



# Super-resolved spatial transcriptomics by deep data fusion

Ludvig Bergenstråhle<sup>1</sup>, Bryan He<sup>1</sup>, Joseph Bergenstråhle<sup>1</sup>, Xesús Abalo<sup>1</sup>, Reza Mirzazadeh<sup>1</sup>, Kim Thrane<sup>1</sup>, Andrew L. Ji<sup>3</sup>, Alma Andersson<sup>1</sup>, Ludvig Larsson<sup>1</sup>, Nathalie Stakenborg<sup>1</sup>, Guy Boeckxstaens<sup>4</sup>, Paul Khavari<sup>1</sup>, James Zou<sup>1</sup>, Joakim Lundeberg<sup>1</sup>✉ and Jonas Maaskola<sup>1,5</sup>

**Current methods for spatial transcriptomics are limited by low spatial resolution. Here we introduce a method that integrates spatial gene expression data with histological image data from the same tissue section to infer higher-resolution expression maps. Using a deep generative model, our method characterizes the transcriptome of micrometer-scale anatomical features and can predict spatial gene expression from histology images alone.**

Experimental methods for spatial transcriptomics fall on a spectrum that trades resolution and molecular sensitivity for multiplexing capacity. On one end of the spectrum, methods based on *in situ* sequencing<sup>1,2</sup> or hybridization<sup>3–5</sup> typically have higher resolution and sensitivity but are more challenging to multiplex over many genes, limiting their usefulness in exploring transcriptome-wide interactions. On the other end, methods based on *in situ* RNA capturing (ISC) using poly(dT) probes<sup>6–8</sup> target all polyadenylated transcripts simultaneously but have lower resolution and sensitivity, limiting their usefulness in studying detailed expression patterns.

To overcome the limitations of current experimental methods, we propose a deep generative model of spatial expression data. Our method fuses ISC data with high-resolution histology images, typically in the form of hematoxylin and eosin stains, to infer transcriptome-wide super-resolved expression maps. The super-resolved expression maps quantify gene expression both within and between the original measurement locations (Fig. 1a). Moreover, we make no assumption about the number or configuration of measurements, thereby allowing prediction from standalone histology images in joint experiments with reference ISC sections from the same tissue (Fig. 1b).

The proposed method casts spatial gene expression and histological image data as observable effects of a latent tissue state. The latent state is modeled over multiple spatial resolutions, capturing both global and local anatomical features. Inference is amortized by a recognition neural network<sup>9,10</sup> that maps the image data to the latent state, allowing the method to scale to arbitrarily large datasets (Fig. 1c and Methods).

As a proof of principle, we construct synthetic expression maps consisting of shapes with three distinct transcriptional subtypes and use the maps to generate observed image and binned expression data (Fig. 1d and Methods). The proposed method classifies pixels into subtypes with almost perfect accuracy (average area under the curve = 1.00; Extended Data Fig. 1a–c). Moreover, errors are associated with higher predictive uncertainty (Fig. 1e).

We verify our findings in a dataset of 12 tissue sections from the mouse olfactory bulb<sup>6</sup>. First, we pool neighboring measurement locations into sets of two and predict their gene-wise expression difference (Extended Data Fig. 1d). We select observations for which one location has higher predicted expression than the other with 95% credibility and find directional correspondence to the ground truth in 95% of cases (Pearson's  $r = 0.82$ ; Extended Data Fig. 1e). Similarly to our finding in synthetic data, errors are associated with higher predictive uncertainty (Extended Data Fig. 1f).

Next, we compare inferred expression to *in situ* hybridization data from the Allen Mouse Brain Atlas<sup>11</sup>. Overall, inferred expression matches the reference data (Fig. 1f and Extended Data Fig. 2). For example, expression of *Ntng1* in the mitral cell layer (MCL) and of *Dusp14* in the MCL and granule layer are closely replicated. In contrast, the ISC data are too coarse to resolve the same expression patterns.

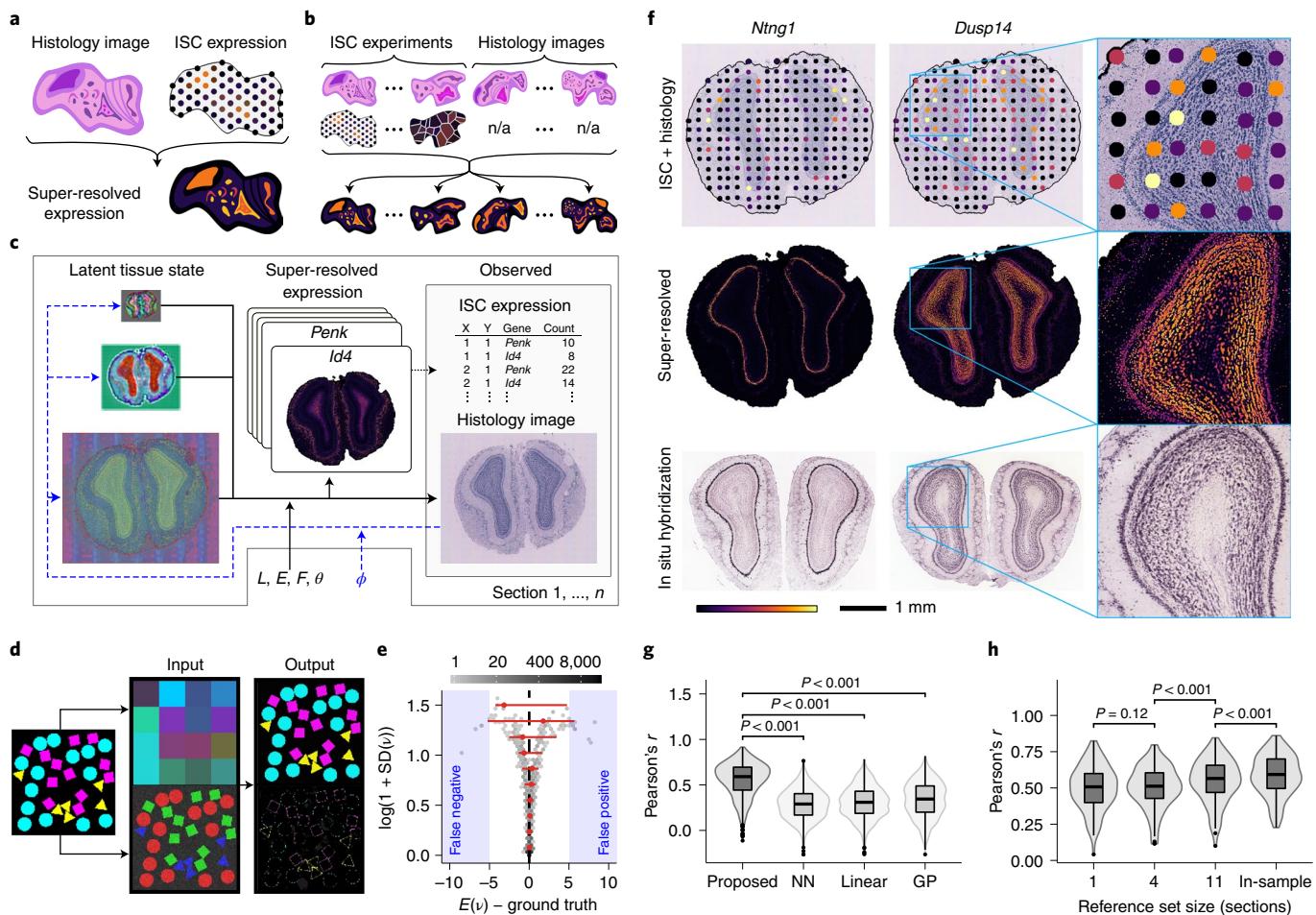
Together, these findings demonstrate that the proposed method decomposes ISC data into higher-resolution subcomponents. Additionally, the dispersion of the predictive distribution is well-calibrated and suitable for quantifying prediction uncertainty.

We benchmark prediction performance against baselines that act only on the expression data. We drop 50% of the spatial gene expression measurements from the mouse olfactory bulb dataset and impute the missing data. The proposed method attains a 0.25 higher gene-wise median Pearson correlation over the best-performing baseline (Fig. 1g and Methods).

To assess the ability of the method to predict expression from histology images, we infer the transcriptome of a holdout section in the mouse olfactory bulb dataset. Ground truth expression patterns are reproduced (Extended Data Fig. 3), and accuracy approaches in-sample performance as more sections are included in the reference set (Fig. 1h). We verify our findings in a human squamous cell carcinoma dataset of four serial tissue sections spaced 150  $\mu\text{m}$  apart (Methods). We predict the expression of the intermediate sections using the outer sections as reference experiments and find broad agreement with the ground truth (Pearson's  $r = 0.72$ ; Extended Data Fig. 4a), outperforming baselines constructed from gene-wise constant and image-based linear regression models (Extended Data Fig. 4b,c). Moreover, the prediction performance is stable over moderate differences in staining despite limited variation in the reference set (Extended Data Fig. 4d).

To assess the robustness of the method, we study a dataset of the human small intestine (Methods). In all experiments, we withhold

<sup>1</sup>SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. <sup>2</sup>Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. <sup>3</sup>Stanford Cancer Institute, Stanford University, Stanford, CA, USA. <sup>4</sup>Department of Chronic Diseases and Metabolism, Katholieke Universiteit te Leuven, Leuven, Belgium. <sup>5</sup>SciLifeLab, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden.  
✉e-mail: joakim.lundeberg@scilifelab.se



**Fig. 1 | Overview and performance evaluation.** **a**, ISC data are fused with histological image data to predict expression both within and between the original measurement locations, producing super-resolved expression maps. **b**, Multiple sections with different measurement configurations can be analyzed jointly, enabling shared feature learning and prediction from histology images without associated expression data. **c**, Model summary. The latent tissue state is mapped through a generator network (black solid arrows) to super-resolved expression maps and the observed histology image. The super-resolved expression maps are linked to the observed ISC data by summation (black dotted arrow). Inference is amortized using a recognition network (blue dashed arrows) that maps the observed histology image to the latent state. Shared variables  $L$ ,  $E$  and  $F$  and parameters  $\theta$  and  $\phi$  enable feature sharing across sections in joint experiments (Methods). **d**, **e**, Synthetic data experiments. **d**, Left: ground truth expression generated by randomly placing shapes of distinct transcriptional subtypes that emit 10 reads per pixel. Middle: observed binned expression (top) and histology image (bottom). Right: inferred pixel-wise posterior expression mean (top) and standard deviation (bottom). **e**, Pixel-level classification error against predictive dispersion. Blue areas indicate false classifications using a mean 5 reads per pixel classification threshold. Error bars indicate 5th, 50th and 95th percentiles in evenly distributed bins.

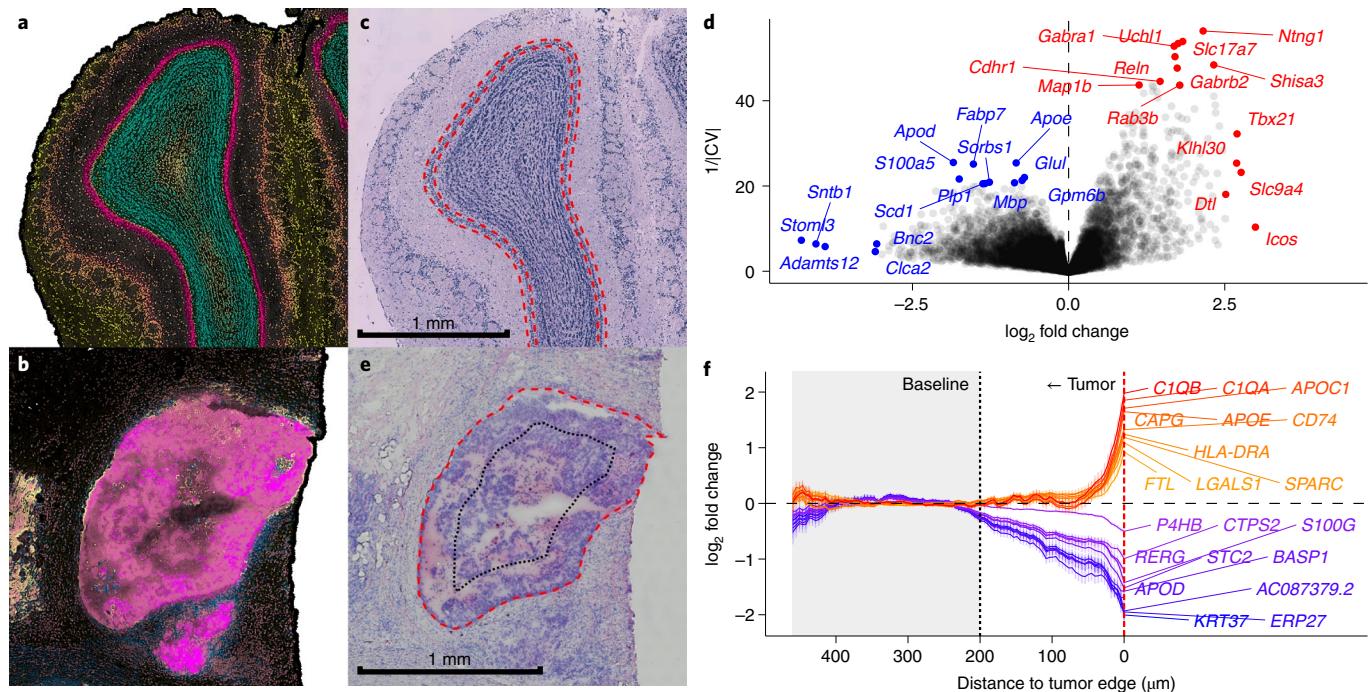
**f–h**, Mouse olfactory bulb experiments. **f**, Comparison among ISC, inferred super-resolved expression and in situ hybridization from the Allen Mouse Brain Atlas. Images show a representative sample of the 12 mouse olfactory bulb sections in the dataset. **g**, **h**, Prediction performance evaluated on the  $n=100$  highest-expressed genes in the mouse olfactory bulb dataset. **g**, Imputation of holdout measurements. Gene-wise Pearson correlation between ground truth and imputed expression over the test locations using the predictive mean (Proposed), nearest neighbor interpolation (NN), linear interpolation (Linear) and Gaussian process regression (GP). **h**, Prediction from histology image. Gene-wise Pearson correlation between ground truth and predictive mean over the test locations on a holdout section over different reference set sizes. Boxes show 25th, 50th and 75th percentiles. Outliers are represented by points and defined as observations further than 1.5 interquartile ranges from the hinges. Whiskers indicate the extent of all non-outlier observations. Pair-wise  $P$  values are based on two-sided Wilcoxon signed-rank tests. n/a, not applicable.

75% of the measurement locations for testing. First, we verify that predictions are consistent across restarts. We find an average correlation between the predicted mean expression vectors over all genes and test locations of 0.98 (Extended Data Fig. 5a). Disagreements are associated with higher prediction uncertainty, indicating that models are mutually consistent (Extended Data Fig. 5b).

Next, we test the sensitivity of the method to misaligned measurement locations by introducing increasing offsets to the training locations. At an offset of 110  $\mu\text{m}$ , there is a 0.040 absolute decrease in median gene-wise correlation between predicted and ground truth expression over the test locations (Extended Data Fig. 6a–c).

Compared to a reference antibody staining of CDHR5 from the Human Protein Atlas<sup>12</sup>, predictions broadly reproduce upregulation in the brush border and are only slightly disturbed when measurements are misaligned (Extended Data Fig. 6e–h). Additionally, we note that offsets of the investigated magnitudes are unlikely in practice.

Finally, we test the sensitivity of the method to histology artifacts by occluding parts of the image data. At a 20% occlusion rate, there is a 0.018 absolute decrease in median gene-wise correlation between predicted and ground truth expression over the test locations. Although expression maps are noticeably disturbed



**Fig. 2 | Characterization of the transcriptome in micrometer-scale anatomical features.** **a, b,** Summarized latent gene expression in the mouse olfactory bulb (**a**) and a DCIS lesion (**b**). Colors indicate anatomical areas with distinct transcriptional phenotypes according to the inferred tissue state (Methods). **c,** Annotation of the MCL profiled in **d**. **a** and **c** show a representative sample of the 12 mouse olfactory bulb sections in the dataset. **d,** Differential expression in the MCL compared to the other layers of the mouse olfactory bulb. **e,** Annotation of the DCIS lesion profiled in **f**. Red dashed line: tumor edge. Black dotted line: baseline boundary, 200  $\mu\text{m}$  from the tumor edge. **b** and **e** show close-ups of the DCIS lesion in a representative sample of the two breast cancer sections in the dataset. **f,** Differential expression compared to baseline as a function of proximity to the tumor edge for the ten most upregulated and downregulated genes at distance zero. Lines show posterior means, and ribbons indicate uncertainty ( $\pm 2$  s.d.). CV, coefficient of variation.

in occluded image regions, visible regions are only mildly affected (Extended Data Fig. 7).

We use the proposed method to study differential expression in micrometer-scale anatomical features of the mouse olfactory bulb and in human breast cancer (Fig. 2a,b and Methods). First, we profile the MCL of the olfactory bulb (Fig. 2c) and find several strongly upregulated and downregulated genes (Fig. 2d). Sorting the genes by the inverted coefficient of variation of their posterior log fold change, we find that 40 of the 100 most upregulated genes are among 229 markers for the MCL identified in a recent single-cell RNA sequencing study<sup>13</sup> (one-sided hypergeometric test  $P$  value =  $1.66 \times 10^{-47}$ ). However, the 20–50- $\mu\text{m}$ -wide MCL is difficult to isolate in the ISC data, which measures expression over 100- $\mu\text{m}$ -diameter areas (Extended Data Fig. 8a). Comparing ISC measurements overlapping with the MCL to non-overlapping measurements in a typical Seurat-based workflow<sup>14</sup> using either a conservative or a liberal selection strategy identifies only, at most, 19 markers (Extended Data Fig. 8b and Methods).

Next, we study spatial dynamics in a ductal carcinoma *in situ* (DCIS) lesion from the breast cancer dataset by profiling transcriptome gradients between the edge and inner area of the tumor (Fig. 2e) and characterizing its cell type composition (Extended Data Fig. 8 and Methods). Although carcinoma cells dominate all parts of the DCIS lesion (Extended Data Fig. 8d,e), we observe several tumor-related genes to be upregulated within 50  $\mu\text{m}$ , half the center-to-center distance between measurements, of the tumor edge (Fig. 2f). For example, the complement component 1q, composed of the C1QA, C1QB and C1QC subcomponents, has been shown to promote angiogenesis and tumor growth<sup>15</sup>. Similarly, CD74 is a known marker for metastatic tumor growth in breast cancer<sup>16</sup> and has been investigated as a potential target for antibody-drug

conjugate therapies in blood cancers. The proximity of CD74 expression to the tumor edge could have important implications for the accessibility of CD74-expressing cells in similar therapies for DCIS. However, further studies are needed to validate this finding.

In summary, we have presented a deep generative model for spatial data fusion. The proposed method combines ISC with histological image data to infer transcriptome-wide super-resolved expression maps. The expression maps resolve micrometer-scale expression signatures that are difficult to isolate in raw ISC data. Identifying and describing such signatures is vital for characterizing small anatomical features and developing effective treatments for disease states.

Moreover, the proposed method can predict spatial gene expression from histology images using reference ISC experiments from the same tissue, thereby providing a means for image-based *in silico* spatial transcriptomics (ISST). We envision future work to enable ISST on a larger scale, addressing the need to train tissue-specific models. Such models need to be able to flag out-of-distribution samples and train on databases spanning a wide range of anatomical conditions. In this context, transfer learning could be leveraged to source reference data from multiple technologies, including bulk and single-cell RNA sequencing. ISST could reduce research costs, unlock spatial gene expression in histology databases or be used as a quality control for experimental data.

#### Online content

Any methods, additional references, Nature Research 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/s41587-021-01075-3>.

Received: 12 March 2020; Accepted: 27 August 2021;  
Published online: 29 November 2021

## References

- Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
- Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
- Fernino, A. M. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
- Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
- Stähli, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
- Rodrigues, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
- Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
- Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. In *2nd International Conference on Learning Representations* (eds Bengio, Y. & LeCun, Y.) <http://arxiv.org/abs/1312.6114> (2014).
- Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31th International Conference on Machine Learning* <http://proceedings.mlr.press/v32/rezende14.html> (2014).
- Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2006).
- Uhlen, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
- Tepe, B. et al. Single-cell RNA-seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. *Cell Rep.* **25**, 2689–2703 (2018).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Bulla, R. et al. C1q acts in the tumour microenvironment as a cancer-promoting factor independently of complement activation. *Nat. Commun.* **7**, 10346 (2016).
- Metodieva, G. et al. CD74-dependent deregulation of the tumor suppressor scribble in human epithelial and breast cancer cells. *Neoplasia* **15**, 660–668 (2013).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## Methods

**Statistical model.** We model the spatial expression data  $X_n$  and histological image data  $I_n$  of each section  $n$  as effects of an underlying spatial tissue state  $Z_n$ . We assume the conditional distribution of the image data  $I$  to be Gaussian, and, following previous work<sup>17</sup> on RNA bulk sequencing data, we assume the conditional distribution of the expression data  $X$  to be negative binomial. The parameters of the conditional distributions are mapped from the latent tissue state  $Z$  through a convolutional generator network  $G$  with learnable parameters  $\theta$ .

Formally, for all sections  $n$ , pixel coordinates  $(x, y)$ , genes  $g$ , metagenes  $m$  and image channels  $c$ , we model the data generating process as follows:

$$Z_n \sim \mathcal{N}(0, \mathbb{I}) \quad (1)$$

$$L_g \sim \mathcal{N}(0, \sigma_{L_g}^2 \mathbb{I}) \quad (2)$$

$$E_g \sim \mathcal{N}(0, \sigma_{E_g}^2 \mathbb{I}) \quad (3)$$

$$F_g \sim \mathcal{N}(0, \sigma_{F_g}^2 \mathbb{I}) \quad (4)$$

$$(s_n, a_n, \mu_n, \sigma_n) \equiv G(Z_n) \quad (5)$$

$$r_{ngxy} \equiv s_{nxy} e^{t_g + \beta_n E_g} \sum_m a_{nmxy} e^{L_{mg}} \quad (6)$$

$$p_{ng} \equiv S(u_g + \beta_n F_g) \quad (7)$$

$$X_{ngxy} | Z_n, L_g, E_g, F_g \sim \text{NB}(r_{ngxy}, p_{ng}) \quad (8)$$

$$\tilde{X}_{ngl} \equiv \sum_{(x,y) \in \mathcal{A}_{nl}} X_{ngxy} \quad (9)$$

$$I_{ncxy} | Z_n \sim \mathcal{N}\left(\mu_{ncxy}, \sigma_{ncxy}^2\right), \quad (10)$$

where  $S$  is the logistic function. The generator network outputs the pixel-wise scaling factors  $s_n$ , metagene activities  $a_n$ , image distribution means  $\mu_n$  and standard deviations  $\sigma_n$ . The gene expression distribution in Eq. (8) is parameterized by the number of failures before stopping  $r_{ngxy}$  defined by Eq. (6), and the success probability  $p_{ng}$  defined by Eq. (7). The sum in Eq. (6) corresponds to a non-negative matrix factorization into the metagene activities  $a_n$  and the exponentiated gene weight matrix  $L$ , which describes the expression program associated with each metagene. The parameters  $t_g$  and  $u_g$  are gene-specific baselines. The fixed effects  $E$  and  $F$  are used to control for condition-wise batch effects. The variable  $\beta_n$  is a row vector of concatenated indicator variables specifying, for each covariate, the condition that section  $n$  belongs to. Equation (9) relates the super-resolved expression  $X$  to the observed expression  $\tilde{X}$  by summation over the pixels  $\mathcal{A}_{nl}$  in each measurement location  $l$ . All variables, parameters and their constraints are summarized in Supplementary Table 1.

During inference, we collapse the model by integrating out  $X$ , which replaces Eqs. (8) and (9) with

$$\tilde{X}_{ngl} | Z_n, L_g, E_g, F_g \sim \text{NB}\left(\sum_{(x,y) \in \mathcal{A}_{nl}} r_{ngxy}, p_{ng}\right). \quad (11)$$

The factorization in Eq. (6) allows us to reduce the first parameter of the negative binomial in Eq. (11) by first summing the scaled metagene activities over the pixels in each measurement location,

$$\sum_{(x,y) \in \mathcal{A}_{nl}} r_{ngxy} = e^{t_g + \beta_n E_g} \sum_m e^{L_{mg}} \sum_{(x,y) \in \mathcal{A}_{nl}} s_{nxy} a_{nmxy}. \quad (12)$$

Equation (12) means that we can avoid computing pixel-wise values of  $r$  for each gene, thus substantially reducing memory requirements during inference.

**Inference.** We use variational inference to approximate the posterior of the latent variables  $p(Z, L, E, F | \tilde{X}, I)$  with a tractable distribution  $q_\phi(Z, L, E, F)$ . Variational methods treat inference as an optimization problem, which makes them suitable for large, high-dimensional models in which exact sampling is difficult<sup>18</sup>. Our approach is analogous to learning in variational autoencoders<sup>9,10</sup>. Similar strategies have been explored in other works for modeling RNA sequencing data<sup>19,20</sup>.

The variational parameters  $\phi$  are found by minimizing the Kullback–Leibler divergence from the variational distribution  $q_\phi$  to the posterior,

$$D_{KL}(q_\phi(Z, L, E, F) || p(Z, L, E, F | \tilde{X}, I)). \quad (13)$$

Equation (13) is not possible to evaluate directly, because the posterior cannot be expressed in closed form. However, subtracting Eq. (13) from the log-evidence  $\log p(\tilde{X}, I)$ , a constant with respect to the variational parameters, yields the alternative objective

$$\mathcal{L} = \mathbb{E}_{q_\phi} [\log p(\tilde{X}, I, Z, L, E, F) - \log q_\phi(Z, L, E, F)], \quad (14)$$

which can be estimated by Monte Carlo sampling. Because Eq. (13) is non-negative, Eq. (14) lower bounds the log-evidence. Therefore, Eq. (14) is commonly known as the evidence lower bound (ELBO).

We update both the variational parameters  $\phi$  and model parameters  $\theta, t, u, \sigma_L^2, \sigma_E^2$  and  $\sigma_F^2$  by gradient ascent of the ELBO (14). This amounts to simultaneously updating the variational distribution  $q_\phi$  to match the posterior while selecting a model  $p$  that maximizes the marginal likelihood.

We use a mean field diagonal Gaussian variational distribution

$$q_\phi(Z, L, E, F) = q_{\phi_L}(L) q_{\phi_E}(E) q_{\phi_F}(F) \prod_n q_{\phi_{Z_n}}(Z_n), \quad (15)$$

where the parameters  $\phi_{Z_n}$  are encoded by a convolutional recognition network  $R$  with weights  $\phi_Z$  applied to the image data:  $\phi_{Z_n} \equiv R_{\phi_Z}(I_n)$ .

Following ref.<sup>9</sup>, gradient estimates are obtained by re-parameterizing the latent variables as a function of auxiliary parameter-free noise. Thus, letting

$$\varepsilon \sim \mathcal{N}(0, \mathbb{I}) \quad (16)$$

$$(Z, L, E, F) \equiv h_\phi(\varepsilon), \quad (17)$$

where  $h_\phi$  is an appropriate shift-and-scale transformation, we can reformulate Eq. (14) as an expectation with respect to  $\varepsilon$  by relying on the law of the unconscious statistician. This makes it straightforward to rewrite the gradient of Eq. (14) as an expectation,

$$\nabla \mathcal{L}(\phi, \theta, t, u, \sigma_L^2, \sigma_E^2, \sigma_F^2) = \mathbb{E}_{q(\varepsilon)} [\nabla \log p(\tilde{X}, I, Z, L, E, F) - \nabla \log q_\phi(Z, L, E, F)]. \quad (18)$$

We approximate (18) using a single Monte Carlo sample and update parameters with the Adam optimizer<sup>21</sup>.

**Data augmentation.** Instead of tiling the data and training on non-overlapping patches, we sample patch coordinates uniformly random from the section surfaces and extract patches directly from the dataset during learning. Patches are distorted with uniformly sampled rotations from  $(0^\circ, 360^\circ)$ , scaling from  $(0.95, 1.05)$  and shearing from  $(0^\circ, 10^\circ)$ . The image data are further distorted with random brightness, contrast, saturation and hue jitter independently sampled from  $(-5\%, 5\%)$ .

**File formats.** Data belonging to each section are stored in the HDF5 file format and read on-demand from disk storage. Only data pertaining to the current mini-batch are kept in memory, allowing the method to be applied to an arbitrary number of sections.

The recognition and generator networks are equivariant to translations by design and learn rotational equivariance from the augmented data distribution. Consequently, sections do not need to be aligned.

**Architecture.** To efficiently capture both global and local anatomical features, the latent tissue state  $Z$  is modeled over multiple resolutions. The recognition and generator networks  $G$  and  $R$  together form an architecture similar to U-Net<sup>22</sup> with the variational distribution of the latent state for each resolution inserted at the corresponding skip connection (Extended Data Fig. 9).

**Metagene selection.** To select the  $M$  metagenes in the model, we implement a drop-and-split strategy that runs in parallel to inference. Briefly, we start out with  $M=1$  metagenes. At fixed intervals, we estimate the ELBO (14) with and without each of the  $M$  metagenes. Metagenes that contribute to the ELBO are split into two new metagenes that inherit parameters from their parent while non-contributing metagenes are dropped.

The drop-and-split strategy serves two purposes. First, it bootstraps new metagenes with sensible defaults, thus speeding up inference. Second, by splitting up already established metagenes, it encourages the model to explore solutions in which some metagenes have highly similar expression profiles. This might, for example, help the model tease apart transcriptional subtypes of closely related cell types.

In our experiments, we limit the maximum number of metagenes to 40 to reduce memory consumption.

**Runtime.** Convergence is assessed qualitatively by inspecting the ELBO during training. Learning is terminated once the ELBO plateaus.

Training the model on one section takes approximately 1 d on an Nvidia GeForce RTX 2080 Ti graphics card. We have observed a roughly linear scaling in

runtime with respect to the number of sections in the model (Extended Data Fig. 10). Using a trained model to predict super-resolved expression maps for 1,000 genes with ten Monte Carlo samples takes 1–2 h per section.

**Prediction.** We estimate the posterior predictive distribution of a quantity  $\chi$  with  $N$  Monte Carlo samples drawn from the variational distribution:

$$\begin{aligned} p(\chi | \tilde{X}, I) &= \int p(\chi | Z, L, E, F) dP(Z, L, E, F | \tilde{X}, I) \\ &\approx \int p(\chi | Z, L, E, F) dQ_\phi(Z, L, E, F) \\ &\simeq \frac{1}{N} \sum_{i=1}^N p(\chi | Z^{(i)}, L^{(i)}, E^{(i)}, F^{(i)}), \end{aligned} \quad (19)$$

where  $i$  is a sample index, and  $P$  and  $Q_\phi$  denote the cumulative distribution functions of the corresponding lowercase densities.

We consider different quantities  $\chi$  for different types of analyses, as described below.

**Gene expression.** To predict spatial gene expression, we consider the conditional mean expression

$$\begin{aligned} \nu_k &\equiv \mathbb{E} \left[ \sum_{(n,x,y) \in \mathcal{A}_k} X_{nxy} | Z, L, E, F \right] \\ &= \sum_{(n,x,y) \in \mathcal{A}_k} \frac{r_{nxy} p_n}{1 - p_n} \end{aligned} \quad (20)$$

in arbitrarily defined areas  $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$ . The posterior distribution of  $(\nu_1, \dots, \nu_K)$  is estimated according to Eq. (19). The resultant mixture components are point mass distributions centered on the Monte Carlo samples  $\{\nu^{(i)}\}_i$ .

**Gene counts.** To predict read counts, we consider the expression vector

$$X_k \equiv \sum_{(n,x,y) \in \mathcal{A}_k} X_{nxy} \quad (21)$$

in arbitrarily defined areas  $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$ . The posterior distribution of  $(X_1, \dots, X_K)$  is estimated according to Eq. (19). The resultant mixture components are negative binomials analogous to Eq. (11) with parameters given by the Monte Carlo samples.

In contrast to  $\nu$ ,  $X$  captures aleatoric uncertainty in the sampling process of the ISC data. It is, therefore, suitable for comparisons to expression measurements.

**Differential gene expression.** To predict differential gene expression, we consider the log<sub>2</sub> fold change of the normalized mean expression of a gene  $g$  between the areas  $\mathcal{A}_1$  and  $\mathcal{A}_2$

$$\eta_g \equiv \log_2 \frac{\nu_{1g}}{\sum_g \nu_{1g}} - \log_2 \frac{\nu_{2g}}{\sum_g \nu_{2g}}. \quad (22)$$

The posterior distribution of  $\eta$  is estimated according to Eq. (19). The resultant mixture components are point mass distributions centered on the Monte Carlo samples  $\{\eta^{(i)}\}_i$ .

**Summarized expression maps.** To visualize transcriptional anatomy, we consider the pixel-wise metagene activity  $a$  and scale  $s$ . The posterior predictive means  $\bar{a}$  and  $\bar{s}$  are estimated according to Eq. (19). We project  $\bar{a}$  onto its first three principal components and append  $-\bar{s}$  along the channel axis. We then apply a channel-wise affine transformation to map all values into (0, 1). The resulting coordinates are used as cyan, magenta, yellow and black-encoded color values.

**In silico spatial transcriptomics.** Sections without expression data can be treated as ordinary sections with an empty set of measurement locations. However, to speed up learning, we have found it advantageous to exclude them during training. To predict expression in a section  $n'$  not included in the model, we approximate

$$\begin{aligned} p(Z_{n'}, L, E, F | \tilde{X}, I, I_{n'}) &= p(Z_{n'} | I_{n'}) p(L, E, F | \tilde{X}, I) \\ &\approx q_{R_\phi(I_{n'})}(Z_{n'}) q_\phi(L, E, F) \end{aligned} \quad (23)$$

and estimate  $p(\chi | \tilde{X}, I, I_{n'})$  analogously to Eq. (19).

**Prediction baselines.** Aside from the proposed method, we consider three baseline methods for spatial expression imputation: nearest neighbor interpolation, linear interpolation and Gaussian process regression. These baselines are described below.

**Nearest neighbor interpolation.** For nearest neighbor interpolation, we construct a Voronoi tessellation from the measurement locations in the training data. Imputed expression is taken as the area-weighted average of the tessellated training locations in each test location.

**Linear interpolation.** For linear interpolation, we consider the center of each measurement location. We construct a Delaunay triangulation from the training locations. Imputed expression is taken as the distance-weighted average of the training locations on the simplex of each test location.

**Gaussian process regression.** In this section, we change notation for consistency with the literature on Gaussian processes, which typically denotes the independent and dependent variables as  $x$  and  $y$ , respectively. Thus, we let  $y_g \in \mathbb{R}^n$  be the expression values of gene  $g$  across  $n$  training locations in a given section and  $x \in \mathbb{R}^{n \times 3}$  be their center coordinates. We model the expression data as

$$y_g \sim \mathcal{GP}(x | m_g, K_g), \quad (24)$$

where  $\mathcal{GP}$  designates a Gaussian process,  $m_g$  is a gene-wise constant and  $K_g$  is a kernel function. We use a squared exponential kernel,

$$K_g(x_i, x_j) = \theta_g \exp \left( -\frac{\|x_i - x_j\|_2^2}{2l_g^2} \right) + (\sigma_g + \epsilon) \delta_{ij}, \quad (25)$$

where  $\delta_{ij}$  is the Kronecker delta,  $l_g$  is a gene-wise length scale,  $\theta_g$  is a gene-wise output scale,  $\sigma_g$  is a gene-wise noise term and  $\epsilon$  is a global noise term, shared across all genes.

Optimal parameters  $m_g^*$  and  $K^*$  are inferred by maximizing the marginal likelihood using gradient ascent, as implemented in ref.<sup>23</sup>. Imputed expression  $\hat{y}_g$  is taken as the posterior predictive mean of the expression  $\tilde{y}_g$  in the test locations  $\tilde{x}$  given the observed data under the optimal parameters,

$$[y_g, \tilde{y}_g] \sim \mathcal{GP}([x, \tilde{x}] | m_g^*, K_g^*) \quad (26)$$

$$\hat{y}_g \equiv \mathbb{E}[\tilde{y}_g | \tilde{x}, x, y_g]. \quad (27)$$

The analysis is repeated independently for each section.

**Differential expression baseline.** To find differentially expressed genes in the MCL of the mouse olfactory bulb using ISC data, we follow a typical workflow based on Seurat<sup>14</sup>. First, we select measurement locations overlapping with the MCL annotation (Extended Data Fig. 8a). Then, we log-normalize the data and compute differentially expressed genes using the FindMarkers function. Genes are sorted by the computed  $P$  values.

**Cell type mapping.** To map cell types in the DCIS dataset, we use anchor-based label transfer as described in ref.<sup>14</sup> and implemented in Seurat. We use pre-annotated single-cell reference data from a HER2-positive breast cancer biopsy<sup>24</sup> and relabel the data into broad cell type classes. First, we log-normalize the reference and query data and extract the 2,000 most variable genes using the vst method. Then, anchors are identified using the FindTransferAnchors function without anchor filtering. Finally, labels are transferred using the TransferData function with the argument k.weight set to 8.

**Datasets.** Datasets and experimental settings are described in Supplementary Table 2.

**Synthetic data.** Ground truth expression maps are constructed by randomly placing up to 32 non-overlapping shapes on a surface of size  $384 \times 384$  pixels. Three different shapes are considered: circles, squares and rectangles. Shapes emit a mean 10 reads per pixel of a marker transcript unique for that shape.

Measurement locations are arranged in a grid of size  $4 \times 4$  covering the data surface. The observed count data are Poisson distributed with a rate equal to the summed mean emission in each location. The observed image data are generated by color-coding the shapes and adding pixel-independent Gaussian noise.

**Human squamous cell carcinoma dataset.** Skin tissue was collected from patients with suspected squamous cell carcinoma after obtaining informed consent. The diagnosis was histologically verified by a board-certified dermatopathologist. The study protocol was approved by the institutional review board at Stanford University Medical Center (protocol no. 21750).

A biopsy containing a squamous cell carcinoma lesion was snap-frozen in isopentane (Sigma-Aldrich) and embedded in optimal cutting temperature (OCT) compound (Sakura). The sample was cryosectioned at 10- $\mu\text{m}$  intervals for a total length of 490  $\mu\text{m}$ . Sections starting at positions 0  $\mu\text{m}$ , 160  $\mu\text{m}$ , 320  $\mu\text{m}$  and 480  $\mu\text{m}$  were collected and placed on a slide from a Visium Spatial Gene Expression Slide & Reagent Kit (10x Genomics). Library preparation, sequencing and read trimming followed the procedure described in ref.<sup>25</sup>. The resulting data were pre-processed using Space Ranger (10x Genomics).

Histology images for the staining stability experiments were created by synthetically separating the hematoxylin and eosin stains in the original image following ref.<sup>26</sup> and remixing them with the proportions shown in Extended Data Fig. 4d.

**Human small intestine dataset.** Small intestinal tissue was collected from patients undergoing colorectal surgery after obtaining informed consent. The study

protocol was approved by the medical ethics committee of University Hospitals Leuven (approval no. S62935).

A biopsy from healthy tissue in the ileum was snap-frozen in isopentane and embedded in OCT compound. The sample was cryosectioned at 10- $\mu\text{m}$  intervals. Library preparation was performed on four adjacent sections using the Visium Spatial Gene Expression Slide & Reagent Kit, following the manufacturer's user guide (CG000239 Rev D). The optimal permeabilization time was determined to be 30 min based on results from an experiment with the Visium Spatial Tissue Optimization Slide & Reagent Kit. Sequencing and read trimming followed the procedure described in ref.<sup>25</sup>. The resulting data were pre-processed using Space Ranger.

**Relationship to previous work.** Our work extends previous research on spatial models of transcriptomics data. Notably, SpatialDE<sup>27</sup> and SPARK<sup>28</sup> model spatial transcriptomics data using Gaussian processes to detect spatially variable genes. However, neither method makes use of histological information or can be used to infer high-resolution expression data. NovoSpaRc<sup>29</sup> reconstructs the spatial organization of single cells by solving an optimal transport problem. Although novoSpaRc can identify zonated genes from single-cell data, accurate inference of spatial expression patterns requires information about the spatial configuration of marker genes. Several other methods<sup>30–33</sup> exist for fusing single-cell with *in situ* sequencing or hybridization data.

Recently, several works<sup>34–36</sup> have explored the possibility of learning the cell type mixes that give rise to observed ISC measurements. Although our method decomposes the ISC data spatially, these methods, instead, decompose the data into cell type components. An interesting future research direction is to combine spatial decomposition with cell type decomposition. In the context of the proposed model, this could, for example, be realized by enforcing cell-type-specific metagene priors.

Concurrently with this work, experimental methods are making progress toward higher-resolution expression measurement. Notably, Slide-seqV2 (ref.<sup>37</sup>) quantifies gene expression at 10- $\mu\text{m}$  resolution with 50% sensitivity compared to single-cell RNA sequencing. However, measurements are limited to a 3-mm-diameter field of view. Moreover, the protocol uses separate sections for expression measurements and histology, making it difficult to associate transcriptome signatures with precise morphological features. DBiT-seq<sup>38</sup> uses microfluidic chips to quantify expression in  $10 \times 10 \times 10\text{-}\mu\text{m}^3$  measurement locations with 15.5% sensitivity compared to single molecule fluorescence *in situ* hybridization. However, the measurements are limited to a  $1 \times 1 \times 1\text{-mm}^3$  field of view and cover only 25% of the surface area.

The contribution of our work is three-fold. First, we have shown that histological image data are highly informative of spatial expression patterns in tissues. Second, we provide an integrative model of spatial gene expression. Our model fuses ISC with high-resolution image data to infer super-resolved expression maps that can be used to study transcriptome-scale gene expression in detailed anatomical features. Third, we have demonstrated the feasibility of predicting spatial gene expression from histology images. We think that image-based *in silico* spatial transcriptomics is a promising future research topic.

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

## Data availability

The mouse olfactory bulb dataset was obtained from the Spatial Research group's website: <https://www.spatialresearch.org/resources-published-datasets/doi-10-1126science-aaf2403>. The breast cancer spatial transcriptomics dataset was obtained from the 10x Genomics data repository: <https://support.10xgenomics.com/spatial-gene-expression/datasets/>. The breast cancer single-cell dataset was obtained from the authors of the original publication<sup>24</sup>. The squamous cell carcinoma dataset is available on Mendeley Data: <https://doi.org/10.17632/2bh5fchcv6.1>. The small intestine dataset is available on Mendeley Data: <https://doi.org/10.17632/v8s9nz948s.1> (folder V19T26-028\_B1).

## Code availability

We have implemented the proposed method in PyTorch<sup>39</sup> and the Pyro probabilistic programming language<sup>40</sup>. The code is available under the MIT license at <https://github.com/ludvb/xfuse>.

## References

17. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
18. Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. Variational inference: a review for statisticians. *J. Am. Stat. Assoc.* **112**, 859–877 (2017).
19. 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).
20. Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **10**, 390 (2019).
21. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *3rd International Conference on Learning Representations* (eds Bengio, Y. & LeCun, Y.) <http://arxiv.org/abs/1412.6980> (2015).
22. Ronneberger, O., Fischer, P. & Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation* 234–241. (Springer, 2015).
23. Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q. & Wilson, A. G. GPY Torch: blackbox matrix–matrix Gaussian process inference with GPU acceleration. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Curran Associates, 2018).
24. Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genetics* **53**, 1334–1347 (2021).
25. Ji, A. L. et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497–514 (2020).
26. Macenko, M. et al. A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro <https://ieeexplore.ieee.org/document/5193250.1109-ISBI.2009.5193250> (2009).
27. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).
28. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* **17**, 193–200 (2020).
29. Nitza, M., Karaikos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).
30. Achim, K. et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
31. 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).
32. Qian, X. et al. Probabilistic cell typing enables fine mapping of closely related cell types *in situ*. *Nat. Methods* **17**, 101–106 (2019).
33. Lopez, R. et al. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. In *ICML Workshop on Computational Biology* (2019).
34. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-00830-w> (2021).
35. Andersson, A. et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**, 565 (2020).
36. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. Spotlight: seeded Nmf regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
37. Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2020).
38. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681 (2020).
39. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32* (eds Wallach, H. et al.) 8024–8035 <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf> (Curran Associates, 2019).
40. Bingham, E. et al. Pyro: deep universal probabilistic programming. *J. Mach. Learn. Res.* **20**, 1–6 (2019).

## Acknowledgements

This work was made possible by generous support from the Knut and Alice Wallenberg Foundation, the Erling-Persson Family Foundation, the Swedish Cancer Society, the Swedish Foundation for Strategic Research, the Swedish Research Council and the Helmsley Charitable Trust.

## Author contributions

L.B. and J.M. designed the method and wrote the paper. B.H., J.B., A.A. and A.L.J. provided valuable feedback and contributed to the analyses. R.M., X.A., K.T., L.L. and N.S. performed the experiments. J.M., J.L., J.Z., P.K. and G.B. supervised the project.

## Competing interests

J.L., R.M., K.T., A.A. and L.L. are scientific consultants for 10x Genomics, which produces spatially barcoded microarrays for *in situ* RNA capturing. The remaining authors declare no competing interests.

## Additional information

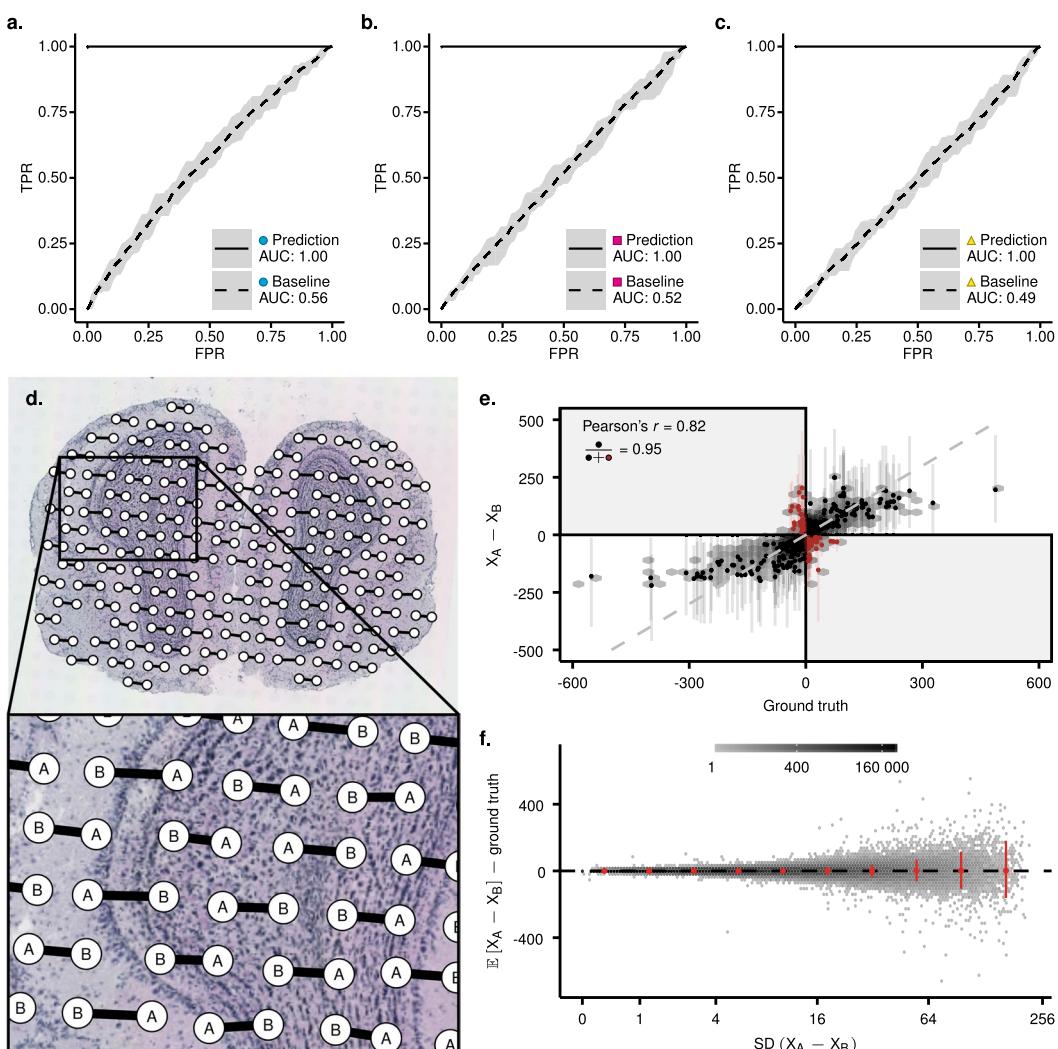
**Extended data** is available for this paper at <https://doi.org/10.1038/s41587-021-01075-3>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41587-021-01075-3>.

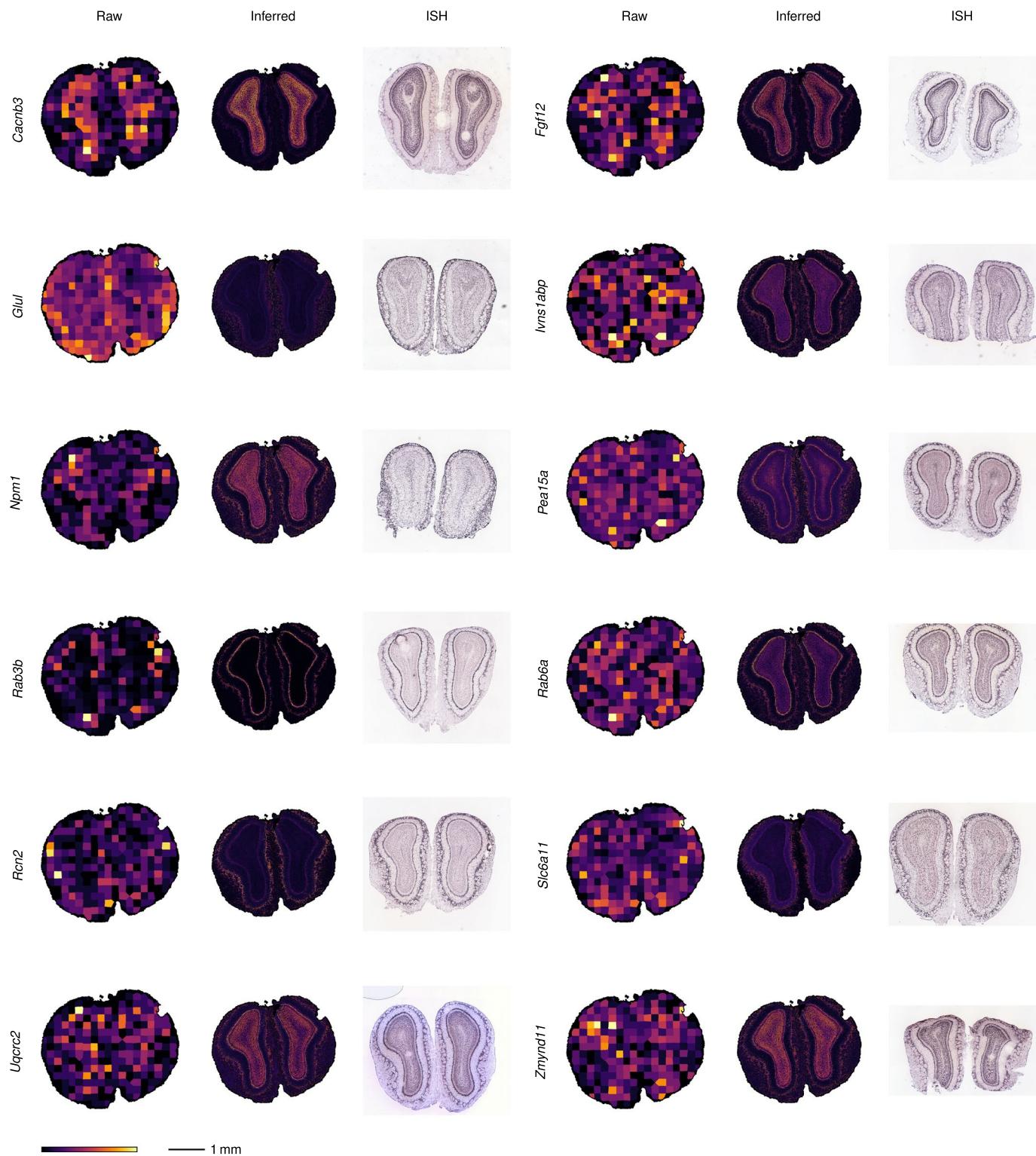
**Correspondence and requests for materials** should be addressed to Joakim Lundeberg.

**Peer review information** *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

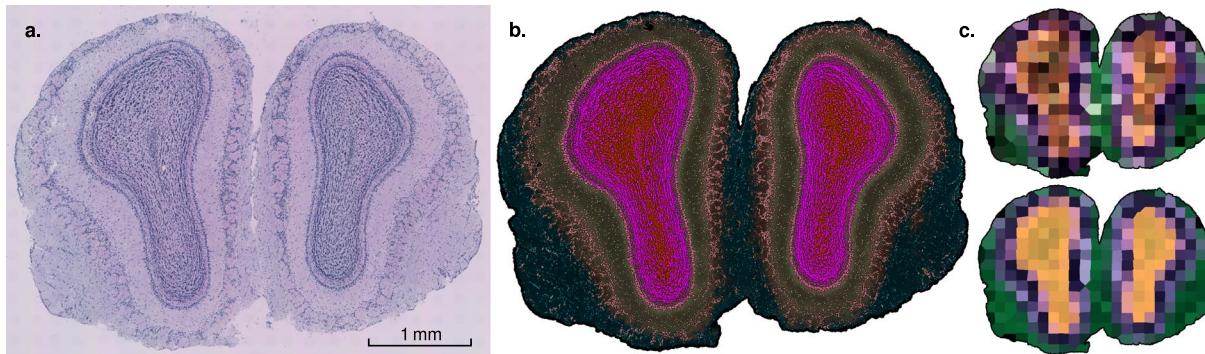
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



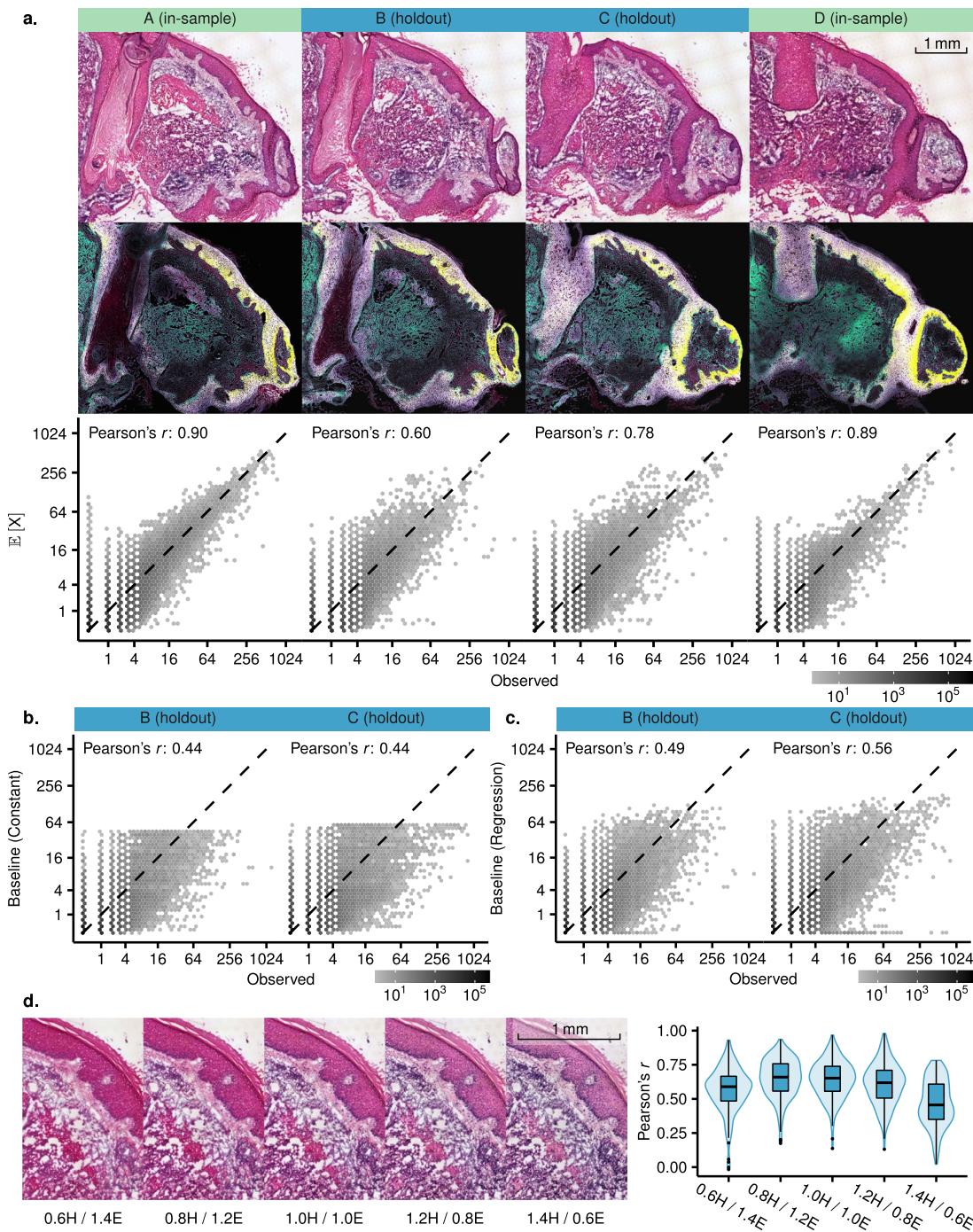
**Extended Data Fig. 1 | Deconvolution experiments.** **a–c.** Synthetic data. Receiver operating characteristic (ROC) curves for pixel-level classification of the three transcriptional subtypes: blue circle (a), red square (b), and yellow triangle (c). Dashed lines show baselines constructed by predicting the observed pixel-average in each measurement location. Ribbons indicate minima and maxima over predictions in 10 random synthetic image patches. **d–f.** Biological data. **d.** The ground truth expression data is downsampled by merging neighboring measurement locations and summing their count values  $X_A + X_B = X$ . The model is trained on the downsampled data  $X$  and used to predict the component counts  $X_A$  and  $X_B$  for each gene. **e.** Predicted direction against ground truth for observations with a 95% credibility of one component having a strictly higher expression than the other. Points show the medians of the predictive distributions, and error bars indicate 90% credibility intervals. Colors indicate if the predicted direction is the same as the ground truth direction. The dashed line indicates identity. For readability, points are only shown for the 10 highest-expressed genes. Hexagonal bins show observations for all genes ( $n = 12\,776$ ). Statistics are based on all genes. **f.** Directional misprediction against prediction uncertainty. In red, points indicate the 50th and error bars the 5th and 95th percentiles in evenly distributed bins.



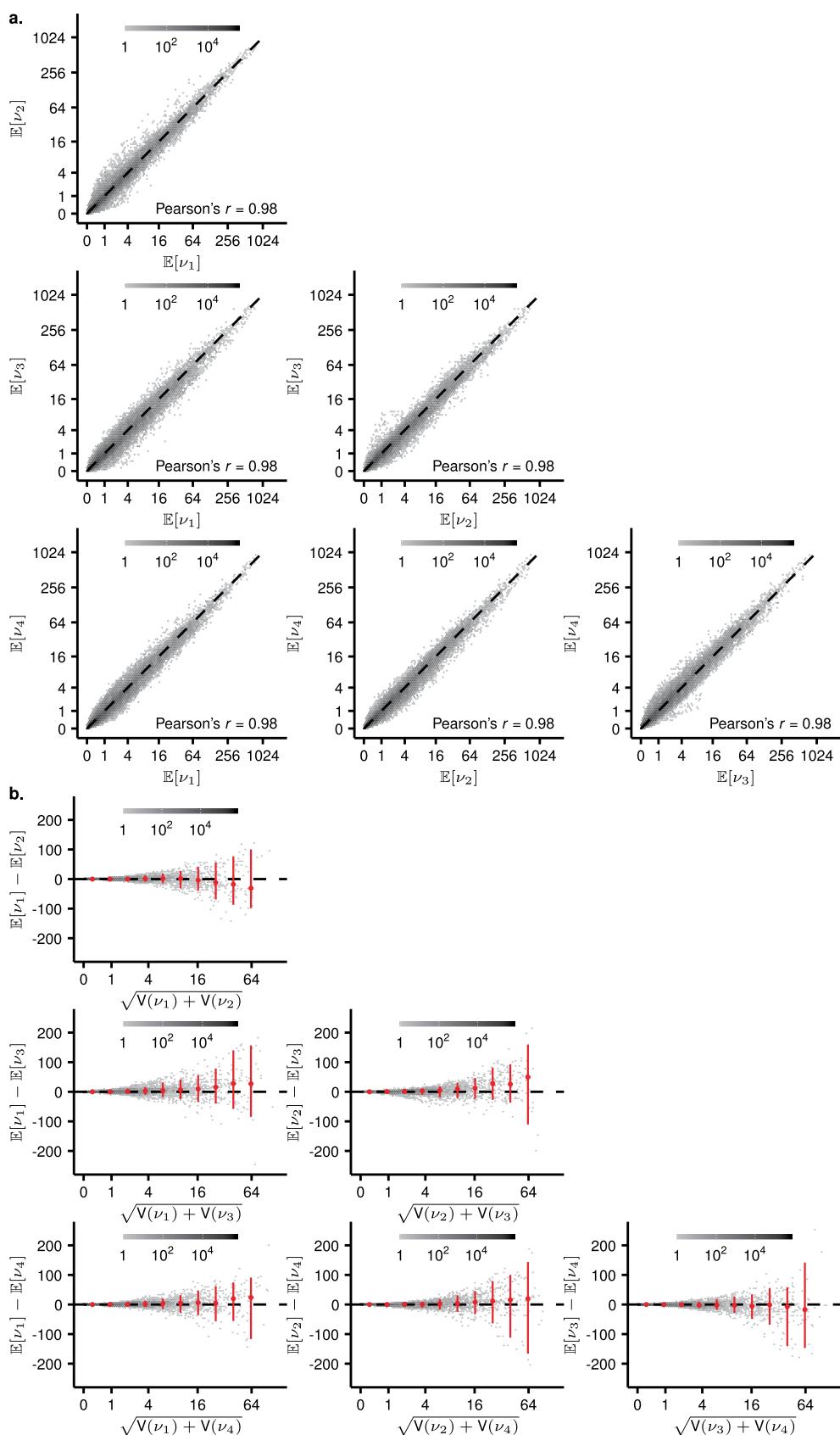
**Extended Data Fig. 2 | Comparison of inferred super-resolved expression maps to in situ hybridization reference data.** Random samples from the 1000 highest-expressed genes. Raw: Raw expression data (Voronoi tessellation). Inferred: Inferred super-resolved expression maps. ISH: In situ hybridization reference data from the Allen Mouse Brain Atlas<sup>11</sup>. Images show a representative sample of the 12 mouse olfactory bulb sections in the dataset.



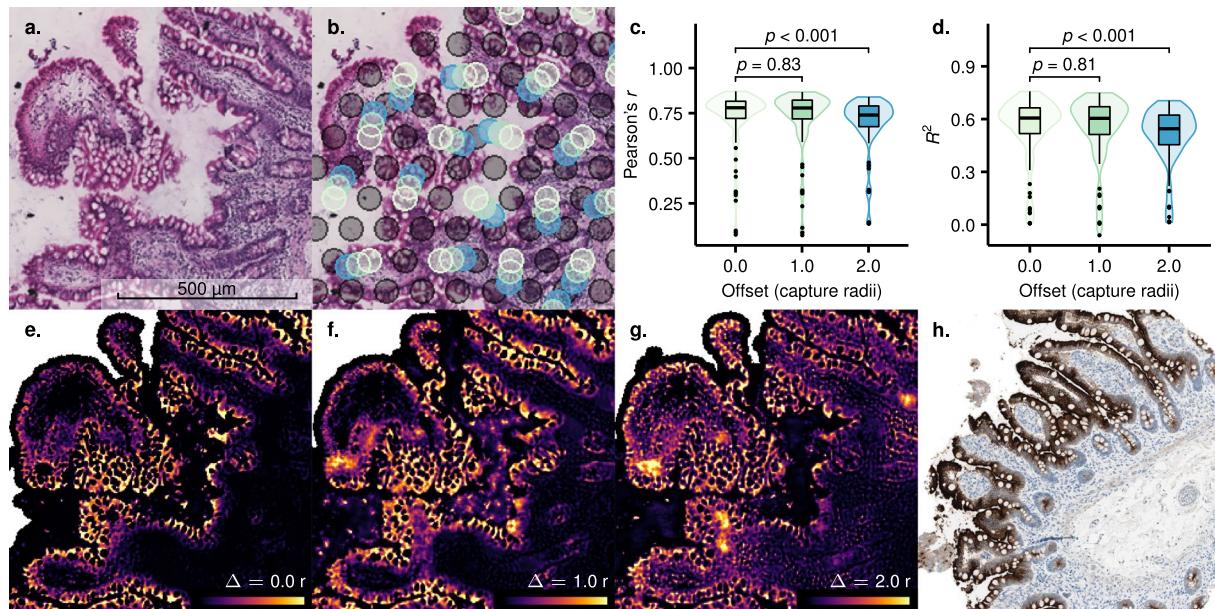
**Extended Data Fig. 3 | Prediction of spatial gene expression from histology images, mouse olfactory bulb experiments.** **a**, Histology image of holdout section (hematoxylin and eosin stain). **b**, Summarized expression map of the predicted metagene expression in the holdout section. **c**, Comparison of summarized expression maps constructed from normalized log ground truth gene expression in the held-out section (top) and normalized log predicted gene expression at the ground truth measurement locations using data from b (bottom). Results are based on an analysis that uses the 12th sample as holdout section (shown in a-c) and the remaining 11 samples as reference experiments.



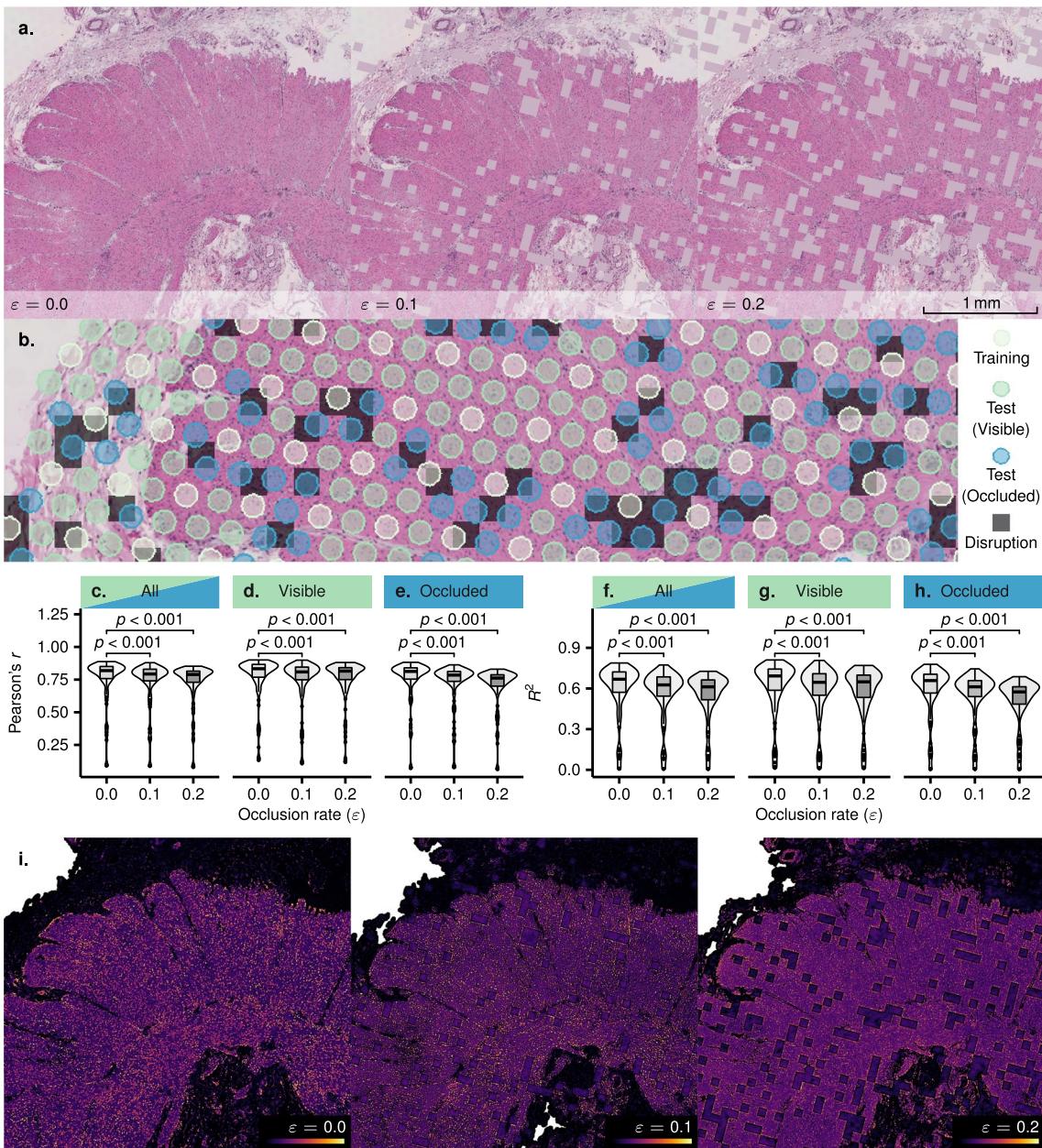
**Extended Data Fig. 4 | Prediction of spatial gene expression from histology images, squamous cell carcinoma experiments.** The dataset consists of four serial tissue sections spaced 150  $\mu\text{m}$  apart. The outer sections A and D are used as reference experiments to predict expression in the intermediate holdout sections B and C. **a.** Top: Histological image data (hematoxylin and eosin stains). Middle: Predicted summarized expression maps. Bottom: Predicted expression against ground truth for all genes ( $n = 11\,025$ ) in 100 randomly sampled test locations. **b,c.** Baseline experiments. **b.** Constant prediction against ground truth for all genes in the same test locations as in (a). For each gene, the prediction is the mean expression in sections A and D. **c.** Image intensity-based linear regression prediction against ground truth for all genes in the same test locations as in (a). The expression  $X_{lg}$  of gene  $g$  in location  $l$  is modeled as  $X_{lg} = \beta_g^0 + \beta_g l_i + \epsilon_{lg}$ , where  $l_i$  is a vector of the channel-wise 5-binned image intensities of location  $l$  and  $\epsilon_{lg}$  a standard normal noise term. The parameters  $\beta_g^0$  and  $\beta_g$  are selected by maximum likelihood estimation with data from sections A and D. Predictions are given by  $X_{lg}^* = \max(0, \beta_g^0 + \beta_g l_i)$ . **d.** Stability to variation in staining intensities. Left: Evaluated hematoxylin (H) and eosin (E) concentrations. Mixes are produced synthetically by rescaling the H and E channels (Methods). Images show representative close-ups from one of the four sections in the dataset. Right: Gene-wise Pearson correlation over all test locations in each holdout section evaluated on the  $n = 100$  highest-expressed genes. Boxes show 25th, 50th, and 75th percentiles. Outliers are represented by points and defined as observations further than 1.5 interquartile ranges from the hinges. Whiskers indicate the extent of all non-outlier observations.



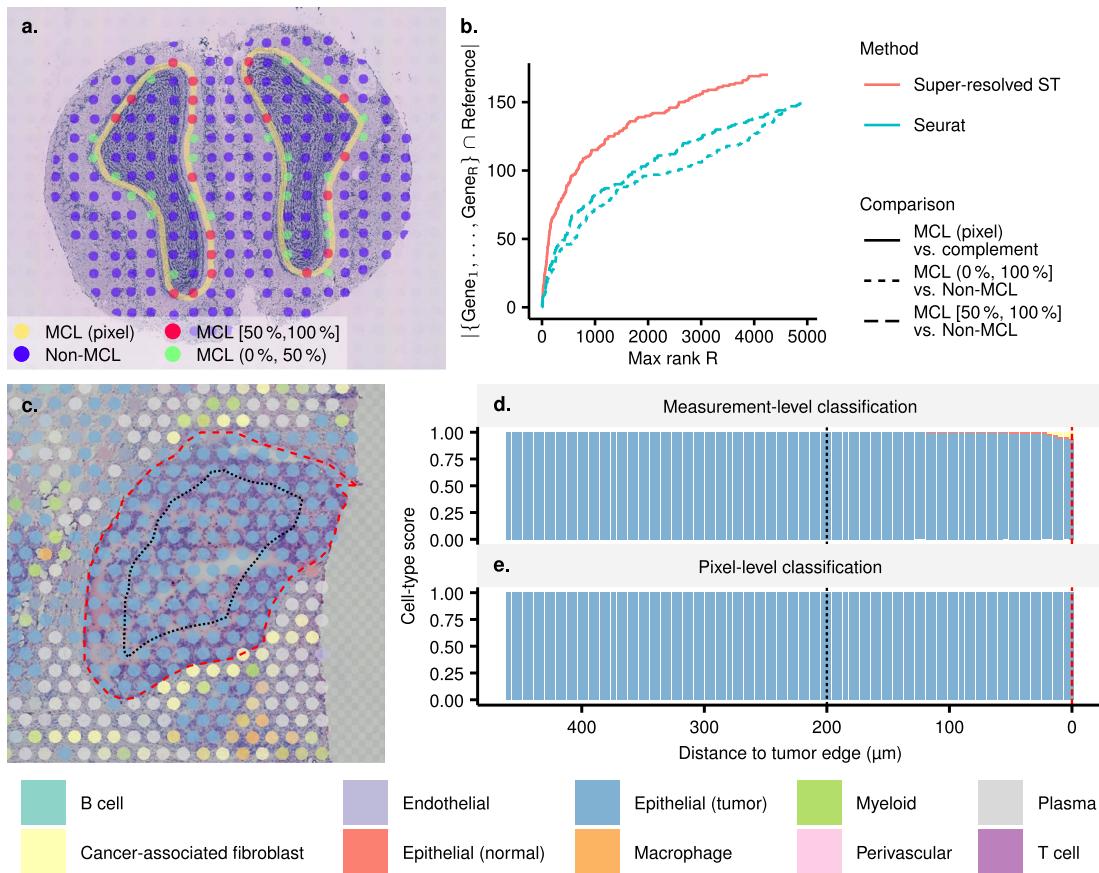
**Extended Data Fig. 5 | Run-to-run variability.** Results are based on four restarts of an analysis of the ileum of the human small intestine. **a**, Correlation plots. Each plot shows the predicted mean count for every gene in 100 test regions in two different runs. The test regions are sampled uniformly over the tissue surface and have the same size as the measurement locations in the original dataset. Correlation coefficients are computed over all genes and test locations ( $n = 6869 \times 100 = 686\,900$ ). **b**, Differences in predicted means  $E[\nu_i] - E[\nu_j]$  against prediction uncertainty  $\sqrt{V(\nu_i) + V(\nu_j)}$  for runs  $i$  and  $j$ . In red, points indicate the 50th and error bars the 5th and 95th percentiles in evenly distributed bins.



**Extended Data Fig. 6 | Robustness to measurement location misalignment.** **a**, Hematoxylin and eosin stain of a section from the ileum of the human small intestine. Representative close-up of a small area of the brush border. The brush border in the section measures approximately 7 mm in length. **b**, Conceptual illustration of the measurement locations on the Visium array. Dark circles indicate test locations withheld during training. Light circles indicate training locations over three misalignment levels: 0.0 (light green), 1.0 (green), and 2.0 (blue) radii of the measurement locations ( $r = 55\mu\text{m}$ ). The direction of the misalignment is uniformly random. **c,d**, Gene-wise Pearson correlation between predicted and ground truth expression (c) and coefficient of determination (d) over the test locations of the  $n = 100$  highest-expressed genes for increasing offsets. Boxes show 25th, 50th, and 75th percentiles. Outliers are represented by points and defined as observations further than 1.5 interquartile ranges from the hinges. Whiskers indicate the extent of all non-outlier observations. Pairwise  $p$ -values are based two-sided Wilcoxon signed-rank tests. Exact  $p$ -values (top to bottom): c,  $2.98 \times 10^{-7}$  and  $8.30 \times 10^{-1}$ ; d,  $3.45 \times 10^{-9}$  and  $8.14 \times 10^{-1}$ . **e-g**, Predicted expression of *CDHR5* when the training set has 0.0 (e), 1.0 (f), or 2.0 (g) radii misalignment. Close-ups of the same area as in a. **h**, Reference antibody staining for *CDHR5* in the small intestine from the Human Protein Atlas<sup>12</sup>.

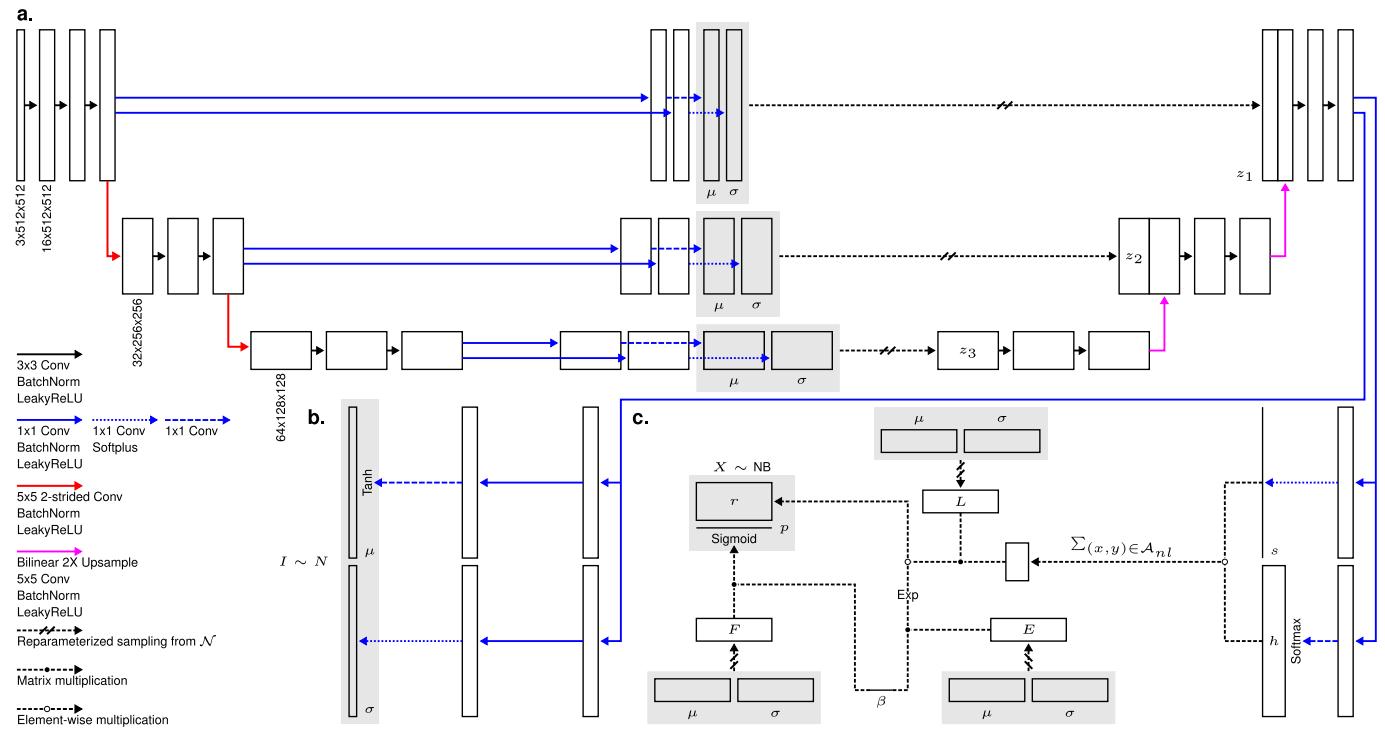


**Extended Data Fig. 7 | Robustness to image disruptions.** **a**, Hematoxylin and eosin stains of a section from the ileum of the human small intestine with increasing levels of occlusion noise ( $\varepsilon$ ). Representative close-ups of the smooth muscle layers. The smooth muscle layers in the section measure approximately  $3 \times 5 \text{ mm}^2$ . Noise is added by randomly sampling a proportion  $\varepsilon$  of tiles from a  $100 \times 100$  grid covering the histology image and replacing them with the mean color intensity of the slide. **b**, Conceptual illustration of the measurement locations on the Visium array. The locations are divided into a training (light green) and test set. The test set is further divided into regions that are fully visible in all experiments (green) and regions that are at least partially occluded in some experiments (blue). **c-h**, Performance under different occlusion levels evaluated using the gene-wise Pearson correlation between predicted and ground truth expression (c-e) and coefficient of determination (f-h) over the test locations of the  $n = 100$  highest-expressed genes. Boxes show 25th, 50th, and 75th percentiles. Outliers are represented by points and defined as observations further than 1.5 interquartile ranges from the hinges. Whiskers indicate the extent of all nonoutlier observations. Pairwise  $p$ -values are based two-sided Wilcoxon signed-rank tests. Exact  $p$ -values (top to bottom): c,  $1.48 \times 10^{-17}$  and  $9.18 \times 10^{-18}$ ; d,  $1.71 \times 10^{-13}$  and  $3.77 \times 10^{-13}$ ; e,  $9.75 \times 10^{-18}$  and  $8.39 \times 10^{-18}$ ; f,  $8.65 \times 10^{-18}$  and  $5.85 \times 10^{-18}$ ; g,  $1.15 \times 10^{-16}$  and  $2.24 \times 10^{-15}$ ; h,  $7.44 \times 10^{-18}$  and  $6.03 \times 10^{-18}$ . **i**, Prediction of ACTG2, a gene coding for gamma-enteric smooth muscle actin, over different occlusion levels.

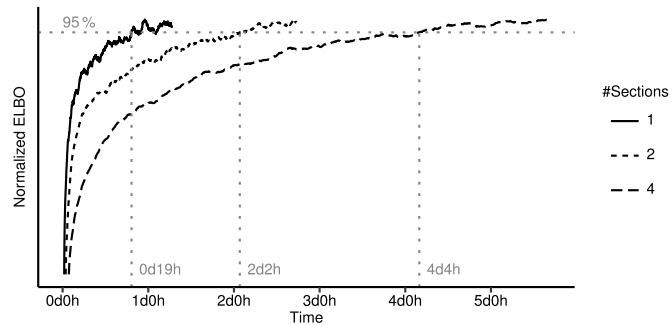


**Extended Data Fig. 8 | Differential gene expression and cell-type composition.** **a,b**, Differential gene expression, mouse olfactory bulb dataset.

**a**, Annotation of the mitral cell layer (MCL). Percentages indicate area overlap with pixel annotation. **b**, Agreement with MCL marker reference list<sup>13</sup> over different set sizes of genes predicted to be differentially expressed. Genes are ranked by the inverted coefficient of variation of their posterior log fold change (Super-resolved ST) or *p*-value (Seurat). Only genes with a predicted positive log fold change are shown. **c–e**, Cell-type composition, ductal carcinoma in situ (DCIS) dataset. **c**, Predicted cell types in each measurement location. Colors correspond to the score-weighted sums of the cell-type labels' RGB coordinates. **d**, Measurement-level classification as a function of proximity to the tumor edge. Bar heights show classification scores across all measurement locations weighted by their overlap with each distance isoline. **e**, Pixel-level classification as a function of proximity to the tumor edge. Bar heights show classification scores based on the predicted expression for the pixel band at each distance. Red dashed line: Tumor edge. Black dotted line: 200 μm isoline.



**Extended Data Fig. 9 | Architecture.** **a**, Fusion network. **b**, Image data decoder. **c**, Expression data decoder. Volume dimensions and number of down- and upsampling steps are exempliative.



**Extended Data Fig. 10 | Runtime.** Normalized ELBO over time for three runs with varying dataset sizes. Vertical lines indicate time points when the runs reached 95% of the maximum attained normalized ELBO for the first time.

## Reporting Summary

Nature Research 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 Research 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	Sequencing data from the squamous cell carcinoma and small intestine datasets were processed with 10X Genomics Space Ranger version 1.0.0.
Data analysis	Analyses in Figs. 1f to 1h and 2 and Extended Data Figs. 2 and 3 were conducted with XFuse [1] version 0.1.1. Analyses in Extended Data Fig. 8 were conducted with XFuse version 0.1.1 and Seurat version 3.2.2. Analyses in Figs. 1d and 1e and Extended Data Figs. 1, 4, 5, 6, 7, and 10 were conducted with XFuse version 0.2.0.  [1]: XFuse is our open-source implementation of the proposed method (see Code availability statement).

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 Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

The mouse olfactory bulb dataset was obtained from the Spatial Research group's website, <https://www.spatialresearch.org/resources-published-datasets/> doi-10.1126/science-aaf2403. The breast cancer dataset was obtained from the 10X Genomics data repository, <https://support.10xgenomics.com/spatial-gene->

# 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://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	No sample size calculation was performed as results are not intended to describe population-wide biological phenomena but to demonstrate the proposed method across a wide range of tissue types. Sample sizes were thus determined either by data availability (external datasets) or with this objective in mind (the squamous cell carcinoma and small intestine datasets).
Data exclusions	In each section, we filter out low-count measurements and measurements outside the tissue. This is done for two reasons: First, it allows us to crop the data volume to the tissue area, thus reducing the size of the data and speeding up inference. Second, while there should be no gene expression outside the tissue, measurements may, in practice, be non-zero due to technical errors and diffusion. Therefore, we instead introduce a virtual measurement of the area outside the tissue which is set to zero for all genes. These preprocessing steps were pre-established.
Replication	The proposed method is evaluated on datasets from different tissues and technologies. Since the method is based on stochastic optimization, learned models may differ slightly in each run. This variability is analyzed in the main text and External Data Figure 5. We have found independent runs on the same dataset to produce qualitatively and quantitatively similar results.
Randomization	Not applicable (no experimental groups)
Blinding	Not applicable (no experimental groups)

## 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 & experimental systems		Methods
n/a	Involved in the study	n/a
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms	
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants	
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data	
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern	

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Datasets are used only for method evaluation. There are no relevant population characteristics.
Recruitment	For the squamous cell carcinoma dataset, tissue was collected from a patient with a confirmed diagnosis after obtaining informed consent. For the small intestine dataset, tissue was collected from a patient undergoing colorectal surgery after obtaining informed consent. Datasets are used only for method evaluation. There is no risk of self-selection bias.
Ethics oversight	The study protocol associated with the squamous cell carcinoma dataset was approved by the Institutional Review Board at Stanford University (protocol #21750). The study protocol associated with the small intestine dataset was approved by the Medical Ethics Committee of the University Hospitals Leuven (approval number S62935).

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