusters for which the (z-scored) ratio between the highest and second-highest expressed gene set is above 1.5 with the most highly expressed organ. To assign ambiguous clusters with ratios below 1.5, we inspected marker expression individually:

- Cluster5: Thymus/thyroid gene set expression is highest, but because lung marker *Ripply3* (ref. 51) and *Irx1* (ref. 93) expression is high (average z-score logged expression, 0.90) while thymus/thyroid marker *Nkx2-1* (refs. 52,94) is low (−0.15), we labeled Cluster5 as 'dorsal lung' (second-highest expressing organ).
- Cluster6: Dorsal lung gene set expression is highest, with pancreas a close second. Because the cluster has minimal *Ripply3* and *Irx1* expression (0.18) but is enriched for pancreas marker *Pdx1* (ref. 95) expression (0.43), we labeled Cluster6 cells as pancreas.
- Cluster7: Pancreas and liver gene set expression is highest and second highest, respectively. Due to high *Pdx1* expression (0.99) and low liver marker *Ppy*<sup>96</sup> expression (−0.12), we kept the pancreas label for this cluster.

We inferred COVET representations for the scRNA-seq gut tube cells using the trained ENVI model and then measured pairwise AOT distances between the conjoined set of seqFISH and scRNA-seq COVET matrices to generate UMAP embeddings and PhenoGraph clusters. The data generated seven COVET clusters (CC0–CC7), which are highly congruent with emerging organs in the scRNA-seq data, indicating their spatial delineation (Extended Data Fig. 4a): thymus/thyroid cells were assigned to CC0 (75%) or the spatially proximal CC1 (17%); dorsal lung cells were assigned to CC1 (52%) or CC0 (36%); ventral lung cells were assigned to CC2 (62%) or the highly related clusters CC1 (12%) or CC3 (19%); liver cells were assigned to CC2 (94%); pancreas cells were assigned to CC3 (58%) or the related cluster CC2 (26%); and intestine cells were assigned almost entirely to CC4–CC7, with only 1% assigned to CC3.

Gut tube cells in the seqFISH data were assigned organ labels via their COVET representations. We fit an AOT metric kNN classifier ( $k = 5$ ) on the scRNA-seq ENVI COVET matrices and their organ labels and used the classifier to assign budding organ labels to seqFISH COVET (Fig. 3a). Projecting labels back onto their seqFISH coordinates reveals the spatial pattern of organogenesis, from thymus/thyroid to the lung compartments, liver, pancreas and intestine and colon from anterior to posterior (Fig. 3b).

To calculate average COVET matrices predicted by ENVI (scRNA-seq organs) or measured directly (seqFISH), we compute the AOT average for the matrix set by calculating the matrix square of mean of their MSQR:

$$\widehat{\text{COVET}} = \left( \frac{1}{n} \sum_{i=1}^n \sqrt{\text{COVET}_i} \right)^2$$

Mean COVET matrices are highly congruent between the two datasets for both dorsal and ventral lung cells, although the scRNA-seq COVET matrices are slightly smoother as they were inferred by ENVI rather than measured (Fig. 3c). To find gene groupings, we performed hierarchical clustering on the 64 genes in each mean COVET matrix, finding that *Dlk1*, *Gata4*, *Gata5*, *Aldh1a2* and *Foxf1* covary in the ventral lung COVET but not in the dorsal lung, whereas *Tagln*, *Six3*, *Thbs1*, *T* and *Epcam1* exhibit the opposite pattern.

We generated clusters of each COVET matrix by plotting their average expression in cells near the anterior gut tube (fewer than 50 distance units away), but not the gut tube itself, and found that ventral niche genes are enriched in the pharyngeal mesoderm, whereas dorsal niche genes localize to brain and cranial mesoderm (Fig. 3d). As pharyngeal mesoderm is ventral to the gut, and brain and cranium are dorsal, the uniquely covarying genes in the COVET matrices allow us to reconstruct each lung compartment's spatial context.

We also assigned budding organs using integration methods that do not model spatial context (gimVI and Tangram) and computed ENVI without COVET to highlight the importance of explicit modeling of microenvironment:

- **gimVI:** We trained gimVI on the full embryogenesis scRNA-seq and seqFISH datasets using defaults in the 'Benchmarking imputation' subsection (10 latent dimension, 200 epochs, NB for spatial and ZINB for single cell). We took the subset of gut tube cells in each modality from the learned latent embedding of scRNA-seq and seqFISH data and similarly learned a kNN classifier ( $k = 5$ ) from the single-cell latent and organ assignment, using it to predict labels on the spatial latent.
- **Tangram:** Using parameters in the 'Benchmarking imputation' subsection (1,000 epochs, uniform density prior, 'cells' mode), we trained Tangram to learn a mapping matrix from scRNA-seq to spatial data. We subset the Tangram matrix to the mapping from scRNA-seq gut tube to seqFISH gut tube cells and re-normalized the columns to sum to 1. We transferred organ labels using Tangram's 'project\_cell\_annotations' function, which uses the subsetted mapping matrix to calculate the probability of each organ being assigned to each seqFISH gut tube cell, and we labeled according to the most probable organ.
- **ENVI without COVET:** We retrained ENVI to solely reconstruct gene expression profiles, excluding any COVET-related information. A kNN classifier ( $k = 5$ ) on the learned latent was used to transfer organ labels from the scRNA-seq gut tube onto seqFISH cells.

Due to the lack of gene vocabulary and small number of gut tube cells, other methods could not assign labels as reliably as ENVI: gimVI failed to delineate dorsal lung from thymus/thyroid cells and missed almost all liver cells, and Tangram's labeling lacked coherent spatial structure (Extended Data Fig. 4b). Without COVET, ENVI was unable to distinguish between ventral lung and liver, although its results most

closely resembled the COVET-based assignment and known organ organization.

### ENVI robustness to neighborhood size

The optimal number of neighbors used to construct COVET depends on dataset features and desired analysis (see the 'Spatial covariance representation' subsection), but ENVI is nevertheless robust to variations of this parameter. For the seqFISH dataset, we calculated COVET matrices with  $k = 6, 8, 10$  or 12 nearest neighbors (original,  $k = 8$ ) and retrained ENVI on each representation. For each of the four ENVI models, we assigned organ labels onto the seqFISH gut tube cells, again using a kNN classifier on COVET matrices in AOT space. Despite doubling neighborhood size and inherent stochasticity in training deep learning models with batch gradient descent, all versions reliably assigned cells to spatial context (Extended Data Fig. 5a). Although there are some differences, even the worst-performing mode ( $k = 6$ ), which mislabeled many dorsal lung cells as thymus/thyroid, is more accurate than competing methods (Extended Data Fig. 4b).

### AP polarity of developing spine and NMP cells

Spinal cord cells and their NMP precursors in the seqFISH data (total, 2,830 cells) span the embryo AP axis and make up a substantial fraction of cells in the scRNA-seq data (1,289 cells, 10% of total). To gauge whether ENVI can correctly map these cells and spatial trends along the AP axis, we first combined empirical seqFISH and ENVI-inferred scRNA-seq COVET matrices from spine and NMP cells and computed DCs via eigendecomposition of the Laplacian of the AOT kNN ( $k = 30$ ) graph in COVET space. We then compared to DCs of spatial coordinates of seqFISH spine and NMP cells, calculated using a kNN ( $k = 30$ ) graph with standard Euclidean distance, finding that pseudo-AP coordinates based on COVET DC are highly congruent with true AP coordinates based on seqFISH DCs (Fig. 4b and Extended Data Fig. 6a) and (logged) expression of known posterior and anterior genes (Fig. 4c,d).

We attempted to reconstruct pseudo-AP axes for gimVI and Scanorama. For gimVI, we used the model trained on the complete embryogenesis datasets and subset the learned gimVI combined latent to only the spine and NMP cells from the spatial and scRNA-seq datasets. We calculated the top three DCs from the latent embeddings and found that DC 2 was most correlated with true AP polarity (seqFISH spine and NMP cells),  $r = 0.76$ . Scanorama is designed for batch integration and uses mutual nearest neighbors to directly correct the gene expression count matrix and remove batch effect. Following scanpy instructions (<https://scanpy.readthedocs.io/en/stable/>), we applied Scanorama to produce integrated count matrices of the seqFISH and scRNA-seq spine and NMP cells. We computed DCs from the combined Scanorama-corrected scRNA-seq and spatial datasets and found that DC 3 is most correlated with true AP,  $r = 0.70$ . Unlike ENVI, both of these methods produced spine and NMP cells in the posterior with low DC values (Extended Data Fig. 6b). We note that, because DC order is arbitrary, we reversed any DC negatively correlated with the true AP. Tangram was excluded from this analysis as it does not calculate a combined embedding from which we can recover a pseudo-AP axis.

To assess the accuracy of pseudo-AP mapping, we ordered scRNA-seq spinal cells by pseudo-AP value and examined expression of canonical markers *Rfx4* (ref. 63) (anterior), *Hoxaas3* (ref. 61) (posterior) and *Hoxb7* (ref. 64) (posterior) (Fig. 4e). Gene expression values were logged and z-scored, and ordered profiles were smoothed with a first-order Savitzky–Golay filter with window size 128 for visual clarity.

To determine the quality of the pseudo-AP axis predicted for scRNA-seq spinal cells by each method, we calculated its correlation with the logged expression of known posterior genes *Hoxaas3*, *Hoxb5os*<sup>61</sup>, *Hoxb9* (ref. 60), *Hoxb7* and *Tlx2* (ref. 97) and anterior genes *Foxa3* (refs. 98,99), *Hoxd3* (ref. 59), *Hoxa2* (refs. 100–102), *Rfx4* and *Hoxd4* (Extended Data Fig. 6c), providing a quantitative recapitulation of pseudo-AP-ordered expression (Fig. 4e).

### Inferring *Sst* neuron cortical depth with MERFISH

We used the BICCN's 252-gene MERFISH primary motor cortex atlas<sup>40</sup> and its matching scRNA-seq reference<sup>70</sup> to demonstrate ENVI in a tissue-wide context. For the single-cell data, we removed cells lacking cell type annotations or labeled as doublets or low quality, leaving 71,183 cells across three samples, and removed genes that both (1) appear in less than 5% of cells and (2) are not in the MERFISH panel. For the MERFISH data, we included all 12 samples, for a total of 276,556 cells from 64 motor cortex slices, and we removed cells lacking a cell type label and the genes *Crispld2* and *Igf2*, as they were absent from the scRNA-seq data, but avoided any additional pre-processing. Both spatial and scRNA-seq datasets were labeled into neuronal and non-neuronal cell types. For brevity and consistency between datasets, we relabeled the MERFISH GABAergic neurons from 'Sst-chodl' to 'Sst' and collapsed the 'PVM', 'macrophage' and 'microglia' labels to 'microglia'.

We used five-fold cross-validation to benchmark ENVI imputation against Tangram, gimVI, NovoSpaRc, uniPort, Harmony and deep-COLOR with default parameters (see the 'Benchmarking imputation' subsection), except that we applied Tangram with 'cell-type' mode, which averages single-cell data per cell type, and ran NovoSpaRc independently for each slice, subsampling scRNA-seq data to 10% of each original size, because these methods do not otherwise scale to these data. ENVI MSSI and Pearson correlations were significantly higher than other methods (Extended Data Fig. 7a), and ENVI imputation of unimaged genes matches ISH from the Allen Brain Atlas (Extended Data Fig. 7b).

The full transcriptome information in scRNA-seq data allowed finer subtyping than the 22 cell types in the MERFISH dataset. Specifically, we further divided the *Sst* interneurons into nine subtypes and extracted gene sets for each subtype using the scanpy 'rank\_genes\_group' function. For the subset of MERFISH genes present in each gene set, we calculated average expression in every MERFISH *Sst* cell. We measured the pairwise correlation between gene sets within each modality and found that each subtype was delineated much more specifically in the single-cell data (Fig. 5b). We quantified this by computing the per-gene-set entropy across the pairwise correlation matrix, after normalizing with 'softmax'. The entropy for each gene set was higher in the MERFISH data, demonstrating the lack of distinction between subtypes.

To map the labeled scRNA-seq *Sst* interneurons to their cortical depth, we embedded ENVI-imputed scRNA-seq COVET matrices and MERFISH COVET matrices into DCs and FDL via a kNN graph ( $k = 100$ ) on AOT distance. The first COVET DC corresponds to pseudodepth and matches the cortical depth of MERFISH cells visualized on a single slice (Fig. 5c) and aligns with the primary axis of the COVET FDL (Fig. 5d). For each scRNA-seq *Sst* neuron, we predicted cortical depth using the pseudodepth axis (COVET DC 1), grouped the results by subtype and plotted their distribution (Fig. 5e).

### osmFISH imaging of somatosensory cortex

We applied ENVI to a 33-gene osmFISH dataset (4,530 cells, one sample) and complementary scRNA-seq dataset (30,005 cells) of the somatosensory cortex<sup>3</sup>, using the authors' cell type annotations and no additional processing besides removing genes expressed in less than 1% of cells in the scRNA-seq data (Extended Data Fig. 8a). As osmFISH data are more dispersed than MERFISH and seqFISH (Extended Data Fig. 2a), we modeled them with the negative binomial instead of the Poisson distribution. Due to the limited size of the scRNA-seq dataset, we changed its parameterizing distribution from NB to ZINB. Because the total sample size is small (fewer than 10,000 cells), we also increased the reliance on the prior latent distribution and increased the regularization to  $\beta = 1.0$ , which is common practice in Bayesian modeling.

We visualized and compared the ENVI and gimVI learned latent spaces with a UMAP embedding labeled by cell type annotations from

the osmFISH and scRNA-seq datasets. The ENVI embedding separates distinct cell types, with similar labels from the two data modalities occupying similar spaces (Extended Data Fig. 8b), whereas gimVI confuses oligodendrocytes and pyramidal neurons and cannot accurately co-embed osmFISH and scRNA-seq endothelial cells (Extended Data Fig. 8c).

We quantified integration quality and calculated the average center-of-mass embedding for each cell type, from both seqFISH and MERFISH datasets, in the gimVI and ENVI embedding spaces. ENVI and gimVI latent dimensions are vastly different in size (512 for ENVI compared to only 10 for gimVI), so we normalized each column in the pairwise distance to a maximum value of 1. In the ENVI latent, the center of mass for each osmFISH cell type is distinctly closer to its counterpart in the scRNA-seq data compared to other cell types, whereas cell types are less well separated in the gimVI latent (Extended Data Fig. 8d). For each cell in the scRNA-seq data, we quantified this as the ratio of its five osmFISH nearest neighbors in the latent space that share its cell type and averaged across the six cell types. The latent cell type agreement was 0.58 for ENVI and 0.38 for gimVI.

Using leave-one-out cross-validation, ENVI outperformed alternative methods on spatial imputation (Extended Data Fig. 8e). We further imputed the expression of three unimaged genes onto the osmFISH dataset using the full ENVI model (Extended Data Fig. 8f) and validated by comparing them to Allen Brain Atlas ISH images of the somatosensory cortex. ENVI imputation and ISH images both specify *Dti4l*, *Rprm* and *Ndst* expression in the L2/3, L5–L6 and CA1 regions, respectively. The Allen Brain Atlas provides both raw ISH images and processed, cell-segmented expression profiles. Because each view is difficult to interpret on its own, we overlaid the processed profiles on top of the raw ISH images for clarity.

### Xenium data analysis of LM

We assayed a slice of mouse brain bearing a LM of melanoma using snRNA-seq and Xenium (see the 'Generation of mouse melanoma LM FFPE-snRNA-seq and Xenium datasets' subsection). Raw Xenium imaging data were processed using the default pipeline provided by 10x Genomics<sup>16</sup> to produce a segmented cell-by-gene count matrix. In brief, nuclear segmentation was applied on DAPI stains, and all RNA molecules in each segmented mask and within a 15- $\mu$ m dilation were assigned to cells to compose a count matrix.

We further filtered the Xenium data by removing cells with library size less than 10 and more than 300 and kept only genes that were in the snRNA-seq data. For the snRNA-seq data, we only kept cells with library size less than 10,000 and removed mitochondrial genes and any gene expressed in less than 5% of cells, unless it was in the Xenium panel. Finally, we removed any doublets predicted by DoubletDetection<sup>103</sup> from either dataset, followed by median library size normalization on the snRNA-seq data. This process resulted in 243 genes captured in 74,132 cells in the Xenium dataset and 9,230 genes sequenced in 9,870 cells by snRNA-seq.

To assign cells to cell types, we independently clustered each dataset with PhenoGraph and searched for per-cluster marker genes using the scanpy 'rank\_genes\_groups' function. We first labeled Xenium data by neuron, endothelium, oligodendrocyte, tumor, astrocyte and immune/fibroblast groups. We then reclustered neurons and annotated into excitatory and inhibitory compartments according to expression of *Slc17a7* and *Gad1* and separated immune/fibroblast into immune cells and fibroblasts. The snRNA-seq data followed a similar hierarchical process, except that fibroblasts and immune cells were distinguished in the first round of clustering. According to an independently curated set of genes for each group<sup>75</sup>, our cell typing matched known transcriptional markers (Extended Data Fig. 9a). We benchmarked ENVI imputation against competing methods as for other methods (see the 'Benchmarking imputation' subsection), finding that ENVI outperforms all methods except for Harmony according to

Pearson correlation but does equally well according to MSS1 (Extended Data Fig. 9b).

To evaluate cell type label transfer from snRNA-seq to Xenium data for ENVI, gimVI, Harmony, deepCOLOR and uniPort, we fitted a kNN ( $k = 5$ ) classifier on the snRNA-seq latent to predict cell type labels and used it to assign labels to the Xenium data. Tangram does not use a latent, so we used its 'project\_cell\_annotations' function and labeled each Xenium cell according to the most probable snRNA-seq cell type mapped to it. NovoSpaRc does not assign cell type labels and was not compared. We measured transfer accuracy with balanced accuracy, the per-cell-type arithmetic mean of precision and recall, averaged across cell types (Extended Data Fig. 9c). ENVI transferred information as accurately as Tangram, uniPort and Harmony and was only slightly superseded by gimVI, validating our cell type annotation label transfer.

To uncover the relationship between phenotype and environment for each cell type in the Xenium dataset, we measured the agreement between clusters derived from expression (phenotype) and COVET representations (environment). Because non-parametric methods are biased by sample size, for each cell type we performed  $k$ -means ( $k = 5$ ) clustering on the logged expression of its cells and separately on its COVET representations. Each clustering was performed 10 times with random starting points. For each cell type, pairwise adjusted Rand index (ARI) was computed between each expression and COVET clustering, for a total of 100 values, and we reported their mean (Fig. 6c).

Unlike excitatory neurons, whose localization pattern is mirrored in their transcriptional profiles, the niche of immune cells in the Xenium dataset (canonically either brain-resident microglia or tumor-colonizing macrophages<sup>75</sup>) was not reflected in their gene expression. We attempted to divide the immune cells into macrophages or microglia (Supplementary Table 4) by computing the average logged expression of each cell type marker gene set in PanglaoDB<sup>75</sup> for every immune cell in the snRNA-seq and Xenium datasets, using only the subset of genes present in the Xenium assay (Fig. 6d). The high degree of overlap between macrophage and microglia genes in the spatial data may explain why, unlike the snRNA-seq data, expression and microenvironment corresponded poorly for immune cells.

We mapped annotated snRNA-seq immune cells to spatial context using the COVET predictions from ENVI. PhenoGraph clustering of snRNA-seq and Xenium immune COVET representations revealed major microenvironment clusters C0, representing immune cells in the cortex surrounded by excitatory neurons, with 80% of snRNA-seq cells annotated as microglia; C1, representing immune cells in the basal ganglia, dominated by inhibitory neuron environments, with 80% of snRNA-seq cells annotated as microglia; and C2, representing cells in and around the tumor, with 90% of snRNA-seq cells annotated as macrophages. These strong associations predict that macrophages are localized to the tumor and its boundary, whereas microglia localize mainly to basal ganglia and cortex, recapitulating the known tendency for brain tumors to recruit bone-marrow-derived macrophages<sup>76,77</sup>.

For further interpretability, ENVI can also invoke the inferred COVET representations and explicitly predict the microenvironment composition of each snRNA-seq cell. For each cell in the Xenium dataset, we counted the instance of each cell type within its  $k = 8$  nearest neighbor microenvironment, resulting in a  $n_{\text{Xenium}}$  by  $|C|$  matrix titled  $M$ , where  $|C| = 8$  is the number of distinct cell types. We then fit a kNN ( $k = 5$ ) regression model to predict  $M$  from COVET representations of the Xenium data. The trained model was applied to the COVET matrices that ENVI predicted for the snRNA-seq data to infer the distribution of cell types in each cell's niche. As for COVET-based clusters, macrophage niches predicted from the snRNA-seq data were highly enriched for tumor cells, whereas microglia niches contained more inhibitory neurons and oligodendrocytes (Fig. 6g).

ENVI can also be extended to identify markers of different macrophage types. Remsik et al.<sup>75</sup> identified *Ccr2*, *Ms4a4c* and *Lst1* as infiltrating monocyte markers based on cellular indexing of transcriptomes

and epitopes by sequencing (CITE-seq) analysis. ENVI imputation of these genes on Xenium immune cells is indeed specific to cells within the tumor (Fig. 6h).

Despite accurately transferring cell type information and imputing missing genes (Extended Data Fig. 9b,c), the absence of direct spatial modeling prevents gimVI and Harmony from reliably inferring subtype-specific microenvironments. We clustered gimVI embeddings of snRNA-seq and Xenium immune cells and found no obvious tumor-related cluster; 90% of snRNA-seq macrophages and 68% of microglia were assigned to gimVI cluster C1, prohibiting clear assignment of subtype to microenvironment (Extended Data Fig. 9d,e). Similarly, gimVI imputation of tumor infiltration genes did not distinctly enrich for immune cells within the tumor, and, despite outperforming ENVI according to Pearson correlation on imaged genes, Harmony also failed to accurately impute the expression of tumor-infiltrating markers (Extended Data Fig. 10).

## Experimental methods

**Whole-mount HCR.** Whole-mount HCR mRNA in situ was performed as described previously<sup>54</sup>, with minor modifications<sup>104</sup>. Mid-gestation embryos at E8.75 were treated with 10  $\mu\text{g ml}^{-1}$  proteinase K for 5 min at room temperature, followed by washing and post-fixation in 4% paraformaldehyde (PFA) for 20 min. Embryos were incubated in hybridization buffer supplemented with 2 pmol of each probe (*Ripply3*, *Nkx2-5* or *Tlx2*) overnight at 37 °C, followed by an amplification step with 60 pmol of each fluorophore-conjugated hairpin for 12–16 h at room temperature. Embryos were then stained with 0.5  $\mu\text{g ml}^{-1}$  DAPI (Thermo Fisher Scientific) and cleared using a modified Ce3D+ clearing protocol<sup>105</sup> for 24–48 h. Images were acquired on a Nikon AIR laser scanning confocal microscope with a  $\times 10$  objective and 3.0- $\mu\text{m}$  z-step size. Image rendering and optical sections were generated using IMARIS (version 9.9.0, BitPlane). All probes, hairpins and buffers were designed by and purchased from Molecular Instruments.

**Generation of mouse melanoma LM FFPE-snRNA-seq and Xenium datasets.** Animal studies were approved by the Memorial Sloan Kettering Cancer Center Institutional Animal Care and Use Committee under protocol 18-01-002. Mice were housed in specific pathogen-free conditions, in an environment with controlled temperature and humidity, on a 12-h light/dark cycle (lights on/off at 6:00/18:00), and with access to regular chow and sterilized tap water ad libitum. For this study, an 8-week-old female C57Bl/6-*Tyr*<sup>c2</sup> mouse (The Jackson Laboratory, 000058, albino C57Bl/6) was injected with 500 B16 LeptoM cells intracranially, as described in Remsik et al.<sup>74</sup>. Two weeks after the injection, the mouse was deeply anaesthetized and transcardially perfused with PBS (MSK Media Core). Tissues, including the brain, were dissected and immediately placed into a tube containing histology-grade PFA (4%; Sigma-Aldrich, HT501128). After overnight incubation, tissue was rinsed with water and submerged in 70% ethanol. The brain was cut coronally into four 2–3-mm-thick sections, placed into a tissue cartridge and embedded in formalin using routine, automated procedure. The embedded tissue was stored at room temperature.

For snRNA-seq-FFPE, a 100- $\mu\text{m}$ -thick section of tissue was pre-processed on a prototype Singulator system. The sample was automatically processed in a NIC+ cartridge (S2 Genomics, 100-215-389) by three 10-min deparaffinization steps (CitriSolv, VWR), rehydrated by successive 1-ml washes of 100%, 100%, 70%, 50% and 30% ethanol and followed by two washes of PBS. The sample was then spun at 1,000g for 3 min and resuspended in 0.5 ml of Nuclei Isolation Reagent (NIR, S2 Genomics, 100-063-396) with 0.1 U  $\text{ml}^{-1}$  RNase inhibitor (Protector, MilliporeSigma, 3335399001); all subsequent solutions had RNase inhibitor. The sample was dissociated to single nuclei in a second NIC+ cartridge with 2 ml of NIR for 10 min, followed by a 2-ml wash with Nuclei Storage Reagent (NSR, S2 Genomics, 100-063-405). The single-nucleus suspension was spun at 500g for 5 min, resuspended

in NSR and counted, and then snRNA-seq was performed on the Chromium instrument (10x Genomics) following the user guide manual for Chromium Fixed RNA Kit, Mouse Transcriptome (SinglePlex). Final libraries were sequenced on an Illumina NovaSeq S4 (R1: 28 cycles; i7: eight cycles; R2: 90 cycles).

To perform Xenium spatial profiling, FFPE mouse brain tissue adjacent to that used for snRNA-seq was sectioned into 5- $\mu$ m-thick slices with a microtome and placed onto the sample area of a Xenium slide (10x Genomics). Profiling was conducted following the 10x Genomics User Guide (CG000578, CG000580 and CG000582). In brief, tissue slices were baked at 42 °C for 3 h and stored overnight in a desiccating chamber. The tissue was then deparaffinized, serially rehydrated and de-crosslinked, before overnight hybridization with gene-specific padlock probes (Mouse Brain Panel, 10x Genomics). After this, the probes were ligated and amplified to generate the rolling circle amplification (RCA) product, which was then prepared for imaging with the Xenium. Before imaging, tissue autofluorescence was suppressed, and DAPI was applied as counterstain. The Xenium was loaded with the necessary reagents for decoding the RCA products, in conjunction with the selection of regions of interest for imaging based on the DAPI images captured by the Xenium.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Raw sequencing data and processed count matrices for snRNA-seq from brain tissue bearing a leptomeningeal metastasis are publicly available in the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE246395. Segmented and processed Xenium data are publicly available through Zenodo (<https://zenodo.org/>) under accession number 10712720.

### Code availability

ENVI and COVET are available as Python packages at <https://github.com/dpeerlab/ENVI> and can be directly installed via 'pip' with the command 'pip install scENVI'. A Jupyter notebook with an ENVI tutorial that reproduces motor cortex MERFISH results is available at [github.com/dpeerlab/ENVI/blob/main/MOP\\_MERFISH\\_tutorial.ipynb](https://github.com/dpeerlab/ENVI/blob/main/MOP_MERFISH_tutorial.ipynb).

### References

79. Andelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J. & Ogden, J. M. Pyramid methods in image processing. *RCA Engineer* [https://persci.mit.edu/pub\\_pdfs/RCA84.pdf](https://persci.mit.edu/pub_pdfs/RCA84.pdf) (1984).
80. Loukas, A. Graph reduction with spectral and cut guarantees. *J. Mach. Learn. Res.* **20**, 1–42 (2018).
81. Kuhn, H. W. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **2**, 83–97 (1955).
82. Villani, C. *Optimal Transport* (Springer, 2009).
83. Horn, R. A. & Johnson, C. R. *Matrix Analysis* (Cambridge Univ. Press, 2012).
84. Van Der Maaten, L., Courville, A., Fergus, R. & Manning, C. Accelerating t-SNE using tree-based algorithms. <https://www.jmlr.org/papers/volume15/vandermaaten14a/vandermaaten14a.pdf> (2014).
85. Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S. & Vert, J.-P. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.* **9**, 284 (2018).
86. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *Proc. 3rd International Conference on Learning Representations (ICLR)* (ICLR, 2015).
87. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
88. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).
89. Shengquan, C., Boheng, Z., Xiaoyang, C., Xuegong, Z. & Rui, J. stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics* **37**, i299–i307 (2021).
90. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
91. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729 (2018).
92. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **9**, e98679 (2014).
93. Becker, M. B., Zülch, A., Bosse, A. & Gruss, P. *Irx1* and *Irx2* expression in early lung development. *Mech. Dev.* **106**, 155–158 (2001).
94. Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N. & Izumo, S. The cardiac homeobox gene *Csx/Nkx2.5* lies genetically upstream of multiple genes essential for heart development. *Development* **126**, 1269–1280 (1999).
95. Offield, M. F. et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983–995 (1996).
96. Yang, Y., Akinci, E., Dutton, J. R., Banga, A. & Slack, J. M. W. Stage specific reprogramming of mouse embryo liver cells to a beta cell-like phenotype. *Mech. Dev.* **130**, 602–612 (2013).
97. Tang, S. J. et al. The *Tlx-2* homeobox gene is a downstream target of BMP signalling and is required for mouse mesoderm development. *Development* **125**, 1877–1887 (1998).
98. Monaghan, A. P., Kaestner, K. H., Grau, E. & Schütz, G. Postimplantation expression patterns indicate a role for the mouse *forkhead/HNF-3*  $\alpha$ ,  $\beta$  and  $\gamma$  genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* **119**, 567–578 (1993).
99. Kaestner, K. H., Hiemisch, H., Luckow, B. & Schütz, G. The HNF-3 gene family of transcription factors in mice: gene structure, cDNA sequence, and mRNA distribution. *Genomics* **20**, 377–385 (1994).
100. Gendron-Maguire, M., Mallo, M., Zhang, M. & Gridley, T. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. *Cell* **75**, 1317–1331 (1993).
101. Rijli, F. M. et al. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. *Cell* **75**, 1333–1349 (1993).
102. Barrow, J. R. & Capecchi, M. R. Compensatory defects associated with mutations in *Hoxa1* restore normal palatogenesis to *Hoxa2* mutants. *Development* **126**, 5011–5026 (1999).
103. Shor, J. DoubletDetection: doublet detection in single-cell RNA-seq data. <https://github.com/JonathanShor/DoubletDetection> (2022).
104. Anderson, M. J., Magidson, V., Kageyama, R. & Lewandoski, M. *Fgf4* maintains *Hes7* levels critical for normal somite segmentation clock function. *eLife* **9**, e55608 (2020).
105. Li, W., Germain, R. N. & Gerner, M. Y. Multiplex, quantitative cellular analysis in large tissue volumes with clearing-enhanced 3D microscopy (Ce3D). *Proc. Natl Acad. Sci. USA* **114**, E7321–E7330 (2017).

### Acknowledgements

We thank C. Burdziak and E. Wershof for their insightful comments and the Single Cell Analytics Innovation Lab at Memorial Sloan Kettering Cancer Center for sample processing and protocol development related to brain metastasis data generation. This work was supported by National Cancer Institute (NCI) Cancer Center Support Grant P30

CA08748, NCI grant U54 CA209975, NCI grant R01 DK127821 (M.G.), NCI Human Tumor Atlas Network grant U2C CA233284 (D.P.), the Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation (D.P.) and the Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center at Memorial Sloan Kettering Cancer Center. D.P. is a Howard Hughes Medical Institute investigator.

### Author contributions

D.H. and D.P. conceived the study and algorithm design. D.H. implemented all algorithms used in this study and carried out testing, application and data analysis. D.H., M.G. and A.-K.H. carried out data analysis and interpretation of the embryogenesis data. M.G. collected the mouse embryogenesis HCR imaging data. D.H., J.R. and A.B. analyzed the leptomeningeal metastasis data. J.R. generated metastatic cell lines and performed animal experiments. N.P., J.B. and S.J. developed and optimized the automated nuclei extraction protocol from FFPE samples. C.N. and M.T. collected the Xenium and snRNA-seq data under the supervision of R.C. D.H., T.N. and D.P. wrote the paper. D.P. supervised the study.

### Competing interests

D.P. is a member of the scientific advisory board of and has equity in Insitro. A.B. holds an unpaid position on the scientific advisory board of Evren Scientific and is an inventor on patents 62/258,044,

10/413,522 and 63/052,139 filed by Memorial Sloan Kettering Cancer Center. A.B. and J.R. are inventors of provisional patent applications 63/449,817 and 63/449,823 filed by Memorial Sloan Kettering Cancer Center. J.B., S.J. and N.P. were employees of S2 Genomics during this work and own company stock. The other authors declare no competing interests.

### Additional information

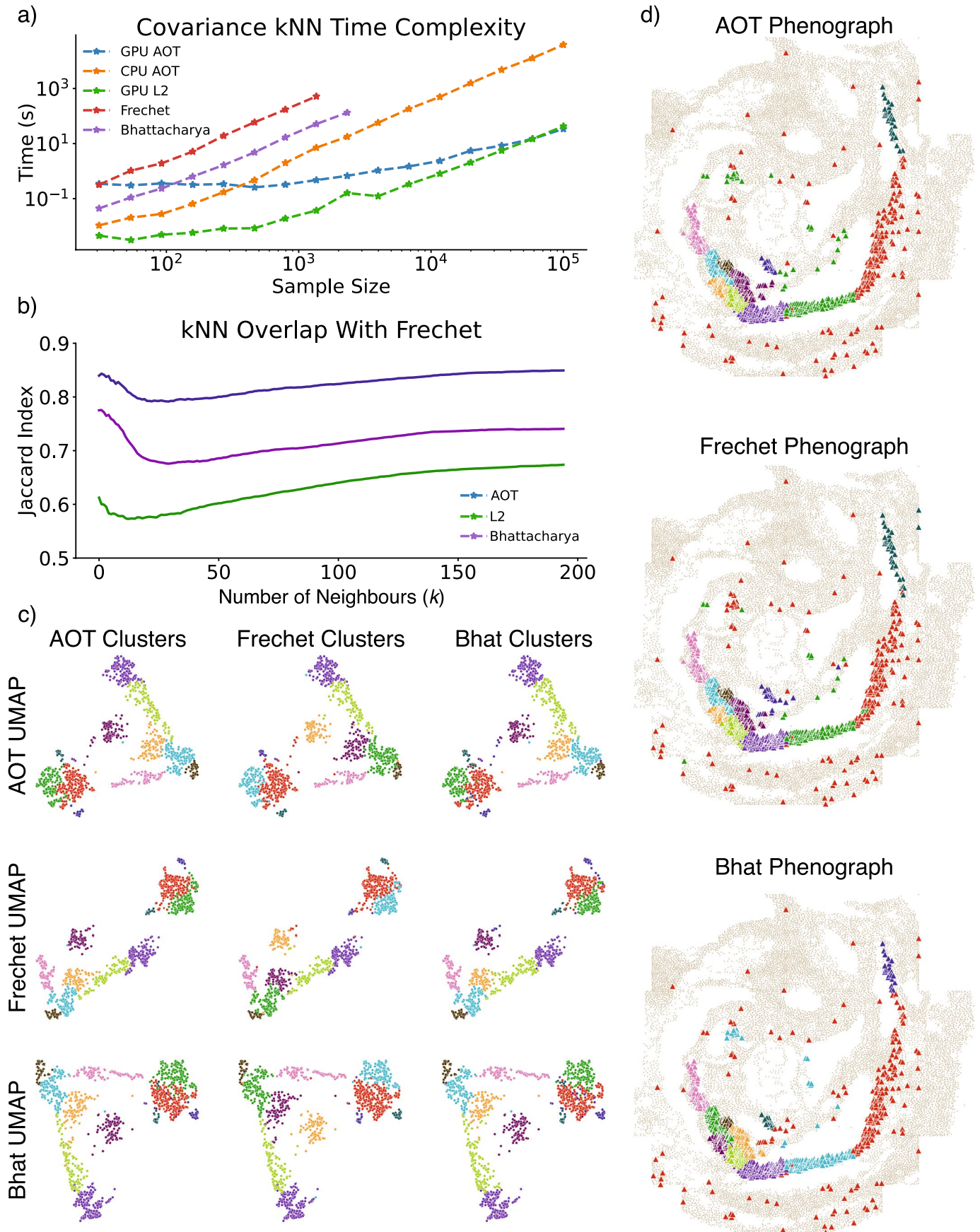
**Extended data** is available for this paper at <https://doi.org/10.1038/s41587-024-02193-4>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41587-024-02193-4>.

**Correspondence and requests for materials** should be addressed to Dana Pe'er.

**Peer review information** *Nature Biotechnology* thanks Tommaso Biancalani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

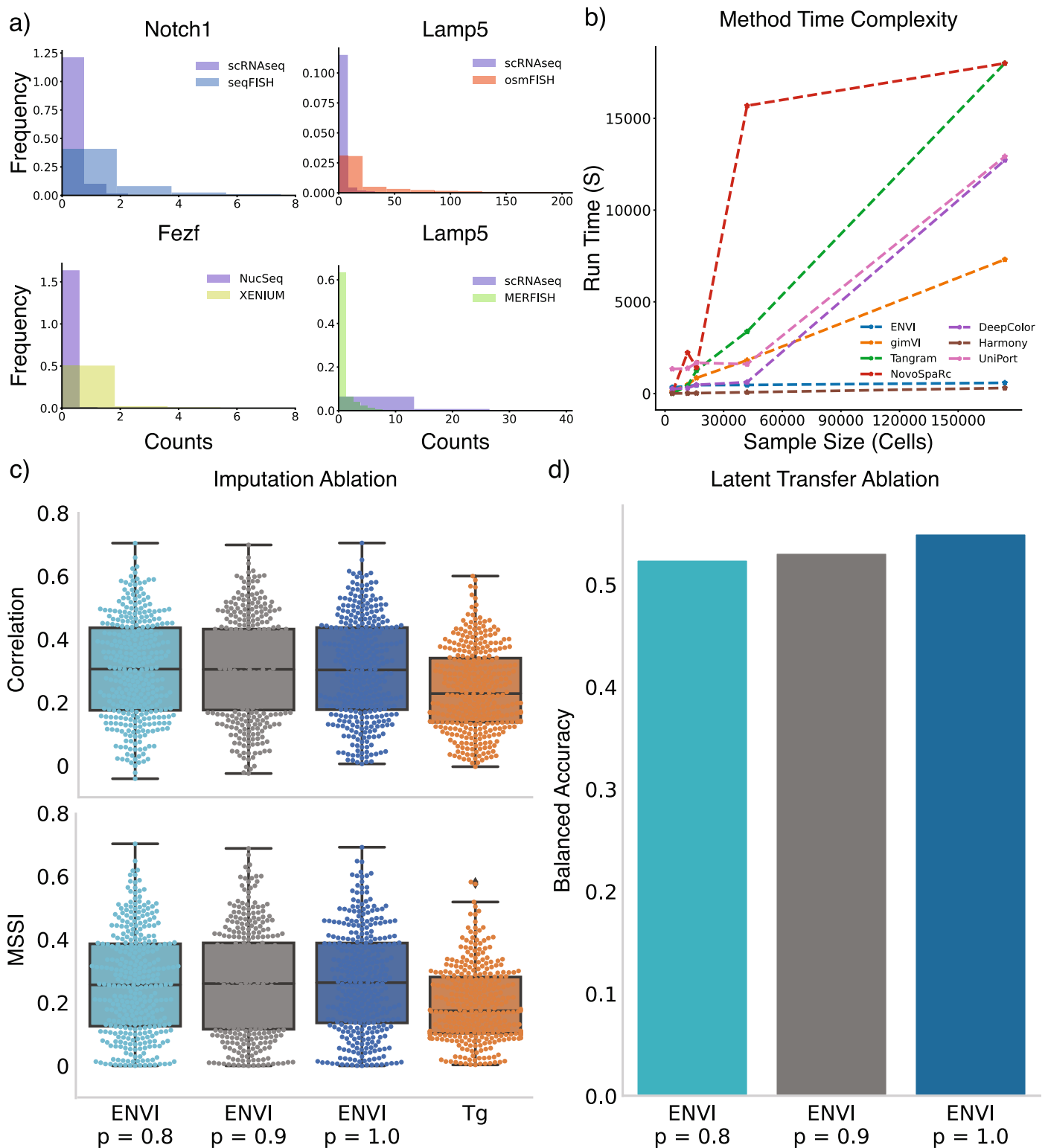


Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Approximate optimal transport (AOT) yields similar results to optimal transport and Bhattacharyya distance, but more efficiently. a**, Run times for computing the kNN graph between sets of randomly generated covariance matrices at various sample sizes. Both axes are in log scale. Fréchet and Bhattacharyya run times are not shown for samples larger than 4,000 cells due to out-of-memory error on a 768-GB, 64-core computing cluster. **b**, Agreement between the true Fréchet and AOT, Bhattacharyya and standard

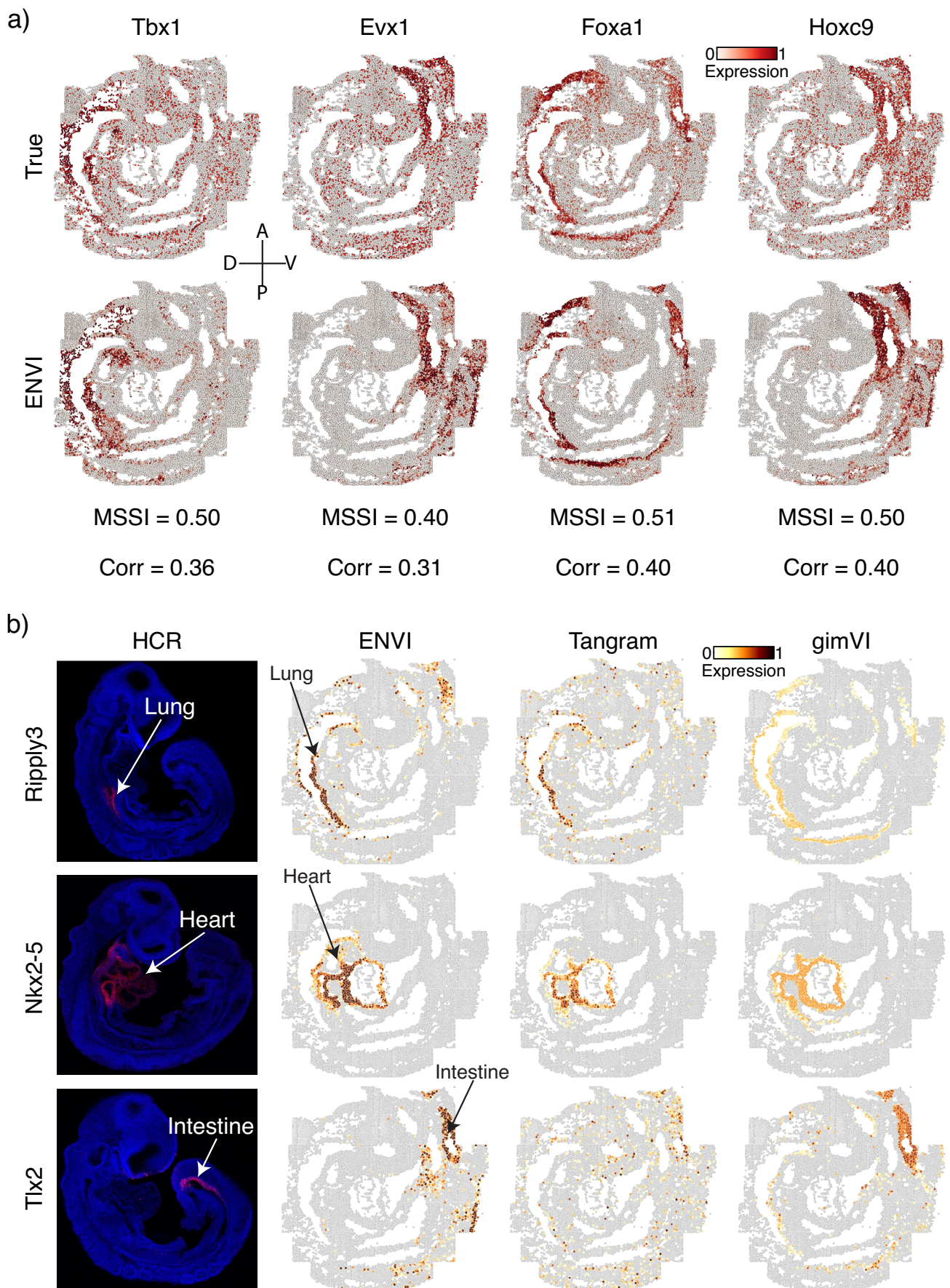
L2 kNN graphs, expressed as Jaccard Index values and computed on COVET matrices of pharyngeal mesoderm cells in the seqFISH embryogenesis dataset. **c**, COVET UMAP embeddings and PhenoGraph clustering of seqFISH pharyngeal mesoderm by different metrics, colored by PhenoGraph clusters of each. **d**, seqFISH data from pharyngeal mesoderm, colored by PhenoGraph clustering of COVET matrices according to each distance metric. Bhat, Bhattacharyya.





**Extended Data Fig. 2 | Modality-specific data features, run times and impact of sparsity on data integration.** **a**, Examples of three genes that exhibit very different expression distributions between four spatial datasets and their matching scRNA-seq data. **b**, Run time of spatial and single-cell integration methods on real datasets of different sizes. Programs were manually terminated at 5 h (18,000 s). **c**, Benchmarking ENVI imputation of the full embryogenesis

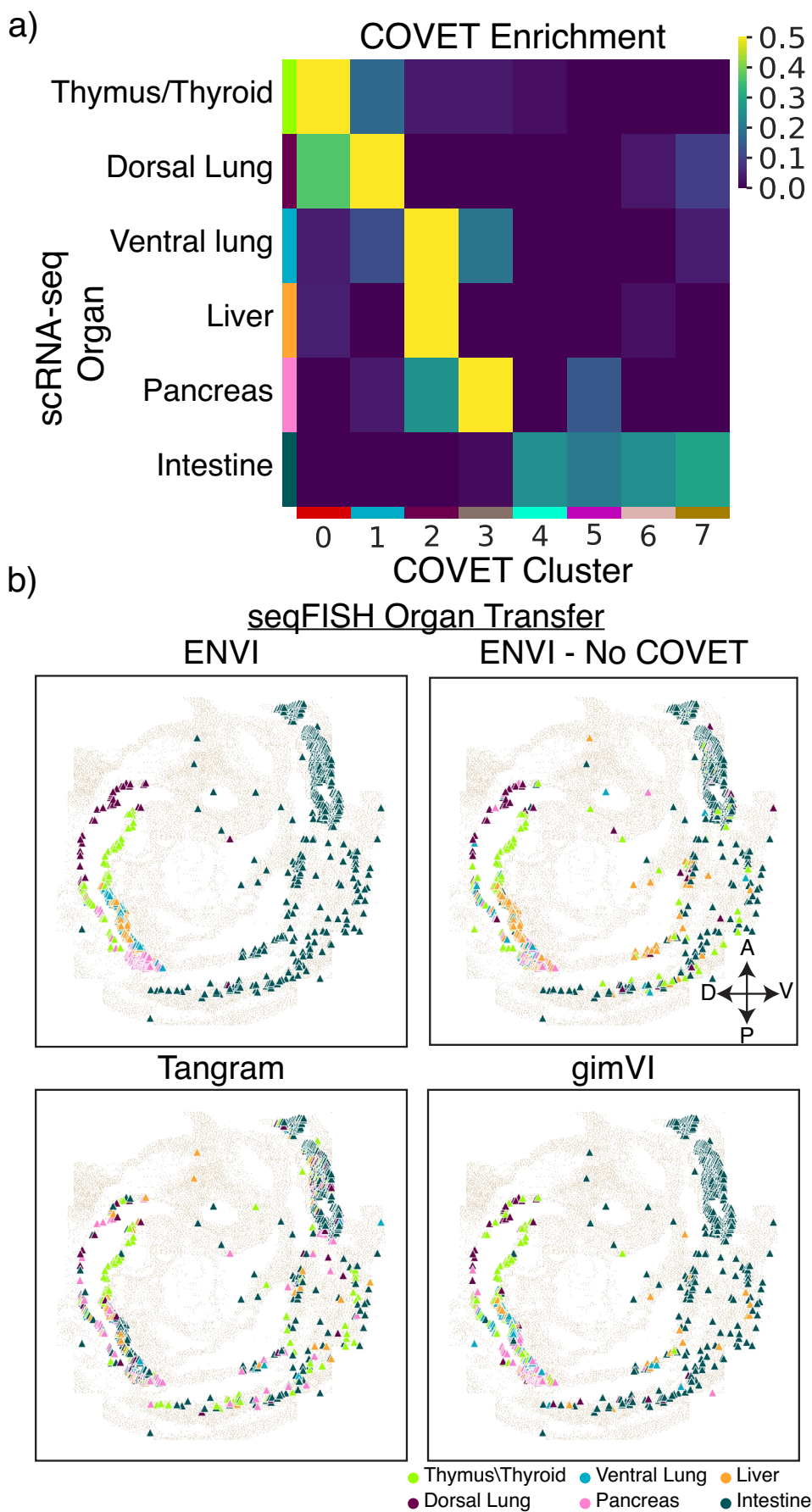
seqFISH and scRNA-seq dataset against 80% and 90% subsampled versions, as well as Tangram (Tg) on the full dataset as reference. Boxes and lines represent interquartile range (IQR) and median, respectively; whiskers represent  $\pm 1.5 \times$  IQR. **d**, Ability of ENVI to transfer cell-type label information from scRNA-seq to spatial data (datasets as in c). ENVI retains cell-type information in the integrated latent even starting from sparser data.



Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | ENVI accurately infers embryogenesis genes missing from seqFISH data. a,** Imputed expression of withheld genes from the seqFISH embryogenesis dataset<sup>37</sup> (bottom) compared to true (measured) expression (top), with corresponding MSI and Pearson correlation reconstruction (Corr)

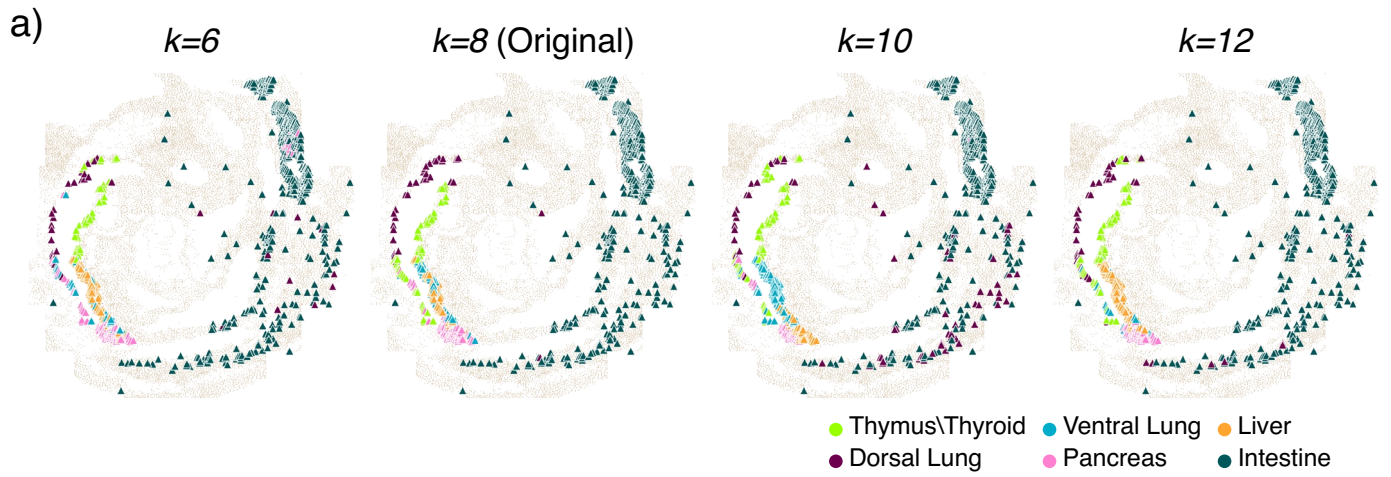
scores. **b,** HCR images of *Ripply3*, *Nkx2-5* and *Tlx2* and their imputation values according to ENVI, Tangram and gimVI. Organs marked by each gene are noted on the HCR and seqFISH images.



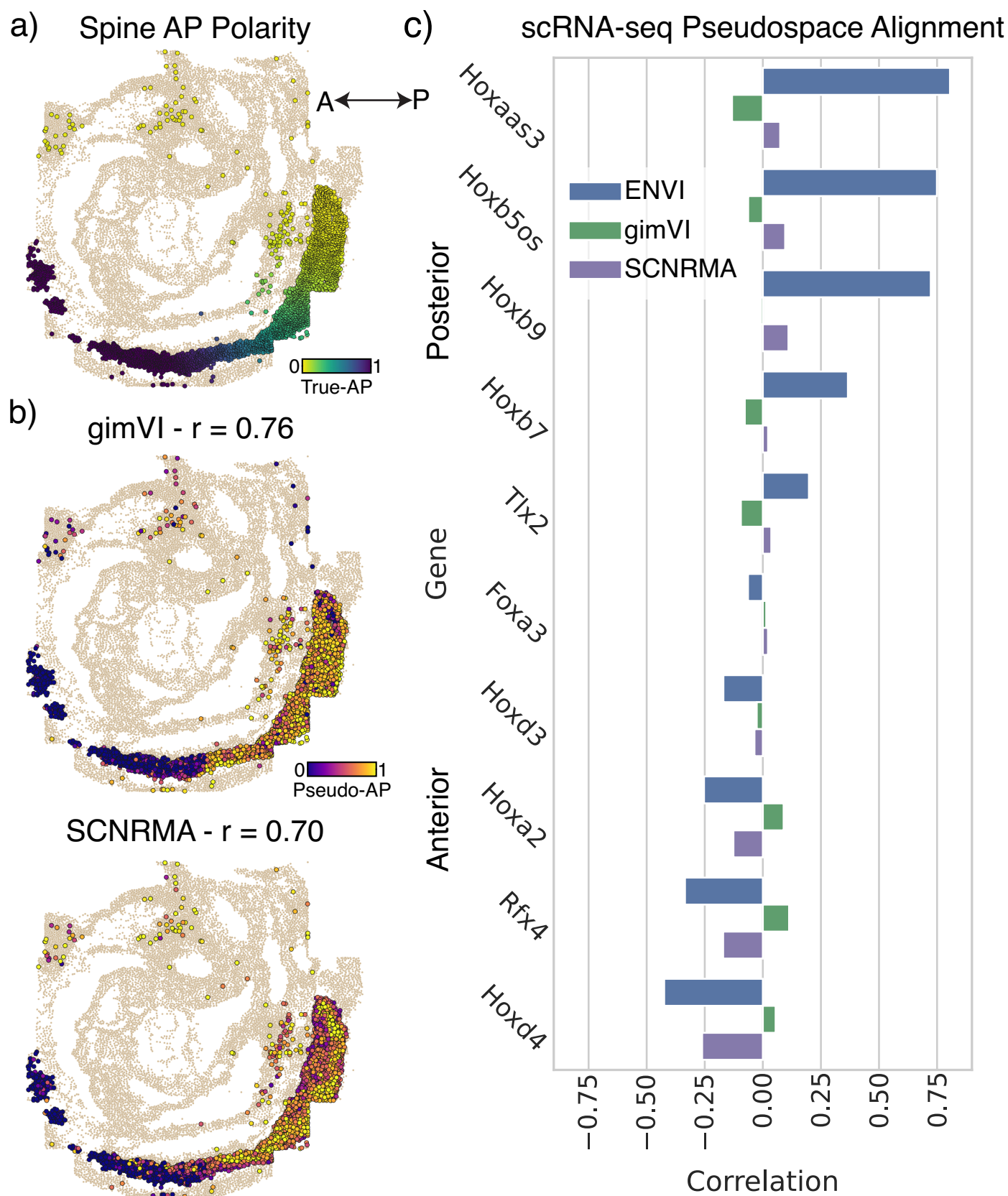
Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | The use of COVET spatial covariance improves cell-type assignment.** **a**, Proportion of scRNA-seq gut tube cells in each organ (row) that fall into each COVET cluster (column), arranged from anterior to posterior. **b**,

Assignment of developing organs to seqFISH gut tube cells via ENVI COVET space, latent space of ENVI when trained without COVET, gimVI latent space and Tangram cell-type mapping.

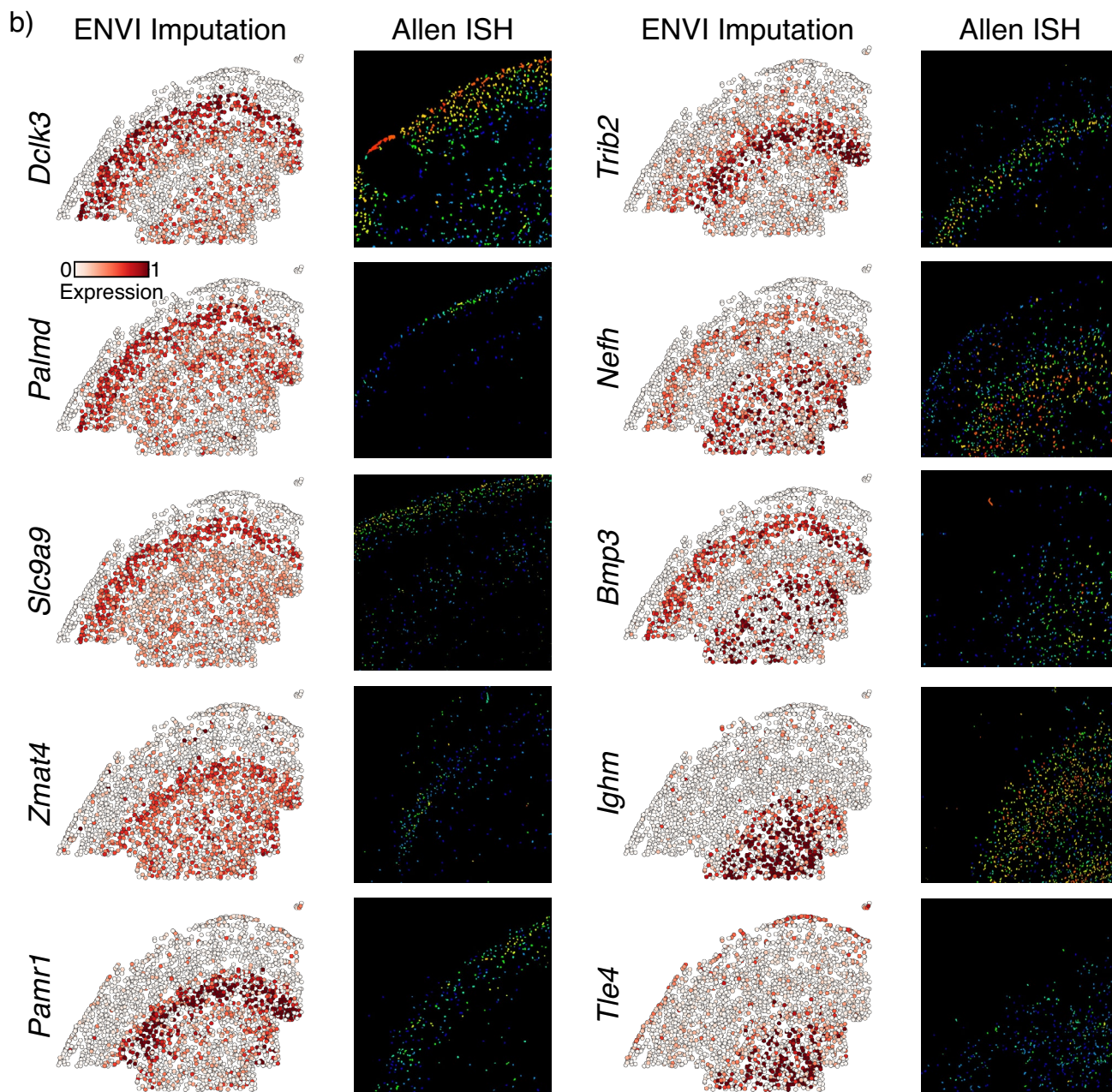
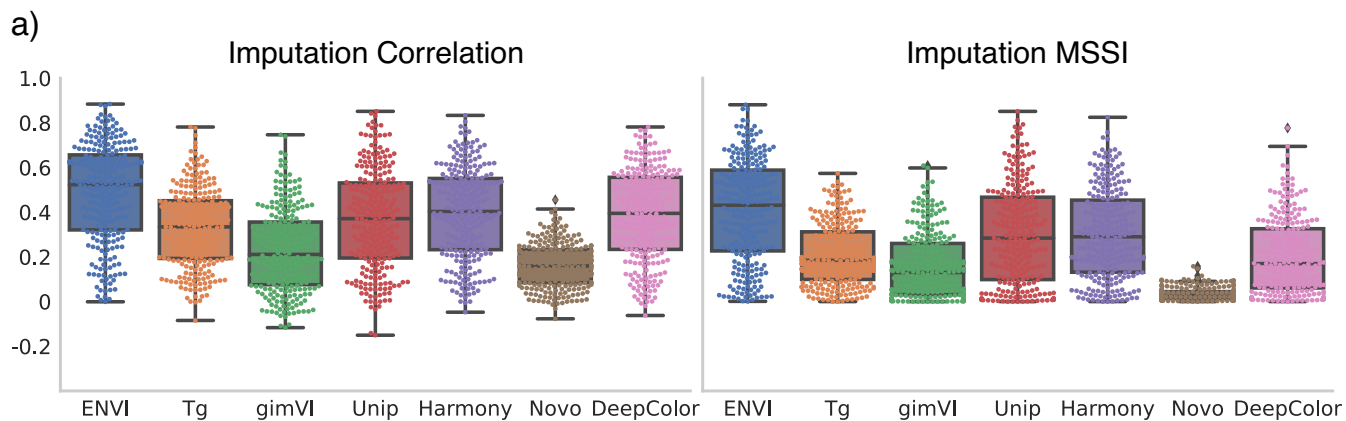


**Extended Data Fig. 5 | ENVI is robust to variation in COVET neighborhood size.** Transfer of organ labels onto seqFISH gut tube cells according to independent instances of ENVI, each trained according to COVET representations based on a different number of nearest spatial neighbors ( $k$ ). Spatial context predictions remain robust across  $k$  values.



**Extended Data Fig. 6 | ENVI reliably recovers the AP axis during spine development.** **a**, Spine and NMP cells from seqFISH data, colored by AP polarity calculated from the first DC of their spatial coordinates. **b**, Pseudo-AP of seqFISH spine and NMP cells from DC analysis of gimVI and Scanorama. Values denote Pearson correlation with the true AP axis. **c**, Pearson correlation of ENVI COVET,

gimVI and Scanorama pseudo-AP of spine and NMP scRNA-seq cells, for five canonical posterior markers (higher is better) and anterior markers (lower is better). Pseudo-AP axis is based on the DC best aligned with true depth (DC 1, 2 and 3 for ENVI, gimVI and Scanorama, respectively).

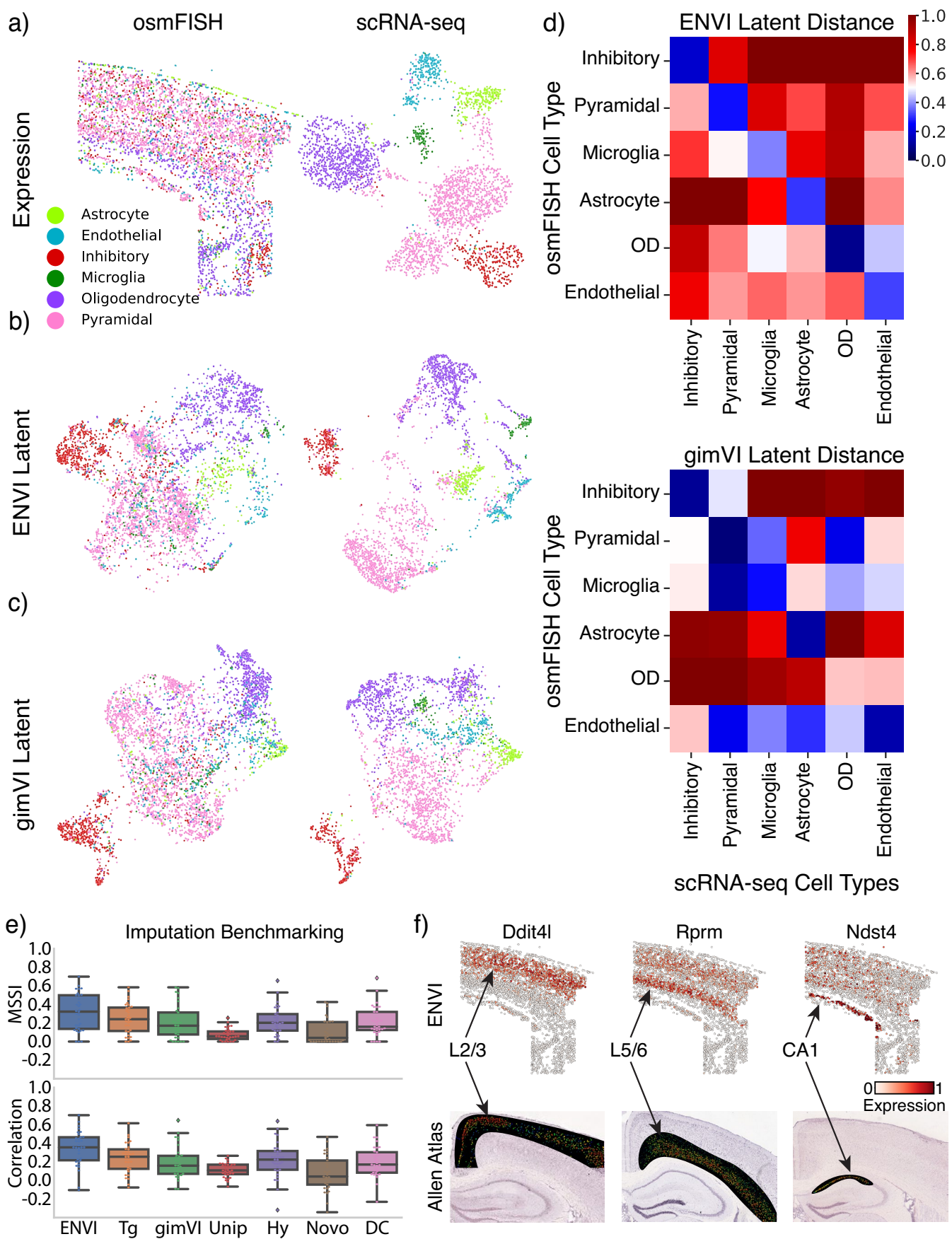


Extended Data Fig. 7 | See next page for caption.



**Extended Data Fig. 7 | ENVI extends cortical tissue gene expression to the entire genome. a**, Five-fold cross-validation of imputation based on a 252-gene MERFISH dataset from the primary motor cortex<sup>40</sup>. Boxes and lines represent IQR and median, respectively; whiskers represent  $\pm 1.5 \times$  IQR. Comparison of MSSI\Pearson correlations between ENVI and other methods (one-sided t-test,  $n = 252$ ) generates p-values, from left (Tangram) to right (deepC), of

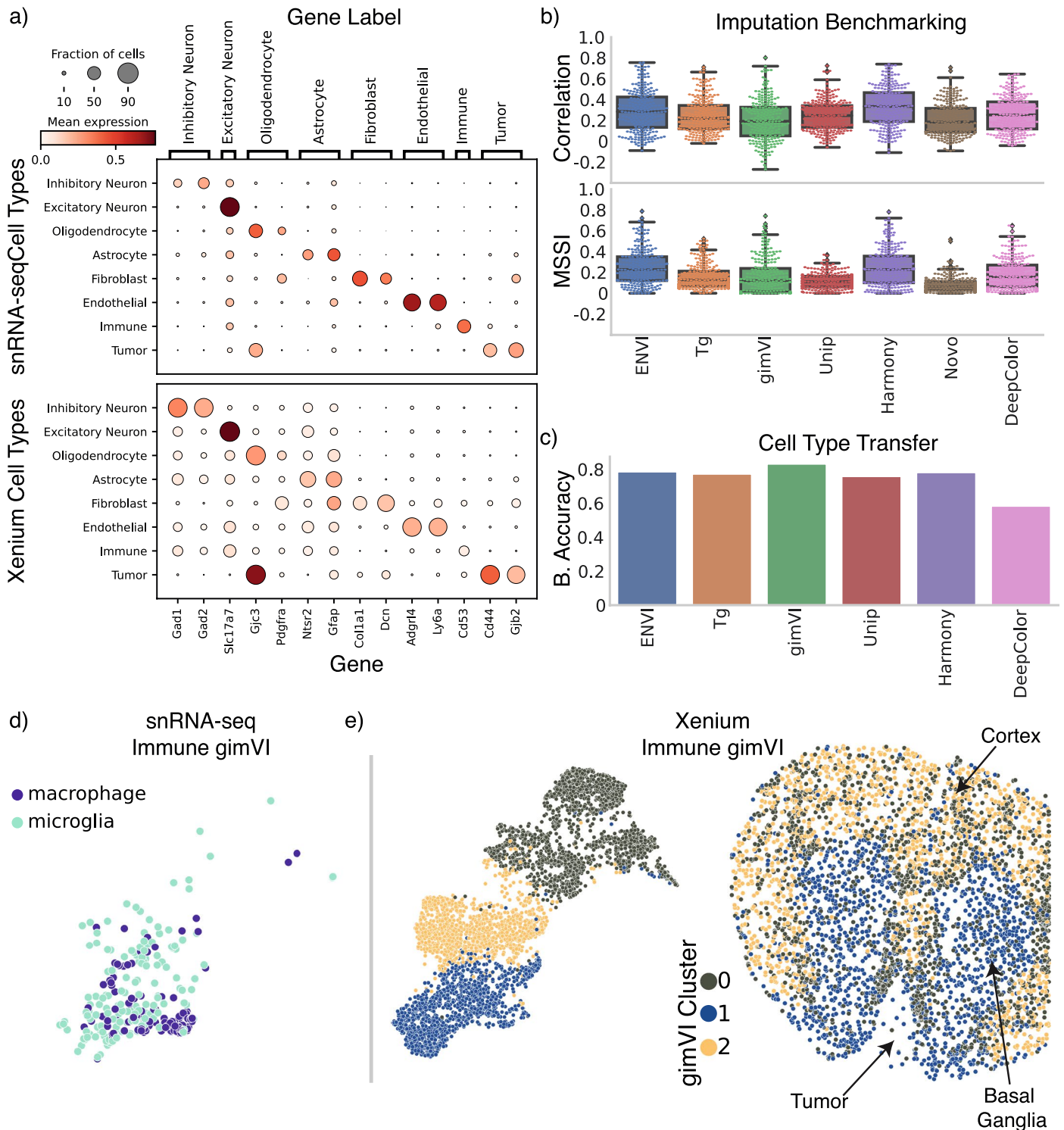
$4.67 \cdot 10^{-29}$ ,  $1.80 \cdot 10^{-19}$ ,  $1.62 \cdot 10^{-72}$ ,  $7.29 \cdot 10^{-66}$ ,  $4.70 \cdot 10^{-36}$ ,  $1.72 \cdot 10^{-50}$ ,  
 $5.25 \cdot 10^{-28}$ ,  $4.10 \cdot 10^{-41}$ ,  $5.23 \cdot 10^{-72}$ ,  $1.35 \cdot 10^{-86}$ ,  $1.87 \cdot 10^{-64}$ ,  $2.38 \cdot 10^{-66}$ .  
**b**, ENVI imputation of genes selected due to their clear in situ hybridization profiles in the Allen Brain Atlas (mouse.brain-map.org), projected onto the MERFISH data, with corresponding ISH expression in the motor cortex. Novo, NovoSpaRc; deepC, deepCOLOR.



Extended Data Fig. 8 | See next page for caption.

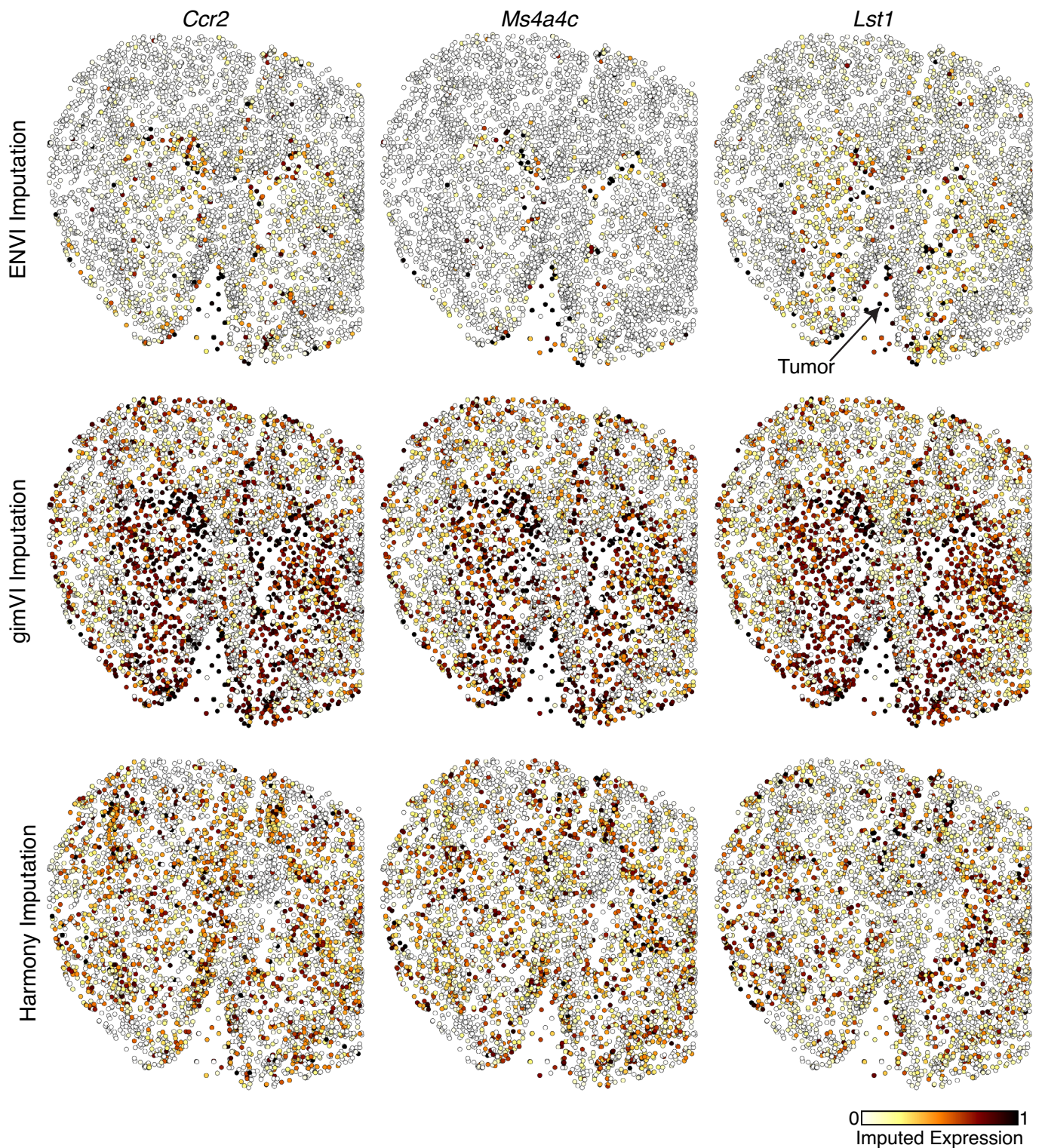
**Extended Data Fig. 8 | ENVI integrates osmFISH and scRNA-seq data from the somatosensory cortex.** **a**, osmFISH with segmented cells and UMAP visualization of scRNA-seq datasets of the mouse somatosensory cortex, colored by cell types as annotated in Codeluppi et al.<sup>3</sup> **b**, UMAP visualizations of the ENVI integrated latent embedding of the osmFISH and scRNA-seq modalities, colored by cell types as in **a**. Latent integration score; bASW = 0.62. **c**, Same as **b**, but with latent embeddings from gimVI. **d**, Normalized distance between the center-of-mass of each cell type according to the ENVI and gimVI latent embeddings. **e**, Benchmarking of imputation based on leave-one-out training, evaluated by

Pearson correlation and MSSl on a 33-gene osmFISH dataset of the somatosensory cortex<sup>3</sup>. Boxes and lines represent IQR and median, respectively; whiskers represent  $\pm 1.5 \times$  IQR. In order, MSSl \ Pearson correlation p-values (one-sided t-test,  $n = 33$ ) are:  $1.41 \cdot 10^{-4}$  \  $6.69 \cdot 10^{-10}$ ,  $9.03 \cdot 10^{-4}$  \  $2.64 \cdot 10^{-6}$ ,  $5.25 \cdot 10^{-9}$  \  $2.03 \cdot 10^{-10}$ ,  $1.21 \cdot 10^{-3}$  \  $1.62 \cdot 10^{-5}$ ,  $4.34 \cdot 10^{-10}$  \  $1.43 \cdot 10^{-9}$ ,  $7.11 \cdot 10^{-4}$  \  $9.08 \cdot 10^{-6}$ . **f**, ENVI-imputed expression of unimaged cortical markers *Ddit4l* (L2/3), *Rprm* (L5/6) and *Ndst* (Hippocampus, CA1) (top) and corresponding expression in the Allen Brain Atlas (mouse.brain-map.org) (bottom). Tg, Tangram; Hy, Harmony; Novo, NovoSpaRc; DC, deepCOLOR.



**Extended Data Fig. 9 | Cell type validation and extended benchmarking of ENVI on Xenium data.** **a**, Expression of marker genes for each annotated cell type in the snRNA-seq and Xenium data. **b**, Extended imputation benchmarking of ENVI against uniPort (Unip), Harmony and deepCOLOR. Boxes and lines represent IQR and median, respectively; whiskers represent  $\pm 1.5 \times$  IQR. In order, MSSI\ Pearson correlation p-values (one-sided t-test,  $n = 243$ ) are:  $1.21 \cdot 10^{-11}$ \  $1.88 \cdot 10^{-3}$ ,  $6.62 \cdot 10^{-16}$ \  $3.43 \cdot 10^{-13}$ ,  $1.20 \cdot 10^{-33}$ \  $5.82 \cdot 10^{-7}$ ,  $3.93 \cdot 10^{-1}$ \  $9.99 \cdot 10^{-1}$ ,  $2.66 \cdot 10^{-45}$ \  $4.89 \cdot 10^{-19}$ ,  $6.92 \cdot 10^{-12}$ \  $1.13 \cdot 10^{-4}$ . **c**, Balanced accuracy for

annotating Xenium cell types from snRNA-seq labels. ENVI, Tangram (Tg), gimVI, uniPort (Unip) and Harmony all perform similarly. **d**, UMAP of gimVI latent space of snRNA-seq immune cells, colored by subtype. **e**, gimVI latent UMAP and PhenoGraph clusters of Xenium immune cells. UMAP and clusters are calculated using both Xenium and snRNA-seq immune cells. Most microglia (68%) and macrophages (90%) are assigned to cluster C1, preventing clear association between subtype and microenvironment.



**Extended Data Fig. 10 | Imputation of tumor-infiltrating macrophage markers at the tumor-immune boundary.** ENVI, gimVI and Harmony-based imputation of three tumor-infiltrating macrophage markers onto Xenium immune cells. Only ENVI correctly predicts the expected pattern of expression, showing enrichment in immune cells inside the tumor region.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw sequencing data and processed count matrices for snRNA-seq from brain tissue bearing a leptomeningeal metastasis is publicly available in GEO ( <https://www.ncbi.nlm.nih.gov/geo/>) under accession GSE246395. Segmented and processed Xenium data is publicly available through Zenodo (<https://zenodo.org/>) under

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

No human data is included in our study

Population characteristics

*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*

Recruitment

*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*

Ethics oversight

*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample sizes (cell numbers) for snRNA-seq and Xenium datasets collected by us were determined based on filtering and processing by Cell Ranger and the Xenium pipeline from 10x Genomics. All other data in the manuscript was published by other groups and study designs are described in their respective publications.

Data exclusions

Data was not excluded. Only individual cells from single-cell and spatial samples were removed based on quality and library size. Full filtering details are available in the Methods.

Replication

ENVI and COVET were shown to be robust to data sparsity and neighborhood size based on replicated in silico ablation studies. In addition, we demonstrated consistency between ENVI imputation of missing genes onto spatial data, and ENVI inference of spatial context onto dissociated single-cell data.

Randomization

For benchmarking the gene imputation performance of different algorithms on spatial datasets with hundreds of genes, we performed k-fold (k = 5) cross validation. For each relevant dataset, genes were randomly divided into 5 groups of approximately equal size. Every algorithm was trained to reconstruct the expression of one group based on the remaining 4 groups, and accuracy was measured between reconstructed and true expression profiles.

Blinding

No blinding was necessary as our study is exploratory.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

## Methods

- n/a  Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern

- n/a  Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

- Laboratory animals
- Wild animals
- Reporting on sex
- Field-collected samples
- Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.