SpaCeNet: Spatial Cellular Networks from omics data

Stefan Schrod[†1], Niklas Lück[†1], Robert Lohmayer[2], Stefan Solbrig[3], Dennis Völkl[3], Tina Wipfler[3], Katherine H. Shutta[4,5], Marouen Ben Guebila[4], Andreas Schäfer[3], Tim Beißbarth[1,6], Helena U. Zacharias[7], Peter J. Oefner[8], John Quackenbush[4,5], and Michael Altenbuchinger[*1]

[1]Department of Medical Bioinformatics, University Medical Center Göttingen, Germany.
[2]Leibniz Institute for Immunotherapy, 93053 Regensburg, Germany.
[3]Institute of Theoretical Physics, University of Regensburg, Germany.
[4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
[5]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
[6]Campus Institute Data Science (CIDAS), University of Göttingen, Germany.
[7]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover Medical School, Hannover, Germany.
[8]Institute of Functional Genomics, University of Regensburg, Regensburg, Germany.

Advances in omics technologies have allowed spatially resolved molecular profiling of single cells, providing a window not only into the diversity and distribution of cell types within a tissue, but also into the effects of interactions between cells in shaping the transcriptional landscape. Cells send chemical and mechanical signals which are received by other cells, where they can subsequently initiate context-specific gene regulatory responses. These interactions and their responses shape the individual molecular phenotype of a cell in a given microenvironment. RNAs or proteins measured in individual cells together with the cells' spatial distribution provide invaluable information about these mechanisms and the regulation of genes beyond processes occurring independently in each individual cell. "SpaCeNet" is a method designed to elucidate both the intracellular molecular networks (how molecular variables affect each other within the cell) and the intercellular molecular networks (how cells affect molecular variables in their neighbors). This is achieved by estimating conditional independence relations between captured variables within individual cells and by disentangling these from conditional independence relations between variables of different cells. A python implementation of SpaCeNet is publicly available at https://github.com/sschrod/SpaCeNet.

[*]Correspondence: michael.altenbuchinger@bioinf.med.uni-goettingen.de. †The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

1 Introduction

Measurements of spatially resolved RNA or protein expression patterns open unprecedented opportunities to study questions in areas such as developmental biology and pathophysiology, where interactions between cells are known to influence a wide range of processes. Methods such as "spatial transcriptomics" [1] and Slide-Seq [2] used molecular barcoding to count mRNAs aggregated in small regions and integrated those with images to produce a transcript map. A drawback of these early methods was that mRNAs from multiple cells in a small region could contribute to the observed signal, masking differences between cells and cell-cell interactions. Subsequent barcoding techniques substantially improved resolution, such as Slide-Seq [3] and Slide-SeqV2 [4] with a resolution of $\sim$ 10µm, as well as, most recently, Seq-Scope [5] and Stereo-Seq [6] with a resolution of $\sim$ 0.5µm. *In situ* hybridization or sequencing methods can measure the expression of many genes with single-cell resolution (and even subcellular resolution) [7, 8, 9, 10], but these methods require complex instrumentation and long imaging times. Three-dimensional intact tissue sequencing in single cells has been achieved by STARmap (spatially-resolved transcript amplicon readout mapping) [11], which is capable of measuring simultaneously the expression of about 1000 genes at single-cell resolution in a three-dimensional thick tissue section. Although spatial omics technologies are in their infancy, the launch of commercial products such as the 10X Genomics Visium platform has led to increased interest in methods that will allow the analysis, interpretation and exploration of data generated [12, 13, 14, 15, 16, 17].

Cells organize themselves spatially within tissues and organisms in order to carry out specific functions. This organization is orchestrated via signals that include physical interactions via cell-cell contact, chemical signals, and exosome-mediated transfer of RNAs between cells. Each cell's individual phenotype together with its location in space relative to other cells captures information about this process. For instance, genes encoding chemokines (chemoattractant cytokines that facilitate intercellular communication) are first transcribed and then translated before the respective proteins pass through the cell membrane into the extracellular domain where they recruit leukocytes. When these signals are received by the leukocytes, they initiate signaling cascades that finally induce a molecular response leading the cells to adapt their individual molecular phenotype and alter their behavior. While such interactions are known to be essential for many biological processes [18], there are no well-established statistical methods to investigate the relationships between spatial organization, gene or protein expression, and cellular phenotype. There are first approaches to infer cellular interactions from spatial omics, although they differ substantially with respect to semantics, modelling approaches, and overall objectives. In seminal work [13], Arnol et al. proposed Spatial Variance Component Analysis (SVCA) to decompose the expression of individual genes into spatial and non-spatial contributions, namely into cell intrinsic effects, general environmental effects, and cell-cell interaction effects. SVCA is motivated by Gaussian processes and models gene/protein expression independently of each other. Thus, it does not capture complex multivariate relationships between genes [13], but detects individual genes which are involved in cell-cell interactions. Another approach is Node-Centric Expression Modeling (NECM) [19]. NECM uses a graph neural network to predict cells' observed gene expression vector from respective cell type label and niche, with the latter resembling cell-cell communication in terms of statistical dependencies between cells. As such, NECM identifies cell-type couplings. However, it does not identify gene pairs involved in cellular interactions. This was addressed by COMMunication analysis by Optimal Transport (COMMOT) using the framework of optimal transport to account for the competition among ligand and receptor species, while taking into account the spatial distances between cells [20], by Mixture of Experts for Spatial Signaling genes Identification (MESSI) considering an additional cell neighbourhood to infer intra-cellular responses [21], and by Multiview Intercellular SpaTial

modeling framework (MISTY) using an aggregated direct neighbourhood to resolve spatial effects [22]. However, rigorous statistical models to infer complex cellular interdependencies from spatially distributed molecular data are still missing.

The statistical inference of correlation-based molecular networks from high-dimensional omics data is based on the assumption that coordinated expression in a cell might provide insight into processes that are activated or inactivated in different phenotypes. Early attempts used pairwise measures of gene-gene coexpression such as mutual exclusivity, Pearson's and Spearman's correlations and identified network edges between genes based on a correlation threshold [23, 24]. Such measures of coexpression can provide insight into active biological mechanisms, but they are vulnerable to identifying spurious associations [25]. These associations can be the consequence of indirect dependencies that cannot be resolved if pairwise relationships between molecular variables are considered only in isolation from all other molecular variables. This stimulated research in high-dimensional statistics, specifically in Probabilistic Graphical Models (PGMs). PGMs resolve the dependency structure of molecular variables and, thus, disentangle direct from indirect associations. This is even possible in high-dimensional settings, where the number of variables is larger than the number of observations. The main concept underlying PGMs is conditional independence. Two variables X and Y are considered as conditionally independent (given all remaining variables), if X does not provide any additional information about Y that is not already covered by the remaining variables. Thus, although X and Y might be pairwise correlated, they can still be conditionally independent; the pairwise correlation could be just the consequence of X's and Y's indirect relationship mediated via other variables and not due to a direct relationship [26]. This powerful concept made PGMs one of the favored approaches to resolve molecular networks from molecular data [24] and, importantly, they can be straightforwardly extended to account for complex single-cell and multi-omics data [27, 28]. It is nevertheless noteworthy that PGMs are undirected graphical models and do not resolve causal relationships, although, intriguingly, lower bounds on causal effects can be provided [29].

Spatial Cellular Networks from omics data (SpaCeNet) is a method for analyzing patterns of correlation in spatial transcriptomics data by extending the concept of conditional independence to spatially distributed information, facilitating reconstruction of both the intracellular and the intercellular interaction networks with single-cell spatial resolution. SpaCeNet was developed to address the diversity of cellular interactions and the various length scales over which they occur. SpaCeNet introduces flexible interaction potentials in combination with appropriate regularization strategies to allow this diversity in cellular state, tissue organisation and spatial communication to be handled effectively. We validate SpaCeNet in extensive simulation studies and illustrate its capacity to augment exploratory data analysis of spatial transcriptomics data from the mouse visual cortex and the *Drosophila* blastoderm.

2 Results

2.1 SpaCeNet infers cell-cell interactions from spatial omics data

SpaCeNet concept The individual molecular phenotype of a cell is shaped by its microenvironment and the underlying processes are diverse, comprising the exchange of signals via direct physical contact or via signalling molecules. Throughout the article, we will refer to any process with which cells affect each others molecular appearance simply as cell-cell or cellular "interactions". Spatial Transcriptomics (ST) involves measuring cellular phenotypes along with cells' location in space, providing information about such processes and serving as a valuable resource to study cell-cell interactions. To develop SpaCeNet, we require single-cell resolved ST data (single-cell profiles + cellular positions) such that cells can be described as interacting

units. In recent years, such data have become increasingly available via both multiplexing and barcoding techniques [11, 5, 6].

Key to SpaCeNet is a statistically sound approach that uses probabilistic graphical models (PGMs) to decompose the observed cellular profiles into contributions arising from ordinary cellular variability and contributions from cellular interactions. Formally, SpaCeNet decomposes each cell's observed molecular profile into (i) a baseline contribution corresponding to the cell's molecular profile in isolation (Figure 1A) and (ii) a residual contribution attributed to its interaction with other cells (Figure 1B and C). Importantly, this decomposition uses a model parametrization such that vanishing parameters encode conditional independence statements, namely intracellular and intercellular spatial conditional independence (SCI) relationships as introduced in the following and as summarized in Suppl. Figure S1.

PGMs are graphical models in which nodes represent variables assumed to be distributed according to a probabilistic model and edges between nodes represent conditional dependence between them, where the conditioning is on the remaining nodes in the network. The absence of edges in a PGM defines a set of Conditional Independence (CI) relations corresponding to direct pairwise independence between variables: for two variables X and Y that are CI, any association between X and Y observed in the data can be attributed to indirect effects of other variables in the system. Conversely, if X and Y are not CI, i.e., there is an edge between the two, this dependence is not mediated by other variables in the system. As such, PGMs identify and remove potential erroneous (false-positive; indirect) associations. This makes PGMs versatile tools to infer gene-association networks, as repeatedly shown in the literature [24, 25].

Here, we develop a similar framework to model intracellular and intercellular CI relations in ST data. We estimate SCI between variable X^a (gene X in cell a) and variable Y^b (gene Y in a neighboring cell b) keeping all other variables in the data fixed; similar to CI, by conditioning on all variables except for X^a and Y^b, we can study the direct dependence of X^a on Y^b and vice versa. As such, SpaCeNet disentangles (1) direct from indirect relationships of variables in individual cells, and (2) direct from indirect relationships between molecular variables in different cells. A direct consequence of both is that each cell's environmental adaptation in its spatial context can be estimated distinctly from its molecular phenotype in isolation. Since we are interested in sample estimates of cellular interaction patterns, SpaCeNet infers SCI simultaneously across all pairs of cells in a ST dataset.

SpaCeNet infers both short and long-range cell-cell interactions while taking into account the molecular diversity of cells SpaCeNet encodes intercellular SCI relationships via interaction potentials mediating the association strength between molecular variables across cells in a distance dependent manner. We denote the interaction potential between X^a (gene X in cell a) and Y^b (gene Y in cell b) at distance r_{ab} as $\rho_{XY}(r_{ab})$. Since $\rho_{XY}(r_{ab})$ depends only on the distance r_{ab} between two cells, the intercellular potential between genes X and Y takes the same functional form across all modeled cells and, moreover, vanishing potentials $\rho_{XY}(r_{ab})$ imply that X^a and Y^b are SCI across all cells a and b. One might ask whether such a one-function-fits-all approach is appropriate to model complex cellular communication between molecularly diverse cells, considering that even related cell types, such as T-cells and B-cells, fulfill very different tasks and are expected to send and receive very different signals. As briefly illustrated, SpaCeNet can encode such diverse cell-cell communication patterns taking into account the individual cell's molecular phenotype without requiring *a priori* specification of involved cell types, variables, and interaction mechanisms. Assuming all variables except X^a and Y^b to be fixed, X^a depends on Y^b via the regression formula $X^a = \mu_X - \frac{1}{\omega_{XX}}\rho_{XY}(r_{ab})(Y^b - \mu_Y) + \epsilon$ (this can be also seen by marginalizing the full density Eq. (1) over all variables except X^a and Y^b). Thus, the effect that molecular variable Y^b in cell b exerts on X^a in cell a via cell-cell interaction is

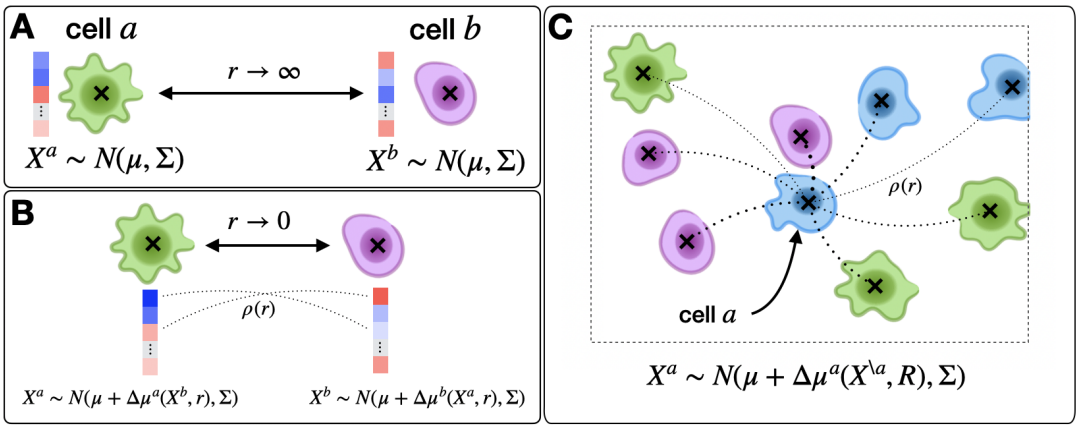


Figure 1: **SpaCeNet concept and how it applies to different cellular contexts.** Figure **A** shows a schematic picture of two isolated cells a and b which are infinitely separated (distance $r \to \infty$). In this case, respective cellular profiles with expression levels illustrated in blue and red are modeled by a multivariate normal distribution with mean expression μ and covariance Σ. The latter encodes the complex co-expression pattern of different RNAs/proteins and facilitates a description of molecularly diverse cells. Figure **B** shows for comparison a scenario of two neighboring cells, where the cells affect each other; the molecular phenotype changes compared to Figure **A** as a consequence of the different cellular context. This molecular adaptation of cell a (analogously for cell b) is modeled by a shifted mean expression vector $\mu + \Delta\mu^a(X^b, r)$, where the molecular adaptation depends on the molecular phenotype of cell b and the cell-cell distance r. The molecular adaptation is parametrized by interaction potentials which directly provide estimates of spatial gene-gene dependencies between gene i and j. Figure **C** sketches the more complex scenario of a set of interacting cells. As such, the expression of cell a is affected by all surrounding cells in a distance-dependent way, as illustrated by thin and bold dotted lines for long- and short-range interactions, respectively. One should note that the mean expression of cell a now depends on the molecular phenotype of all other cells $X^{\setminus a}$ as well as all respective cell-cell distances summarized in R.

directly related to the individual molecular profile of cell b, which means that the molecular phenotype of a cell determines how it communicates with other cells. One should also note that SpaCeNet infers the $\rho_{XY}(r)$ without assuming a particular functional dependence on r_{ab} via series expansion in powers of $1/r_{ab}$. This series expansion is motivated by the fact that infinitely separated cells cannot communicate through the release or absorbance of particular signaling molecules and the drop in concentration of these molecules with distance suggests that $\rho_{XY}(r_{ab})$ should vanish for $r_{ab} \to \infty$.

To show that SpaCeNet can reconstruct diverse cell-cell interactions, we performed four illustrative simulations shown in Figure 2a to d, where we simulated radial dependencies corresponding to long-distance interactions (such as via paracrine signaling, Figure 2a and b), short-distance interactions (such as via cell-cell contact) using an exponentially decreasing potential with short range (Figure 2c), and an interaction where the potential first grows with r_{ab}, peaks at average distances, and then goes to zero for large r_{ab} (Figure 2d). The precise radial dependencies are provided in the caption of Figure 2. In principle, SpaCeNet can model interaction potentials to arbitrary orders in $1/r_{ab}$. We present the corresponding radial dependencies estimated via SpaCeNet for expansions up to order $(1/r_{ab})^L$ for $L = 1$ (solid lines), $L = 3$ (dashed lines), and $L = 10$ (dotted lines). As expected, the approximation of the true underlying interaction improves with increasing L. However, increasing the order in $1/r_{ab}$ results in higher model

complexity, which in turn comes at the cost of additional parameters, increasing the risk of overfitting and computational burden.

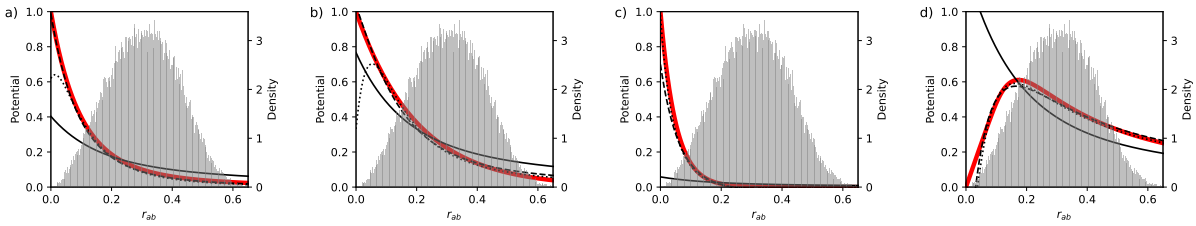


Figure 2: **SpaCeNet can reconstruct diverse cellular interaction patterns.** Figures a to d show different potentials used to simulate data in red in dependence of the distance between cells r_{ab}. Potentials inferred by SpaCeNet are shown for three different expansion orders: $L = 1$ (black solid lines), $L = 3$ (black dashed lines), and $L = 10$ (black dotted lines), illustrating that with higher order, increasingly complex cellular interactions can be reconstructed. The corresponding ground-truth potentials (red lines) are (a) a standard decreasing potential corresponding to the second-order term in the series expansion, $\rho(r_{ab}) = (1 - \exp(-r_{ab}/r_0))^2 (r_0/r_{ab})^2$ with $r_0 = 1/10$, (b) a long-range exponential potential $\rho(r_{ab}) = \exp(-5\, r_{ab})$, (c) a short-range exponential potential $\rho(r_{ab}) = \exp(-20\, r_{ab})$, and (d) a potential $\rho(r_{ab}) = 10\,(1 - \exp(-r_{ab}/2))\,(1 - \exp(-1/(5r_{ab})^2))$ which increases for small r_{ab} and then decreases again. The corresponding densities of the pairwise distances r_{ab} are shown as histograms in the background of the figures with the respective y-axis depicted at the figures' right axis. For more details about data simulation see Suppl. Sect. 1.

SpaCeNet recovers inter-and intracellular interactions from *in silico* generated tissues We systematically evaluated SpaCeNet's capacity to reconstruct intra- and intercellular networks defined in terms of SCI relationships. To this end, we generated *in silico* tissues from the full probability density Eq. (1) and tested if the ground truth intra- and intercellular network edges are correctly recovered (see also Suppl. Section 2). Since cellular interactions occur at various length scales, we performed three different studies involving (1) long-range interactions, (2) medium-range interactions, (3) short-range interactions, and (4) a mixture of long-, medium- and short-range interactions. For each study, we simulated a total of $n \cdot S \in \{10^3, 10^4, 10^5\}$ cells, where n is the number of cells per ST slide and $S \in \{1, 10, 100\}$ the number of ST slides. Performance was assessed using the area under the precision recall curve (AUPRC) and the area under the receiver operating characteristic curve (AUROC) (see Suppl. Table S1).

In our first simulation study we used an exponentially decreasing interaction potential $\rho_{ij}(r_{ab}) = \Delta\rho_{ij} \cdot \exp(-\phi_{ij} r_{ab})$ with a comparatively large range of $1/\phi_{ij} = 1/5$. Suppl. Figure S4A gives AUPRC versus $n \cdot S$ to recover the correct intracellular networks (the intracellular precision matrix $\mathbf{\Omega}$, left figure) and the intercellular interaction networks (the cell-cell interaction parameters $\mathbf{\Delta\rho}$, right figure). We find that SpaCeNet is capable of reconstructing intracellular networks across the full range of simulated cells, $n \cdot S$, with a median AUROC larger than ~ 0.85 throughout all settings. Edge recovery of cell-cell interactions was achieved with a median AUPRC ranging from 0.28 for $n = 10^3$ and $S = 1$ (blue) to 0.92 for $n = 10^3$ and $S = 100$. It is worth noting that edge recovery for both the intra- and intercellular networks was slightly better if cells are distributed across measurements, as seen by comparison of the AUPRC for $n = 10^3$ and $S = 1$ (blue), for $n = 100$ and $S = 10$ (orange), and for $n = 10$ and $S = 100$ (green), that all share the same total number of cells $n \cdot S = 10^3$. This trend also persisted

for $n \cdot S = 10^4$. We also tested the same class of potentials with medium range $1/\phi_{ij} = 1/10$ (Suppl. Figure S4B) and short range $1/\phi_{ij} = 1/20$ (Suppl. Figure S4C). The results were similar to the first simulation study. The edge recovery in terms of AUPRC improved with an increasing number of cells $n \cdot S$, the more so when cells were distributed across multiple samples S. Finally, we studied how diverse cell-cell interactions impair model inference (Suppl. Figure S4D). For this, we repeated the previous simulations, but created flexible potentials $\rho_{ij}(r_{ab}) = \Delta\rho_{ij} \exp(-\phi_{ij} r_{ab})$ with $\phi_{ij} \sim \text{Unif}(5, 20)$ and $\Delta\rho_{ij} \sim \text{Unif}(-1, 1)$ between molecular variables i and j that differ with respect to interaction range and strength. For illustration purposes, these different potential ranges are shown in Suppl. Figure S2 for $\phi_{ij} = 5$ (solid line), $\phi_{ij} = 10$ (dashed line), and $\phi_{ij} = 20$ (dotted line). In summary, we found that SpaCeNet achieves an edge-recovery performance that is almost as good as in the more controlled setting of the previous simulation studies, in particular when the total number of cells $n \cdot S$ is increased. An exemplary SpaCeNet model obtained in this study together with its ground truth is shown in Suppl. Figure S3 for two different model regularization parameters β. Results considering AUROC for all four simulation studies are shown in Suppl. Figure S4, supporting our previous findings.

To further illustrate the relevance of distance dependent interactions, we compared SpaCeNet to two alternative approaches, namely an ordinary spatial correlation and its naive partial extension implemented as follows. For each individual cell i and gene j, we summed up the expression of the top K nearest neighbors, yielding an additional set of "regional expression variables". Then, we correlated the cell's expression of gene j with the regional expression of gene k for all j and k, yielding a matrix of correlation coefficients. The performance evaluation shown in Suppl. Figure S5 and Suppl. Table S2 contrasts the performance of this naive correlation based approach for $K = 1, 5, 10, 20, 50$ with SpaCeNet. We observe that, first, overall this approach cannot compete with SpaCeNet and, second, that its performance depends on K, with long-range interactions typically better captured by large K (simulation A) and short range interaction better captured for small K. In contrast, SpaCeNet dynamically adjusts for the interaction range and does not require any *a priori* specification of interacting neighbors. As a further benchmark, we included an extension of the naive spatial correlation measure to a partial correlation measure. Let j' denote the regional expression variable which represents the expression of gene j of the top K neighbors. A remedy to estimate spatial associations between genes j and k has to take into account that both might also correlate to each other within individual cells. Therefore, we decided to calculate partial correlations between variables j and k' conditioning on j' and k. Performance analysis of this additional baseline strongly supports our finding that SpaCeNet, as a dynamic approach that intrinsically infers complex distance dependencies, is much more suited to infer cell-cell interactions (Suppl. Figure S6 and Suppl. Table S3).

SpaCeNet recovers intra- and intercellular interactions from *in silico* tissues generated via mechanistic modeling Former simulations illustrate that SpaCeNet can reliably recover SCI relationships. Next, we aimed to verify SpaCeNet in more realistic settings. Tanevski et al. [22] used a comprehensive *in silico* tissue model to mimic the interactions of different cell types through ligand binding and subsequent signaling events. They generated two *in silico* tissues with pre-specified cell-cell interactions between four cell types, capturing interactions of 29 molecular species including 5 ligands, 5 receptors, and 19 intracellular signaling proteins. The authors suggested Multiview Intercellular SpaTial modeling framework (MISTy) to derive importance scores for each pair of markers to infer intracellular and intercellular molecular interactions. We followed their work and considered an intracellular interaction as correct if there is a direct interaction between the markers in the model's networks and an intercellular interaction as correct if one of the markers is responsible for a ligand production and the other

Table 1: Performance in terms of AUPRC to recover intra- (left) and intercellular (right) interactions from *in silico* tissues generated via mechanistic modeling for MISTy and SpaCeNet.

AUPRC (intracellular)	MISTy	SpaCeNet
Tissue 1	0.21	0.58
Tissue 2	0.19	0.54
Tissue 1+2	0.21	0.50

AUPRC (intercellular)	MISTy	SpaCeNet
Tissue 1	0.07	0.21
Tissue 2	0.06	0.22
Tissue 1+2	0.07	0.22

one is activated by it [22]. Respective results in terms of AUPRC are summarized in Table 1 for MISTy and SpaCeNet. To evaluate SpaCeNet's performance, we restricted ourselves to leading order potentials to ensure comparability. We observe that SpaCeNet consistently outperformed MISTy with respect to both the recovery of intra- and intercellular interactions. This is also supported by a respective analysis considering AUROC (Suppl. Table S4). One should note that MISTy does not disentangle direct from indirect relationships. As a consequence, most false positive interactions derived by MISTy might be the result of indirect or higher-order interactions [22].

2.2 SpaCeNet resolves molecular mechanisms involved in cellular interactions in the mouse visual cortex

In the following, we present an example analysis using SpaCeNet for spatial transcriptomics data from the mouse visual cortex generated by STARmap [11]. STARmap labels and amplifies cellular RNAs. These amplicons are then transferred to a hydrogel while lipids and proteins are removed. The hydrogel is optically transparent and can be sequentially imaged through multiple cycles with a low probability of errors and miscodings. STARmap measurements of the mouse visual cortex were downloaded from https://www.starmapresources.org/data. The data consist of measurements for 28 RNAs in ∼30,000 cells together with their respective locations in a $0.1 \times 1.4 \times 1.7$ mm^3 tissue section.

We used SpaCeNet to estimate a global, intracellular network (the precision matrix $\mathbf{\Omega}$) and a network of spatial molecular interactions (the spatial interaction parameters $\mathbf{\Delta\rho}^{(\cdot)}$), where we set $L = 3$ and split the data into four equally sized batches of which three served for model building and one for model validation and hyper-parameter calibration (Suppl. Figure S7). We tuned α, $\beta \in [10^{-5}, 10]$ on a 4×4 grid that was refined 6 times (Suppl. Figure S8) and selected the best set of hyper-parameters based on the highest validation pseudo-log-likelihood. A final model was estimated using the full data set.

For the intracellular precision matrix $\mathbf{\Omega}$ we obtained a complete matrix with weights summarized in Suppl. File 1. For the spatial interactions $\mathbf{\Delta\rho}^{(\cdot)}$, SpaCeNet selected a set of 134 out of 406 possible spatial associations (Suppl. Figure S9, Suppl. File 1). We ranked edges according to $\Delta\rho_{ij} = ||\mathbf{\Delta\rho}_{ij}^{(\cdot)}||_2 = \sqrt{\sum_l \Delta\rho_{ij}^{(l)^2}}$, for which the greatest edge weight was between myelin basic protein gene (*Mbp*) and FMS-related receptor tyrosine kinase 1 (*Flt1*) with $\Delta\rho_{Mbp,\ Flt1} = 0.328$ and the leading-order contribution $\mathbf{\Delta\rho}^{(1)}_{Mbp,\ Flt1} = 0.327$. Negative/positive values of $\Delta\rho_{ij}^{(l)}$ correspond to positive/negative spatial associations in analogy to the definition of the precision matrix $\mathbf{\Omega}$.

First, we verified the absolute residuals between the observed data matrix $\mathbf{X}$ and its SpaCeNet reconstruction $\hat{\mathbf{X}}$ using either (a) the intracellular edges only ($\mathbf{\Delta\rho}^{(\cdot)} = \mathbf{0}$), or (b) the full model (see also Suppl. Sect. 2). The results corresponding to this analysis are shown in Figure 3a and

b, respectively. We observed that the spatial associations improve model building substantially, as illustrated by the band of large residuals in the upper left quadrant for both *Flt1* and *Mbp* (Figure 3a and d), that were reduced substantially via the spatial interactions (Figure 3b and e). Figure 3c and f show the corresponding contributions from the spatial interactions $\boldsymbol{\Delta\rho}^{(\cdot)}$.

The connection between *Mbp* and *Flt1* was interesting for a number of reasons. Myelin basic protein is the second most abundant protein, after proteolipid protein, of the myelin membrane in the central nervous system (CNS), making up approximately 30% of the total protein of myelin [30] and *Mbp* mutants lack compact myelin in the CNS [31]. Myelin surrounds nerve cell axons to insulate them and to increase the conduction of electric impulses [32]. Demyelinating diseases of the CNS, of which multiple sclerosis (MS) is the most common, are characterized by damaged myelin sheaths [33]. *Flt1* encodes a member of the vascular endothelial growth factor receptor (VEGFR) family. VEGFRs mediate diverse cellular communication signals controlling developmental processes, such as neurogenesis or gliogenesis [34]. VEGFRs recognize vascular endothelial growth factors (VEGFs), whose expression has been shown to be upregulated in both acute and chronic MS plaques [35]. Our SpaCeNet model resolves that *Flt1* levels are negatively associated with *Mbp* levels in neighboring cells. A possible interpretation is that cells expressing *Flt1* accumulate in the spatial vicinity of cells which lack *Mbp*. This would be in line with the observation that VEGF-A (which is recognized by Flt1) promotes migration of oligodendrocyte precursor cells (OPCs) in a concentration-dependent manner [36], as shown by anti-Flk-1 (not by anti-Flt-1) receptor-blocking antibody. Moreover, it was shown in the medulla oblongata of the adult mouse that OPCs contribute to focal remyelination and that VEGF signaling might be required for their proliferation [37]. It has been long known that oligodendrocytes are the myelinating cells of the CNS derived from OPCs [38], therefore our finding could be a hint that Flt1 mediates signals necessary to guide myelinating cells such as OPCs and mature oligodendrocytes to axons that lack myelin. Interestingly, in a study by Vaquie et al. [39], injured axons were shown to communicate with Schwann cells to trigger the formation of actin spheres. These spheres constrict the axons, leading to their fragmentation and faster removal of debris after injury. This process was shown to be controlled by VEGFR1 activity, and can be acquired by oligodendrocytes through enforced expression of VEGFR1 [39].

It should be noted that SpaCeNet models are associative, not causal. Thus, we can not elucidate whether decreased *Mbp* levels imply increased *Flt1* in neighboring cells or if increased *Flt1* levels imply decreased *Mbp* in neighboring cells. Moreover, since only a selected set of 28 genes was spatially assessed by STARmap in a thick tissue section, and since numerous signaling steps may be involved in establishing this relationship, the underlying mechanisms remain to be fully elucidated.

The spatial association *Flt1-Mbp* was followed in strength by *Ctgf-Gja1*, *Ctgf-Pcp4*, and *Reln-Sst* associations. Connective tissue growth factor (*Ctgf*), also known as Cellular Communication Network Factor 2 (*Ccn2*), belongs to the CCN family. CCN proteins are a family of extracellular matrix proteins involved in intercellular signaling [40]. We obtained $\Delta\rho_{Ctgf,\ Gja1} = 0.103$, with leading-order contribution $\boldsymbol{\Delta\rho}^{(1)}_{Ctgf,\ Gja1} = 0.1023$. *Ctgf* has been reported to facilitate gap junction intercellular communication in chondrocytes through up-regulation of connexin 43 (*Gja1*) expression [41]. Purkinje Cell Protein 4 (Pcp4) regulates calmodulin activity and might contribute to neuronal differentiation through the activation of calmodulin-dependent kinase signaling pathways [42]. The *Ctgf-Pcp4* potential was estimated with $\Delta\rho_{Ctgf,\ Pcp4} = 0.083$ and $\boldsymbol{\Delta\rho}^{(1)}_{Ctgf,\ Pcp4} = 0.0831$; we are not aware of a mechanism that might explain the spatial association observed between *Ctgf* and *Pcp4*. This association was followed by *Reln* (Reelin) - *Sst* (Somatostatin) with parameters $\Delta\rho_{Reln,\ Sst} = 0.083$ and $\boldsymbol{\Delta\rho}^{(1)}_{Reln,\ Sst} = 0.0780$. The gene *Reln* encodes a secreted extracellular matrix protein that might play a role in cell-cell interactions critical for cell positioning. Somatostatin affects transmission rates of neurons in the CNS and

cell proliferation. GABAeric cortical interneurons can be delineated to 95% by the markers Pv, Sst (co-expressed with Reelin), Reelin (without Sst), and Vip [43]. The medial ganglionic eminence gives rise to the population of interneurons that co-express Reelin and Somatostatin, while the caudal ganglionic eminence gives rise to interneurons that express Somatostatin but lack Reelin [43]. The observed spatial *Reln-Sst* association could be a hint that the distribution of these two populations is spatially well organized and not random.

SpaCeNet assumes multivariate normality. To further ensure that the derived results are robust and to explore the extent to which they are biased by data preprocessing, we repeated the former analysis but used log-transformed data instead. Log-transformed data are believed to make data more normally distributed, but are also prone to inflating the variance for lowly expressed genes [44]. Thus, it is not *a priori* clear whether SpaCeNet can better deal with data on a natural scale or on log scale. Respective top edges obtained by SpaCeNet are summarized in Suppl. Table S5, illustrating that the top hits are still recovered with confidence, although the ordering slightly changed; among the top genes, the pairs *Mbp-Flt1*, *Ctgf-Gja1*, and *Ctgf-Pcp4* were confirmed. The pair *Sst-Pvalb* newly appeared among the top hits, which might be attributed to spatial delineation of Parvalbumin-positive (*Pvalb+*) and Somatostatin-positive (*Sst+*) cells. As a final consistency check, we explored the potential role of redundant variables and how they potentially compromise model inference. For this purpose, we replaced *Mbp* in the data matrix by two noisy copies of itself and repeated our analysis with results shown in Suppl. Table S5. We observed that the edges to *Mbp* are still robustly established, but edge strength was lowered by a factor of approximately two. This is in line with recent findings in mixed graphical models [45].

Comparative analysis of spatial associations using pair-wise and partial spatial correlations: For comparison and illustration purposes, we performed the ordinary pair-wise spatial correlation analysis outlined above on the mouse visual cortex data, selecting $K = 5$ in line with [13]. The largest absolute spatial correlation value between different genes was observed for the pair *Mbp-Flt1* with $\bar{r}_{Mbp,\ Flt1} = 0.82$, which was also the top hit in the SpaCeNet analysis. However, we observed the opposite sign for the association; SpaCeNet suggests a negative, while this analysis suggests a positive association. To further resolve this contradictory result, we compared the estimated pair-wise spatial correlation matrix with an ordinary correlation matrix derived across all single-cell profiles (see Suppl. Figures S10A and S11A), respectively. Comparison of both matrices reveals a strong coincidence with differences distributed as shown in Suppl. Figures S10 and Suppl. Figure S11.

Given the high similarity between both, we hypothesized the presence of strong auto-correlations between neighboring cells, meaning that high spatial correlations are just the consequence of similar expression patterns in adjacent cells. For the pair *Mbp-Flt1*, the ordinary correlation analysis revealed a correlation coefficient of 0.87, supporting this hypothesis for this specific case.

To explore if such spatial auto-correlations deteriorate estimates of spatial associations, we applied additionally the outlined spatial partial correlation measure with $K = 5$. Table 2 compares the estimated spatial partial associations with the results of SpaCeNet for the outlined top pairs. Both approaches now suggest a negative association for all four pairs. Moreover, the relative association strength is very similar with *Mbp-Flt1* being the strongest association, and *Ctgf-Pcp4* and *Reln-Sst* having similar association strengths. The pair *Ctgf-Gja1* shows a relatively weak association in contrast to the SpaCeNet analysis. One should note, however, that previous analysis was based on a nearest neighbor estimate of gene expression, ignoring potential complex radial dependencies, and did not explore the full capacity of PGMs to disentangle direct

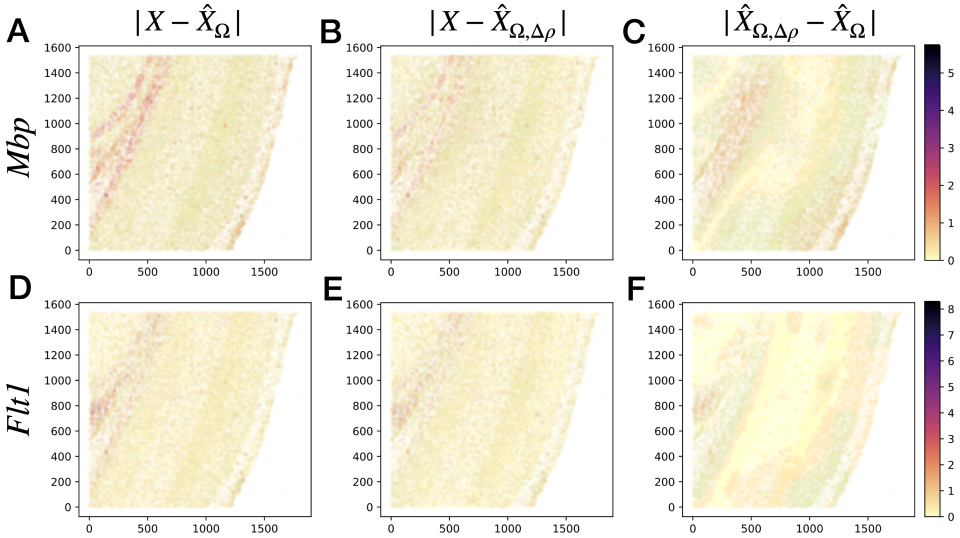


Figure 3: **Spatial interactions improve on goodness of fit.** The figures show absolute prediction residuals for spatially associated genes *Mbp* (top row) and *Flt1* (bottom row) in the mouse visual cortex data from [11] with respect to their position in space. The left column (Figures A and C) shows the absolute residuals between the ground truth data and the predictions based on the intracellular model parameters only, $|\mathbf{X} - \hat{\mathbf{X}}_{\boldsymbol{\Omega}}|$. The middle column Figures B and E display the residuals if both intra- and intercellular interactions are considered, $|\mathbf{X} - \hat{\mathbf{X}}_{\boldsymbol{\Omega},\boldsymbol{\Delta\rho}}|$. The right column Figures C and F show the contributions from spatial interactions only, $|\hat{\mathbf{X}}_{\boldsymbol{\Omega},\boldsymbol{\Delta\rho}} - \hat{\mathbf{X}}_{\boldsymbol{\Omega}}|$.

from indirect relationships by considering all molecular variables simultaneously, which might explain the observed differences. Interestingly, ranking from highest to lowest absolute spatial partial correlations also yielded the pair *Mbp-Flt1* as top hit, suggesting that disentangling the intra- from inter-cellular correlations might be one of the main ingredients necessary to identify the most interesting spatial associations.

Table 2: Estimated spatial partial correlations versus the association strength determined by SpaCeNet for the four top-ranked associations discussed in the main text. Here, jk' corresponds to spatial partial correlations between variable j and k' conditioning on j' and k. An analoguous definition holds for $j'k$ and the respective average is given by $\frac{1}{2}(j'k + jk')$. SpaCeNet's leading order association strength is given by $\boldsymbol{\Delta\rho}^{(1)}$ with positive signs corresponding to negative associations in line with the definition of the precision matrix of ordinary GGMs.

	jk'	$j'k$	$\frac{1}{2}(j'k + jk')$	$\boldsymbol{\Delta\rho}^{(1)}$
Flt1 - Mbp	-0.560	-0.557	-0.558	0.327
Ctgf - Gja1	-0.118	-0.119	-0.119	0.102
Ctgf - Pcp4	-0.227	-0.234	-0.231	0.083
Reln - Sst	-0.174	-0.205	-0.190	0.078

2.3 SpaCeNet infers spatial gene-gene association patterns in the *Drosophila* blastoderm from low-resolution spatial omics

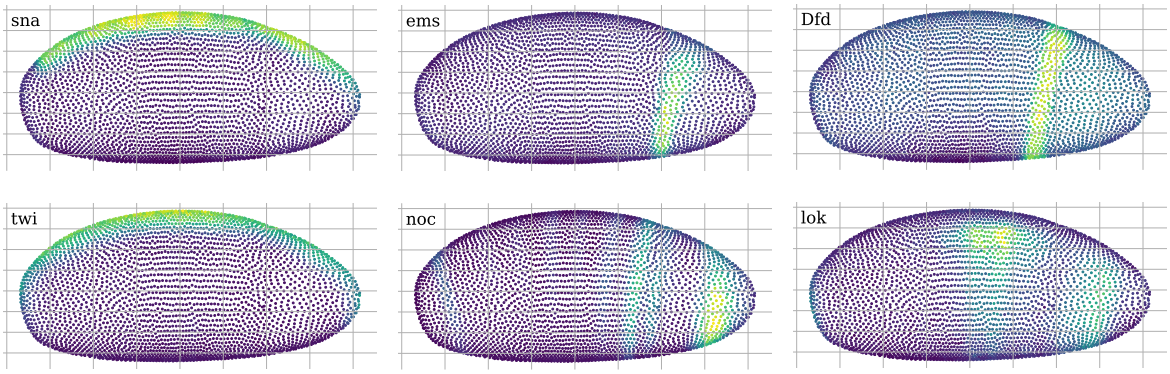
The previous analysis used data generated by STARmap [11], an in-situ sequencing technology for dense measurements at single-cell resolution, which, however, is limited in throughput.

In contrast, omics readouts together with spatial information (the transcriptome in a given spatially defined region of a specimen) have become more and more available. Here, we illustrate that SpaCeNet augments their analysis, even though the data do not resolve single cells.

The Berkeley Drosophila Transcription Network Project used a registration technique that uses image-based data from hundreds of *Drosophila* blastoderm embryos, each co-stained for a reference gene and one gene out of a preselected gene set, to generate a virtual *Drosophila* embryo [46]. We retrieved these virtual embryo data consisting of 84 genes whose expression levels were measured at 3039 embryonic locations from http://bimsbstatic.mdc-berlin.de/rajewsky/DVEX/. We then used SpaCeNet to estimate the intracellular network (the precision matrix $\mathbf{\Omega}$) and the network of spatial molecular interactions ($\mathbf{\Delta\rho}^{(\cdot)}$) for $L = 3$. We performed a hyper-parameter grid search, where we trained the model on 70% of the data and validated it on the remaining 30% (Suppl. Figure S12 and S13). The best set of hyper-parameters was selected based on the highest validation pseudo-log-likelihood and was used to fit a final model on the full data (Suppl. File 2). The intracellular network $\mathbf{\Omega}$ is a full matrix and so there is a rich dependency structure among variables not related to their spatial context. The spatial-interaction network, $\mathbf{\Delta\rho}^{(\cdot)}$, in contrast, is sparse with 238 out of 3570 possible interactions (Suppl. Figure S14). Note that the spatial molecular interactions improved the goodness-of-fit on validation data and thus also the generalizability of the model, which highlights the need to include spatial interactions.

Figure 4 contrasts the expression of the gene pairs *sna-twi*, *ems-noc* and *Dfd-lok*, that showed the highest spatial association in the SpaCeNet analysis with $\Delta\rho_{twi,\ sna} = 0.0368$, $\Delta\rho_{ems,\ noc} = 0.0306$, and $\Delta\rho_{Dfd,\ lok} = 0.0211$. We observed *twi* and *sna* to be active in the same spatial regions, which is consistent with both the positive spatial association suggested by SpaCeNet in leading order ($\Delta\rho^{(1)}_{twi,\ sna} = -0.0361$) and with the joint activation of *twi* and *sna* in the differentiation of the *Drosophila* mesoderm in localized (ventral) regions of early embryos [47]. In contrast, the genes *ems* and *noc* (for which SpaCeNet estimated a negative spatial association at leading order ($\Delta\rho^{(1)}_{ems,\ noc} = 0.0301$)) are active in adjacent but different areas of the *Drosophila* embryo. A similar observation can be made for the genes *Dfd* and *lok*, for which SpaCeNet also estimated a negative leading-order association ($\Delta\rho^{(1)}_{Dfd,\ lok} = 0.0210$). To further explore the robustness of former results with respect to pre-processing biases, we re-performed our analysis but additionally applied a log-transformation (Suppl. Table S6). Again, these analyses support the robustness of SpaCeNet as almost all leading edges coincide between both approaches.

SpaCeNet as an inferential tool to predict spatial gene expression in the *Drosophila* blastoderm SpaCeNet infers a joint density function describing spatially distributed, potentially high-dimensional molecular features. Thus, it can be also utilized as an inferential tool, predicting gene expression of a cell given its cellular context. For illustration purposes, consider the gene *Kr* that encodes the Krüppel protein, a transcriptional repressor expressed in the center of the embryo during the cellular blastoderm stage [48]. First, we predicted *Kr* expression in a leave-one-position-out approach that provides levels of *Kr* based on a cell's environment. The ground truth is shown in Suppl. Figure S15A and the corresponding predictions in Suppl. Figure S15B. We found that an accordingly trained SpaCeNet model can predict each cell's individual expression, given its location in space (mean squared error (MSE) $= 4.36 \cdot 10^{-3}$). However, slight deviations were observed in the highlighted region (red arrow). Next, we tested whether this discrepancy could be resolved by specifying *a priori* the expression levels of the remaining genes at the position of interest (Suppl. Figure S15C). This reduced the MSE to $2.88 \cdot 10^{-3}$, and the highlighted region better agrees with the ground truth, as can be seen by comparing Suppl. Figure S15A to C.



(a) $\Delta\rho^{(1)}_{twi,\ sna} = -0.036$ (b) $\Delta\rho^{(1)}_{ems,\ noc} = 0.030$ (c) $\Delta\rho^{(1)}_{Dfd,\ lok} = 0.021$

Figure 4: **Visualization of genes with strong spatial associations in the *Drosophila* blastoderm as identified by SpaCeNet.** Figures show gene expression levels of the top ranked gene pairs as identified by SpaCeNet with *sna* and *twi* in Figures A and B, *ems* and *noc* in Figures C and D, and *Dfd* and *lok* in Figures E and F, respectively. The genes *sna* and *twi* (Figures A and B) are expressed in the same areas in concordance with $\Delta\rho^{(1)}_{twi,\ sna} = -0.036$ as inferred by SpaCeNet. Genes *ems* and *noc* (Figures C and D) are expressed in adjacent but different areas, which is consistent with the inferred SpaCeNet association $\Delta\rho^{(1)}_{ems,\ noc} = 0.030$. This also holds true for the gene pair *Dfd* and *lok* (Figures E and F), which are also expressed in adjacent but different areas, consistent with $\Delta\rho^{(1)}_{Dfd,\ lok} = 0.021$. For all illustrations, the three-dimensional spatial coordinates were projected into a two-dimensional plane using a principal component analysis.

SpaCeNet analysis of high-throughput spatial transcriptomics Finally, we illustrate that SpaCeNet can deal with high-throughput data, capturing hundreds or even thousands of variables and measurements for tens of thousands of cells. In [49], the mouse organogenesis spatiotemporal transcriptomic atlas (MOSTA) was generated, which provides gene-expression maps with single-cell resolution of mouse organogenesis. We downloaded pre-processed data of the coronal hemibrain section, capturing in total measurements of 56,731 individual cells together with spatial locations[1]. We further filtered the gene space to genes which are at least expressed in 30% (Mosta A) and 10% of cells (Mosta B), yielding 315 and 1741 genes, respectively. We then applied SpaCeNet to each of these datasets (using $L = 1$ for computational efficiency), and performed a grid search for model selection. The best model was selected based on smallest Akaike Information Criterion (AIC). The top spatial associations are summarized in Suppl. Table S7 for. The total computation time for the final model was approximately 30 and 150 minutes, respectively, computed on a single A100 GPU, highlighting that SpaCeNet is computationally highly efficient. We want to highlight that model findings might be affected by low sequencing depths, challenging SpaCeNet's model assumptions. However, despite this limitation given by the data, we want to discuss the findings. For both datasets Mosta A and B, the top hits predominantly showed a positive spatial association ($\Delta\rho_{ij} < 0$). For instance, among the top 5 hits in both analysis, we observed the gene pairs *Mt1-Mt2* and *Mobp-Mbp*. *Mt1* and *Mt2* are both melatonin receptors. Interestingly, analyses in adult rat brain showed differences in the distribution of MT1 and MT2 proteins, and the labeling often appeared complementary in regions displaying both receptors [50]. The positive spatial association observed by SpaCeNet would be in line with this finding; a high expression of MT1 in a cell implies a high expression of MT2 in its neighbors and vice versa. Myelin-associated oligodendrocytic basic protein

[1] `https://ftp.cngb.org/pub/SciRAID/stomics/STDS0000058/Cell_bin_matrix/Mouse_brain_Adult_GEM_CellBin.tsv.gz`

(Mobp) is a myelin constituent exlusively expressed by oligodendrocytes [51] and shares physiochemical and biological properties with Mbp. Both were expressed in similar regions (Suppl. Figure S17), but were shown to differ with respect to localization and expression timing [51], with Mobp occurring significantly later than Mbp at the late stages of myelination. Finally, considering negative spatial associations, we detected a strong connection between *Apoe* and *Ptgds*. Prostaglandin-H2 D-isomerase is an enzyme encoded by the *Ptgds* gene. Prostaglandins are key players in neuroinflammatory and neurodegenerative diseases [52]. Apolipoprotein E (*Apoe*) influences various biological processes, and both the *Apoe* gene and its encoded proteins hold promise as targets for therapies against Alzheimer disease (AD) [53]. In summary, even in scenarios where SpaCeNet's model assumptions are challenged, it suggests promising findings for further analyses.

3 Discussion

SpaCeNet is a network inference method that determines intracellular correlation networks and cell-cell associations from spatial molecular data. SpaCeNet is based on probabilistic graphical modeling and extends the concept of conditional independence (CI) by spatial information through estimating spatial conditional independencies (SCI). These intracellular and intercellular SCI relationships encode information about how molecular variables affect each other across space. We verified SpaCeNet in comprehensive simulation studies and demonstrated the information that SpaCeNet can extract from spatial transcriptomics data in two example data sets: an expression map of the mouse visual cortex and a virtual RNA map of the *Drosophila* blastoderm. The analysis of the mouse visual cortex allowed us to generate hypotheses about the spatial organization of cell populations, such as *Flt1*-mediated signals which could be involved in the recruitment of myelinating cells towards axons which lack myelin basic protein (*Mbp*). The analysis of the *Drosophila* blastoderm showed that SpaCeNet can also yield insights if data do not have cellular resolution. In the latter case, SpaCeNet resolved spatial association patterns between molecular variables, as observed between the gene-pairs *twi-sna* and *ems-noc*.

Modern spatial transcriptomics techniques are capable of measuring RNA transcript levels with single-cell resolution in a three-dimensional space. Many molecular processes that take place within and between cells, such as translation to protein, possible post-translational modifications, and subsequent signaling cascades comprising secretion of molecules and their recognition by other cells, are not directly measured by these technologies. Although laboratory methods under development, such as single-cell proteomics and metabolomics [54, 55] or single-cell epigenetic or chromatin confirmation measurements may one day bridge this gap, at present one can only infer such associations with appropriate analytical methods. SpaCeNet is an important first step in inferring complex, multi-omic, intracellular and intercellular association networks that takes advantage of high-dimensional omics data and spatial information. The present implementation of SpaCeNet uses single-cell spatial transcriptomics data and so can only infer associations between RNA expression levels within and, more importantly, between different cells – without knowledge about any potentially complex intermediate processes. SpaCeNet is based on a robust Gaussian graphical model framework based on partial correlations, which can scale to the increasingly large and complex spatial multi-omics data sets that new laboratory technologies provide. Importantly, hidden variables can prohibit the inference of direct relationships. If they remain hidden, we are not able to infer their role in mediating dependencies, potentially leading to erroneous direct associations. However, with new technologies capturing more and more molecular variables, methods such as SpaCeNet might be key to identify the direct statistical dependencies in favor of those mediated by other variables. In this context, it is also worth mentioning that SpaCeNet's current implementation is limited to modeling

continuous variables. While many omics measurements such as gene expression levels can be reasonably modeled as continuous, this is not the case for all genomic readouts, and variables such as mutation status and chromatin accessibility might be better described as categorical variables. SpaCeNet could be easily adapted to include those using mixed graphical modeling suggested by [56]. As is generally the case for correlation-based methods, SpaCeNet does not identify causal interactions, although lower bounds on causal effects can be derived from observational data [29], and techniques such as directed acyclic graphs and respective equivalence classes may be adapted to SCI relationships, which we intend to explore in the future.

Nevertheless, SpaCeNet represents an important step forward in the analysis of spatial expression data, allowing us to move from a simple atlas of expression values and cell types to models that capture complex patterns of interactions that allow tissues to function and guide cellular growth, development, and disease processes. As new experimental techniques deliver larger and more complex multi-omics data combined with higher resolution information on the location of individual cells, techniques like SpaCeNet will become increasingly important for integrating spatial and biological contexts.

4 Materials and Methods

4.1 Data structure

SpaCeNet is designed to model continuous, potentially high-dimensional molecular variables (e.g., gene-expression levels) which are measured together with spatial information. Formally, denote the molecular variables as X_i with $i = 1, \ldots, p$ and let $\mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^n]^\top \in \mathbb{R}^{n \times p}$ be a data matrix with n measurements (molecular profiles) in its rows. Then each of these measurements $\mathbf{x}^a$ is annotated with a position r_a of cell a in space. Throughout the section, we identify the $\mathbf{x}^a$ as molecular profiles of individual cells and r_a with the respective point-position in space.

4.2 Concept

SpaCeNet considers the observed spatially distributed molecular profiles of single cells within a statistical model which decomposes the observed gene expression into two components (see Figure 1). The first contribution accounts for the statistical diversity of isolated cells, meaning that we assume that the molecular profiles of individual cells follow a p-dimensional multivariate normal distribution, $X \sim N(\mu, \Sigma)$ with mean profiles μ and covariance Σ (see Figure 1A). Note that this is a reasonable assumption if data are approximate normal or if they are transformed accordingly (e.g., via non-paranormal transformations) [57, 58]. The second contribution accounts for the fact that cells are not isolated but observed in a complex environment where cells can interact with each other (see Figure 1B and C). This effect is encoded as shifts of the distribution's mean profiles μ and the amount by which the individual profiles are shifted depends on the molecular phenotypes of the surrounding cells and their distances, as schematically illustrated for two interacting cells in Figure 1B and for multiple cells in 1C. The functional dependence of these shifts on the neighbor's gene expression and distance is not a priori clear. Key to SpaCeNet is a parameterization which accounts for diverse distance dependencies and which can be straightforwardly interpreted in terms of conditional independence relationships, as outlined in the following.

4.3 Spatial conditional independence

Estimates of Conditional Independence (CI) relationships are key to the inference of molecular networks, as they allow the disentanglement of direct and indirect statistical relationships and as such reduce the number of false positive associations [25]. If the spatial context is ignored, CI between variables X_i and X_j given all remaining variables can be expressed as $X_i \perp X_j | \{\text{rest}\}$, where "rest" is the set of all variables in X except X_i and X_j. We extend this language to intracellular and intercellular spatial conditional independence (SCI) relationships:

- *Intracellular SCI* relations between variables X_i^a and X_j^a measured within one cell a are expressed as $X_i^a \perp X_j^a | \{\text{rest}\}$ with the term "rest" referring to all other variables of cell a and to all variables of all other cells.
- *Intercellular SCI* relations between variables X_i^a and X_j^b measured in different cells a and b, conditioned on all remaining variables, are denoted as $X_i^a \perp X_j^b | \{\text{rest}\}$ with $a \neq b$, where "rest" refers to all variables of cell a except X_i^a, all variables of cell b except X_j^b, and all variables of all remaining cells.

Thus, such as CI is designed to distinguish direct from indirect relationships between molecular variables, SCI relationships are designed to do so for the intra- and intercellular relationships between molecular variables.

4.4 Full probability density

Gaussian Graphical Models (GGMs) are PGMs which assume multivariate normally distributed data. The model parameters of GGMs are collected in the precision matrix $\mathbf{\Omega} = (\omega_{ij})$, which directly parameterizes CI relationships; $\omega_{ij} = 0$ if and only if there is CI between variables i and j given all other variables. In the following, we will develop a joint probability density which directly encodes SCI relationships. Specifically, we assume that data follow an np-dimensional multivariate normal distribution with precision matrix $\mathbf{\Lambda}$, where the entries in $\mathbf{\Lambda}$ capture both intracellular and intercellular SCI relations. The *full probability density* of SpaCeNet is

$$f_{\text{full}}(\mathbf{X}|\mathbf{R}) = \sqrt{\frac{\det \mathbf{\Lambda}}{(2\pi)^{np}}} \exp\left\{-\frac{1}{2}(\boldsymbol{\xi} - \mathbf{m})^\top \mathbf{\Lambda} (\boldsymbol{\xi} - \mathbf{m})\right\}, \tag{1}$$

where we stack the individual cells' profiles $\mathbf{x}^a \in \mathbb{R}^p, a = 1, \ldots, n$, vertically in $\boldsymbol{\xi} = \text{vec}(\mathbf{X}^\top)$ and use a global, location-agnostic mean vector $\boldsymbol{\mu}$ for all $\mathbf{x}^a$ such that $\mathbf{m} = \mathbf{1}_n \otimes \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p, \mu_1, \ldots, \mu_p, \ldots)^\top$. All pairwise cell-cell distances are collected in a matrix $\mathbf{R} = (r_{ab}) \in \mathbb{R}^{n \times n}$, where $r_{ab} = r_{ba}$ denotes the Euclidean distance between cells a and b. The space-agnostic conditional independence (CI) relations for gene expression levels of a single cell a can be recovered from Eq. (1) by marginalizing over all variables of all other cells $b \neq a$.

Next, we decompose the precision matrix $\mathbf{\Lambda} \in \mathbb{R}^{np \times np}$ into

$$\mathbf{\Lambda} = \mathbf{\Lambda}_{\text{within}} + \mathbf{\Lambda}_{\text{between}} \tag{2}$$

with a matrix for intracellular (within-cell) associations

$$\mathbf{\Lambda}_{\text{within}} = \mathbb{I}_n \otimes \mathbf{\Omega} = \begin{pmatrix} \mathbf{\Omega} & \mathbf{0}_{p\times p} & \cdots & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \mathbf{\Omega} & & \vdots \\ \vdots & & \ddots & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \cdots & \mathbf{0}_{p\times p} & \mathbf{\Omega} \end{pmatrix} \tag{3}$$

and a matrix for intercellular (between-cell) associations

$$\mathbf{\Lambda}_{\text{between}} = \begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{\Lambda}^{12}_{\text{between}} & \cdots & \mathbf{\Lambda}^{1n}_{\text{between}} \\ \mathbf{\Lambda}^{21}_{\text{between}} & \mathbf{0}_{p\times p} & & \vdots \\ \vdots & & \ddots & \mathbf{\Lambda}^{(n-1)n}_{\text{between}} \\ \mathbf{\Lambda}^{n1}_{\text{between}} & \cdots & \mathbf{\Lambda}^{n(n-1)}_{\text{between}} & \mathbf{0}_{p\times p} \end{pmatrix}. \tag{4}$$

Due to the parametrization of $\mathbf{\Lambda}_{\text{within}}$ in Eq. (3), the same precision matrix $\mathbf{\Omega} = (\omega_{ij}) \in \mathbb{R}^{p\times p}$ parametrizes the within-cell associations for all cells, where $\omega_{ij} = 0$ corresponds to intracellular SCI between X_i and X_j. The conventional (space-agnostic) GGM is recovered for $\mathbf{\Lambda}_{\text{between}} = \mathbf{0}_{np\times np}$.

We further assume that the intercellular association of gene i in any cell a with gene j in any other cell b is described by some function of the cells' Euclidean distance r_{ab}. We denote this radial cell-cell interaction potential by $\rho_{ij}(r_{ab})$ and write $\mathbf{\Lambda}^{ab}_{\text{between}} = (\rho_{ij}(r_{ab})) \in \mathbb{R}^{p\times p}$, which we require to be symmetric with respect to both $i \leftrightarrow j$ and $a \leftrightarrow b$. The $(np \times np)$-dimensional precision matrix $\mathbf{\Lambda}$ is required to be positive definite for f_{full} to be a valid probability density.

4.5 Cell-cell interaction potentials

From the definition of the probability density function, Eq. (1), we see that

$$X^a_i \perp X^b_j | \{\text{rest}\} \quad \text{with} \quad a \neq b \quad \Leftrightarrow \quad \rho_{ij}(r_{ab}) = 0, \tag{5}$$

where the set "rest" refers to all variables of cell a except X^a_i, all variables of cell b except X^b_j, and all variables of all other cells. Thus, all intercellular SCI relations are encoded in the set of $p(p+1)/2$ independent functions $\rho_{ij}(r) = \rho_{ji}(r)$. To ensure that cells that are infinitely separated do not interact, we require $\rho_{ij}(r) = 0$ for $r \to \infty$. To approximate the potential $\rho_{ij}(r)$, we use a power-series in $(1 - e^{-r/r_0})\frac{r_0}{r}$,

$$\rho^*_{ij}(r) = \sum_{l=1}^{L} \Delta\rho^{(l)}_{ij} \left(1 - e^{-r/r_0}\right)^l \left(\frac{r_0}{r}\right)^l, \tag{6}$$

where l is the order in the series expansion and $\mathbf{\Delta\rho}^{(l)} = (\Delta\rho^{(l)}_{ij}) \in \mathbb{R}^{p\times p}$ the corresponding coefficient matrix. The latter matrices are required to be symmetric, $\Delta\rho^{(l)}_{ij} = \Delta\rho^{(l)}_{ji}$. This leads to an approximation for the intercellular precision matrix $\mathbf{\Lambda}_{\text{between}}$ given by

$$\mathbf{\Lambda}^*_{\text{between}} = \sum_{l=1}^{L} \mathbf{\Theta}^{(l)} \otimes \mathbf{\Delta\rho}^{(l)}, \tag{7}$$

with

$$\Theta^{(l)}_{ab} = \Theta^{(l)}_{ba} = \begin{cases} 0 & \text{for } a = b, \\ (1 - e^{-r_{ab}/r_0})^l \left(\frac{r_0}{r_{ab}}\right)^l & \text{else.} \end{cases} \tag{8}$$

The coefficients $\Delta\rho^{(l)}_{ij}$ are model parameters which are estimated using a regularized pseudo-log-likelihood, as outlined below. The expansion (6) naturally fulfills $\lim_{r\to\infty} \rho^*_{ij}(r) = 0$ and $\lim_{r\to 0} \rho^*_{ij}(r) = c_{ij}$ with constants $c_{ij} = \sum_{l=1}^{L} \Delta\rho^{(l)}_{ij}$. Note, an expansion in $(1 - e^{-r/r_0})\frac{r_0}{r}$ has the advantage that terms do not diverge for $r \to 0$, which is in contrast to an expansion in $1/r$. Thus, the factor $(1-e^{-r/r_0})$ smooths the divergence and the amount of smoothing is determined by the additional parameter r_0.

4.6 Pseudo-log-likelihood

The precision matrix $\mathbf{\Lambda}$ is of size $np \times np$, which makes a naive maximum-likelihood-based estimate (using, for example, a gradient descent) intractable for reasonably large p and n. We address this issue by using a pseudo-log-likelihood approach [59, 56], which is a computationally efficient and consistent estimator formed by products of all the conditional distributions. Let $\mathbf{X}^{\setminus a}$ denote all gene expression levels in all cells except cell a, and $\mathbf{x}^a_{\setminus j}$ denote all gene expression levels in cell a except gene j. We consider the conditional density

$$f_{\mathrm{c}}(x^a_j|\mathbf{R}, \mathbf{X}^{\setminus a}, \mathbf{x}^a_{\setminus j}) = \sqrt{\frac{\omega_{jj}}{2\pi}} \exp\Bigg\{ -\frac{1}{2}\omega_{jj}\Bigg(x^a_j - \mu_j + \frac{1}{\omega_{jj}} \times \Bigg[\sum_{k\neq j}^{p} \omega_{jk}(x^a_k - \mu_k) + \sum_{b=1}^{n}\sum_{l=1}^{L}\sum_{k=1}^{p}(\Theta^{(l)}_{ab}\Delta\rho^{(l)}_{jk})(x^b_k - \mu_k)\Bigg]\Bigg)^2\Bigg\}, \tag{9}$$

obtained from the full density given in Eq. (1). This yields the pseudo-log-likelihood

$$\ell(\mathbf{\Omega}, \mathbf{\Delta\rho}^{(\cdot)}, \boldsymbol{\mu}) = \sum_{a=1}^{n}\sum_{j=1}^{p} \log\left(f_{\mathrm{c}}(x^a_j|\mathbf{R}, \mathbf{X}^{\setminus a}, \mathbf{x}^a_{\setminus j})\right). \tag{10}$$

4.7 Regularization:

Parameter regularization has been repeatedly shown to improve the inference of GGMs using approaches such as node-wise lasso regression [60], the graphical lasso [61], or covariance shrinkage [62]. This is particularly the case if the number of variables exceeds or is of the same order of magnitude as the number of measurements [25]. For S independent measurements (spatial transcriptomics slides), the full, regularized pseudo-log-likelihood-based optimization problem of SpaCeNet is given by

$$\underset{\mathbf{\Omega}, \mathbf{\Delta\rho}^{(\cdot)}, \boldsymbol{\mu}}{\text{minimize}} \Bigg\{ -\sum_{s=1}^{S} \ell_s(\mathbf{\Omega}, \mathbf{\Delta\rho}^{(\cdot)}, \boldsymbol{\mu}) + nS\left(\alpha \sum_{i<j}|\omega_{ij}| + \beta \sum_{i\leq j}\sqrt{\sum_{l=1}^{L}\left(\Delta\rho^{(l)}_{ij}\right)^2}\right)\Bigg\}. \tag{11}$$

Eq. (11) penalizes the off-diagonal elements of the intracellular precision matrix $\mathbf{\Omega}$ via L_1 regularization [63]. The intercellular interactions are regularized via group-lasso terms [64], where the group contains the coefficients of the spatial interactions at different orders $l = 1, \ldots, L$ of the series expansion. This regularization has the advantage that sparseness is induced simultaneously across all orders in the expansion, so we induce sparseness in the potentials ρ_{ij} and not just in the different orders of its expansion. The hyper-parameters α and β calibrate the regularization strength of the intracellular and intercellular associations, respectively. In our analysis, we standardized the expression data prior to model learning to ensure comparable penalization of the variables.

4.8 Implementation:

Eq. (11) is a convex optimization problem with lasso and group-lasso regularization terms, which can be efficiently solved via proximal gradient descent [65]. SpaCeNet uses a proximal gradient descent with Nesterov acceleration [66, 67]. The optimization terminates when the loss in Eq. (11) improves less than a user-specified threshold. The computational requirements for

training a SpaCeNet depend on the number of considered cells and genes. For example, training SpaCeNet on a Nvidia A100 GPU takes around one minute for the Drosophila data (3.039 cells and 84 genes), less than two minutes on the StarMap data (33.598 cells and 28 genes), around 30 minutes on the Mosta data with more than 30% non-zero entries (56.731 cells and 315 genes) and around 150 minutes on the Mosta data with more than 10% non-zero entries (56.731 cells and 1741 genes).

5 Supplementary Data

Supplementary Data are available online.

6 Funding

The work of HUZ and MA was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grants no. 01ZX1912A and 01ZX1912C). SS, NL, DV and MA were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [AL 2355/1-1 "Digital Tissue Deconvolution - Aus Einzelzelldaten lernen"]. KHS was supported by the US National Institutes of Health, including the National Cancer Institute (R35CA220523) and the National Heart, Lung, and Blood Institute (P01HL114501 and T32HL007427). JQ and MBG were supported by grants from the National Cancer Institute (U24CA231846 and R35CA220523), JQ was additionally supported by the National Human Genome Research Institute (R01HG011393) of the US National Institutes of Health.

References

[1] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[2] Samuel G Rodriques, Robert R Stickels, Aleksandrina Goeva, Carly A Martin, Evan Murray, Charles R Vanderburg, Joshua Welch, Linlin M Chen, Fei Chen, and Evan Z Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

[3] Samuel G. Rodriques, Robert R. Stickels, Aleksandrina Goeva, Carly A. Martin, Evan Murray, Charles R. Vanderburg, Joshua Welch, Linlin M. Chen, Fei Chen, and Evan Z. Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

[4] Robert R Stickels, Evan Murray, Pawan Kumar, Jilong Li, Jamie L Marshall, Daniela J Di Bella, Paola Arlotta, Evan Z Macosko, and Fei Chen. Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqv2. *Nature biotechnology*, 39(3):313–319, 2021.

[5] Chun-Seok Cho, Jingyue Xi, Yichen Si, Sung-Rye Park, Jer-En Hsu, Myungjin Kim, Goo Jun, Hyun Min Kang, and Jun Hee Lee. Microscopic examination of spatial transcriptome using seq-scope. *Cell*, 184(13):3559–3572.e22, 2021.

[6] Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, Xin Wang, Huifang Lu, Xi Chen, Xing Liu, Xin Huang, Zhao Li, Yan Hong, Yujia Jiang, Jian Peng, Shuai Liu, Mengzhe Shen, Chuanyu Liu, Quanshui Li, Yue Yuan, Xiaoyu Wei, Huiwen Zheng, Weimin Feng, Zhifeng Wang, Yang Liu, Zhaohui Wang, Yunzhi Yang, Haitao Xiang, Lei Han, Baoming Qin, Pengcheng Guo, Guangyao Lai, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792.e21, 2022.

[7] Kok Hao Chen, Alistair N. Boettiger, Jeffrey R. Moffitt, Siyuan Wang, and Xiaowei Zhuang. Spatially resolved, highly multiplexed rna profiling in single cells. *Science*, 348(6233):aaa6090, 2015.

[8] Sheel Shah, Eric Lubeck, Wen Zhou, and Long Cai. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron*, 92(2):342–357, 2016.

[9] Je Hyuk Lee, Evan R. Daugharthy, Jonathan Scheiman, Reza Kalhor, Joyce L. Yang, Thomas C. Ferrante, Richard Terry, Sauveur S. F. Jeanty, Chao Li, Ryoji Amamoto, Derek T. Peters, Brian M. Turczyk, Adam H. Marblestone, Samuel A. Inverso, Amy Bernard, Prashant Mali, Xavier Rios, John Aach, and George M. Church. Highly multiplexed subcellular rna sequencing in situ. *Science*, 343(6177):1360–1363, 2014.

[10] Daniel Gyllborg, Christoffer Mattsson Langseth, Xiaoyan Qian, Eunkyoung Choi, Sergio Marco Salas, Markus M Hilscher, Ed S Lein, and Mats Nilsson. Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Research*, 48(19):e112–e112, 09 2020.

[11] Xiao Wang, William E Allen, Matthew A Wright, Emily L Sylwestrak, Nikolay Samusik, Sam Vesuna, Kathryn Evans, Cindy Liu, Charu Ramakrishnan, Jia Liu, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, 361(6400):eaat5691, 2018.

[12] Valentine Svensson, Sarah A Teichmann, and Oliver Stegle. Spatialde: identification of spatially variable genes. *Nature methods*, 15(5):343–346, 2018.

[13] Damien Arnol, Denis Schapiro, Bernd Bodenmiller, Julio Saez-Rodriguez, and Oliver Stegle. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell reports*, 29(1):202–211, 2019.

[14] Zhen Fan, Runsheng Chen, and Xiaowei Chen. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Research*, 48(D1):D233–D237, 11 2019.

[15] Joseph Bergenstråhle, Ludvig Larsson, and Joakim Lundeberg. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics*, 21(1):482, July 2020.

[16] Chang Xu, Xiyun Jin, Songren Wei, Pingping Wang, Meng Luo, Zhaochun Xu, Wenyi Yang, Yideng Cai, Lixing Xiao, Xiaoyu Lin, Hongxin Liu, Rui Cheng, Fenglan Pang, Rui Chen, Xi Su, Ying Hu, Guohua Wang, and Qinghua Jiang. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Research*, 50(22):e131–e131, 10 2022.

[17] Zhiwei Fan, Yangyang Luo, Huifen Lu, Tiangang Wang, YuZhou Feng, Weiling Zhao, Pora Kim, and Xiaobo Zhou. SPASCER: spatial transcriptomics annotation at single-cell resolution. *Nucleic Acids Research*, 51(D1):D1138–D1149, 10 2022.

[18] Erick Armingol, Adam Officer, Olivier Harismendy, and Nathan E Lewis. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics*, 22(2):71–88, 2021.

[19] David S. Fischer, Anna C. Schaar, and Fabian J. Theis. Modeling intercellular communication in tissues using spatial graphs of cells. *Nature Biotechnology*, 41(3):332–336, 2023.

[20] Zixuan Cang, Yanxiang Zhao, Axel A. Almet, Adam Stabell, Raul Ramos, Maksim V. Plikus, Scott X. Atwood, and Qing Nie. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nature Methods*, 20(2):218–228, 2023.

[21] Dongshunyi Li, Jun Ding, and Ziv Bar-Joseph. Identifying signaling genes in spatial single-cell expression data. *Bioinformatics*, 37(7):968–975, 2021.

[22] Jovan Tanevski, Ricardo Omar Ramirez Flores, Attila Gabor, Denis Schapiro, and Julio Saez-Rodriguez. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome biology*, 23(1):1–31, 2022.

[23] Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics*, volume 7, pages 1–15. BioMed Central, 2006.

[24] Daniel Marbach, James C Costello, Robert Küffner, Nicole M Vega, Robert J Prill, Diogo M Camacho, Kyle R Allison, Manolis Kellis, James J Collins, and Gustavo Stolovitzky. Wisdom of crowds for robust gene network inference. *Nature methods*, 9(8):796–804, 2012.

[25] Michael Altenbuchinger, Antoine Weihs, John Quackenbush, Hans Jörgen Grabe, and Helena U Zacharias. Gaussian and mixed graphical models as (multi-) omics data analysis tools. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1863(6):194418, 2020.

[26] Katherine H Shutta, Roberta De Vito, Denise M Scholtens, and Raji Balasubramanian. Gaussian graphical models with applications to omics analyses. *Statistics in Medicine*, 41(25):5150–5187, 2022.

[27] Qiuyu Wu and Xiangyu Luo. Estimating heterogeneous gene regulatory networks from zero-inflated single-cell expression data. *The Annals of Applied Statistics*, 16(4):2183–2200, 2022.

[28] Katherine H Shutta, Deborah Weighill, Rebekka Burkholz, Marouen Ben Guebila, Dawn L DeMeo, Helena U Zacharias, John Quackenbush, and Michael Altenbuchinger. DRAGON: Determining Regulatory Associations using Graphical models on multi-Omic Networks. *Nucleic Acids Research*, 51(3):e15–e15, 12 2022.

[29] Marloes H Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A):3133–3164, 2009.

[30] JM Boggs. Myelin basic protein: a multifunctional protein. *Cellular and Molecular Life Sciences CMLS*, 63(17):1945–1961, 2006.

[31] Carol Readhead, Naoki Takasashi, H David Shine, Raul Saavedra, Richard Sidman, and Leroy Hood. Role of myelin basic protein in the formation of central nervous system myelin. *Annals of the new York Academy of Sciences*, 605(1):280–285, 1990.

[32] Bruce P Bean. The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, 8(6):451–465, 2007.

[33] Seth Love. Demyelinating diseases. *Journal of clinical pathology*, 59(11):1151–1159, 2006.

[34] Ina M Wittko-Schneider, Fabian T Schneider, and Karl H Plate. Brain homeostasis: Vegf receptor 1 and 2—two unequal brothers in mind. *Cellular and Molecular Life Sciences*, 70(10):1705–1725, 2013.

[35] Martin A Proescholdt, Steven Jacobson, Nancy Tresser, Edward H Oldfield, and Marsha J Merrill. Vascular endothelial growth factor is expressed in multiple sclerosis plaques and can induce inflammatory lesions in experimental allergic encephalomyelitis rats. *Journal of Neuropathology & Experimental Neurology*, 61(10):914–925, 2002.

[36] Kazuhide Hayakawa, Loc-Duyen D Pham, Angel T Som, Brian J Lee, Shuzhen Guo, Eng H Lo, and Ken Arai. Vascular endothelial growth factor regulates the migration of oligodendrocyte precursor cells. *Journal of Neuroscience*, 31(29):10666–10670, 2011.

[37] Daishi Hiratsuka, Erkin Kurganov, Eriko Furube, Mitsuhiro Morita, and Seiji Miyata. Vegf-and pdgf-dependent proliferation of oligodendrocyte progenitor cells in the medulla oblongata after lpc-induced focal demyelination. *Journal of Neuroimmunology*, 332:176–186, 2019.

[38] Steve E Pfeiffer, Art E Warrington, and Rashmi Bansal. The oligodendrocyte and its many cellular processes. *Trends in cell biology*, 3(6):191–197, 1993.

[39] Adrien Vaquie, Alizée Sauvain, Mert Duman, Gianluigi Nocera, Boris Egger, Felix Meyenhofer, Laurent Falquet, Luca Bartesaghi, Roman Chrast, Christophe Maurice Lamy, et al. Injured axons instruct schwann cells to build constricting actin spheres to accelerate axonal disintegration. *Cell Reports*, 27(11):3152–3166, 2019.

[40] Joon-Il Jun and Lester F Lau. Taking aim at the extracellular matrix: Ccn proteins as emerging therapeutic targets. *Nature reviews Drug discovery*, 10(12):945–963, 2011.

[41] Zuping Wu, Chenchen Zhou, Quan Yuan, Demao Zhang, Jing Xie, and Shujuan Zou. Ctgf facilitates cell-cell communication in chondrocytes via pi3k/akt signalling pathway. *Cell Proliferation*, 54(3):e13001, 2021.

[42] François Mouton-Liger, Sophie Thomas, Revital Rattenbach, Laetitia Magnol, Vanessa Larigaldie, Aurelie Ledru, Yann Herault, Catherine Verney, and Nicole Créau. Pcp4 (pep19) overexpression induces premature neuronal differentiation associated with ca2+/calmodulin-dependent kinase ii-δ activation in mouse models of down syndrome. *Journal of Comparative Neurology*, 519(14):2779–2802, 2011.

[43] Goichi Miyoshi, Jens Hjerling-Leffler, Theofanis Karayannis, Vitor H Sousa, Simon JB Butt, James Battiste, Jane E Johnson, Robert P Machold, and Gord Fishell. Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. *Journal of Neuroscience*, 30(5):1582–1594, 2010.

[44] Wolfgang Huber, Anja Von Heydebreck, Holger Sültmann, Annemarie Poustka, and Martin Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18(suppl_1):S96–S104, 2002.

[45] Robin Kosch, Katharina Limm, Annette M. Staiger, Nadine S. Kurz, Nicole Seifert, Bence Oláh, Stefan Solbrig, Marita Ziepert, Emil Chteinberg, Rainer Spang, Reiner Siebert, Helena U. Zacharias, German Ott, Peter J. Oefner, and Michael Altenbuchinger. Priomics: integration of high-throughput proteomic data with complementary omics layers using mixed graphical modeling with group priors. *bioRxiv*, 2023.

[46] Charless C Fowlkes, Cris L Luengo Hendriks, Soile VE Keränen, Gunther H Weber, Oliver Rübel, Min-Yu Huang, Sohail Chatoor, Angela H DePace, Lisa Simirenko, Clara Henriquez, et al. A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm. *Cell*, 133(2):364–374, 2008.

[47] Y Tony Ip, Ronald E Park, David Kosman, Karina Yazdanbakhsh, and Michael Levine. dorsal-twist interactions establish snail expression in the presumptive mesoderm of the drosophila embryo. *Genes & development*, 6(8):1518–1530, 1992.

[48] Jonathan D Licht, Martha J Grossel, James Figge, and Ulla M Hansen. Drosophila krüppel protein is a transcriptional represser. *Nature*, 346(6279):76–79, 1990.

[49] Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

[50] Baptiste Lacoste, Debora Angeloni, Sergio Dominguez-Lopez, Sara Calderoni, Alessandro Mauro, Franco Fraschini, Laurent Descarries, and Gabriella Gobbi. Anatomical and cellular localization of melatonin mt 1 and mt 2 receptors in the adult rat brain. *Journal of pineal research*, 58(4):397–417, 2015.

[51] Andreas Holz, Bibiana Bielekova, Roland Martin, and Michael B. A. Oldstone. Myelin-Associated Oligodendrocytic Basic Protein: Identification of an Encephalitogenic Epitope and Association with Multiple Sclerosis1. *The Journal of Immunology*, 164(2):1103–1109, 01 2000.

[52] Isabel Vieira de Assis Lima, Leandro Francisco Silva Bastos, Marcelo Limborço-Filho, Bernd L Fiebich, Antonio Carlos Pinheiro De Oliveira, et al. Role of prostaglandins in neuroinflammatory and neurodegenerative diseases. *Mediators of inflammation*, 2012, 2012.

[53] Mirna Safieh, Amos D Korczyn, and Daniel M Michaelson. Apoe4: an emerging therapeutic target for alzheimer's disease. *BMC medicine*, 17:1–17, 2019.

[54] Vivien Marx. A dream of single-cell proteomics. *Nature Methods*, 16(9):809–812, 2019.

[55] Kyle D Duncan, Jonas Fyrestam, and Ingela Lanekoff. Advances in mass spectrometry based single-cell metabolomics. *Analyst*, 144(3):782–793, 2019.

[56] Jason D Lee and Trevor J Hastie. Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics*, 24(1):230–253, 2015.

[57] Han Liu, John Lafferty, and Larry Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(80):2295–2328, 2009.

[58] Rong Zhang, Zhao Ren, and Wei Chen. Silggm: An extensive r package for efficient statistical inference in large-scale gene networks. *PLOS Computational Biology*, 14(8):1–14, 08 2018.

[59] Julian Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 24(3):179–195, 1975.

[60] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, 34(3):1436–1462, 2006.

[61] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[62] Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1), 2005.

[63] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[64] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

[65] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

[66] Yurii E Nesterov. A method for solving the convex programming problem with convergence rate o (1/kˆ 2). *Dokl. akad. nauk Sssr*, 269:543–547, 1983.

[67] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. *International conference on machine learning*, pages 1139–1147, 2013.