Super-resolved spatial transcriptomics by deep data fusion

Ludvig Bergenstråhle[1,2,*], Bryan He[3], Joseph Bergenstråhle[1,2], Alma Andersson[1,2], Joakim Lundeberg[1,2,*], James Zou[3], and Jonas Maaskola[1,4]

[1]SciLifeLab, Stockholm, Sweden
[2]KTH Royal Institute of Technology, Stockholm, Sweden
[3]Stanford University, California, United States
[4]Stockholm University, Stockholm, Sweden
[*]Corresponding authors: {ludvig.bergenstrahle, joakim.lundeberg}@scilifelab.se

2020-02-28

In situ RNA capturing has made it possible to record histology and spatial gene expression from the same tissue section, but methods to jointly analyze both kinds of data are still missing. Here, we present XFuse, a scalable deep generative model for spatial data fusion. XFuse can infer high-resolution, full-transcriptome spatial gene expression from histological image data and be used to characterize transcriptional heterogeneity in detailed anatomical structures.

Spatial transcriptomics allows researchers to study cell behavior in the spatial domain and has been used to describe cellular organization in the hippocampus [1], to characterize intra-tumor heterogeneity in human breast [2], pancreatic [3], and prostate cancer [4], to analyze spatial dynamics during embryonic cardiogenesis [5], and in many other contexts.

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 [6, 7] or hybridization [8, 9, 10, 11] typically have high resolution and high sensitivity but are difficult 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 [2, 12, 13] target all poly-adenylated transcripts simultaneously but have lower resolution and sensitivity, limiting their usefulness in studying detailed expression patterns.

To overcome the limitations of current spatial transcriptomics methods, we propose XFuse, a deep generative model of spatial expression data. XFuse models spatial gene expression and histological image data as observable effects of a latent tissue state (Fig. 1a, Methods). By fusing low-sensitivity, low-resolution ISC expression data with high-resolution histological image data, XFuse can infer accurate full-transcriptome spatial gene expression at the same resolution as the image data.

Inspired by recent work on deep segmentation networks [14], we define the latent tissue state over multiple spatial resolutions, allowing XFuse to capture both global and local anatomical contexts. Inference of the latent state and corresponding high-resolution expression data is based on ideas from the literature on variational autoencoders [15, 16]. Importantly, while optimizing model parameters, we jointly learn a neural network (typically referred to as a *recognition* or *inference* network) that maps the image data to the variational parameters of the latent state. As a result, the inferred posterior of the latent state is not kept in memory but recomputed for each mini-batch during training, allowing XFuse to scale to arbitrarily large datasets.

To evaluate the performance of XFuse, we study a dataset [2] consisting of 12 sections from the mouse olfactory bulb. First, we test in-sample performance by dropping 50% of all measurement locations and use XFuse to impute the missing expression data. We compare the results to a pixel-wise interpolation scheme that fills in missing data with the expression of the closest non-missing location and find that XFuse achieves a 23 % lower median root-mean-square error (Fig. 1b).

The recognition network allows XFuse to predict expression on unseen samples using only their histological image data. To assess this ability, we next test out-of-sample performance by holding out an entire section from the training set and use XFuse to predict its expression data. We find that XFuse faithfully reproduces ground truth expression patterns (Fig. S1) and that accuracy approaches in-sample performance as more sections are included in the training set (Fig. 1c).

Finally, we compare inferred gene expression to in situ hybridization data from the Allen Mouse Brain Atlas [17] and find clear correspondences (Fig. 1d). Crucially, the inferred expression data appears considerably more infor-

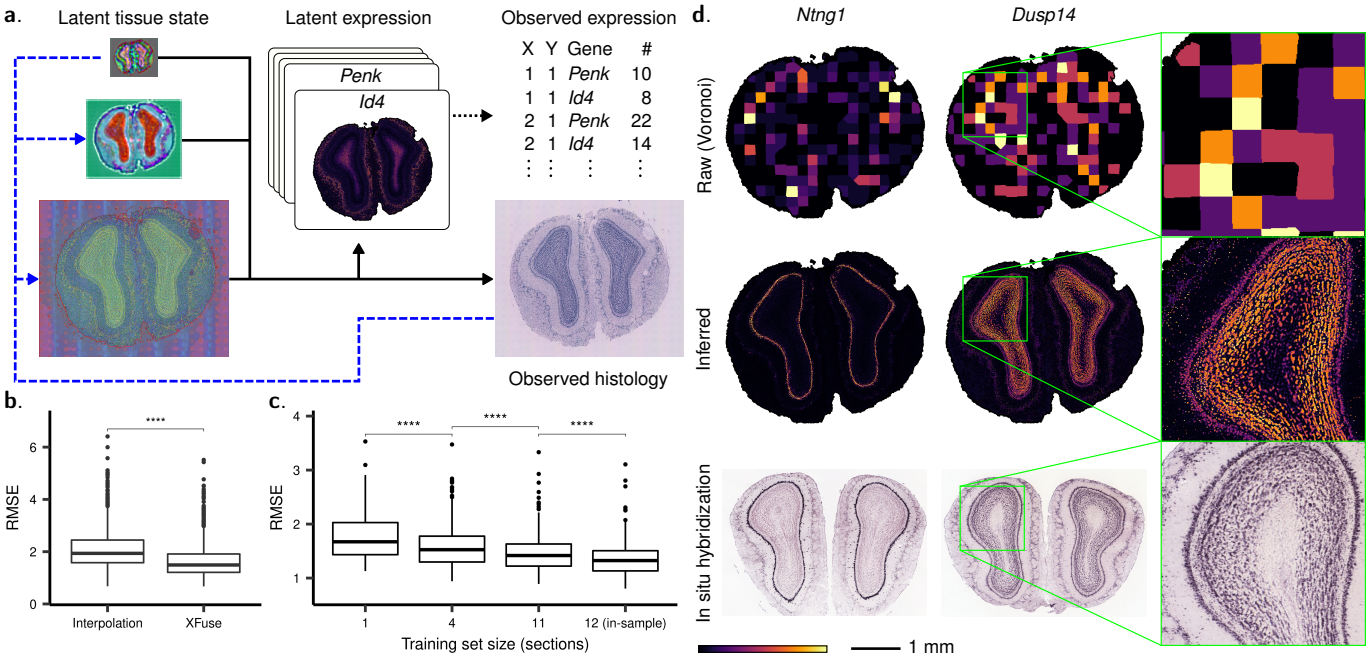

Figure 1: Conceptual overview and performance evaluation. (a) Histological image and expression data are modeled as effects of a latent spatial tissue state. The tissue state has multiple resolutions, capturing both global and local anatomical contexts, and is mapped through a generator network (black solid arrows) to histological image and high-resolution, latent expression data. The latent expression data is linked to the observed expression data by summation (black dotted arrow). Inference is amortized using a recognition network (blue dashed arrows) that maps the observed image data to the latent tissue state. (b), (c) Root-mean-square error (RMSE) of (b) imputation compared to pixel-wise, zero-order interpolation and (c) out-of-sample prediction over different training set sizes. Count values are normalized to mean one in each measurement location. Asterisks (****) indicate significance at the $p \leq 0.0001$ level using a two-sided Wilcoxon signed-rank test. (d) Comparison of inferred high-resolution expression data to in situ hybridization reference data from the Allen Mouse Brain Atlas.

mative about the transcriptional anatomy of the tissue than the raw spatial transcriptomics data (Fig. 1d and Figs. S2 and S3).

We demonstrate the applicability of our method by studying detailed anatomical structures in the mouse olfactory bulb and in human breast cancer. In both datasets, XFuse finds clear patterns of fine-grained expression heterogeneity (Figs. 2a and 2b), which can be quantified in terms of differential expression (Methods).

We first profile the mitral cell layer of the olfactory bulb (Fig. 2c) and find several strongly up- and down-regulated genes (Fig. 2d). To verify our results, we sort the genes by the inverted coefficient of variation of their posterior $\log_2$ fold change and find that 40 out of the 100 most upregulated genes are among 229 markers for the mitral cell layer identified in a recent single-cell RNA-sequencing study [18] (one-sided hypergeometric test p-value: 1.66×10^{-47}). Meanwhile, the raw data is too coarse-grained to resolve the mitral cell layer on its own (Fig. S3). We conclude that XFuse successfully deconvolves mixed expression signals by integrating expression patterns across anatomical areas that share similar morphology.

Next, we study spatial dynamics in a ductal carcinoma in situ (DCIS) lesion from the breast cancer dataset by profiling transcriptome gradients between the inner area of the tumor and its outermost edge (Fig. 2e). We find several genes related to immune activity and tumor progression to be upregulated at the border of the tumor (Fig. 2f). For example, the complement component 1q, composed of the C1QA, C1QB, and C1QC subcomponents, have been shown to promote angiogenesis and tumor growth in the tumor microenvironment [19]. Similarly, CD74 is a known marker for metastatic tumor growth in breast cancer [20], and antibody-drug conjugate therapies targeting CD74 expressing cells are currently being developed for 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 malignant DCIS. However, further studies are needed to validate this finding.

Consistent with the above results, the pathways that are enriched for the 100 most upregulated genes at the tumor border include, for example, extracellular structure organization (p-value: 2.10×10^{-18}), immune system processes (p-value: 1.37×10^{-11}), blood vessel development (p-value: 8.51×10^{-6}), and cell migration regulation (p-value: 8.98×10^{-6}) (Table S1). In contrast, none of these pathways are enriched for the 100 most downregulated genes (Table S2).

Critically, while the distance between measurement locations in the raw expression data is 100 µm, several differentially expressed genes become upregulated first within 50 µm of the tumor border (Fig. 2f). We conclude that it is only by learning a high resolution state of the underlying transcriptional anatomy of the tissue that it becomes

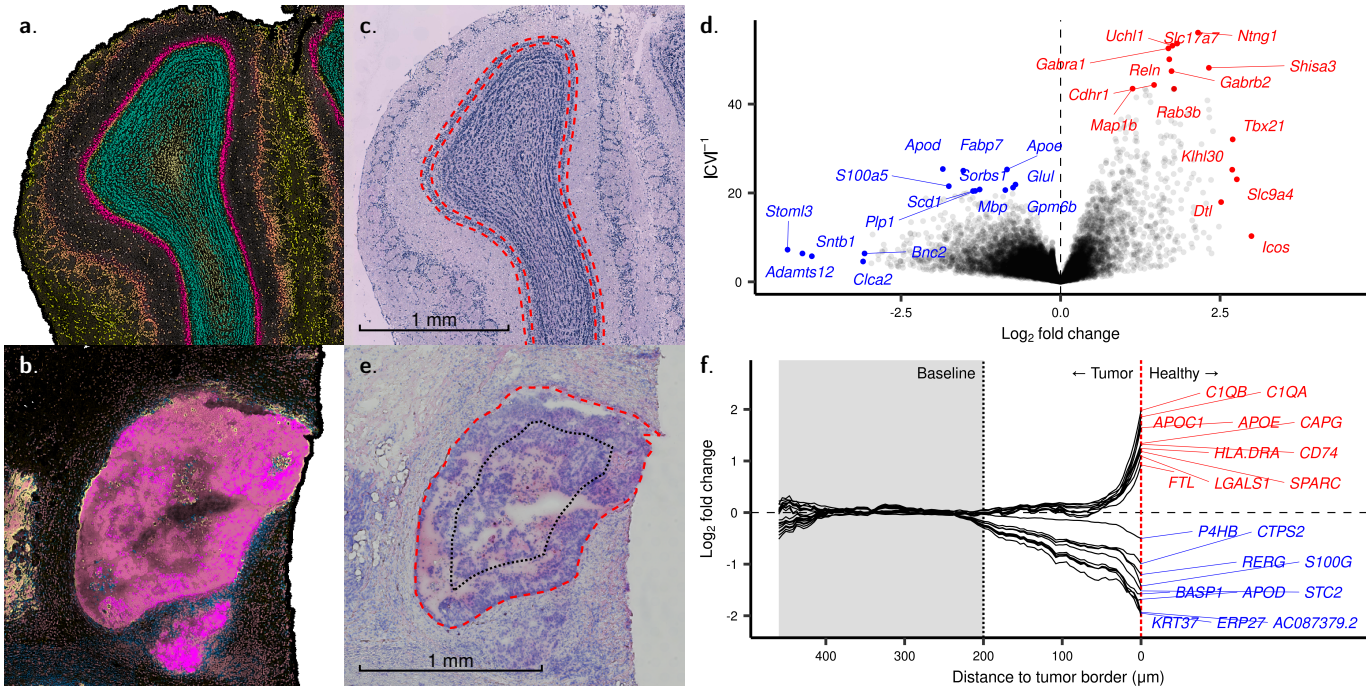

Figure 2: Characterization of transcriptional heterogeneity in detailed anatomical structures. (a), (b) Summarized latent gene expression in (a) the mouse olfactory bulb and (b) a human ductal carcinoma in situ (DCIS) lesion. Colors indicate anatomical areas with distinct transcriptional phenotypes according to the inferred tissue state (Methods). (c) Annotation of the mitral cell layer profiled in (d). (d) Differential expression in the mitral cell layer compared to the other parts of the mouse olfactory bulb. (e) Annotation of the DCIS lesion profiled in (f). Red dashed line: Tumor border. Black dotted line: Baseline boundary, 200 µm from the tumor border. (f) Differential expression compared to baseline as a function of distance to the tumor border.

possible to fully resolve the detailed expression landscape describing these genes. Determining its precise topology is paramount to understanding cellular interactions at the micro-scale and in developing effective treatments for a wide range of diseases.

In summary, we have presented XFuse, a deep generative model for spatial data fusion. XFuse combines ISC expression data with histological image data to infer accurate full-transcriptome spatial gene expression at an unprecedented resolution. We have found that XFuse exposes spatial contingencies that are difficult to discern in the raw expression data and can characterize differential expression in detailed anatomical structures. Moreover, once model parameters have been learned, XFuse can predict spatial expression from the histological image data of unseen samples, effectively providing a means for in silico spatial transcriptomics.

We envision future work to enable in silico spatial transcriptomics on a larger scale. By learning accurate recognition and generator networks across diverse tissues, it may be possible to accurately predict spatial expression without sequencing. In silico spatial transcriptomics could provide tremendous cost savings for large spatial transcriptomics projects or act as a way to verify the integrity of sequencing data from experimental methods.

Acknowledgments

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, and the Swedish Research Council.

Author contributions

L.B. and J.M. designed the method. L.B. implemented the method and wrote the paper. B.H., J.B., and A.A. provided valuable feedback and contributed to the analysis. J.M., J.Z., and J.L. supervised the project.

Competing interests

J.L. is a scientific advisor at 10x Genomics, which produces spatially barcoded microarrays for in situ RNA capturing.

References

[1] Sheel Shah et al. "In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus". In: *Neuron* 92.2 (2016), pp. 342–357. DOI: 10.1016/j.neuron.2016.10.

001. URL: https://doi.org/10.1016/j.neuron.2016.10.001.

[2] Patrik L. Ståhl et al. "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics". In: *Science* 353.6294 (2016), pp. 78–82. DOI: 10.1126/science.aaf2403. URL: https://doi.org/10.1126/science.aaf2403.

[3] Reuben Moncada et al. "Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas". In: *Nature Biotechnology* nil.nil (2020), nil. DOI: 10.1038/s41587-019-0392-8. URL: https://doi.org/10.1038/s41587-019-0392-8.

[4] Emelie Berglund et al. "Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity". In: *Nature Communications* 9.1 (2018), p. 2419. DOI: 10.1038/s41467-018-04724-5. URL: https://doi.org/10.1038/s41467-018-04724-5.

[5] Michaela Asp et al. "A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart". In: *Cell* 179.7 (2019), 1647–1660.e19. DOI: 10.1016/j.cell.2019.11.025. URL: https://doi.org/10.1016/j.cell.2019.11.025.

[6] Rongqin Ke et al. "In situ sequencing for RNA analysis in preserved tissue and cells". In: *Nature Methods* 10.9 (2013), pp. 857–860. DOI: 10.1038/nmeth.2563. URL: https://doi.org/10.1038/nmeth.2563.

[7] Je Hyuk Lee et al. "Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues". In: *Nature Protocols* 10.3 (2015), pp. 442–458. DOI: 10.1038/nprot.2014.191. URL: https://doi.org/10.1038/nprot.2014.191.

[8] A. M. Femino. "Visualization of Single RNA Transcripts in Situ". In: *Science* 280.5363 (1998), pp. 585–590. DOI: 10.1126/science.280.5363.585. URL: https://doi.org/10.1126/science.280.5363.585.

[9] Eric Lubeck et al. "Single-cell in situ RNA profiling by sequential hybridization". In: *Nature Methods* 11.4 (2014), pp. 360–361. DOI: 10.1038/nmeth.2892. URL: https://doi.org/10.1038/nmeth.2892.

[10] K. H. Chen et al. "Spatially resolved, highly multiplexed RNA profiling in single cells". In: *Science* 348.6233 (2015), aaa6090–aaa6090. DOI: 10.1126/science.aaa6090. URL: https://doi.org/10.1126/science.aaa6090.

[11] Chee-Huat Linus Eng et al. "Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+". In: *Nature* 568.7751 (2019), pp. 235–239. DOI: 10.1038/s41586-019-1049-y. URL: https://doi.org/10.1038/s41586-019-1049-y.

[12] Samuel G. Rodriques et al. "Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution". In: *Science* 363.6434 (2019), pp. 1463–1467. DOI: 10.1126/science.aaw1219. URL: https://doi.org/10.1126/science.aaw1219.

[13] Sanja Vickovic et al. "High-definition spatial transcriptomics for in situ tissue profiling". In: *Nature Methods* 16.10 (2019), pp. 987–990. DOI: 10.1038/s41592-019-0548-y. URL: https://doi.org/10.1038/s41592-019-0548-y.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Lecture Notes in Computer Science*. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28. URL: https://doi.org/10.1007/978-3-319-24574-4_28.

[15] Diederik P Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *CoRR* (2013). arXiv: 1312.6114 [stat.ML]. URL: http://arxiv.org/abs/1312.6114v10.

[16] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: *CoRR* (2014). arXiv: 1401.4082 [stat.ML]. URL: http://arxiv.org/abs/1401.4082v3.

[17] Ed S. Lein et al. "Genome-wide atlas of gene expression in the adult mouse brain". In: *Nature* 445.7124 (2006), pp. 168–176. DOI: 10.1038/nature05453. URL: https://doi.org/10.1038/nature05453.

[18] Burak Tepe et al. "Single-Cell RNA-seq of Mouse Olfactory Bulb Reveals Cellular Heterogeneity and Activity-Dependent Molecular Census of Adult-Born Neurons". In: *Cell Reports* 25.10 (2018), 2689–2703.e3. DOI: 10.1016/j.celrep.2018.11.034. URL: https://doi.org/10.1016/j.celrep.2018.11.034.

[19] Roberta Bulla et al. "C1q acts in the tumour microenvironment as a cancer-promoting factor independently of complement activation". In: *Nature Communications* 7.1 (2016), p. 10346. DOI: 10.1038/ncomms10346. URL: https://doi.org/10.1038/ncomms10346.

[20] Gergana Metodieva et al. "CD74-dependent Deregulation of the Tumor Suppressor Scribble in Human Epithelial and Breast Cancer Cells". In: *Neoplasia* 15.6 (2013), 660–IN21. DOI: 10.1593/neo.13464. URL: https://doi.org/10.1593/neo.13464.

Methods

Statistical model

XFuse models the spatial expression data, X_n, and histological image data, I_n, of each sample 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 [1] on RNA bulk sequencing data, we assume the conditional distribution of the expression data X to be negative binomial. The rate of the latter is factorized into M metagenes, parameterized by a gene loading matrix L. The parameters of the conditional distributions are mapped from the latent tissue state Z through a convolutional generator network G with learnable parameters θ.

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

$$Z_n \sim \mathcal{N}(0, \mathbb{I}) \tag{1}$$
$$L_g \sim \mathcal{N}(0, \sigma_{L_g}^2 \mathbb{I}) \tag{2}$$
$$E_g \sim \mathcal{N}(0, \sigma_{E_g}^2 \mathbb{I}) \tag{3}$$
$$F_g \sim \mathcal{N}(0, \sigma_{F_g}^2 \mathbb{I}) \tag{4}$$
$$(s_n, a_n, \mu_n, \sigma_n) \equiv G_\theta(Z_n) \tag{5}$$
$$r_{ngxy} \equiv s_{nxy} e^{t_g + \beta_n E_g} \Sigma_m a_{nmxy} e^{L_{mg}} \tag{6}$$
$$p_{ng} \equiv S\left(u_g + \beta_n F_g\right) \tag{7}$$
$$X_{ngxy} \,|\, Z_n, L_g, E_g, F_g \sim \mathrm{NB}\left(r_{ngxy}, p_{ng}\right) \tag{8}$$
$$\tilde{X}_{ngl} \equiv \Sigma_{(x,y)\in\mathcal{A}_n(l)} X_{ngxy} \tag{9}$$
$$I_{ncxy} \,|\, Z_n \sim \mathcal{N}\left(\mu_{ncxy}, \sigma_{ncxy}^2\right), \tag{10}$$

where S is the logistic function, β_n is a row vector of indicator variables specifying group membership, and $\tilde{X}$ is the observed expression data at location l covering the area $\mathcal{A}_n(l)$. The fixed effects E and F can be used to control for batch effects or to characterize differential expression between sample groups.

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

$$\tilde{X}_{ngl} \,|\, Z_n, L_g, E_g, F_g \sim \mathrm{NB}\left(\Sigma_{(x,y)\in\mathcal{A}_n(l)} r_{ngxy}, p_{ng}\right). \tag{11}$$

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 $q_\phi(Z, L, E, F)$. The variational parameters ϕ and the parameters θ of the generator network are found by minimizing the Kullback-Leibler divergence from q_ϕ to the posterior, which is equivalent to maximizing the evidence lower bound (ELBO),

$$\mathcal{L}(\phi, \theta, t, u, \sigma_L^2, \sigma_E^2, \sigma_F^2) = \mathbb{E}_{q_\phi}\Big[\log p_\theta(\tilde{X}, I, Z, L, E, F) - \log q_\phi(Z, L, E, F)\Big]. \tag{12}$$

We use a mean-field diagonal Gaussian variational distribution

$$q_\phi(Z, L, E, F) = q_{\phi_L}(L) q_{\phi_E}(E) q_{\phi_F}(F) \prod_n q_{\phi_{Z_n}}(Z_n), \tag{13}$$

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

We update the parameters ϕ, θ, t, u, σ_L^2, σ_E^2, and σ_F^2 by gradient ascent on the objective (12) using the Adam optimizer [2]. Following [3], gradient estimates are obtained by reparameterizing the latent variables as a function of parameter-free noise. Briefly, letting

$$\varepsilon \sim \mathcal{N}(0, \mathbb{I}) \tag{14}$$
$$(Z, L, E, F) \equiv h_\phi(\varepsilon), \tag{15}$$

where h_ϕ is a shift-and-scale transformation such that Eqs. (1) to (4) are satisfied, we can reformulate Eq. (12) as an expectation with respect to ε by relying on the law of the unconscious statistician. This makes it straightforward to rewrite the gradient of Eq. (12) as an expectation,

$$\nabla\mathcal{L}(\phi, \theta, t, u, \sigma_L^2, \sigma_E^2, \sigma_F^2) = \mathbb{E}_{p(\varepsilon)}\Big[\nabla \log p_\theta(\tilde{X}, I, Z, L, E, F) - \nabla \log q_\phi(Z, L, E, F)\Big]. \tag{16}$$

We approximate (16) using a single Monte Carlo sample for each update step and train on patches extracted from the dataset.

The dataset is augmented with random rotations, scaling, and shearing. The image data is further augmented with random color jitter.

Architecture

To efficiently capture both global and local anatomical contexts, we model the latent tissue state Z over multiple resolutions. The recognition and generator networks G and R together form an architecture similar to U-Net [4] with the variational distribution of the latent state for each resolution inserted at the corresponding skip connection (Fig. S4).

Model selection

To select the number of metagenes M 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 (12) 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.

High-resolution gene expression

We infer denoised latent gene expression by estimating the posterior distribution of

$$\nu \equiv \mathbb{E}[X \mid Z, L, E, F] = \frac{rp}{1-p} \tag{17}$$

with N Monte Carlo samples drawn from the variational distribution:

$$\begin{aligned} p(\nu \mid \tilde{X}, I) &= \int p(\nu \mid Z, L, E, F) dP(Z, L, E, F \mid \tilde{X}, I) \\ &\approx \int p(\nu \mid Z, L, E, F) dQ_\phi(Z, L, E, F) \\ &\approx \frac{1}{N} \sum_{i=1}^{N} \delta_{\nu^{(i)}}(\nu), \end{aligned} \tag{18}$$

where P and Q_ϕ denote the cumulative distribution functions of the corresponding lower-case densities and $\delta_{\nu^{(i)}}$ is the Dirac delta function centered at the i:th sample of ν.

We compute gene expression maps as the mean of the point-mass mixture (18),

$$\mathbb{E}[\nu \mid \tilde{X}, I] \approx \frac{1}{N} \sum_{i=1}^{N} \nu^{(i)}. \tag{19}$$

To predict latent gene expression in an unseen sample n', we approximate

$$\begin{aligned} p(Z_{n'}, L, E, F \mid \tilde{X}, I, I_{n'}) &= p(Z_{n'} \mid I_{n'}) p(L, E, F \mid \tilde{X}, I) \\ &\approx q_{R_\phi(I_{n'})}(Z_{n'}) q_\phi(L, E, F) \end{aligned} \tag{20}$$

and estimate $\mathbb{E}[\nu_{n'} \mid \tilde{X}, I, I_{n'}]$ similar to Eq. (19).

Differential expression analysis

We consider the $\log_2$ conditional mean expression of an area $\mathcal{A}_i$,

$$\begin{aligned} \epsilon_i &\equiv \log_2 \mathbb{E}\left[\sum\nolimits_{(n,x,y)\in\mathcal{A}_i} X_{nxy} \mid Z, L, E, F\right] \\ &= \log_2 \sum_{(n,x,y)\in\mathcal{A}_i} \nu_{nxy}. \end{aligned} \tag{21}$$

The posterior distribution of the normalized $\log_2$ fold change of a gene g between the areas $\mathcal{A}_1$ and $\mathcal{A}_2$,

$$\eta_g \equiv \epsilon_{1g} - \epsilon_{2g} - \log_2 \sum_{g'} 2^{\epsilon_{1g'}} + \log_2 \sum_{g'} 2^{\epsilon_{2g'}}, \tag{22}$$

is estimated analogous to Eq. (18). Mean and variance estimates are computed on the resultant point-mass mixture:

$$\mathbb{E}[\eta_g \mid \tilde{X}, I] \approx \frac{1}{N} \sum_{i=1}^{N} \eta_g^{(i)} \tag{23}$$

$$\mathrm{Var}(\eta_g \mid \tilde{X}, I) \approx \frac{1}{N} \sum_{i=1}^{N} \left(\eta_g^{(i)}\right)^2 - \left(\frac{1}{N} \sum_{i=1}^{N} \eta_g^{(i)}\right)^2. \tag{24}$$

Summarized expression maps

To visualize transcriptional anatomy, we estimate the posterior mean metagene activity a and pixel-wise scale s similar to Eq. (19). We project a onto its first three principal components and append $-s$ 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 CMYK-encoded color values.

Pathway analysis

Pathway analyses are conducted using g:Profiler [5] with the GO:BP database [6, 7]. Reported p-values are adjusted with the g:SCS procedure provided by g:Profiler.

Relationship to prior work

Our work extends previous research on spatial models of transcriptomics data. Notably, SpatialDE [8] and SPARK [9] 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 [10] reconstructs the spatial organization of single cells by solving an optimal transport problem. While 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 [11, 12, 13, 14] exist for fusing single cell with in situ sequencing or hybridization data.

The contribution of our work is threefold: First, we have shown that histological image data is highly informative of spatial expression patterns in tissues. Second, we provide an integrative model of in situ capturing spatial transcriptomics. Our model fuses spatial gene expression data with high-resolution image data, thereby making it possible to study full-transcriptome expression heterogeneity in detailed anatomical structures. Third, we have demonstrated the feasibility of predicting expression in unsequenced samples using only their histological image data. We believe image-based in silico spatial transcriptomics to be a promising future research topic.

Data availability

The mouse olfactory bulb dataset was obtained from the spatial research group's website: https://www.spatialresearch.org. The breast cancer dataset was obtained from 10X Genomics: https://support.10xgenomics.com/spatial-gene-expression/datasets/.

Code availability

We have implemented XFuse in the Pyro probabilistic programming language [15]. The code is available under the MIT license at https://github.com/ludvb/xfuse.

References

[1] Michael I Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12 (2014), p. 550. DOI: 10.1186/s13059-014-0550-8. URL: https://doi.org/10.1186/s13059-014-0550-8.

[2] Diederik P. Kingma and Jimmy Ba. "Adam: a Method for Stochastic Optimization". In: *International Conference on Learning Representations (ICLR)*. 2015.

[3] Diederik P Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *CoRR* (2013). arXiv: 1312.6114 [stat.ML]. URL: http://arxiv.org/abs/1312.6114v10.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Lecture Notes in Computer Science*. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28. URL: https://doi.org/10.1007/978-3-319-24574-4_28.

[5] Uku Raudvere et al. "g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)". In: *Nucleic Acids Research* 47.W1 (2019), W191–W198. DOI: 10.1093/nar/gkz369. URL: https://doi.org/10.1093/nar/gkz369.

[6] Michael Ashburner et al. "Gene Ontology: tool for the unification of biology". In: *Nature Genetics* 25.1 (2000), pp. 25–29. DOI: 10.1038/75556. URL: https://doi.org/10.1038/75556.

[7] Gene Ontology Consortium. "The Gene Ontology resource: 20 years and still GOing strong". In: *Nucleic acids research* 47.D1 (2019), pp. D330–D338. DOI: 10.1093/nar/gky1055. URL: https://doi.org/10.1093/nar/gky1055.

[8] Valentine Svensson, Sarah A Teichmann, and Oliver Stegle. "SpatialDE: identification of spatially variable genes". In: *Nature Methods* 15.5 (2018), pp. 343–346. DOI: 10.1038/nmeth.4636. URL: https://doi.org/10.1038/nmeth.4636.

[9] Shiquan Sun, Jiaqiang Zhu, and Xiang Zhou. "Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies". In: *Nature Methods* nil.nil (2020), nil. DOI: 10.1038/s41592-019-0701-7. URL: https://doi.org/10.1038/s41592-019-0701-7.

[10] Mor Nitzan et al. "Gene expression cartography". In: *Nature* 576.7785 (2019), pp. 132–137. DOI: 10.1038/s41586-019-1773-3. URL: https://doi.org/10.1038/s41586-019-1773-3.

[11] Kaia Achim et al. "High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin". In: *Nature Biotechnology* 33.5 (2015), pp. 503–509. DOI: 10.1038/nbt.3209. URL: https://doi.org/10.1038/nbt.3209.

[12] Rahul Satija et al. "Spatial reconstruction of single-cell gene expression data". In: *Nature Biotechnology* 33.5 (2015), pp. 495–502. DOI: 10.1038/nbt.3192. URL: https://doi.org/10.1038/nbt.3192.

[13] Xiaoyan Qian et al. "Probabilistic cell typing enables fine mapping of closely related cell types in situ". In: *Nature Methods* 17.1 (2019), pp. 101–106. DOI: 10.1038/s41592-019-0631-4. URL: https://doi.org/10.1038/s41592-019-0631-4.

[14] Romain Lopez et al. "A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements". In: *CoRR* (2019). arXiv: 1905.02269 [cs.LG]. URL: http://arxiv.org/abs/1905.02269v1.

[15] Eli Bingham et al. "Pyro: Deep Universal Probabilistic Programming". In: *Journal of Machine Learning Research* 20.28 (2019), pp. 1–6. URL: http://jmlr.org/papers/v20/18-403.html.