Spatio-relational inductive biases in spatial cell-type deconvolution

Ramon Viñas* [1] **Paul Scherer*** [1] **Nikola Simidjievski** [1] **Mateja Jamnik** [1] **Pietro Liò** [1]

Abstract

Spatial transcriptomic technologies profile gene expression *in-situ*, facilitating the spatial characterisation of molecular phenomena within tissues, yet often at multi-cellular resolution. Computational approaches have been developed to infer fine-grained cell-type compositions across locations, but they frequently treat neighbouring spots independently of each other. Here we present GNN-C2L, a flexible deconvolution approach that leverages proximal inductive biases to propagate information along adjacent spots. In performance comparison on simulated and semi-simulated datasets, GNN-C2L achieves increased deconvolution performance over spatial-agnostic variants. We believe that accounting for spatial inductive biases can yield improved characterisation of cell-type heterogeneity in tissues.
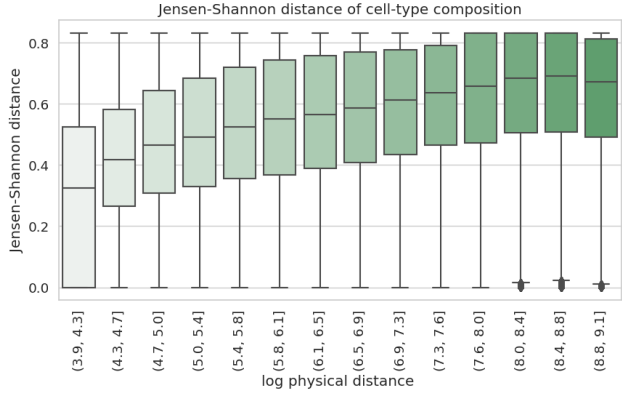
Figure 1: Jensen-Shannon distance of cell-type proportions by spot distance in Xenium dataset (breast cancer, convolved spots of size 50μm). Closer spots tend to exhibit similar cell-type composition.

1. Introduction

Analysing the spatial organisation of cells within a tissue can shed light on fundamental biological processes, including intercellular communication (Fischer et al., 2023) and organogenesis (Lohoff et al., 2022), and mechanisms of diseases like cancer, diabetes, and autoimmune disorders (Solinas et al., 2007; Wang et al., 2013; Vlahopoulos et al., 2015). Spatial transcriptomics technologies have recently enabled gene expression profiling *in situ*, but they often lack single-cell resolution, impeding fine-grained characterisation of cellular heterogeneity and effective reconstruction of tissue architectures.

Computational approaches for cell-type deconvolution in spatial transcriptomics offer a scalable solution to these challenges. These strategies often identify resident cell types from the RNA sequencing of dissociated single cells, yielding cell-type-specific gene expression signatures, and then infer the cell-type composition of every profiled spot (Elosua-Bayes et al., 2021; Cable et al., 2022; Dong & Yuan, 2021). A cutting-edge method in this family is Cell2Location (Kleshchevnikov et al., 2022), a Bayesian deconvolution approach that captures cell-type relationships through a hierarchical model and handles technical sources of variation like differences in mRNA detection sensitivity. Despite numerous benefits, however, existing deconvolution approaches treat spots independently of each other.

In this study, we investigate whether incorporating spatio-relational information leads to improved cell-type mapping. Building on the observation that neighbouring spots often exhibit similar cell-type compositions (Figure 1), we extend Cell2Location (C2L) to incorporate spatial inductive biases. Our approach, named GNN-C2L, propagates learnable messages on the proximity graph of spot transcripts, effectively leveraging the spatial relationships between spots and exploiting the co-location of cell-types (Figure 2). We conduct an extensive ablation study on synthetic and real spatial transcriptomics datasets and show improved deconvolution performance of GNN-C2L over spatial-agnostic variants. Altogether, our work leverages proximal inductive biases to facilitate enhanced reconstruction of tissue architectures.

---

*Equal contribution [1]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. Correspondence to: Paul Scherer <paul.scherer@cl.cam.ac.uk>, Ramon Viñas Torné <rv340@cam.ac.uk>.
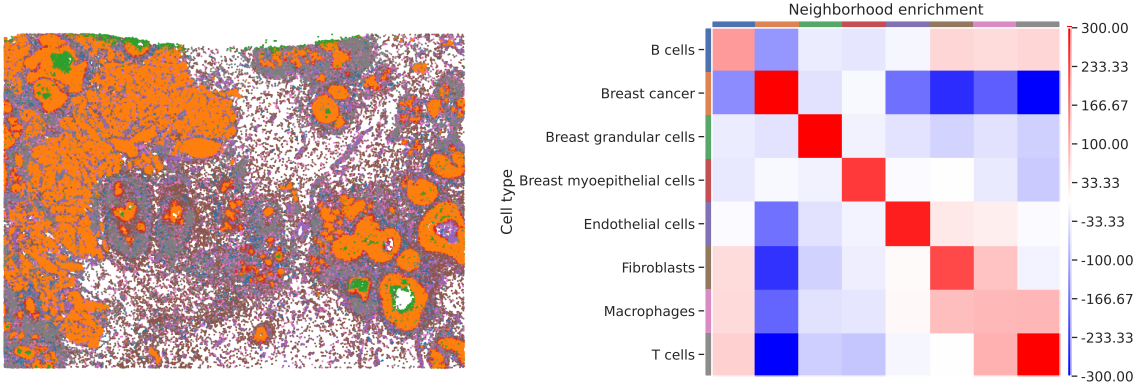
Figure 2: Neighbourhood enrichment analysis on the Xenium dataset (breast cancer). The color legend is given by the y-axis of the neighbourhood heatmap. (*left*) Spatial transcriptomics data colored by cell-type. (*right*) Neighbourhood enrichment z-scores (red and blue indicate enrichment and depletion in the neighbourhood of nearest neighbours, respectively). Cells from the same cell-type tend to co-locate (e.g. breast cancer cells). Immune cells — including T cells, B cells, and macrophages — work in conjunction to modulate the anti-cancer immune response (Gonzalez et al., 2018). Utilising relational inductive biases could therefore enhance the effectiveness of spatial deconvolution models, thereby improving the characterisation of tumor microenvironments at different stages of cancer progression.

2. Methodology

Problem formulation Let $D \in \mathbb{R}^{S \times G}$ denote a count matrix of RNA reads captured at S spots for G genes, using one or multiple batches (e.g. 10x Visium slides or Slide-seq pucks). Let $d_{s,g}$ be the entry of this matrix with the number of reads for gene g in spot s. Let $C \in \mathbb{R}^{F \times G}$ denote a matrix of F reference cell-type signatures for the same set of G genes (e.g. these signatures can be obtained from dissociated single-cell RNA-seq, Appendix A). Denote by $c_{f,g}$ the expression of gene g in signature f. Given the count matrix D and cell-type signatures C, our goal is to infer the cell-type composition $X \in \mathbb{R}^{S \times F}$ of every spot.

Cell2Location Our relational approach builds on Cell2Location (Kleshchevnikov et al., 2022, Appendix A), which models the per-spot read counts D as Negative Binomial (NB) distributed:

$$d_{s,g} \sim \mathrm{NB}(\mu_{s,g}, \alpha_{e,g})$$

where $\alpha_{e,g}$ is an experiment- and gene-specific over-dispersion parameter and the unobserved expression rate $\mu_{s,g}$ is modelled as a linear function of the reference cell-type signatures $c_{f,g}$:

$$\mu_{s,g} = \left(m_g \cdot \sum_f w_{s,f} c_{f,g} + s_{e,g} \right) \cdot y_s$$

where $w_{s,f}$ corresponds to the abundance of cell-type f at location s, m_g is a scaling parameter specific to gene g, $s_{e,g}$ is an experiment- and gene-specific additive shift, and y_s is the detection sensitivity at spot s.

GNN-C2L We propose a hierarchical model for cell-type composition inference that incorporates proximal relationships between spots. Let $\mathcal{N}(s)$ be the set of neighbour indices for spot s. This set of neighbours can be adapted to various spatial arrangements (e.g. hexagonal neighbourhoods for 10X Visium data) and k-hop neighbourhoods. To account for the neighbourhood information, we introduce a latent variable $\gamma_{s,f}$ representing the neighbour-aware cell-type abundances:

$$\gamma_{s,f} \sim \mathrm{Gamma}(\kappa_{s,f}, 1)$$
$$\boldsymbol{\kappa}_s = \psi\big(\boldsymbol{w}_s, \{\{\boldsymbol{w}_j \mid j \in \mathcal{N}(s)\}\}\big)$$

where the shape parameter $\kappa_{s,f}$ depends on the latent variables $\boldsymbol{w}_s$ and $\{\{\boldsymbol{w}_j \mid j \in \mathcal{N}(s)\}\}$ of spot s and its neighbours through a transformation $\psi(\cdot)$. Unlike Cell2Location, this effectively adds graphical dependencies between the neighbour-informed variables $\gamma_{s,f}$ and the latent variables $w_{s,f}$ (Appendix A). Importantly, computing $\gamma_{s,f}$ as a function of $\boldsymbol{w}_s$ allows capturing cell-type co-location patterns.

We then compute mean parameter $\mu_{s,g}$ of the Negative Binomial $\mathrm{NB}(\mu_{s,g}, \alpha_{e,g})$ likelihood using the neighbour-aware cell-type abundances $\gamma_{s,f}$:

$$\mu_{s,g} = \left(m_g \cdot \sum_f \gamma_{s,f} c_{f,g} + s_{e,g} \right) \cdot y_s$$

For all parameters, we utilise the validated hierarchical priors and hyperpriors of Cell2Location (Appendix A).

Incorporating spatial inductive biases The form of $\psi(\cdot)$ determines the inductive biases of the model. In this study, we construct a proximity graph of spatially localised spots,

i.e. we consider physically adjacent spots, allowing for different spatial arrangements (e.g. hexagonal neighbourhoods for 10X Visium data) and k-hop neighbourhoods. We consider several graph neural network architectures for $\psi(\cdot)$, starting with simple graph convolutional network (Wu et al., 2019) to validate whether homophily (enacted by feature propagation) is a useful inductive bias, and introducing other GNN operators to allow for a more expressive use of the available spatio-relational data. We also consider a standard multi-layer perceptron as a baseline to assess whether performance changes can be attributed to similarly parametrised spatial-agnostic transformations. We next describe the alternatives for ψ in greater detail.

MLP-C2L As a spatial-agnostic control, we model $\psi(\cdot)$ with an MLP, i.e. $\boldsymbol{\kappa}_s = \text{MLP}(\boldsymbol{w}_s)$, using a softplus activation function. This model does not utilise any spatial relationships between the spots and, alongside Cell2Location, serves as a control for our hypothesis.

SGC-C2L We construct a GNN-C2L variant using Simple Graph Convolutional (SGC) layers (Wu et al., 2019; Scherer et al., 2019). Let $d_s = |\mathcal{N}(s)|$ be the node degree of spot s. A SGC layer computes the neighbour-aware features $\boldsymbol{\kappa}_s$ using a weighted average of the latent variables $\boldsymbol{w}_s$ in the local neighbourhood:

$$\boldsymbol{\kappa}_s = \text{Linear}(\boldsymbol{h}_s)$$
$$\boldsymbol{h}_s = \frac{1}{d_s + 1}\boldsymbol{w}_s + \sum_{j \in \mathcal{N}(s)} \frac{1}{\sqrt{(d_s + 1)(d_j + 1)}}\boldsymbol{w}_j$$

The feature propagation mechanism biases the representations $\boldsymbol{\kappa}_s$ of neighbouring spots to become more similar to each other, using a degree-normalised adjacency matrix with self-loops. Thus, this simple MLP extension encourages homophilous latent cell-type distributions. Optionally, we can apply an activation function after the linear transformation and stack several SGC layers to expand the receptive field.

GAT-C2L We increase the expressivity of $\psi(\cdot)$ by utilising graph attention networks, specifically GATv2 (Brody et al., 2022). Unlike the constant, degree-dependant neighbouring contribution in the SGC-C2L model, the GATv2-C2L variant employs a learnable attention mechanism with increased control of contribution strengths, allowing to capture both homophilic and cell-type co-location patterns:

$$\boldsymbol{\kappa}_s = \alpha_{s,s}\phi(\boldsymbol{w}_s) + \sum_{j \in \mathcal{N}(s)} \alpha_{s,j}\phi(\boldsymbol{w}_j)$$

where ϕ is an MLP with a softplus activation function. We define the attention coefficient $\alpha_{s,j}$ as:

$$\alpha_{s,j} = \frac{\exp e(\boldsymbol{w}_s, \boldsymbol{w}_j)}{\sum_{k \in \mathcal{N}(s) \cup \{s\}} \exp e(\boldsymbol{w}_s, \boldsymbol{w}_k)}$$
$$e(\boldsymbol{w}_s, \boldsymbol{w}_j) = \mathbf{a}^T \text{LeakyReLU}(\boldsymbol{W}[\boldsymbol{w}_s || \boldsymbol{w}_j])$$

where $||$ is the concatenation operation and $\mathbf{a}$ and $\boldsymbol{W}$ are learnable parameters shared across spots, allowing the neural network to mix signals over the different cell types.

Training and inference We approximate the model parameters through variational inference. For every latent variable, we use a univariate normal distribution to approximate the posterior and utilise a softplus activation to ensure a positivity. Minimisation of the ELBO jointly trains the parameters of the model (and the incorporated GNNs) as well as the variational distribution. After optimisation, we estimate the cell-type abundances of every spot s by averaging $\gamma_{s,f}$ over 1000 samples of the variational distribution.

3. Experimental setup

We study whether incorporating spatial relationships via graph neural networks leads to enhanced cell-type mapping.

Datasets To quantitatively benchmark the baselines, we utilised a synthetic dataset introduced in Cell2Location (Kleshchevnikov et al., 2022) for which we knew the "true" cell-type abundances of each spot. The construction of this dataset is detailed extensively in (Kleshchevnikov et al., 2022). Moreover, we evaluated the methods using two real datasets, MPOA (Moffitt et al., 2018) and Xenium (breast cancer) (Janesick et al., 2022), that have single-cell resolution (yet fewer genes are profiled). To simulate real spots, we divided the tissues into squared spots of size $100\mu m$ and summed the expression of all cells within every square (Appendix B). For the Xenium and MPOA datasets, we constructed the cell-type signatures by averaging the read counts of all cells from every given cell-type (Appendix C).

Hyperparameter settings We used the same hyperparameters for every baseline where applicable. We set the hidden dimensions of each layer to 64 and used a single GNN layer (i.e. 1-hop receptive field). We conducted an ablation study using more graph layers in Appendix D. We minimised the variational lower bound using Adam (Kingma & Ba, 2014) with learning rate of 0.001 for 25,000 epochs in all datasets.

Evaluation metrics For all datasets we assessed performance using the average Pearson R correlation, Jensen-Shannon Divergence (JSD), and the area under precision-recall curve (AUPRC) (macro-averaged over cell-types) between the ground-truth and inferred cell-type proportions. We computed Pearson R over the flat ground-truth and inferred cell-type proportions. We calculated the Jensen-Shannon Divergence between the per-spot ground-truth and inferred cell-type proportions. We binarised the true cell abundance matrix to show which cell types were present in which locations, and then used the inferred cell-type proportions to compute the AUPRC.

Table 1: Average Pearson R, Avg. Jensen Shannon divergence (JSD), and AUPRC scores and standard deviation of 5 seeded runs of each model over all spots. For the synthetic dataset, scores for subcategories of cell types exhibiting distinct cell abundance patterns are also provided. Bold numbers indicate best-performing method for each category of cell types being evaluated for each metric. Overall, the GNN-C2L spatial-aware variants attained equal or superior deconvolution scores than spatial-agnostic baselines.

| | | | Simulated | | | Semi-simulated | | |
Method	ALL	UHCA	ULCA	RHCA	RLCA	MPOA	Xenium	Metric
Cell2location	0.683 ± 0.002	0.882 ± 0.001	0.519 ± 0.007	0.836 ± 0.004	0.422 ± 0.003	0.929 ± 0.001	0.928 ± 0.002	
GNN-C2L (MLP)	0.672 ± 0.024	0.866 ± 0.008	0.661 ± 0.021	0.865 ± 0.007	0.404 ± 0.040	0.920 ± 0.005	$\mathbf{0.929 \pm 0.000}$	R
GNN-C2L (SGC)	0.699 ± 0.023	0.876 ± 0.008	$\mathbf{0.708 \pm 0.020}$	0.883 ± 0.006	0.439 ± 0.041	$\mathbf{0.936 \pm 0.001}$	0.928 ± 0.000	
GNN-C2L (GAT)	$\mathbf{0.737 \pm 0.013}$	$\mathbf{0.885 \pm 0.018}$	0.695 ± 0.032	$\mathbf{0.888 \pm 0.004}$	$\mathbf{0.492 \pm 0.032}$	$\mathbf{0.936 \pm 0.001}$	0.928 ± 0.000	
Cell2location	0.468 ± 0.001	$\mathbf{0.202 \pm 0.002}$	0.496 ± 0.001	0.421 ± 0.002	0.509 ± 0.001	0.204 ± 0.001	0.213 ± 0.005	
GNN-C2L (MLP)	0.457 ± 0.006	0.230 ± 0.012	0.473 ± 0.007	0.387 ± 0.006	0.503 ± 0.009	0.199 ± 0.004	0.211 ± 0.001	Avg. JSD
GNN-C2L (SGC)	0.446 ± 0.006	0.224 ± 0.011	0.460 ± 0.007	$\mathbf{0.368 \pm 0.005}$	0.493 ± 0.009	0.189 ± 0.001	$\mathbf{0.211 \pm 0.001}$	
GNN-C2L (GAT)	$\mathbf{0.435 \pm 0.003}$	0.209 ± 0.021	$\mathbf{0.458 \pm 0.014}$	0.369 ± 0.001	$\mathbf{0.482 \pm 0.006}$	$\mathbf{0.188 \pm 0.001}$	0.212 ± 0.000	
Cell2location	0.591 ± 0.003	0.932 ± 0.006	0.477 ± 0.005	0.783 ± 0.003	0.591 ± 0.003	$\mathbf{0.956 \pm 0.001}$	0.873 ± 0.003	
GNN-C2L (MLP)	0.675 ± 0.002	0.963 ± 0.006	0.590 ± 0.004	0.804 ± 0.004	0.675 ± 0.002	0.951 ± 0.001	0.883 ± 0.001	AUPRC
GNN-C2L (SGC)	0.719 ± 0.002	0.977 ± 0.004	0.646 ± 0.006	$\mathbf{0.861 \pm 0.001}$	0.719 ± 0.002	0.955 ± 0.000	0.884 ± 0.000	
GNN-C2L (GAT)	$\mathbf{0.722 \pm 0.002}$	$\mathbf{0.978 \pm 0.004}$	$\mathbf{0.664 \pm 0.004}$	0.858 ± 0.003	$\mathbf{0.722 \pm 0.002}$	0.952 ± 0.001	$\mathbf{0.884 \pm 0.000}$	

4. Results and discussion

We benchmark spatial-agnostic (Cell2location, GNN-C2L MLP) and spatial-aware GNN-C2L (SGC, GAT) baselines on simulated and semi-simulated (MPOA and Xenium) spatial transcriptomics data (Table 1).

Results on simulated dataset We studied deconvolution performance on the synthetic data over: 1) **ALL**: all cell types, 2) ubiquitious high cell abundance (**UHCA**): 3 high-abundance cell types spatially distributed in uniform manner across the tissue, 3) ubiquitious low cell abundance (**ULCA**): 5 low-abundance cell types spatially distributed in uniform manner across the tissue, 4) regional high cell abundance (**RHCA**): 9 cell types with local distribution patterns, that is, cell types are clustered in specific locations with high abundance and exhibit 0 abundance elsewhere, 5) regional low cell abundance (**RLCA**): 32 low-abundance cell types that have local distribution patterns.

Overall, GNN-C2L consistently outperformed the spatial-agnostic baselines on the synthetic data (Table 1). We observed a marked increase in performance through the utilisation of proximal relations across different metrics and subtasks. Spatial-aware baselines achieved the best scores in 13 out of 15 cases, especially for cell types with low cell abundance (ULCA and RLCA). The performance difference was particularly apparent from the overall scores of the MLP variant of GNN-C2L (ALL R: 0.672 ± 0.024, JSD: 0.457 ± 0.006, AUPRC: 0.675 ± 0.002) and GNN-C2L SGC (ALL R: 0.699 ± 0.023, JSD: 0.446 ± 0.006, AUPRC: 0.719 ± 0.002) — both baselines utilised the number amount of learnable parameters, yet only GNN-C2L (SGC) propagates information across spots. It is also

worth noting that using additional parameters may result in degraded performance, i.e. compared to Cell2Location (ALL R : 0.683 ± 0.002), GNN-C2L (MLP) attained reduced Pearson R correlation and increased variance (ALL R : 0.672 ± 0.024). Altogether, our results highlight the superior ability of GNN-C2L to perform cell-type deconvolution.

Results on semi-simulated datasets In performance comparison on the semi-simulated datasets (MPOA and Xenium), the spatial-aware GNN-C2L variants achieved equal or better deconvolution performance than the spatial-agnostic baselines (Table 1). On MPOA, all baselines performed well — it is worth noting that this is a considerably smaller dataset with larger spot sizes (per-spot average of 18 cells) compared to the synthetic ($\sim$9 cells per spot) and Xenium ($\sim$10 cells per spot) datasets. This may have an effect on the specificity of the transcript readings as well as the usefulness of local information considering the size of micro-architectures in the tissue. We observed that GAT-C2L had the best scores in 2 out of 3 metrics (R: 0.492 ± 0.032, JSD: 0.188 ± 0.001), while Cell2Location was superior in terms of AUPRC (0.956 ± 0.001). In the Xenium dataset, all baselines attained comparable results (e.g. Cell2location R: 0.928 ± 0.000, GAT R: 0.928 ± 0.000; Cell2location AUPRC: 0.873 ± 0.003, MLP AUPRC: 0.883 ± 0.001, SGC AUPRC: 0.884 ± 0.000).

Collectively, our results suggest that spatial deconvolution can benefit from spatio-relational inductive biases, with potential for enhanced reconstruction of tissue architectures.

Code availability GNN-C2L is publicly available at https://github.com/paulmorio/GNN-C2L.

References

Alon, U. and Yahav, E. On the bottleneck of graph neural networks and its practical implications. *ArXiv*, abs/2006.05205, 2020.

Badia-i Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C. H., Ramirez Flores, R. O., et al. decoupler: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, 2 (1):vbac016, 2022.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D. Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 2018.

Brody, S., Alon, U., and Yahav, E. How attentive are graph attention networks? In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=F72ximsx7C1.

Cable, D. M., Murray, E., Zou, L. S., Goeva, A., Macosko, E. Z., Chen, F., and Irizarry, R. A. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology*, 40(4):517–526, Apr 2022. ISSN 1546-1696. doi: 10.1038/s41587-021-00830-w. URL https://doi.org/10.1038/s41587-021-00830-w.

Dong, R. and Yuan, G.-C. Spatialdwls: accurate deconvolution of spatial transcriptomic data. *Genome Biology*, 22(1):145, May 2021. ISSN 1474-760X. doi: 10.1186/s13059-021-02362-7. URL https://doi.org/10.1186/s13059-021-02362-7.

Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I., and Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Research*, 49(9):e50–e50, 02 2021. ISSN 0305-1048. doi: 10.1093/nar/gkab043. URL https://doi.org/10.1093/nar/gkab043.

Fischer, D. S., Schaar, A. C., and Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nature Biotechnology*, 41(3):332–336, Mar 2023. ISSN 1546-1696. doi: 10.1038/s41587-022-01467-z. URL https://doi.org/10.1038/s41587-022-01467-z.

Gonzalez, H., Hagerling, C., and Werb, Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes & development*, 32(19-20): 1267–1284, 2018.

Hamilton, W. L. Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 14(3):1–159.

Heumos, L., Schaar, A. C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M. D., Strobl, D. C., Henao, J., Curion, F., Aliee, H., Ansari, M., Badia-i Mompel, P., Büttner, M., Dann, E., Dimitrov, D., Dony, L., Frishberg, A., He, D., Hediyeh-zadeh, S., Hetzel, L., Ibarra, I. L., Jones, M. G., Lotfollahi, M., Martens, L. D., Müller, C. L., Nitzan, M., Ostner, J., Palla, G., Patro, R., Piran, Z., Ramírez-Suástegui, C., Saez-Rodriguez, J., Sarkar, H., Schubert, B., Sikkema, L., Srivastava, A., Tanevski, J., Virshup, I., Weiler, P., Schiller, H. B., Theis, F. J., and Consortium, S.-c. B. P. Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, Mar 2023. ISSN 1471-0064. doi: 10.1038/s41576-023-00586-w. URL https://doi.org/10.1038/s41576-023-00586-w.

Janesick, A., Shelansky, R., Gottscho, A. D., Wagner, F., Rouault, M., Beliakoff, G., de Oliveira, M. F., Kohlway, A., Abousoud, J., Morrison, C. A., et al. High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of ffpe tissue. *bioRxiv*, pp. 2022–10, 2022.

Kingma, D. and Ba, J. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H. W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., Jain, M. S., Park, J. S., Ramona, L., Tuck, E., Arutyunyan, A., Vento-Tormo, R., Gerstung, M., James, L., Stegle, O., and Bayraktar, O. A. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology*, 40(5): 661–671, May 2022. ISSN 1546-1696. doi: 10.1038/s41587-021-01139-4. URL https://doi.org/10.1038/s41587-021-01139-4.

Lohoff, T., Ghazanfar, S., Missarova, A., Koulena, N., Pierson, N., Griffiths, J., Bardot, E., Eng, C.-H., Tyser, R., Argelaguet, R., et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nature biotechnology*, 40(1):74–85, 2022.

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058, Dec 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0229-2. URL https://doi.org/10.1038/s41592-018-0229-2.

Miller, B. F., Huang, F., Atta, L., Sahoo, A., and Fan, J. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nature Communications*, 13(1): 2339, Apr 2022. ISSN 2041-1723. doi: 10.1038/

s41467-022-30033-z. URL https://doi.org/10.1038/s41467-022-30033-z.

Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., Rubinstein, N. D., Hao, J., Regev, A., Dulac, C., and Zhuang, X. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, 362(6416):eaau5324, 2018. doi: 10.1126/science.aau5324. URL https://www.science.org/doi/abs/10.1126/science.aau5324.

Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., Rybakov, S., Ibarra, I. L., Holmberg, O., Virshup, I., et al. Squidpy: a scalable framework for spatial omics analysis. *Nature methods*, 19(2):171–178, 2022.

Scherer, P., Andrés-Terré, H., Liò, P., and Jamnik, M. Decoupling feature propagation from the design of graph auto-encoders. *CoRR*, abs/1910.08589, 2019. URL http://arxiv.org/abs/1910.08589.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1):1 – 28, 2017. doi: 10.1214/16-STS576. URL https://doi.org/10.1214/16-STS576.

Solinas, G., Vilcu, C., Neels, J. G., Bandyopadhyay, G. K., Luo, J.-L., Naugler, W., Grivennikov, S., Wynshaw-Boris, A., Scadeng, M., Olefsky, J. M., and Karin, M. Jnk1 in hematopoietically derived cells contributes to diet-induced inflammation and insulin resistance without affecting obesity. *Cell Metabolism*, 6(5):386–397, 2007. ISSN 1550-4131. doi: https://doi.org/10.1016/j.cmet.2007.09.011. URL https://www.sciencedirect.com/science/article/pii/S1550413107002926.

Vlahopoulos, S. A., Cen, O., Hengen, N., Agan, J., Moschovi, M., Critselis, E., Adamaki, M., Bacopoulou, F., Copland, J. A., Boldogh, I., Karin, M., and Chrousos, G. P. Dynamic aberrant NF-κB spurs tumorigenesis: a new model encompassing the microenvironment. *Cytokine Growth Factor Rev.*, 26(4):389–403, August 2015.

Wang, K., Grivennikov, S. I., and Karin, M. Implications of anti-cytokine therapy in colorectal cancer and autoimmune diseases. *Annals of the Rheumatic Diseases*, 72(suppl 2):ii100–ii103, 2013. ISSN 0003-4967. doi: 10.1136/annrheumdis-2012-202201. URL https://ard.bmj.com/content/72/suppl_2/ii100.

Wolf, F. A., Angerer, P., and Theis, F. J. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15, Feb 2018. ISSN 1474-760X. doi: 10.1186/s13059-017-1382-0. URL https://doi.org/10.1186/s13059-017-1382-0.

Wu, F., Zhang, T., Jr., A. H. S., Fifty, C., Yu, T., and Weinberger, K. Q. Simplifying graph convolutional networks. *CoRR*, abs/1902.07153, 2019.

A. Cell2Location and GNN-C2L model description

In order to describe our model we will first go over the construction of Cell2Location (Kleshchevnikov et al., 2022). Cell2Location is a Bayesian inference model built in a hierarchical manner to account for different sources of confounding experimental information.

Let $D \in \mathbb{R}^{S \times G}$ denote a mRNA count matrix with its entries corresponding to mRNA count at spot $s \in \{1, ..., S\}$ from one or multiple batches (i.e. 10x Visium slides or SlideSeq pucks) for genes $g \in \{1, ..., G\}$. Let $\mathcal{C} \in \mathbb{R}^{F \times G}$ denote a matrix of reference cell type signatures obtained from learning on the scRNA data (see Section A.1.1). Note that D and $\mathcal{C}$ need to be aligned such that they cover the same set of genes G. Cell2Location models the elements of D as Negative Binomial distributed (NB), given an unobserved expression level (rate) $\mu_{s,g}$ and a gene- and experiment- specific over-dispersion parameter $\alpha_{e,g}$:

$$d_{s,g} \sim \text{NB}(\mu_{s,g}, \alpha_{e,g}) \tag{1}$$

This can be equivalently expressed as a Gamma-Poisson mixture with a Poisson likelihood (count measurement model) and a Gamma-distributed mean (expression dispersion model):

$$d_{s,g} \sim \text{Poisson}(\text{Gamma}(\alpha_{e,g}, \frac{\alpha_{e,g}}{\mu_{s,g}})) \tag{2}$$

The expression level of genes $\mu_{s,g}$ in the mRNA count space is modelled as a linear function of the reference cell type expression signatures:

$$\mu_{s,g} = \left(m_g \cdot \sum_f w_{s,f} g_{f,g} + s_{e,g} \right) \cdot y_s \tag{3}$$

- Here $w_{s,f}$ denotes regression weight of each reference signature f at location s, which can be regarded as the abundance or proportion of cells expressing reference cell type signature f at s. This is the latent variable that we care about and intend to infer.

- m_g denotes a gene-specific scaling parameter, which adjusts for global differences in expression estimates between technologies.

- $s_{e,g}$ captures gene specific additive shift (due to free-floating RNA).

- y_s denotes a location-specific scaling parameter, which models variation in RNA detection sensitivity across locations and experiments. This parameter scales the contributions of the cell types and the gene specific additive shift $s_{e,g}$.

We dive into the derivation of the prior distributions for each of the latent variables.

Cell abundance across locations $w_{s,f}$ This is Gamma distributed according to

$$w_{s,f} \sim \text{Gamma}(\mu_{s,f}^w v^w, v^w), \tag{4}$$

where v^w is a fixed hyperparameter denoting prior strength, the prior mean parameter is modelled in a hierarchical fashion, decomposing the regression weights into R latent groups of cell types $r = \{1, ..., R\}$ (by default Cell2Location uses $R = 50$) accounting for linear dependencies in spatial abundance of cell types:

$$\mu_{s,f}^w = \sum_r z_{s,r} x_{r,f} \tag{5}$$

Intuitively, R can be considered as the number of cellular compartments or zones in the tissue that are characterised by shared cell type composition. The authors observed that the sensitivity of mapping cell types with small transcriptional

difference increases when accounting for these dependencies. For our purposes we are not interested in this parameter and utilise the default for all experiments.

$z_{s,r}$ and $x_{r,f}$ are prior distributions defined to control absolute scale of the cell type abundance estimates, guiding $w_{s,f}$ to the scale of the number of the abundance of cells expressing reference cell type signature f at location s. These priors are important because there is a non-identifiability between m_g, y_s, $w_{s,f}$ unless informative priors are constructed for each of them. Moreover, the prior distributions help control the sparsity of how many cell types f are expected at each spot s, facilitating application of Cell2Location to tissues and technologies with varying numbers of cells and cell types per location. The hyperparameters controlling the $w_{s,f}$ prior can be estimated from a paired histology image or a literature based estimate. The hyperparameters controlling y_s can be estimated based on total RNA counts in the input data and the quality of the experiment. The prior distribution $z_{s,r}$ is defined as follows:

$$z_{s,r} \sim \text{Gamma}(\frac{B_s}{R}, \frac{1}{\frac{N_s}{B_s}}) \tag{6}$$

$$N_s \sim \text{Gamma}(\hat{N} \cdot v^n, v^n) \tag{7}$$

$$B_s \sim \text{Gamma}(\hat{B}, 1) \tag{8}$$

where N_s is associated to the latent average number of cells in each location, and B_s is the latent number of groups r expected in each spot s. $\hat{N}$ is a user-defined estimate of the expected number of cells per location (see end of this section). $\hat{B}$ is the expected average number of cellular components or zones per location; by default it is initialised to 7. v^n denotes a prior strength. The construction is done such that $\sum_r z_{s,r} \geq N_s$. In other words, the expectation of the sum over $z_{s,r}$ equals the expected number of cells per location N_s and that on average each location has a high value of $z_{s,r}$ for B_s expected cell type groups.

$x_{r,f}$ represents the contribution of each latent cell type group r to the abundance of each cell type f and is Gamma distributed in the following manner:

$$x_{r,f} \sim \text{Gamma}(\frac{K_r}{R}, K_r) \tag{9}$$

$$K_r \sim \text{Gamma}(\frac{\hat{A}}{\hat{B}}, 1) \tag{10}$$

K_r represents the unobserved number of cell types for each group r. This prior controls the absolute values of $x_{r,f}$ such that on average $\sum_r x_{r,f} = 1$. $\hat{A}$ and $\hat{B}$ intuitively represent the expected number of cell types per spot, and the expected number of cellular components per spot respectively. By default both $\hat{A}$ and $\hat{B}$ are initialised at 7, indicating the prior belief that the spatial abundance of each cell type f is independent from other cell types. Conversely, each group r has have a large value of $x_{r,f}$ for many cell types f when $\hat{A} > \hat{B}$.

Gene specific multiplicative scaling factor m_g This is modelled as Gamma distributed with hierarchical prior μ^m and α^m which provide regularisation:

$$m_g \sim \text{Gamma}(\alpha^m, \frac{\alpha^m}{\mu^m}) \tag{11}$$

$$\alpha^m = \frac{1}{(o^m)^2} \tag{12}$$

$$o^m \sim \text{Exponential}(3) \tag{13}$$

$$\mu^m \sim \text{Gamma}(1,1) \tag{14}$$

The prior on detection efficiency y_s per location This prior is selected to discourage over normalisation, such that unless data has evidence of strong within-experiment variability in RNA detection sensitivity across locations, it is assumed to be small and close to the mean sensitivity for each experiment or batch y_e:

$$y_s \sim \text{Gamma}(\alpha^y, \frac{\alpha^y}{y_e}) \tag{15}$$

$$y_e \sim \text{Gamma}(10, \frac{10}{\mu^y}) \tag{16}$$

where α^y is a user defined hyperparameter that regularises within experiment variation; and y_e is a latent detection efficiency for each batch or experiment e. mu^y is estimated using observed variables ($d_{s,g}$ and $g_{f,g}$) and the hyperparameter $\hat{N}$ in the following manner:

$$\mu^y = \frac{\frac{\sum_s \sum_g \frac{d_{s,g}}{S}}{\hat{N}}}{\sum_f \sum_g \frac{g_{f,g}}{F}} \tag{17}$$

where we remind ourselves that S is the total number of spots, and F is the total number of cell types.

Overdispersion containment prior $\alpha_{e,g}$ A containment prior (Simpson et al., 2017) is used to model the latent variance of the negative binomial distribution modelling $d_{s,g}$. This prior is intended to encourage simplicity of the NB model making it closer to the Poisson distribution (via larger $\alpha_{e,g}$ values producing bigger probability masses). Thus the prior expresses a belief that most genes have low overdispersion:

$$\alpha_{e,g} = \frac{1}{o_{e,g}^2} \tag{18}$$

$$o_{e,g} \sim \text{Exponential}(\beta^o) \tag{19}$$

$$\beta^o \sim \text{Gamma}(9,3) \tag{20}$$

where the constants 9 and 3 correspond to values of $\alpha_{e,g}$ observed in previous modelling studies (Kleshchevnikov et al., 2022; Lopez et al., 2018).

Additive shift bias $s_{e,g}$ This latent variable accounts for confounding effects on RNA counts for every gene g for every experiment e caused by phenomena such as free-floating RNA in the tissue sample. This additive shift is modelled using a Gamma distribution again with hierarchical experiment specific priors α_e^s and β_e^s which provide regularisation:

$$s_{e,g} \sim \text{Gamma}(\alpha_e^s, \frac{\alpha_e^s}{\mu_e^s}) \tag{21}$$

$$\mu_e^s \sim \text{Gamma}(1, 100) \tag{22}$$

$$\alpha_e^s = \frac{1}{(o_e)^2} \tag{23}$$

$$o_e \sim \text{Exponential}(\beta^s) \tag{24}$$

$$\beta^s \sim \mathrm{Gamma}(9,3) \tag{25}$$

The hierarchical priors α_e^s and β_e^s model the variation across experiments e. β^s serves as a hyperparameter that allows the model to learn α_e^s rather than requiring a user to define it.

This leaves Cell2Location with two hyperparameters whose values have to be considered based on the dataset and how the spatial transcriptomics experiment was performed:

1. $\hat{N}$: the expected number of cells per spot. This is a tissue-level global estimate, which can be derived from paired histology images (see the H&E stained image in Figure 3). An estimate may be obtained by manually counting nuclei in a set of random spots using the appropriate software from the measurement device (e.g. 10x Loupe Browser for the Visium slides outputs). When this is not available one can also use the size of the captured regions relative to an expected cell size.

2. α^y: the regularising hyperparameter for within-experiment variation of RNA detection sensitivity. In the default setting it is assumed that there is little variability in the RNA detection sensitivity so α^y is set to $\alpha^y = 200$ which results in values of y_s close to the mean sensitivity for each experiment y_e. A lower value would enforce a stronger normalisation to the sensitivity, and a correspondingly lower regularisation toward the mean sensitivity.

A.1. Description of Cell2Location deconvolution pipeline

The desired cell type compositions of the spots in Cell2Location are obtained using two main steps:

1. Computing cell type specific gene expression *signatures* using reference single-cell RNA-seq data.

2. Using variational inference to sample latent posterior distributions for the cell type proportions.

A.1.1. COMPUTING REFERENCE CELL TYPE SIGNATURES

Cell type signatures are obtained by performing regularised Negative Binomial regression. The motivation behind using this model is that it would robustly derive the reference expression values of cell types $g_{f,g}$ using input data composed of different batches $e = \{1, ..., E\}$ and technologies $t = \{1, ..., T\}$ that may affect the results (though for our case studies this is actually not utilised) (Kleshchevnikov et al., 2022). Here the expression count matrix $J = \{j_{c,g}\}, c \in C, g \in G$ follows a Negative Binomial distribution with unobserved expression levels (rates) $\mu_{c,g}$ and a gene-specific over-dispersion α_g:

$$J_{c,g} \sim \mathrm{NB}(\mu_{c,g}, \frac{1}{\alpha_g^2}) \tag{26}$$

$\mu_{c,g}$ is modelled as a linear function of the reference cell type signatures and the batch/technical effects:

$$\mu_{c,g} = (g_{f,g} + b_{e,g}) h_e p_{t,g} \tag{27}$$

where h_e is a global scaling parameter between experiments e (for example difference in sequencing depth, or the number of times a given nucleotide has been read in an experiment (Lopez et al., 2018)). $p_{t,g}$ accounts for multiplicative gene-specific difference in sensitivity between technologies, $b_{e,g}$ accounts for additive background shift of each gene in each experiment e caused by free-floating RNA.

The priors of these variables are specified in hierarchical manner using similar constructions as we have seen previously with Cell2Location:

$$g_{f,g} \sim \mathrm{Gamma}(1,1) \tag{28}$$

$$h_e \sim \mathrm{Gamma}(1,1) \tag{29}$$

$$p_{t,g} \sim \text{Gamma}(200, 200) \tag{30}$$

The prior on the additive shift variable $b_{e,g}$ is specified in the same manner as $s_{e,g}$ (Equations 7.21-7.25). The prior on α_g is specified similarly to $\alpha_{e,g}$ (Equations 7.18-7.20):

$$\alpha_g = \frac{1}{o_g^2} \tag{31}$$

$$o_g \sim \text{Exponential}(\beta^o) \tag{32}$$

$$\beta^o \sim \text{Gamma}(9, 3) \tag{33}$$

All model parameters are constrained to be positive. A weak L2 regularisation of $g_{f,g}$, $b_{e,g}$, and α_g is set alonside a strong penalty for deviations of h_e and $p_{t,g}$ from 1. The average for each cell type f is used to initialise $g_{f,g}$. $b_{e,g}$ is initialised at the average expression of each gene g in each experiment e divided by 10 (Lopez et al., 2018; Kleshchevnikov et al., 2022).

A.1.2. INFERENCE

Stochastic variational Bayesian inference is used to approximate the posterior distributions, enacted through Pyro (Bingham et al., 2018) and its autodiff variational inference framework (ADVI). Briefly, the latent posterior distributions of the model are approximated using univariate normal distributions which are softplus transformed to ensure positive scaling. The parameters of the variational distributions are chosen through minimisation of the KL divergence between the variational approximation and the true posterior distribution. This is equivalently achieved by maximising the evidence lower bound (ELBO objective).

After this optimisation, the posterior mean, standard deviation, 5% and 95% quantiles for each parameter are computed using 1000 samples from the variational posterior distribution. The mean was used for all of the results we show in the results section.

A.2. GNN-C2L: spatially aware spatial cell deconvolution

Our methods build upon the Cell2Location pipelines, introducing relational inductive biases enacted by different instances of the message passing neural network framework. This primarily consists of two major additions. The first is the construction of the underlying graph that establishes proximal relationships between each of the observations we are interested in: the spots. The second is the augmentation of the Cell2Location generative model and the introduction of graph neural networks into them.

A.2.1. CONSTRUCTING A SPATIAL PROXIMITY GRAPH ON THE SPATIAL RNA-SEQ OUTPUT

As with many GRL methods, the underlying graph is an important factor behind actualising the assumption we intend to incorporate with relational inductive biases and obtaining performance gains (Hamilton). However, first and foremost we explore the utilisation of proximity and the assumption that proximal spots exhibit similar cell type compositions or specific cell type relationships between spots. To utilise this we want a graph neural network to operate on the proximity graph of spots after they have been selected by standard preprocessing pipelines (Heumos et al., 2023; Lopez et al., 2018; Kleshchevnikov et al., 2022). Depending on the positional arrangement of spots dictated by the spatial transcriptomics technology used — for example, hexagonal arrangement in 10x Visium slides as seen in Figure 4, or the grid arrangement found in our pseudo-synthetic dataset — a different number of neighbours is specified alongside the size of the local perceptive field we want a single layer of a GNN to operate over.

A.2.2. GNN-C2L

We propose a hierarchical model for cell-type composition inference that incorporates proximal relationships between spots. Let $\mathcal{N}(s)$ be the set of neighbour indices for spot s. This set of neighbours can be adapted to various spatial arrangements
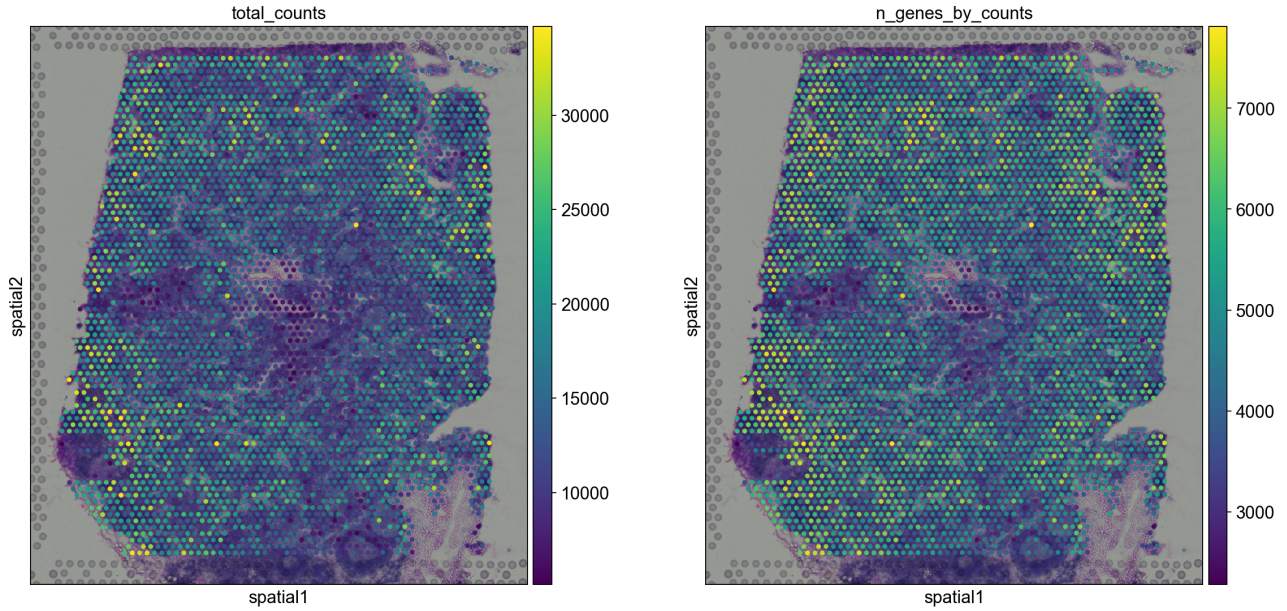
Figure 3: Visium slide of a human lymph node tissue publicly available at the 10x Genomics dataset portal, accessed through scanpy (Wolf et al., 2018). In these images the circular spots are overlaid over the Hematoxylin and eosin stain (H&E) image of the tissue sample. Note the non-uniform dispersion of the spots. The left figure highlights the total number of counts read for every gene, and the right figure summarises the number of genes with at least 1 count in a cell, highlighting the diversity of gene expression patterns spatially across the tissue.
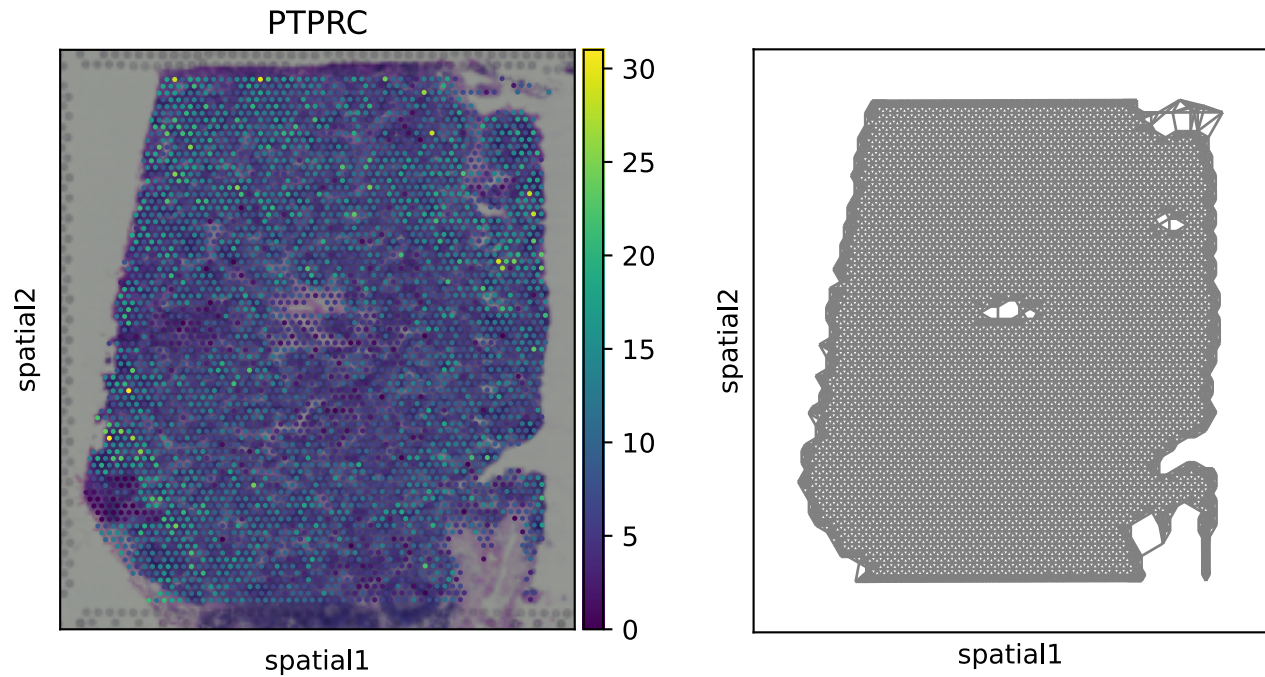

Figure 4: An example of the proximity graph computed using 6 neighbours for each spot in the hexagonal arrangement utilised in the 10X Visium protocol for a human lymph node sample (publicly available through the 10x Genomics dataset portal, accessed through scanpy (Wolf et al., 2018)). On the left is the output of the spatial transcriptomics sequencing with colouring of the spots based on the transcript counts for the gene PTPRC. The right shows the underlying proximity graph between each of the spots with its 6 closest neighbours.

(e.g. hexagonal neighbourhoods for 10X Visium data) and k-hop neighbourhoods. To account for the neighbourhood information, we introduce a latent variable $\gamma_{s,f}$ representing the neighbour-aware cell-type abundances:

$$\gamma_{s,f} \sim \text{Gamma}(\kappa_{s,f}, 1)$$
$$\boldsymbol{\kappa}_s = \psi\big(\boldsymbol{w}_s, \{\{\boldsymbol{w}_j \mid j \in \mathcal{N}(s)\}\}\big)$$

where the shape parameter $\kappa_{s,f}$ depends on the latent variables $\boldsymbol{w}_s$ and $\{\{\boldsymbol{w}_j \mid j \in \mathcal{N}(s)\}\}$ of spot s and its neighbours through a transformation $\psi(\cdot)$. Unlike Cell2Location, this effectively adds graphical dependencies between the neighbour-informed variables $\gamma_{s,f}$ and the latent variables $w_{s,f}$ (Appendix A). Importantly, computing $\gamma_{s,f}$ as a function of $\boldsymbol{w}_s$ allows capturing cell-type co-location patterns.

We then compute mean parameter $\mu_{s,g}$ of the Negative Binomial $\text{NB}(\mu_{s,g}, \alpha_{e,g})$ likelihood using the neighbour-aware cell-type abundances $\gamma_{s,f}$:

$$\mu_{s,g} = \big(m_g \cdot \sum_f \gamma_{s,f} c_{f,g} + s_{e,g}\big) \cdot y_s$$

For all parameters, we utilise the validated hierarchical priors and hyperpriors of Cell2Location (Appendix A).

Incorporating spatial inductive biases The form of $\psi(\cdot)$ determines the inductive biases of the model. In this study, we construct a proximity graph of spatially localised spots, i.e. we consider physically adjacent spots, allowing for different spatial arrangements (e.g. hexagonal neighbourhoods for 10X Visium data) and k-hop neighbourhoods. We consider several graph neural network architectures for $\psi(\cdot)$, starting with simple graph convolutional network (Wu et al., 2019) to validate whether homophily (enacted by feature propagation) is a useful inductive bias, and introducing other GNN operators to allow for a more expressive use of the available spatio-relational data. We also consider a standard multi-layer perceptron as a baseline to assess whether performance changes can be attributed to similarly parametrised spatial-agnostic transformations. We next describe the alternatives for ψ in greater detail.

MLP-C2L As a spatial-agnostic control, we model $\psi(\cdot)$ with an MLP, i.e. $\boldsymbol{\kappa}_s = \text{MLP}(\boldsymbol{w}_s)$, using a softplus activation function. This model does not utilise any spatial relationships between the spots and, alongside Cell2Location, serves as a control for our hypothesis.

SGC-C2L We construct a GNN-C2L variant using Simple Graph Convolutional (SGC) layers (Wu et al., 2019; Scherer et al., 2019). Let $d_s = |\mathcal{N}(s)|$ be the node degree of spot s. A SGC layer computes the neighbour-aware features $\boldsymbol{\kappa}_s$ using a weighted average of the latent variables $\boldsymbol{w}_s$ in the local neighbourhood:

$$\boldsymbol{\kappa}_s = \text{Linear}(\boldsymbol{h}_s)$$
$$\boldsymbol{h}_s = \frac{1}{d_s + 1}\boldsymbol{w}_s + \sum_{j \in \mathcal{N}(s)} \frac{1}{\sqrt{(d_s + 1)(d_j + 1)}}\boldsymbol{w}_j$$

The feature propagation mechanism biases the representations $\boldsymbol{\kappa}_s$ of neighbouring spots to become more similar to each other, using a degree-normalised adjacency matrix with self-loops. Thus, this simple MLP extension encourages homophilous latent cell-type distributions. Optionally, we can apply an activation function after the linear transformation and stack several SGC layers to expand the receptive field.

GAT-C2L We increase the expressivity of $\psi(\cdot)$ by utilising graph attention networks, specifically GATv2 (Brody et al., 2022). Unlike the constant, degree-dependant neighbouring contribution in the SGC-C2L model, the GATv2-C2L variant employs a learnable attention mechanism with increased control of contribution strengths, allowing to capture both homophilic and cell-type co-location patterns:

$$\boldsymbol{\kappa}_s = \alpha_{s,s}\phi(\boldsymbol{w}_s) + \sum_{j \in \mathcal{N}(s)} \alpha_{s,j}\phi(\boldsymbol{w}_j)$$

where ϕ is an MLP with a softplus activation function. We define the attention coefficient $\alpha_{s,j}$ as:

$$\alpha_{s,j} = \frac{\exp e(\boldsymbol{w}_s, \boldsymbol{w}_j)}{\sum_{k \in \mathcal{N}(s) \cup \{s\}} \exp e(\boldsymbol{w}_s, \boldsymbol{w}_k)}$$
$$e(\boldsymbol{w}_s, \boldsymbol{w}_j) = \mathbf{a}^T \text{LeakyReLU}(\boldsymbol{W}[\boldsymbol{w}_s || \boldsymbol{w}_j])$$

where $||$ is the concatenation operation and $\mathbf{a}$ and W are learnable parameters shared across spots, allowing the neural network to mix signals over the different cell types.

Training and inference   We approximate the model parameters through variational inference. For every latent variable, we use a univariate normal distribution to approximate the posterior and utilise a softplus activation to ensure a positivity. Minimisation of the ELBO jointly trains the parameters of the model (and the incorporated GNNs) as well as the variational distribution. After optimisation, we estimate the cell-type abundances of every spot s by averaging $\gamma_{s,f}$ over 1000 samples of the variational distribution.

B. Dataset details

B.1. Purely synthetic dataset

Briefly, this simulated spatial transcriptomic dataset introduced in Cell2Location (Kleshchevnikov et al., 2022) consists of 2,500 spots by sampling cells from 49 reference cell types, using a combination of ubiquitous abundance patterns and cell types distributed according to regional tissue zones as observed in semi-simulated data (Kleshchevnikov et al., 2022). Reference cell type signatures were adapted from a mouse brain scRNA-seq dataset sequenced by Kleshchevnikov et al. (2022). The dataset reflects diverse cell abundance patterns across ubiquitous and spatially restricted cell types, which allow us to evaluate our methods within different cell abundance scenarios in addition to just overall accuracy. The paired reference scRNA-seq dataset consists of 8,111 cells exhibiting 12,080 genes whose expression is measured (recall this number is an intersection of the set of transcripts measured in the scRNA-seq and spatial transcriptomic experiments). This dataset is preprocessed and utilised to compute reference cell type signatures matrix $\mathcal{C}$ as detailed in Appendix C.1. More information on the construction of this dataset is detailed extensively in Section 5 of the supplementary materials in (Kleshchevnikov et al., 2022).

B.2. MPOA

Briefly, the MPOA dataset is a simulated spatial transcriptomic dataset constructed by aggregating the gene expression of cells from single-cell resolution multiplex error robust fluorescence in situ hybridisation (MERFISH) data of the mouse medial preoptic area (Moffitt et al., 2018). The availability of both single-cell resolution gene expression in-situ makes it an ideal candidate for creating simulated spatial transcriptomic data. It contains measurements of 135 genes selected to distinguish between major non-neuronal cell types as well as neuronal subtypes. Transcriptional clustering analysis on the scRNA-seq measurements identified 9 major cell types which are used to form the underlying target proportions we wish to infer. To simulate the multi-cellular spatial transcriptomic data, the single-cell resolution MERFISH data is aggregated into $100\mu m^2$ pixels which we proxy as spot transcripts. More information on this dataset and its availability can be found in the supplementary materials in (Miller et al., 2022).

B.3. Xenium

The Xenium breast cancer dataset (Janesick et al., 2022) jointly profiled the expression and spatial location of 167,885 cells (313 genes) from a formalin-fixed, paraffin-embedded (FFPE) human breast cancer section. We utilised `squidpy` (Palla et al., 2022) to download, process the data, and detect Leiden communities (`https://squidpy.readthedocs.io/en/stable/external_tutorials/tutorial_xenium.html`). We then employed `decoupler-py` (Badia-i Mompel et al., 2022) to perform overrepresentation analysis (Figure 5), using breast cancer marker genes (Janesick et al., 2022). In total, we identified 8 broad cell types including immune cells (e.g. T cells, B cells, and macrophages) and breast cells (e.g. grandular cells, myoepithelial cells, and cancer cells). We aggregated (i.e. summed) the single-cell resolution Xenium data into $100\mu m^2$ pixels as a proxy for convolved spot transcripts.

C. Additional experimental setup details

C.1. Inference of cell type signatures

As in Cell2Location (Kleshchevnikov et al., 2022), two approaches can be taken to compute cell type signatures matrix $C \in \mathbb{R}^{F \times G}$:
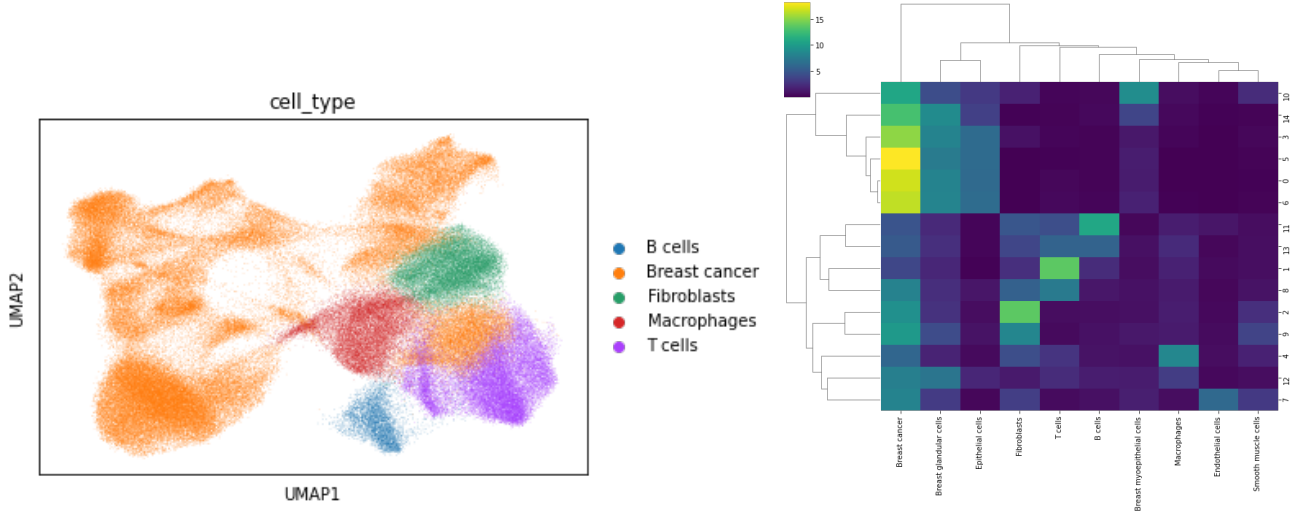
Figure 5: (left) UMAP plot of Xenium colored by Leiden clusters. (right) Overrepresentation analysis scores for each Leiden cluster.

1. Inference using Negative Binomial regression   Given the raw untransformed and unnormalised count matrix $\mathbf{J} \in \mathbb{R}^{|C| \times |G|}$ of the reference scRNA-seq data, several preprocessing steps were taken before it is input to the NB regression to find cell type signatures. The data is filtered at 2 cut-offs: i) selecting genes detected that have more than 0 mRNA counts for at least 5% of the cells, and ii) selecting genes with a mean expression greater than 1.1 and a greater than 0 mRNA count in at least 0.05% of the cells. Subsequently it is given to the NB regression algorithm to infer the cell type signatures.

To infer the cell type signatures, the variational parameters of the Negative Binomial regression model were trained with stochastic gradient descent on the ELBO objective using a batch size of 1024 cells and an Adam optimiser with a learning rate set at 0.001 for 500 epochs. The posterior cell signature values were sampled 1000 times and the mean values were used. To control and simplify comparison between the deconvolution models, the cell type signatures are shared between each of the methods we evaluate.

This approach is used in the purely synthetic dataset, and is typically the recommended approach in Cell2Location as the regression model is robust to different technology and batch effects.

2. Using mean counts of genes for each cell type   Given the raw untransformed and unnormalised count matrix $\mathbf{J} \in \mathbb{R}^{|C| \times |G|}$ of the reference scRNA-seq data, we compute the average count of genes $g \in G$ for each of the cell types in F to construct $\boldsymbol{C} \in \mathbb{R}^{F \times G}$. This approach is computationally efficient and can provide good mappings when the scRNA-seq data is obtained from a single batch or better yet is paired with the spatial transcriptomic data, as is the case in the MPOA and Xenium datasets. This approach was utilised in our evaluation for MPOA and Xenium.

D. Effect of increasing neighbourhood size

The increased performance through the utilisation of the 1-hop proximal neighbourhoods begs the question of how the size of the receptive field can influence performance. Recall from Section 2.4.2 that this is as simple as adding more layers to the graph neural network layers. We present the table of results examining increase of receptive field with SGC-C2L and GAT-C2L in Tables 2, 3 and 4. Whilst all of the models still consistently outperform the original Cell2Location and MLP-C2L, we can see certain performs patterns that are in line with GNN theory. Specifically, in all but the average JSD metric for ULCA we see that the best performing models exist in the first models exhibiting up to 3 layers, exhibiting a drop in performance after the best performance. This pattern is common GNN based methods due to the oversmoothing and oversquashing phenomenon (Alon & Yahav, 2020) that prevents GNNs from effectively incorporating information from distant neighbours as the aggregation of messages into fixed size vectors creates an information bottleneck. In addition to the mechanical limitation of the GNNs, we also have to consider the relationship between the growing receptive field and its absolute distance away from the target spot we want to influence in terms of the sizes of cell colocation patterns. Depending

Table 2: Average Pearson R correlation and standard deviation of 5 seeded runs of each model over all spots. Correlation values for subcategories of cell types exhibiting distinct cell abundance patterns are also provided. Bold numbers indicate best performing method for each category of cell types being evaluated.

Methods	ALL	UHCA	ULCA	RHCA	RLCA
SGC-C2L1	0.699 ± 0.023	0.876 ± 0.008	0.708 ± 0.020	0.883 ± 0.006	0.439 ± 0.041
SGC-C2L2	0.711 ± 0.036	0.890 ± 0.015	0.682 ± 0.027	0.878 ± 0.01	0.458 ± 0.050
SGC-C2L3	0.684 ± 0.063	$\mathbf{0.897 \pm 0.019}$	0.689 ± 0.030	0.883 ± 0.006	0.421 ± 0.086
SGC-C2L4	0.704 ± 0.025	0.883 ± 0.022	0.673 ± 0.043	0.881 ± 0.009	0.445 ± 0.043
SGC-C2L5	0.701 ± 0.016	0.884 ± 0.015	0.665 ± 0.032	0.882 ± 0.007	0.443 ± 0.034
SGC-C2L6	0.701 ± 0.016	0.884 ± 0.015	0.665 ± 0.032	0.882 ± 0.007	0.443 ± 0.034
GAT-C2L1	$\mathbf{0.737 \pm 0.013}$	0.885 ± 0.018	0.695 ± 0.032	0.888 ± 0.004	$\mathbf{0.492 \pm 0.032}$
GAT-C2L2	0.722 ± 0.022	0.879 ± 0.020	0.710 ± 0.042	$\mathbf{0.889 \pm 0.004}$	0.473 ± 0.029
GAT-C2L3	0.679 ± 0.039	0.872 ± 0.021	$\mathbf{0.723 \pm 0.016}$	0.887 ± 0.007	0.425 ± 0.052
GAT-C2L4	0.709 ± 0.047	0.878 ± 0.016	0.695 ± 0.024	0.883 ± 0.004	0.474 ± 0.070
GAT-C2L5	0.713 ± 0.050	0.857 ± 0.015	0.698 ± 0.027	0.878 ± 0.009	0.478 ± 0.082
GAT-C2L6	0.715 ± 0.050	0.858 ± 0.016	0.699 ± 0.025	0.878 ± 0.009	0.480 ± 0.082

on the cell-types, tissue architecture, and fidelity of the spatial transcriptomic technology, differing receptive field sizes over spots will be biologically relevant to capturing cell colocation patterns.

Table 3: Average of average Jensen-Shannon divergence (JSD) along with standard deviation of 5 seeded runs of each model. JSD values for subcategories of cell types exhibiting distinct cell abundance patterns are also provided. Bold numbers indicate best performing method for each category of cell types being evaluated.

Methods	ALL	UHCA	ULCA	RHCA	RLCA
SGC-C2L1	0.446 ± 0.006	0.224 ± 0.011	0.460 ± 0.007	0.368 ± 0.005	0.493 ± 0.009
SGC-C2L2	0.443 ± 0.007	0.208 ± 0.021	0.467 ± 0.010	0.371 ± 0.009	0.489 ± 0.007
SGC-C2L3	0.447 ± 0.011	$\mathbf{0.199 \pm 0.017}$	0.463 ± 0.006	0.369 ± 0.007	0.499 ± 0.015
SGC-C2L4	0.448 ± 0.006	0.216 ± 0.019	0.472 ± 0.014	0.375 ± 0.008	0.494 ± 0.009
SGC-C2L5	0.448 ± 0.005	0.207 ± 0.022	0.473 ± 0.010	0.375 ± 0.007	0.493 ± 0.008
SGC-C2L6	0.448 ± 0.005	0.207 ± 0.022	0.473 ± 0.010	0.375 ± 0.007	0.493 ± 0.008
GAT-C2L1	$\mathbf{0.435 \pm 0.003}$	0.209 ± 0.021	0.458 ± 0.014	0.369 ± 0.001	$\mathbf{0.482 \pm 0.006}$
GAT-C2L2	0.438 ± 0.006	0.223 ± 0.017	0.458 ± 0.014	0.363 ± 0.002	0.486 ± 0.005
GAT-C2L3	0.447 ± 0.008	0.222 ± 0.025	0.450 ± 0.009	$\mathbf{0.356 \pm 0.004}$	0.496 ± 0.011
GAT-C2L4	0.441 ± 0.010	0.215 ± 0.018	0.452 ± 0.010	0.358 ± 0.004	0.487 ± 0.013
GAT-C2L5	0.445 ± 0.015	0.243 ± 0.017	$\mathbf{0.448 \pm 0.009}$	0.362 ± 0.007	0.492 ± 0.020
GAT-C2L6	0.444 ± 0.015	0.242 ± 0.017	0.448 ± 0.009	0.362 ± 0.007	0.491 ± 0.020

Table 4: Average AUPRC scores and standard deviation of 5 seeded runs of each model over all spots. Scores for subcategories of cell types exhibiting distinct cell abundance patterns are also provided. Bold numbers indicate best performing method for each category of cell types being evaluated.

Methods	ALL	UHCA	ULCA	RHCA	RLCA
SGC-C2L1	0.719 ± 0.002	0.977 ± 0.004	0.646 ± 0.006	0.861 ± 0.001	0.719 ± 0.002
SGC-C2L2	0.716 ± 0.003	0.978 ± 0.001	0.644 ± 0.006	0.860 ± 0.001	0.716 ± 0.003
SGC-C2L3	0.710 ± 0.002	$\mathbf{0.979 \pm 0.002}$	0.649 ± 0.005	0.852 ± 0.001	0.710 ± 0.002
SGC-C2L4	0.701 ± 0.004	0.972 ± 0.003	0.639 ± 0.007	0.845 ± 0.005	0.701 ± 0.004
SGC-C2L5	0.701 ± 0.007	0.975 ± 0.003	0.633 ± 0.009	0.848 ± 0.005	0.701 ± 0.007
SGC-C2L6	0.701 ± 0.007	0.975 ± 0.003	0.633 ± 0.009	0.848 ± 0.005	0.701 ± 0.007
GAT-C2L1	0.722 ± 0.002	0.978 ± 0.004	0.664 ± 0.004	0.858 ± 0.003	0.722 ± 0.002
GAT-C2L2	$\mathbf{0.726 \pm 0.001}$	0.977 ± 0.003	0.665 ± 0.007	0.865 ± 0.001	$\mathbf{0.726 \pm 0.001}$
GAT-C2L3	0.721 ± 0.003	0.970 ± 0.003	$\mathbf{0.679 \pm 0.006}$	$\mathbf{0.870 \pm 0.002}$	0.721 ± 0.003
GAT-C2L4	0.710 ± 0.003	0.968 ± 0.002	0.670 ± 0.006	0.867 ± 0.001	0.710 ± 0.003
GAT-C2L5	0.700 ± 0.002	0.959 ± 0.001	0.652 ± 0.010	0.865 ± 0.001	0.700 ± 0.002
GAT-C2L6	0.702 ± 0.003	0.961 ± 0.003	0.652 ± 0.009	0.865 ± 0.001	0.702 ± 0.003