NICHEVI: A PROBABILISTIC FRAMEWORK TO EMBED CELLULAR INTERACTION IN SPATIAL TRANSCRIPTOMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatial transcriptomics has the potential to reveal cellular interactions by measuring gene expression in situ while maintaining the tissue context of each cell. Existing deep learning methods for non-spatial single-cell omics optimize cellular embeddings of gene expression. They enable the harmonization between experimental batches while embedding the variation of the cell state. Spatial transcriptomics allows one to study the cell state composition of a spatial neighborhood. These cellular niches confine the tissue organization and encompass functional units of an organ. However, computational methods for encoding meaningful low-dimensional representations of both gene expression and cell states of neighboring cells a are currently lacking. Here, we introduce NicheVI, a deep learning model that decodes gene expression, niche cell-type composition, and variation in cell state of other cells within a niche. In case studies, NicheVI uncovered additional fine-grained heterogeneity of cell-types not captured by non-spatial and other spatially aware models and corresponding to the cellular niche.

1 INTRODUCTION

Cells in our bodies do not function in isolation. Instead, biological systems rely on cell-cell interactions to enable high-level tissue functions. The interaction of cells with their surroundings orchestrates high-order phenomena, such as the immune response or the maintenance of homeostasis. Therefore, the characterization of the complex cellular interplay has long been a scientific endeavor (Wagner et al., 2016; Palla et al., 2022).

Methods have been developed to quantify transcripts in situ. These spatial transcriptomics technologies (ST) unravelled a deeper understanding of how tissues are organized by revealing the cellular transcriptional state and its spatial context. Image-based ST technologies, such as Vizgen MERSCOPE, 10X XENIUM, and Nanostring CosMx, can target up to a few thousand genes while retaining cellular resolution.

These technologies enable researchers to study individual cell contributions to higher-level tissue and organ functions, such as the mechanisms by which tumor-infiltrating leukocytes are imprinted by the local tumor microenvironment to drive tolerance against the tumor. The location of a cell within a tissue shapes cell states through cellular interactions and metabolic gradients. Our aim is to learn a low-dimensional representation of each cell that captures both cell-type heterogeneity and niche variation. Such a representation can help reveal dependencies between gene expression and the tissue microenvironment, defined as the immediate neighbors of a cell. Clustering this spatially aware cell representation reveals subsets of cells with slight difference in transcriptome but strong differences in their microenvironment. Uncovering these variations in expression correlated with the spatial niche helps to uncover novel functional cell states and might improve our understanding of cell phenotypes in disease.

Related work. Various methods have been developed to learn representations of non-spatial single-cell RNA-seq data, such as scVI (Lopez et al., 2018) and related models (Gayoso* et al., 2022). The low-dimensional latent space learned by these variational auto-encoders (VAE) is used to extract information from single-cell omics, and its embeddings can be used for downstream

analysis, e.g. to define clusters of cell types. It also provides a principled way of harmonizing information from different studies. The generative part of the VAE is used for hypothesis testing in the data domain, e.g., for differential gene expression analysis or gene-gene coexpression.
A few methods have recently adapted this VAE scheme to spatial transcriptomics, allowing learning a spatially aware latent space. Haviv et al. (2023) introduced ENVI, a model that jointly embeds paired scRNA sequencing and image-based spatial transcriptomics data in a single latent space. One network decodes the transcriptome-wide gene expression, and a second network decodes the environmental context, modeled as the local gene covariance matrix. In this way, ENVI latent space incorporates spatial context and full transcriptome information. Dong et al. (2023) presented SIMVI, which embeds the gene expression of a cell and the gene expression of its niche in intrinsic and spatially induced latent spaces. It aims to disentangle intrinsic and spatial variations by minimizing the mutual information between both embeddings.

Our contributions. Here, we present NicheVI, a probabilistic deep-learning tool that dissects spatial transcriptomics data into spatially aware niche phenotypes at the single cell level. We showcase NicheVI's performance to detect cells differing in their spatial niche (nichetypes) by dissecting murine microglia into fine-grained regional subsets that correspond to microglia within human annotated regions of the murine brain. In a second case study, NicheVI detected two subsets of cytotoxic T cells in breast cancer tissues located close to highly invasive or relatively benign tumor subtypes, respectively. Differential gene expression analysis between T cells close to invasive and non-invasive tumor types revealed up-regulation of programs in CD8 T cells leading to immune evasion.

2 RESULTS

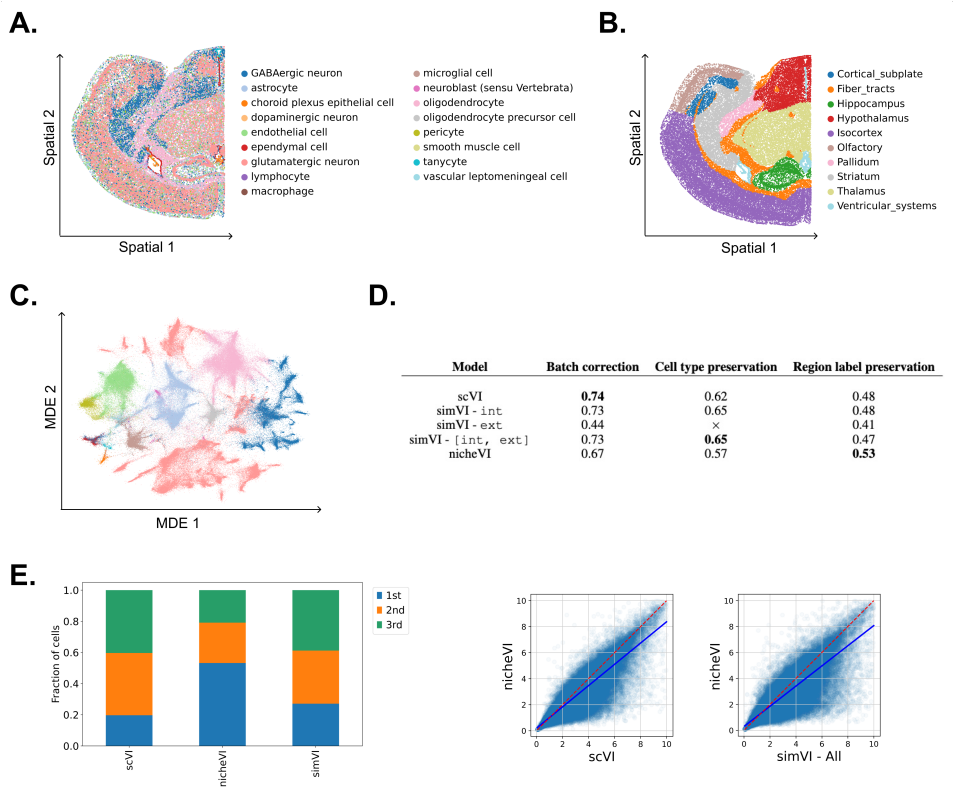
2.1 NICHEVI LEARNS A CELLULAR EMBEDDING REFLECTING GENE EXPRESSION AND CELLULAR NICHE.

We consider a single cell resolution spatial transcriptomics data set, which provides the following measurements for N cells and G genes. For each cell n, we denote by $x_n \in \mathbb{N}^G$ the vector of observed gene expression in the cell. We assume that each cell is annotated and we denote by c_n its discrete type assignment, taking values in $\mathcal{T} = \{1, ..., T\}$, where T is the total number of cell types in the data set. We also assume knowledge of batch annotations (donor, sample ID, etc.) stored in a vector s_n.
Furthermore, the experiment provides cell coordinates for each cell, $y_n \in \mathbb{R}^2$. We take the K nearest neighbors of a cell to define its niche using the Euclidean distance in physical space. We characterize the niche by its cell-type composition and gene expression. We denote by α_n the T dimensional vector of cell type proportions among the K nearest neighbors of the cell n. Its values are in the probability simplex. The matrix $\eta_n \in \mathbb{R}^{T \times D}$ characterizes the gene expression profiles of each cell type, embedded in a D-dimensional space. We assume that $D \ll G$, and obtain η_n via dimensionality reduction (PCA, scVI or similar). See A.1 for details.

The proposed graphical model is shown in Figure 4c. The latent variable z_n is a vector of dimension P following a normal distribution, $\mathcal{N}(0, I_P)$. Here, P is a hyperparameter that represents the dimension of the latent embedding. NicheVI encodes cell expression x_n and decodes the expression and composition of the niche α_n and the activation state of all cell types in the niche η_n, while treating the batch s_n as a covariate. We condition the decoder on s_n similar to the approach in scVI. Therefore, the latent variable z_n captures the intrinsic and spatial sources of variation of the cell in a batch-independent manner.
Finally, we model the observed gene expressions x_n as samples from the Poisson distribution whose rate is determined by a neural network taking both z_n and s_n as input. We estimate the posterior distribution of z_n using amortized variational inference, allowing us to approximate this latent variable for each individual cell. All parameters of the model and of the variational distributions are learned via evidence lower bound maximization (Kingma & Welling, 2013).

2.2 DELINEATING REGION-SPECIFIC MICROGLIA POPULATIONS IN THE MOUSE BRAIN

To demonstrate the ability of NicheVI to learn meaningful spatially aware embeddings, we first consider the brain data set of Zhang et al. (2023), which imaged a panel of 1100 genes across the entire adult mouse brain using MERFISH. The slides are whole hemispheres annotated with main cell types (Figure 1a) and anatomical regions (Figure 1b). We use scVI to learn expression embeddings eta_n (Figure 1c). We evaluated NicheVI against scVI and simVI, using the *single cell integration benchmark* developed by Luecken et al. (2022). The metric details can be found in A.3. For the benchmark, scVI acts as a non-spatial baseline, and we considered three flavors of simVI: we first used the intrinsic and spatial variation embeddings separately and then used a concatenation of the two. Considering individual slides as batch covariates, we find that learning neighborhood properties makes batch alignment harder, as scVI is the top-performing method for batch correction. simVI has the best performance for capturing cell-type, while NicheVI has the best performance in anatomical region label preservation, while maintaining the cell type information (Figure 1d). We also introduce the niche composition divergence to quantify to what extent cells within similar micro-environments are embedded together and show the best performance for NicheVI (Figure 1e).



Model	Batch correction	Cell type preservation	Region label preservation
scVI	**0.74**	0.62	0.48
simVI - `int`	0.73	0.65	0.48
simVI - `ext`	0.44	×	0.41
simVI - `[int, ext]`	0.73	**0.65**	0.47
nicheVI	0.67	0.57	**0.53**

Figure 1: Benchmarking on the MERFISH mouse brain. **A.** Example tissue slice colored by cell type labels. **B.** Slice annotated by major regions. **C.** We use scVI to embed cell expression. Cell types are separated in the Minimum Distorsion Embedding space (Agrawal et al., 2021). **D.** SCIB metrics benchmarking scVI, nicheVI and 3 flavours of simVI: `int` = intrinsinc embedding, `ext` = spatial-induced, `[int, ext]` = concatenation of the two. Scores range between 0 and 1. In the following, we use the `[int, ext]` version of simVI as it performed best in the benchmark. **E.** Latent neighbors niche composition divergence. We consider the 100-nn neighborhood of each cell in NicheVI latent space and compute the average divergence between niche compositions (Section A.3). Smaller values indicate better integration of cells with similar context. *Left:* Summary of method ranks, we see that the divergence computed from nicheVI latent is lowest for 53% of the cells. *Right:* Scatter plot of NicheVI against scVI and simVI. We cap the divergence to 10 in the figure for better visualization. Dashed line is identity, solid blue line is fitted to the data. Slopes smaller than one indicate a trend for smaller values for NicheVI.

We then evaluated to what extent the nicheVI embedding, when restricted to a single cell type, can be clustered by region. To do so, we computed the Normalized Mutual Information (NMI) between Leiden clustering and region class assignments. Overall, NicheVI exhibited superior performance (Figure 5). Next, we set out to study whether capturing niche information allows novel insights into cell states. We used microglia to further dissect this, as this cell state is present throughout the murine brain and it has been shown that this cell type has regional-dependent gene expression. Indeed, microglia cell states are resolved by all three methods (Figure 2a), while NicheVI detected additional nichetypes. For example, NicheVI identified a microglia niche type (group 7) that was located primarily in the hippocampus (Figure 2c). This cluster was characterized by the expression of *Rreb1*, which has been demonstrated to be downregulated in microglia under LPS stimulation (Freilich et al., 2013) and the expression of *Tgfbr2*, which has been shown to be associated with a quiescent state in microglia. Therefore, we find that microglia in the hippocampus are least activated compared to other regions of the brain. Studying these microglia cells during pathological aging could reveal how these cells become pro-inflammatory in Alzheimer's disease. Another cluster delineated microglia in the striatum (cluster 9) and showed upregulation of (*Nfatc1*), which is important for the reaction of microglia to LPS stimulation demonstrating a more activated state (Nagamoto-Combs & Combs, 2010). In addition, we find an up-regulation of *Col27a1* in cluster 5, which is associated with surveillant microglia (Uyar et al., 2022).
We have demonstrated here that nicheVI can uncover novel cell states associated with spatial distribution of microglia. Studying these cell states during disease might help uncover their functional role.

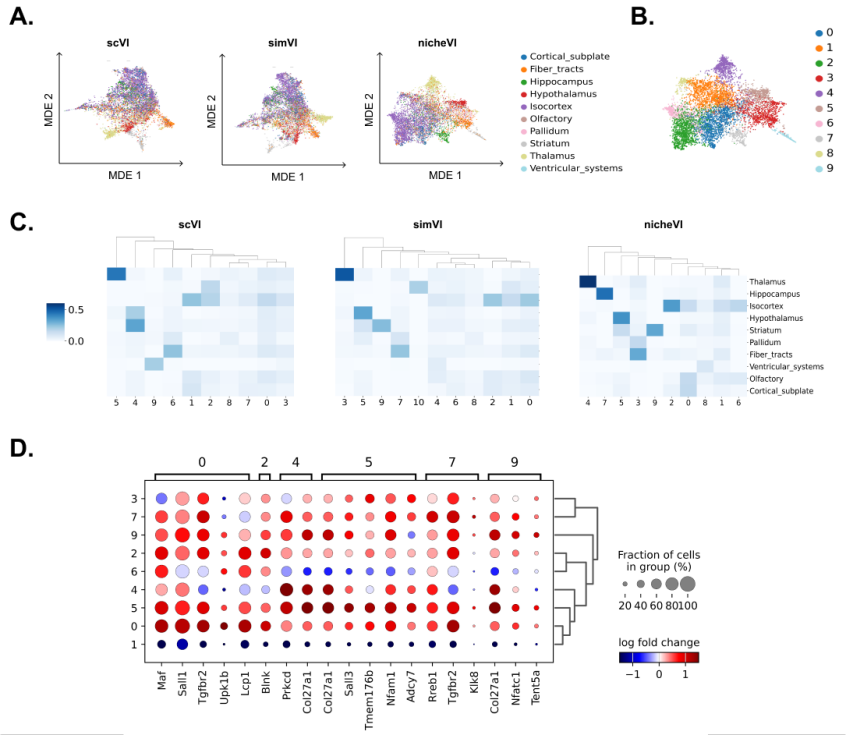


Figure 2: Study of microglial cells in the mouse brain. **A.** Separation of microglia into main regions in the MDE space. **B.** Leiden clustering of NicheVI microglia embedding. **C.** Confusion matrix showing overlap of Leiden clusters with main brain regions, measured by Jaccard index. NicheVI identifies more region-sepcific clusters. **D.** Differential expression analysis of each microglia cluster from NicheVI.

2.3 CHARACTERIZATION OF T CELLS SUBTYPES IN BREAST CANCER

Next, we applied NicheVI to a 10X Xenium data set of human breast cancer sections (Figure 3a). Janesick et al. (2023) identified distinct tumor domains, corresponding to in situ ductal carcinoma (DCIS1 / 2) and invasive tumor (Figure 3b).

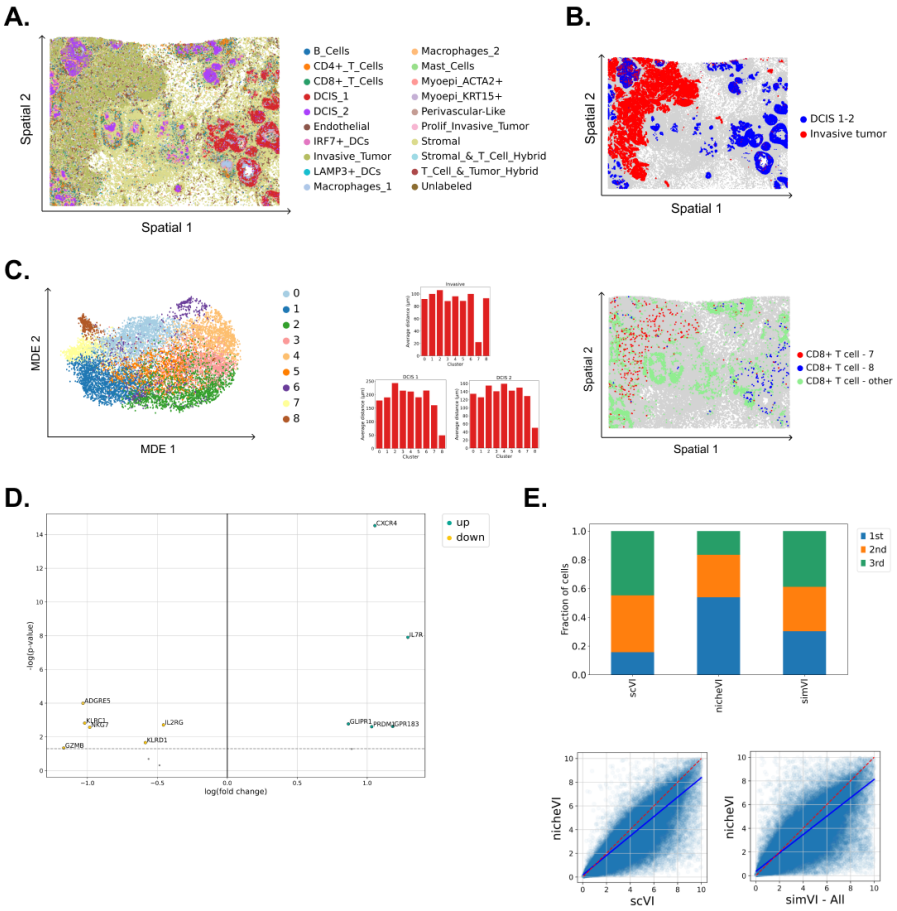


Figure 3: Analysis of the XENIUM breast cancer samples. **A.** Tissue colored by cell type labels. **B.** We highlight the main two cancer regions - ductal carcinoma in situ and invasive. **C.** *Left:* Leiden clustering of nicheVI CD8+T cell embedding. *Center:* Average distance to the 3 tumor regions for cells in each cluster. We identify cluster 7 as close to invasive tumor, while cluster 8 is close to DCIS 1 and 2. *Right:* Spatial plot of the Leiden clusters 7 and 8 overlap with the two cancer regions. **D.** Diffential expression analysis betwen the T cells in the DCIS regions and the T cells in the invasive region. **E.** Latent neighbors niche composition divergence. We consider the 100-nn neighborhood of each cell in nicheVI latent space and compute the average divergence between niche compositions (Section A.3). Smaller values indicate better integration of cells with similar context. *Top:* Summary of method ranks, we see that the divergence computed from nicheVI latent is lowest for 54% of the cells. *Bottom:* Scatter plot of nicheVI against scVI and simVI. We cap the divergence to 10 in the figure for better visualization. Dashed line is identity, solid blue line is fitted to the data. Slopes smaller than one indicate a trend for smaller values for nicheVI.

We focussed our analysis on CD8 T cells, which are essential for antitumor response. NicheVI dissected cytotoxic T cells into meaningful niche types characterized by unique tissue localization and gene expression programs. Specifically, NicheVI identified two clusters of CD8 T cells that were located close to two distinct molecular tumor subtypes: cluster 8 resided in close proximity to DCIS1 and DCIS2 pre-malignant cells (Figure 3c), a cancer precursor stage that can develop into invasive disease (Janesick et al., 2023). In contrast, cells of cluster 7 were closely associated with invasive

tumor cells. Differential gene expression between cluster 7 and cluster 8 (Figure 3d) revealed that cytotoxic T cells in the vicinity of precursor tumor cells DCIS1/2 demonstrated increased expression of genes encoding for the cytotoxic molecules *NKG7*, *KLRG1*, *GZMB* and *IL2RG*, which are essential for a productive antitumor response (Wang et al., 2021; Li et al., 2022b). On the contrary, CD8 T cells of cluster 7 close to invasive tumor cells demonstrated increased expression of the chemokine receptor *CXCR4* and *PRDM1*, which has been associated with tumor escape mechanisms (Li et al., 2022a). Pharmacological inhibition of *CXCR4* is currently being evaluated for the treatment of breast cancer (Steele et al., 2023). Thus, NicheVI enable fine-grained dissection of CD8 T cells and may provide a causal link between the identified niche types and anti-tumor response.

Finally, we compare NicheVI to our benchmark methods in terms of their ability to embed cells in the breast cancer TME in a meaningful manner. Using the niche composition divergence criterion as in the brain, we find that the NicheVI latent space indeed better reflects the immediate microenvironments of the cells (Figure 3e).

3 Discussion

Spatial transcriptomics allows one to study cellular niches and understand spatially-confined units inside tissues. However, using embeddings that are not aware of the spatial context obscures those intricacies of gene expression that is dependent on neighboring cells. We have introduced here a novel method to embed cells into a latent space that is not only informative of the gene expression of a specific cell but also the composition and expression of cell-types surrounding a specific cell. We have demonstrated that NicheVI can uncover novel nichetypes in the mouse brain and found several changes in transcription associated with those niches.

NicheVI uses the gene expression of the center cell to learn the latent embedding of a cell by an encoder. Therefore, it will not learn the spatial niche of a cell for which gene expression and niche are independent. NicheVI's clusters can therefore be interpreted as gene expression changes of the centering cell correlating with niche structure, and the network learns the mapping of gene expression to the local spatial composition. Instead, simVI uses the cellular composition of neighboring cells to encode a spatially aware latent space and therefore learns the relationship of the cellular composition of the neighborhood and the cell itself. Therefore, NicheVI does not provide insight into the effect of neighboring cells on gene expression, but instead uncovers subtle changes in gene expression between cells in different niches. Identifying these cell types will improve our understanding of cell biology. To underline this, we identified that CD8 T cells close to in situ carcinoma are highly cytotoxic, while cells in established cancers show signs of immune dysfunction. Further studying this effect will provide insights into how cancers can suppress and bypass effective antitumor immunity.

REFERENCES

Akshay Agrawal, Alnur Ali, and Stephen Boyd. Minimum-distortion embedding. *arXiv*, 2021.

Pierre Boyeau, Stephen Bates, Can Ergen, Michael I Jordan, and Nir Yosef. Calibrated identification of feature dependencies in single-cell multiomics. *bioRxiv*, pp. 2023–11, 2023.

Mingze Dong, Harriet Kluger, Rong Fan, and Yuval Kluger. Simvi reveals intrinsic and spatial-induced states in spatial omics data. *bioRxiv*, 2023.

Robert W Freilich, Maya E Woodbury, and Tsuneya Ikezu. Integrated expression profiles of mrna and mirna in polarized primary murine microglia. *PLOS ONE*, 8(11), 2013.

Adam Gayoso*, Romain Lopez*, Galen Xing*, Pierre Boyeau, Katherine Wu, Michael Jayasuriya, Edouard Mehlman, Maxime Langevin, Yining Liu, Jules Samaran, Gabriel Misrachi, Achille Nazaret, Oscar Clivio, Chenling Xu, Tal Ashuach, Mariano Gabitto, Mohammad Lotfollahi, Valentine Svensson, Eduardo da Veiga Beltrame, Vitalii Kleshchevnikov, Carlos Talavera-Lopez, Lior Pachter, Fabian J Theis, Aaron Streets, Michael I. Jordan, Jeffrey Regier, and Nir Yosef. A python library for probabilistic analysis of single-cell omics data. *Nature Biotechnology*, March 2022.

Doron Haviv, Mohamed Gatie, Anna-Katerina Hadjantonakis, Tal Nawy, and Dana Pe'er. The covariance environment defines cellular niches for spatial inference. *bioRxiv*, 2023.

Amanda Janesick, Robert Shelansky, Andrew D Gottscho, Florian Wagner, Stephen R Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A Morrison, Michelli F Oliveira, Jordan T Sicherman, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nature Communications*, 14(1):8353, 2023.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Q Li, L Zhang, W You, et al. Prdm1/blimp1 induces cancer immune evasion by modulating the usp22-spi1-pd-l1 axis in hepatocellular carcinoma cells. *Nature Communications*, 13(1):7677, 2022a.

Xian-Yang Li, Dillon Corvino, Bianca Nowlan, Amelia Roman Aguilera, Susanna S Ng, Matthias Braun, Anthony R Cillo, Tobias Bald, Mark J Smyth, and Christian R Engwerda. Nkg7 is required for optimal antitumor t-cell immunity. *Cancer Immunology Research*, 10(2):154–161, 2022b.

Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, December 2018.

Malte D Luecken, Maren Büttner, Kridsadakorn Chaichoompu, Anna Danese, Marta Interlandi, Michaela F Müller, Daniel C Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, et al. Benchmarking atlas-level data integration in single-cell genomics. *Nature methods*, 19(1):41–50, 2022.

Kumi Nagamoto-Combs and Colin K Combs. Microglial phenotype is regulated by activity of the transcription factor, NFAT (nuclear factor of activated T cells). *The Journal of Neuroscience*, 30 (28):9641–9646, 2010.

Giovanni Palla, David S Fischer, Aviv Regev, and Fabian J Theis. Spatial components of molecular tissue biology. *Nature Biotechnology*, 40(3):308–318, 2022.

MM Steele, A Jaiswal, I Delclaux, et al. T cell egress via lymphatic vessels is tuned by antigen encounter and limits tumor control. *Nature Immunology*, 24:664–675, 2023.

O Uyar, JM Dominguez, M Bordeleau, L Lapeyre, FG Ibáñez, L Vallières, ME Tremblay, J Corbeil, and G Boivin. Single-cell transcriptomics of the ventral posterolateral nucleus-enriched thalamic regions from hsv-1-infected mice reveal a novel microglia/microglia-like transcriptional response. *Journal of Neuroinflammation*, 19(1):81, Apr 2022.

Allon Wagner, Aviv Regev, and Nir. Yosef. Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology*, 34(11):1145–1160, 2016.

Huiling Wang, Yong Huang, Jian He, Liping Zhong, and Yongxiang Zhao. Dual roles of granzyme b. *Scandinavian Journal of Immunology*, 94(3):e13086, 2021.

Meng Zhang, Xingjie Pan, Won Jung, Aaron Halpern, Stephen W Eichhorn, Zhiyun Lei, Limor Cohen, Kimberly A Smith, Bosiljka Tasic, Zizhen Yao, et al. A molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature*, pp. 343–354, 2023.

A APPENDIX

A.1 NICHEVI DETAILS

Descriptive model. We propose a latent variable model for spatial transcriptomics assay aiming to capture both gene expression heterogeneity and spatial variation resulting from the microenvironment.

In particular, we assume these two sources of variability are both captured by a P-dimensional latent variable

$$z_n \sim \mathcal{N}(0, I_P), \tag{1}$$

with $P \ll G$. As our model is general for any cell n, we drop the index notation such that $\{z_n, s_n, \alpha_n, x_n, \eta_n\} = \{z, s, \alpha, x, \eta\}$. We model the observed expression profile x as follows:

$$x|z, s \sim \mathbf{Poisson}\left(g(z, s)\right). \tag{2}$$

Here, g is a non-linear function with output in $(\mathbb{R}+)^T$, where T is the total number of cell-types in the data.

We also allow z to capture *compositional* and *expression* variability of neighboring cells. In particular, we assume that the cell-type proportions of the cell's K nearest neighbors are obtained as

$$\alpha|z \sim \mathbf{Dirichlet}\left(f(z)\right), \tag{3}$$

where f is a non-linear function with output in $(\mathbb{R}+)^T$. Last, we assume that the neighboring cells' average expression profiles are obtained as

$$\eta_t|z, s, \alpha \sim \begin{cases} \mathcal{N}\left(h_t^1(z, s), diag(h_t^2(z, s))\right), & \text{if } \alpha_t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $t = 1, ..., T$ and h^1, h^2 are non-linear functions with outputs in $\mathbb{R}^D$ and $\mathbb{R}_+^{D\times D}$ respectively.

Variational inference recipe. We present the evidence lower bound decomposition for one cell. The joint probability distribution factorizes as:

$$p(z, \alpha, x, \eta \mid s) = p(\alpha \mid z)p(x \mid z, s)p(\eta \mid z, s, \alpha)p(z). \tag{5}$$

We use variational inference to approximate the posterior $p(z|x, s)$. The variational approximation $q\left(z \mid x, s\right)$ is chosen to be Gaussian with a diagonal covariance matrix, with parameters given by an encoder network applied to (x, s). The variational lower bound is:

$$\begin{aligned}
\log p\left(\alpha, x, \eta \mid s\right) = \log \mathbb{E}_{q(z|x,s)}&\left[\frac{p\left(z, \alpha, x, \eta \mid s\right)}{q\left(z \mid x, s\right)}\right] \\
\geq \mathbb{E}_{q(z|x,s)}&\left[\log \frac{p\left(z, \alpha, x, \eta \mid s\right)}{q\left(z \mid x, s\right)}\right] \\
\geq -D_{KL}(q(z|x, s)\|p(z)) &+ \mathbb{E}_{q(z|x,s)}\left[\log p(\alpha \mid z)\right] + \mathbb{E}_{q(z|x,s)}\left[\log p(x \mid z, s)\right] \\
&+ \mathbb{E}_{q(z|x,s)}\left[\log p(\eta \mid z, s, \alpha)\right].
\end{aligned}$$

Assuming $\{\eta_t|z, s, \alpha\}_{t\in\mathcal{T}}$ are mutually independent, we have

$$\log p(\eta \mid z, s, \alpha) = \sum_{t\in\mathcal{T}} \log p(\eta_t \mid z, s, \alpha). \tag{6}$$

We use analytical expression for the Kullback-Leibler divergences as well as the log-likelihoods to optimize the lower bound.

The standard VAE terms are

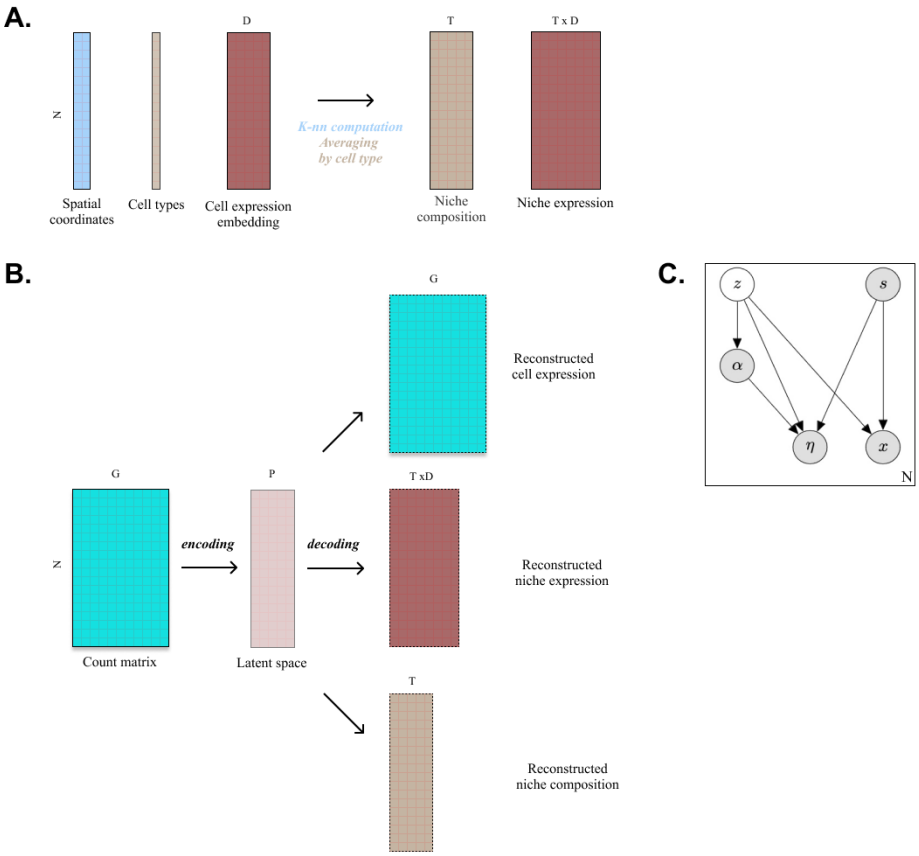


Figure 4: NicheVI model. **A.** In addition to the cell gene expression vector, we have access to spatial coordinates and assume the availability of cell type annotation and gene expression embedding. We define the cell niche by K nearest neighbors and compute the vector of cell type proportions within the niche. We also retrieve the K gene expression embeddings of the neighbors and average them by cell type, so that we get a $T \times D$ matrix. If a cell type is not present in the niche, its entry is set to zero. **B.** NicheVI uses a probabilistic auto-encoder architecture to encode the cell gene expression, and reconstruct the cell expression and the niche properties. **C.** The underlying graphical model. Shaded nodes represent observed random variables. Empty represent latent random variables. Edges signify conditional dependency. The rectangle represent independent replication across cells.

$$\mathcal{L}_{ELBO} = -D_{KL}(q(z|x,s)\|p(z)) + \mathbb{E}_{q(z|x,s)}\left[\log p(x \mid z, s)\right], \tag{7}$$

while the spatial contribution is

$$\mathcal{L}_{niche} = \mathbb{E}_{q(z|x,s)}\left[\log p(\alpha \mid z)\right] + \mathbb{E}_{q(z|x,s)}\left[\log p(\eta \mid z, s, \alpha)\right]. \tag{8}$$

Introducing $\beta > 0$ as the spatial regularization strength, we write the total loss as

$$\mathcal{L} = \mathcal{L}_{ELBO} + \beta\, \mathcal{L}_{niche}. \tag{9}$$

A.2 Model architecture and training

We report below the architectures of the modes used. The architectures are constant across datasets.

scVI The model has 1 hidden layer with 128 neurons and a 10-dimensional latent space. It is trained for 400 epochs, with a batch size of 128 cells. The learning rate is 0.001 and we use Adam as the optimizer (Kingma & Ba, 2014). We keep a constant KL weight of 1.

NicheVI The niche cell type composition and average expression are computed using the K nearest neighbors of each cell, in our experiments we fixed $K = 20$. We add to the scVI architecture a niche composition decoder as well as a niche expression decoder. Both have 1 hidden layer of size 128. During training, we keep a constant KL weight of 1 and the spatial regularization strength is set to $\beta = 10$. We use scVI to provide cell embeddings.

simVI The spatial graph is also built using a K-nn scheme with 20 neighbors. We kept default settings for the architecture: two hidden layers for the gene expression encoder and one for the decoder, one graph attention layer for the spatial encoder. Hidden layer size is 128. During training, mutual information regularization strength is kept to its default value of 5 and the two KL terms are kept to 1.

A.3 METRICS

Batch correction and label preservation. We evaluate NicheVI integration performance using the Jax implementation of Luecken et al. (2022) benchmark methodology. Specifically, the batch correction score is the average of the *iLISI, KBET, Silhouette batch* and *Graph connectivity* metrics, and the biological conservation score is computed from the *cLISI, Silhouette label, KMeans NMI/ARI*. We benchmark our embeddings with the cell type, and then with the region annotations as labels.

Divergence in niche composition between latent neighbors. A desirable property of our model is the embedding of cells with similar spatial contexts. Let us first define $\mathcal{E}_{z_n}$ to be the set of K' nearest neighbors in the latent space. One characterization of a cell spatial context is the vector of niche cell type proportions. For a cell n, we denote it α_n. We therefore want to compare this quantity with the niche cell type proportions of each of the latent nearest neighbors $\mathcal{E}_{z_n}$. Our metric ζ_n is defined as the average KL divergence between α_n and the proportions of niche cell types of each of the cells in $\mathcal{E}_n$:

$$\zeta_n = \text{mean}(\mathbb{KL}(\alpha_n, \alpha_i), i \in \mathcal{E}_{z_n}). \tag{10}$$

.

A.4 DATASETS

All data sets used in this study are publicly available. We provide download links and processing steps below.

MERFISH mouse brain atlas. Zhang et al. (2023) imaged a panel of 1100 genes across the entire adult mouse brain with the MERFISH technology. For two different mice, we selected three consecutive slices, gathering in total six slices and 250k cells. We used the provided cell type and major brain region annotations. The data can be downloaded from the CellXGene portal. Our subset corresponds to slide indices $'C57BL6J-1.074','C57BL6J-1.077','C57BL6J-1.079','C57BL6J-2.035','C57BL6J-2.036','C57BL6J-2.037'$.

XENIUM human breast cancer. Using the 10X XENIUM In Situ platfrom in human breast cancer samples, Janesick et al. (2023) studied tumor invasion in ductal carcinoma in situ (DCIS). Two tumor samples were analyzed with a panel of 313 genes, for a total of 282k cells. The authors defined three different tumor domains: two types of DCIS and invasive tumor. The data can be downloaded from 10x Genomics.

A.5 DIFFERENTIAL EXPRESSION ANALYSIS DETAILS

In both the brain and breast cancer studies, we conducted a differential expression analysis between different clusters of a single cell type (microglia, T cell). However, the limitation in cellular segmentation of our technologies implies that some genes not expressed in a given cell type might still be detected. Following Boyeau et al. (2023), we restrict our analysis to genes with higher expression in the cell type of interest than other cell types. We identified these genes by one-vs-all differential expression. For the microglia analysis, we selected genes with log-fold change higher than 2 and adjusted p-value for mean expression t-test higher than 0.05. This gives a list of 49 genes. For the CD8+ T cells, we selected genes with log-fold change higher than 0.3 and adjusted p-value higher than 0.05, for a total of 65 genes.

B SUPPLEMENTARY FIGURES

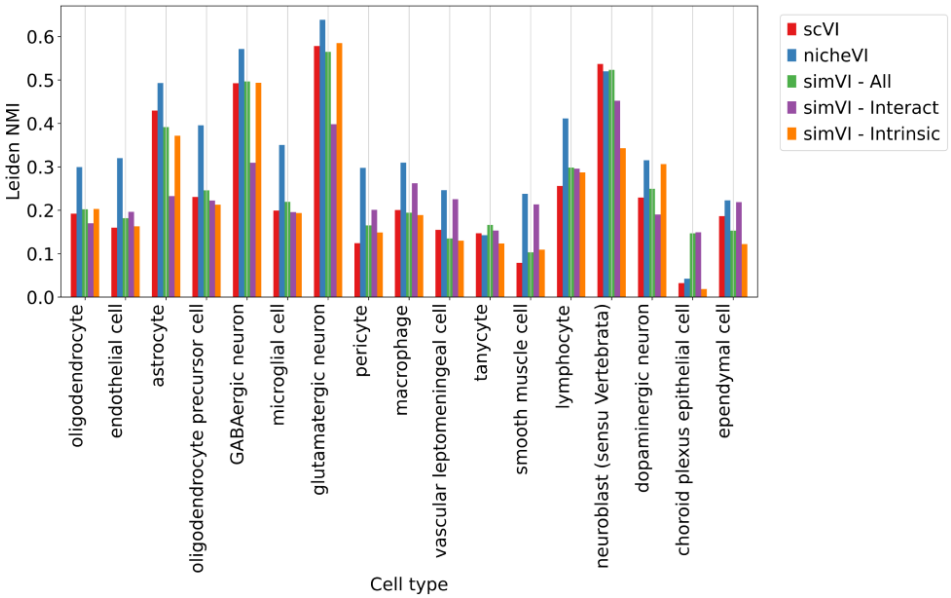


Figure 5: **A.** Normalised Mutual Information between Leiden clustering and main brain region annotations. We compared NicheVI with scVI and three flavours of simVI. For most of the cell types, NicheVI peforms best. We tested the significance using a related t-test between nicheVI and each method. $p_{scVI} = 1.33e-05$, $p_{simVI-All} = 1.44e-4$, $p_{simVI-Int} = 4.42e-07$, $p_{simVI-Ext} = 7.34e-4$.