Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

Method

Spatial Cellular Networks from omics data with SpaCeNet

Stefan Schrod,[1,9] Niklas Lück,[1,9] 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, 37077 Göttingen, Germany;* [2]*Leibniz Institute for Immunotherapy, 93053 Regensburg, Germany;* [3]*Institute of Theoretical Physics, University of Regensburg, 93053 Regensburg, Germany;* [4]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA;* [5]*Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA;* [6]*Campus Institute Data Science (CIDAS), University of Göttingen, 37077 Göttingen, Germany;* [7]*Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover Medical School, 30625 Hannover, Germany;* [8]*Institute of Functional Genomics, University of Regensburg, 93053 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 (CI) relations between captured variables within individual cells and by disentangling these from CI relations between variables of different cells.

[Supplemental material is available for this article.]

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" (Ståhl et al. 2016) and Slide-seq (Rodriques et al. 2019) 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 (Rodriques et al. 2019) and Slide-seqV2 (Stickels et al. 2021) with a resolution of ∼10 μm, as well as, most recently, Seq-Scope (Cho et al. 2021) and Stereo-seq (Chen et al. 2022) with a resolution of ∼0.5 μm. In situ hybridization or sequencing methods can measure the expression of many genes with single-cell resolution (and even subcellular resolution) (Lee et al. 2014; Chen et al. 2015; Shah et al. 2016; Gyllborg et al. 2020), 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) (Wang et al. 2018), which is capable of measuring simultaneously the expression of ∼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 (Svensson et al. 2018; Arnol et al. 2019; Fan et al. 2019, 2022; Bergenstråhle et al. 2020; Xu et al. 2022).

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

[9]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Corresponding author:
michael.altenbuchinger@bioinf.med.uni-goettingen.de
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.279125.124.

© 2024 Schrod et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

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. Although such interactions are known to be essential for many biological processes (Armingol et al. 2021), 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, modeling approaches, and overall objectives. In seminal work, Arnol et al. (2019) proposed spatial variance component analysis (SVCA) to decompose the expression of individual genes into spatial and nonspatial contributions, namely, into cell-intrinsic effects, general environmental effects, and cell–cell interaction effects. SVCA is motivated by Gaussian processes and models the expression of genes independently of each other. Thus, it does not capture complex multivariate relationships between genes (Arnol et al. 2019), but detects individual genes which are involved in cell–cell interactions. Another approach is node-centric expression modeling (NCEM) (Fischer et al. 2023). NCEM 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, NCEM 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 (Cang et al. 2023), by Mixture of Experts for Spatial Signaling genes Identification (MESSI) considering an additional cell neighborhood to infer intracellular responses (Li et al. 2021), and by Multiview Intercellular SpaTial modeling framework (MISTy) using an aggregated direct neighborhood to resolve spatial effects (Tanevski et al. 2022). 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 correlation, and identified network edges between genes based on a correlation threshold (Margolin et al. 2006; Marbach et al. 2012). Such measures of coexpression can provide insight into active biological mechanisms, but they are vulnerable to identifying spurious associations (Altenbuchinger et al. 2020). 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). An instance of PGMs is Gaussian graphical models (GGMs), also known as partial correlations. GGMs resolve the dependency structure of molecular variables under the assumption of multivariate normality. GGMs 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 (Altenbuchinger et al. 2020). The main concept underlying GGMs is conditional independence (CI). 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 (Shutta et al. 2022). This powerful concept made GGMs one of the favored approaches to resolve molecular networks from molecular data (Marbach et al. 2012) and, importantly, they can be straightforwardly extended to account for complex single-cell and multiomics data (Wu and Luo 2022; Shutta et al. 2023). It is nevertheless noteworthy that GGMs are undirected graphical models and do not resolve causal relationships, although, in principle, lower bounds on causal effects can be derived from observational data (Maathuis et al. 2009).

Spatial Cellular Networks from omics data (SpaCeNet) is a method for analyzing patterns of correlation in spatial transcriptomics data by extending the concept of CI to spatially distributed information, facilitating the 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 the cellular state, tissue organization, 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, mouse organogenesis, and the *Drosophila* blastoderm.

Results

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 signaling molecules. Throughout the article, we will refer to any process with which cells affect each other's 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 (Wang et al. 2018; Cho et al. 2021; Chen et al. 2022).

Key to SpaCeNet is a statistically sound approach that uses PGMs to decompose the observed cellular profiles into contributions arising from ordinary cellular variability and contributions from cellular interactions (Schrod et al. 2024). Formally, SpaCeNet decomposes each cell's observed molecular profile into (1) a baseline contribution corresponding to the cell's molecular profile in isolation (Fig. 1A) and (2) a residual contribution attributed to its interaction with other cells (Fig. 1B,C). Importantly, this decomposition uses a model parametrization such that vanishing parameters

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

encode CI statements, namely, intracellular and intercellular spatial conditional independence (SCI) relationships as introduced in the following and as summarized in Supplemental Figure S1.

Undirected 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 defines a set of 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, that is, 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 (Marbach et al. 2012; Altenbuchinger et al. 2020).

Here, we develop a similar framework to model intracellular and intercellular CI relations in ST data. We estimate SCI between variable X_i^a (gene i in cell a) and variable X_j^b (gene j in another cell b) keeping all other variables in the data fixed; similar to CI, by conditioning on all variables except for X_i^a and X_j^b, we can study the direct dependence of X_i^a on X_j^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. Because we are interested in sample estimates of cellular interaction patterns, SpaCeNet infers SCI simultaneously across all pairs of cells in an ST data set.

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_i^a (gene i in cell a) and X_j^b (gene j in cell b) at distance r_{ab} as $\rho_{ij}(r_{ab})$. Because $\rho_{ij}(r_{ab})$ depends only on the distance r_{ab} between two cells, the intercellular potential between genes X_i^a and X_j^b takes the same functional form across all modeled cells a and b. Moreover, vanishing potentials $\rho_{ij}(r_{ab})$ imply that X_i^a and X_j^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. Considering the conditional distribution of X_i^a given the "rest" (Eq. 9), we see that X_i^a can be modeled by a linear regression involving X_j^b as predictor variable, where the contribution of X_j^b to the mean expression of X_i^a is given by $-\frac{1}{\omega_{ii}}\rho_{ij}(r_{ab})(X_j^b - \mu_j)$. Thus, the effect that the molecular variable X_j^b in cell b exerts on X_i^a in cell a via cell–cell interaction is 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 $\rho_{ij}(r)$ 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 $\rho_{ij}(r_{ab})$ 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–D, where we simulated radial dependencies corresponding to long-distance interactions (via cell signaling) (Fig. 2A,B), short-distance interactions (such as via cell–cell contact) using an exponentially decreasing potential with short range (Fig. 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} (Fig. 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

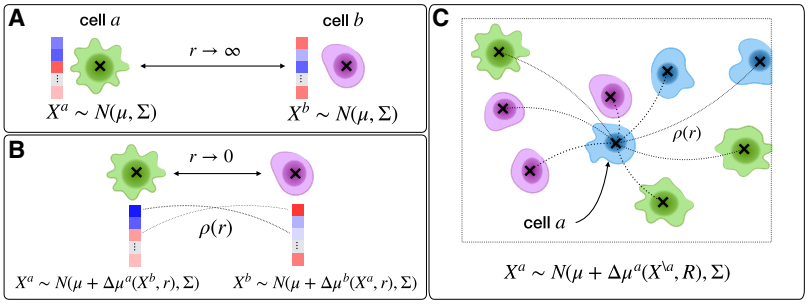


Figure 1. SpaCeNet concept and how it applies to different cellular contexts. (*A*) A schematic picture of two isolated cells a and b which are infinitely separated (distance $r \to \infty$). In this case, respective cellular profiles with expression levels illustrated in blue and red are modeled by a multivariate normal distribution with mean expression μ and covariance Σ. The latter encodes the complex coexpression pattern of different RNAs/proteins and facilitates a description of molecularly diverse cells. (*B*) A scenario of two neighboring cells, where the cells affect each other; the molecular phenotype changes compared to *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 parameterized by interaction potentials which directly provide estimates of spatial gene–gene dependencies between gene i and j. (*C*) A sketch of the more complex scenario of a set of interacting cells. Here, the expression of the 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 the 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.

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

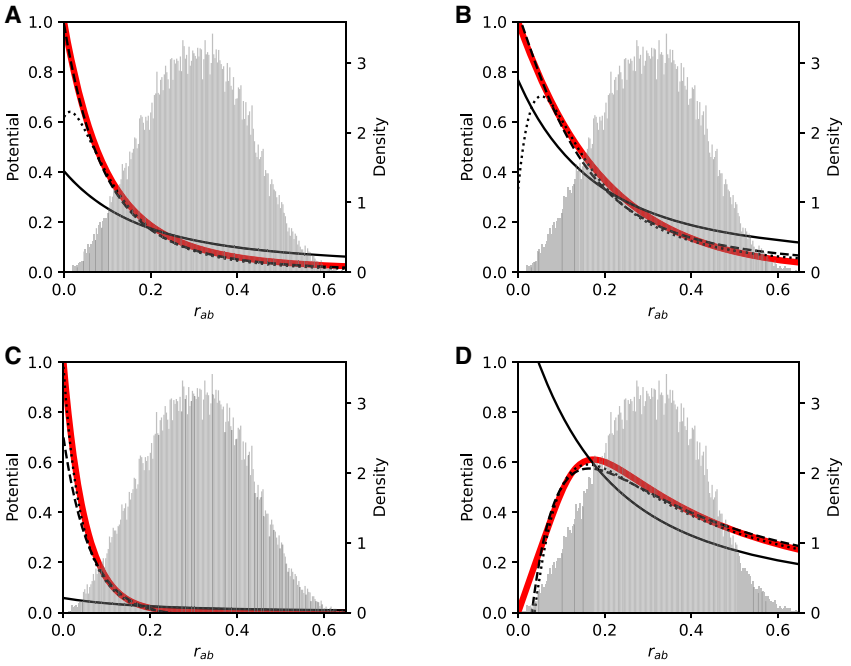


Figure 2. SpaCeNet can reconstruct diverse cellular interaction patterns. (*A–D*) Different potentials are used to simulate data in red in dependence on the distance between cells r_{ab}. Potentials inferred by SpaCeNet are shown for three different expansion orders: $L=1$ (black solid lines), $L=3$ (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 $\rho(r_{ab})=\exp(-5r_{ab})$, (*C*) a short-range exponential potential $\rho(r_{ab})=\exp(-20r_{ab})$, 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 Supplemental Section 1.

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

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 Equation 1 and tested if the ground truth intra- and intercellular network edges are correctly recovered (see also Supplemental Section 2). Because cellular interactions occur at various length scales, we performed four 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 $S\in\{1, 10, 100\}$ is 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 Supplemental 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 $1/\phi_{ij}=1/5$. Supplemental Figure S4A gives AUPRC versus $n\cdot S$ to recover the correct intracellular networks (the intracellular precision matrix $\boldsymbol{\Omega}$, left figure) and the intercellular interaction networks (the cell–cell interaction parameters $\Delta\boldsymbol{\rho}$, 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 AUPRC 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 $1/\phi_{ij}=1/10$ (Supplemental Fig. S4B) and short range $1/\phi_{ij}=1/20$ (Supplemental Fig. 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 (Supplemental Fig. 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 Supplemental 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 Supplemental Figure S3 for two different model regularization parameters β. Results considering AUROC for all four simulation studies are shown in Supplemental Table S1, 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 l for all j and l, yielding a matrix of correlation coefficients. The performance evaluation shown in Supplemental Figure S5 and Supplemental Table S2 contrasts the performance of this naive correlation-based approach for $K=1$, 5, 10, 20, 50 nearest neighbors with SpaCeNet. We observe that, first, overall

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

this approach cannot compete with SpaCeNet and, second, that its performance depends on the number of considered nearest neighbors 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 l 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 l' conditioning on j' and l. 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 (Supplemental Fig. S6; Supplemental 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. (2022) 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 prespecified cell–cell interactions between four cell types, capturing interactions of 29 molecular species including 5 ligands, 5 receptors, and 19 intracellular signaling proteins, capturing $n = 3953$ cells each. The authors suggested 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 one is activated by it (Tanevski et al. 2022). 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 (Supplemental 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 (Tanevski et al. 2022).

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

	MISTy	SpaCeNet
AUPRC (intracellular)		
Tissue 1	0.21	0.58
Tissue 2	0.19	0.54
Tissue 1 + 2	0.21	0.50
AUPRC (intercellular)		
Tissue 1	0.07	0.21
Tissue 2	0.06	0.22
Tissue 1 + 2	0.07	0.22

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 (Wang et al. 2018). 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. 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 $\Delta\boldsymbol{\rho}^{(\cdot)}$), 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 (Supplemental Fig. S7). We tuned $\alpha, \beta \in [10^{-5}, 10]$ on a 4×4 grid that was refined 6 times (Supplemental Fig. 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 Supplemental File 1. For the spatial interactions $\Delta\boldsymbol{\rho}^{(\cdot)}$, SpaCeNet selected a set of 134 out of 406 possible spatial associations (Supplemental Fig. S9; Supplemental File 1). We ranked edges according to $\Delta\rho_{ij} = ||\Delta\boldsymbol{\rho}_{ij}^{(\cdot)}||_2 = \sqrt{\sum_l \Delta\rho_{ij}^{(l)2}}$, 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 $\Delta\boldsymbol{\rho}^{(1)}_{Mbp,\,Flt1} = 0.327$. 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 (1) the intracellular edges only ($\Delta\boldsymbol{\rho}^{(\cdot)} = \mathbf{0}$), or (2) the full model (see also Supplemental Section 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* (Fig. 3A,D), that were reduced substantially via the spatial interactions (Fig. 3B,E). Figure 3C and F shows the corresponding contributions from the spatial interactions $\Delta\boldsymbol{\rho}^{(\cdot)}$.

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 ~30% of the total protein of myelin (Boggs 2006), and *Mbp* mutants lack compact myelin in the CNS (Readhead et al. 1990). Myelin surrounds nerve cell axons to insulate them and to increase the conduction of electric impulses (Bean 2007). Demyelinating diseases of the CNS, of which multiple sclerosis (MS) is the most common, are characterized by damaged myelin sheaths (Love 2006). *Flt1* encodes a member of the vascular endothelial growth factor receptor (VEGFR) family. VEGFRs mediate diverse cellular communication signals

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

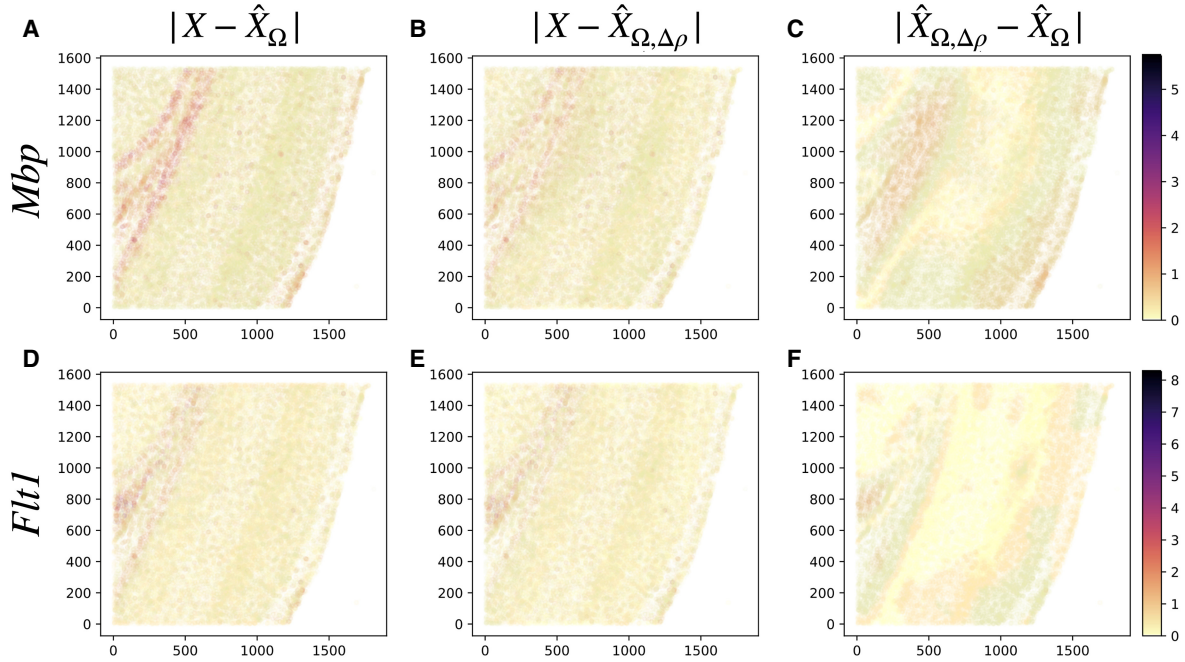


Figure 3. Spatial interactions improve on goodness of fit. The figures show absolute prediction residuals for spatially associated genes *Mbp* (*top* row) and *Flt1* (*bottom* row) in the mouse visual cortex data from Wang et al. (2018) with respect to their position in space. The *left* column (*A*,*D*) shows the absolute residuals between the ground truth data and the predictions based on the intracellular model parameters only, $|\mathbf{X} - \hat{\mathbf{X}}_{\Omega}|$. The *middle* column (*B*,*E*) displays the residuals if both intra- and intercellular interactions are considered, $|\mathbf{X} - \hat{\mathbf{X}}_{\Omega,\Delta\boldsymbol{\rho}}|$. The *right* column (*C*,*F*) shows the contributions from spatial interactions only, $|\hat{\mathbf{X}}_{\Omega,\Delta\boldsymbol{\rho}} - \hat{\mathbf{X}}_{\Omega}|$.

controlling developmental processes, such as neurogenesis or gliogenesis (Wittko-Schneider et al. 2013). VEGFRs recognize vascular endothelial growth factors (VEGFs), whose expression has been shown to be upregulated in both acute and chronic MS plaques (Proescholdt et al. 2002). 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 VEGFA (which is recognized by FLT1) promotes the migration of oligodendrocyte precursor cells (OPCs) in a concentration-dependent manner (Hayakawa et al. 2011), as shown by anti-KDR (also known as FLK1) (not by anti-FLT1) 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 (Hiratsuka et al. 2019). It has been long known that oligodendrocytes are the myelinating cells of the CNS derived from OPCs (Pfeiffer et al. 1993); 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. In a study by Vaquié et al. (2019), 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 FLT1 (also known as VEGFR1) activity, and can be acquired by oligodendrocytes through enforced expression of VEGFR1 (Vaquié et al. 2019).

It should be noted that SpaCeNet models are associative, not causal. Thus, we cannot elucidate whether decreased *Mbp* levels imply increased *Flt1* in neighboring cells or if increased *Flt1* levels imply decreased *Mbp* in neighboring cells. Moreover, because only a selected set of 28 genes was spatially assessed by STARmap in a thick tissue section, and because 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 (Jun and Lau 2011). We obtained $\Delta\rho^{(1)}_{Ctgf,Gja1} = 0.103$, 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 upregulation of gap junction protein, alpha 1 (*Gja1*, also known as connexin 43) expression (Wu et al. 2021). Purkinje cell protein 4 (PCP4) regulates calmodulin activity and might contribute to neuronal differentiation through the activation of calmodulin-dependent kinase signaling pathways (Mouton-Liger et al. 2011). The *Ctgf–Pcp4* potential was estimated with $\Delta\rho_{Ctgf,Pcp4} = 0.083$ and $\boldsymbol{\Delta\rho}_{Ctgf,Pcp4} = 0.0831$; 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 $\boldsymbol{\Delta\rho}_{Reln,Sst} = 0.0780$. The gene *Reln* encodes a secreted extracellular matrix protein that might play a role in cell–cell interactions critical for cell positioning. Somatostatin affects the transmission rates of neurons in the CNS and cell proliferation. GABAeric cortical interneurons can be delineated to 95% by the markers PV, SST (coexpressed with reelin), reelin (without SST), and VIP (Miyoshi et al. 2010). The medial ganglionic eminence gives rise to the population of interneurons that coexpress reelin and somatostatin, whereas the caudal ganglionic eminence gives rise to interneurons that express somatostatin but lack reelin (Miyoshi et al. 2010). The observed spatial *Reln–Sst* association could be a hint that the

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

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 (Huber et al. 2002). 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 Supplemental 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 the 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 Supplemental 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 (Kosch et al. 2023).

Comparative analysis of spatial associations using pairwise and partial spatial correlations

For comparison and illustration purposes, we performed the ordinary pairwise spatial correlation analysis outlined above on the mouse visual cortex data, using the $K=5$ nearest neighbors in line with Arnol et al. (2019). 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, whereas this analysis suggests a positive association. To further resolve this contradictory result, we compared the estimated pairwise spatial correlation matrix with an ordinary correlation matrix derived across all single-cell profiles (see Supplemental Figs. S10A, S11A), respectively. Comparison of both matrices reveals a strong coincidence with differences distributed as shown in Supplemental Figures S10 and 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 $K=5$ nearest neighbors. 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 the previous analysis was based on the nearest neighbor estimate of gene expression, ignoring potential complex radial dependencies, and did not explore the full capacity of PGMs to disentangle direct from indirect relationships by considering all molecular variables simultaneously, which might explain the observed differences. Ranking from highest to lowest absolute spatial partial correlations also yielded the pair *Mbp–Flt1* as the top hit, suggesting that disentangling the intra- from intercellular 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

	jk'	$j'k$	$\frac{1}{2}(j'k + jk')$	$\Delta\boldsymbol{\rho}^{(1)}$
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

Here, jk' corresponds to spatial partial correlations between variable j and k' conditioning on j' and k. An analogous 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 $\Delta\boldsymbol{\rho}^{(1)}$ with positive signs corresponding to negative associations in line with the definition of the precision matrix of ordinary GGMs.

SpaCeNet infers spatial gene–gene-association patterns in the Drosophila blastoderm from low-resolution spatial omics data

The previous analysis used data generated by STARmap (Wang et al. 2018), 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 costained for a reference gene and one gene out of a preselected gene set, to generate a virtual *Drosophila* embryo (Fowlkes et al. 2008). We retrieved these virtual embryo data consisting of 84 genes whose expression levels were measured at 3039 embryonic locations. We then used SpaCeNet to estimate the intracellular network (the precision matrix $\boldsymbol{\Omega}$) and the network of spatial molecular interactions ($\Delta\boldsymbol{\rho}^{(\cdot)}$) 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% (Supplemental Figs. S12, 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 (Supplemental File 2). The intracellular network $\boldsymbol{\Omega}$ is a full matrix and so there is a rich dependency structure among variables not related to their spatial context. The spatial-interaction network, $\Delta\boldsymbol{\rho}^{(\cdot)}$, in contrast, is sparse with 238 out of 3570 possible interactions (Supplemental Fig. 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.

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

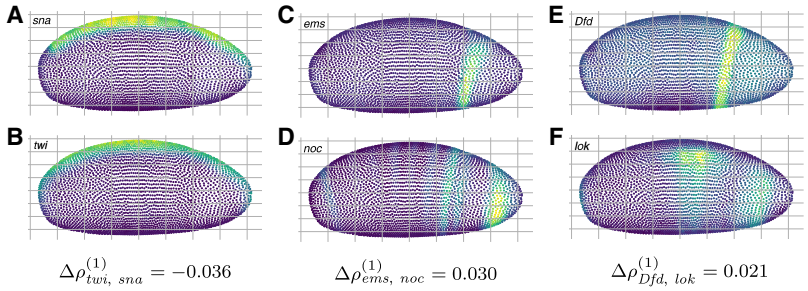


Figure 4. Visualization of genes with strong spatial associations in the *Drosophila* blastoderm as identified by SpaCeNet. Figures show gene expression levels of the top ranked gene pairs as identified by SpaCeNet with *sna* and *twi* in *A* and *B*, *ems* and *noc* in C and *D*, and *Dfd* and *lok* in *E* and *F*, respectively. The genes *sna* and *twi* (*A*,*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* (*C*,*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* (*E*,*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.

Figure 4A–F contrasts the expression of the gene pairs *sna–twi*, *ems–noc*, and *Dfd–lok*, respectively. The analysis 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 (Ip et al. 1992). 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 preprocessing biases, we reperformed our analysis but additionally applied a log-transformation (Supplemental 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 used as an inferential tool, predicting the 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 (Licht et al. 1990). 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 Supplemental Figure S15A and the corresponding predictions in Supplemental 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] = 4.36×10^{-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 (Supplemental Fig. S15C). This reduced the MSE to 2.88×10^{-3}, and the highlighted region better agrees with the ground truth, as can be seen by comparing Supplemental Figure S15A–C.

SpaCeNet analysis of high-throughput spatial transcriptomics data

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 Chen et al. (2022), the mouse organogenesis spatiotemporal transcriptomic atlas (MOSTA) was generated, which provides gene expression maps with single-cell resolution of mouse organogenesis. We downloaded preprocessed data of the coronal hemibrain section, capturing, in total, measurements of 56,731 individual cells together with their spatial location. 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 data sets (using $L = 1$ for computational efficiency), and performed a grid search for model selection (Supplemental Fig. S16). The best model was selected based on the lowest Akaike information criterion (AIC). The total computation time for the final model was ~30 and 150 minutes, respectively, computed on a single A100 GPU, highlighting that SpaCeNet is computationally highly efficient. The top spatial associations are summarized in Supplemental Table S7, illustrating a high coincidence between the top associated genes of settings A and B. Further, the activated gene regions of the top associated genes for *Mobp–Mbp* are shown in Supplemental Figure S17 and for *Mt2–Mt1* in Supplemental Figure S18.

We want to highlight that model findings might be affected by low sequencing depths, challenging SpaCeNet's model assumptions. However, despite this limitation, the findings are worth discussing, as they do not only demonstrate the power of SpaCeNet in finding valid CI relations between genes expressed in different cells, but they also highlight general limitations of (spatial) transcriptomics in reconstructing biology. A comprehensive discussion can be found in Supplemental Section 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 CI by spatial information through estimating 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 three example data sets: an expression map of the mouse visual cortex, a virtual RNA map of the *Drosophila* blastoderm, and an analysis of high-throughput ST data of mouse organogenesis. The analysis of the mouse visual cortex allowed us to generate hypotheses about the spatial organization of cell populations, such as *Flt1*-mediated

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

signals which could be involved in the recruitment of myelinating cells toward 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 a 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*. Our analysis of the mouse organogenesis data finally showed SpaCeNet's capability to investigate large-scale problems involving 1000s of genes in 10,000s of cells.

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 posttranslational modifications, and subsequent signaling cascades comprising the 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 (Duncan et al. 2019; Marx 2019) 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, multiomic, intracellular, and intercellular association networks that take 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 GGM framework, which can scale to the increasingly large and complex spatial multiomics 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. Although 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 Lee and Hastie (2015). 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 (Maathuis et al. 2009), 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 multiomics 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.

Methods

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 this section, we identify the $\mathbf{x}^a$ as molecular profiles of individual cells and r_a with the respective point-position in space.

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 Fig. 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 Fig. 1A). Note that this is a reasonable assumption if data are approximately normal or if they are transformed accordingly (e.g., via nonparanormal transformations) (Liu et al. 2009; Zhang et al. 2018). 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 Fig. 1B,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 Figure 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 CI relationships, as outlined in the following.

Spatial conditional independence

Estimates of 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 (Altenbuchinger et al. 2020). 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 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, 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.

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

Full probability density

GGMs are PGMs that 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 CI relations for gene expression levels of a single cell a can be recovered from Equation 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 Equation 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.

Cell–cell interaction potentials

From the definition of the probability density function, Equation 1, we see that

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

where the set "rest" refers to all variables of cell a except X_i^a, all variables of cell b except X_j^b, 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 $(1 - e^{-r/r_0})\frac{r_0}{r}$,

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

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.

Pseudo-log-likelihood

The precision matrix $\mathbf{\Lambda}$ is of size $np \times np$, which makes a naive maximum-likelihood-based estimate (using, e.g, a gradient descent) intractable for reasonably large p and n. We address this issue by using a pseudo-log-likelihood approach (Besag 1975; Lee and Hastie 2015), 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

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

obtained from the full density given in Equation 1. This yields the pseudo-log-likelihood

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

Regularization

Parameter regularization has been repeatedly shown to improve the inference of GGMs using approaches such as node-wise LASSO regression (Meinshausen and Bühlmann 2006), the graphical LASSO (Friedman et al. 2008), or covariance shrinkage (Schäfer and Strimmer 2005). This holds particularly true when the number of variables exceeds or is of the same order of magnitude as the number of measurements (Altenbuchinger et al. 2020). For S independent measurements (spatial transcriptomics slides), the full,

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

regularized pseudo-log-likelihood-based optimization problem of SpaCeNet is given by

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

Equation 11 penalizes the off-diagonal elements of the intracellular precision matrix Ω via L_1 regularization (Tibshirani 1996). The intercellular interactions are regularized via group-LASSO terms (Yuan and Lin 2006), 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 before model learning to ensure comparable penalization of the variables.

Implementation

Equation 11 is a convex optimization problem with LASSO and group-LASSO regularization terms, which can be efficiently solved via proximal gradient descent (Parikh and Boyd 2014). SpaCeNet uses a proximal gradient descent with Nesterov acceleration (Nesterov 1983; Sutskever et al. 2013). The optimization terminates when the loss in Equation 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 an Nvidia A100 GPU takes ~1 min for the *Drosophila* data (3039 cells and 84 genes), <2 minutes on the STARmap data (33,598 cells and 28 genes), ~30 minutes on the Mosta data with more than 30% nonzero entries (56,731 cells and 315 genes) and ~150 minutes on the Mosta data with more than 10% nonzero entries (56,731 cells and 1741 genes).

Hyper-parameter calibration

SpaCeNet is a statistical model derived by a regularized maximum-likelihood maximization. Consequently, generalization performance can be assessed by evaluating the model log-likelihood on unseen data. For the presented analyses, our strategy was to either use separate batches of ST slides for model evaluation (see Supplemental Section 1, "Simulation studies") or to cut ST slides into training and validation compartments. Moreover, we implemented the AIC, which serves as an alternative for model selection. The chosen analysis strategies are provided in the main text. Respective hyper-parameter surfaces are given in Supplemental Figures S8, S13, and S16. Further details on hyper-parameter calibration in the context of the Simulation studies can be found in Supplemental Section 1.

Data sets

All ST data used in this article are publicly available. The STARmap measurements of the mouse visual cortex were generated by Wang et al. (2018) and downloaded from GitHub (https://github.com/drieslab/spatial-datasets/tree/master/data/2018_starmap_3D_cortex). The *Drosophila* virtual embryo data (Fowlkes et al. 2008) were obtained from GitHub (https://github.com/rajewsky-lab/novosparc/tree/master/novosparc/datasets/bdtnp). The mouse organogenesis spatiotemporal transcriptomic atlas (MOSTA) data (Chen et al. 2022) were obtained from https://ftp.cngb.org/pub/SciRAID/stomics/STDS0000058/Cell_bin_matrix/Mouse_brain_Adult_GEM_CellBin.tsv.gz.

Software availability

Our implementation of SpaCeNet is publicly available at GitHub (https://github.com/sschrod/SpaCeNet) and as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

The work of H.U.Z. and M.A. was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grants nos. 01ZX1912A and 01ZX1912C). S.S., N.L., D.V., and M.A. were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (AL 2355/1-1 "Digital Tissue Deconvolution—Aus Einzelzelldaten lernen"). K.H.S. was supported by the US National Institutes of Health, including the National Cancer Institute (R35CA220523) and the National Heart, Lung, and Blood Institute (T32HL007427). J.Q. and M.B.G. were supported by grants from the National Cancer Institute (U24CA231846 and R35CA220523), J.Q. was additionally supported by the National Human Genome Research Institute (R01HG011393) of the US National Institutes of Health. Finally, P.J.O. was supported by the Wilhelm Sander-Stiftung (grant no. 2018.037.1).

Author contributions: M.A. and R.L. conceptualized the study and together with S.Sc., N.L., T.W., S.So., and A.S. developed the SpaCeNet model; S.Sc. and N.L. implemented, verified, and applied the SpaCeNet model; D.V. improved the implementation and applied SpaCeNet to real-world data; S.Sc., N.L., and M.A. wrote the main manuscript; K.H.S., M.B.G., J.Q., H.U.Z., T.B., and P.J.O. suggested improvements and contributed to writing the manuscript.

References

Altenbuchinger M, Weihs A, Quackenbush J, Grabe HJ, Zacharias HU. 2020. Gaussian and mixed graphical models as (multi-) omics data analysis tools. *Biochim Biophys Acta Gene Regul Mech* **1863:** 194418. doi:10.1016/j.bbagrm.2019.194418

Armingol E, Officer A, Harismendy O, Lewis NE. 2021. Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet* **22:** 71–88. doi:10.1038/s41576-020-00292-x

Arnol D, Schapiro D, Bodenmiller B, Saez-Rodriguez J, Stegle O. 2019. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell Rep* **29:** 202–211.e6. doi:10.1016/j.celrep.2019.08.077

Bean BP. 2007. The action potential in mammalian central neurons. *Nat Rev Neurosci* **8:** 451–465. doi:10.1038/nrn2148

Bergenstråhle J, Larsson L, Lundeberg J. 2020. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21:** 482. doi:10.1186/s12864-020-06832-3

Besag J. 1975. Statistical analysis of non-lattice data. *J R Stat Soc Ser D (Stat)* **24:** 179–195. doi:10.2307/2987782

Boggs J. 2006. Myelin basic protein: a multifunctional protein. *Cell Mol Life Sci* **63:** 1945–1961. doi:10.1007/s00018-006-6094-7

Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, Atwood SX, Nie Q. 2023. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* **20:** 218–228. doi:10.1038/s41592-022-01728-4

Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. 2015. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348:** aaa6090. doi:10.1126/science.aaa6090

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, Qiu X, Yang J, Xu J, Hao S, et al. 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185:** 1777–1792.e21. doi:10.1016/j.cell.2022.04.003

Cho CS, Xi J, Si Y, Park SR, Hsu JE, Kim M, Jun G, Kang HM, Lee JH. 2021. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184:** 3559–3572.e22. doi:10.1016/j.cell.2021.05.010

Duncan KD, Fyrestam J, Lanekoff I. 2019. Advances in mass spectrometry based single-cell metabolomics. *Analyst* **144:** 782–793. doi:10.1039/C8AN01581C

Fan Z, Chen R, Chen X. 2019. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res* **48:** D233–D237. doi:10.1093/nar/gkz934

Fan Z, Luo Y, Lu H, Wang T, Feng Y, Zhao W, Kim P, Zhou X. 2022. SPASCER: spatial transcriptomics annotation at single-cell resolution. *Nucleic Acids Res* **51:** D1138–D1149. doi:10.1093/nar/gkac889

Fischer DS, Schaar AC, Theis FJ. 2023. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat Biotechnol* **41:** 332–336. doi:10.1038/s41587-022-01467-z

Fowlkes CC, Hendriks CLL, Keränen SV, Weber GH, Rübel O, Huang MY, Chatoor S, DePace AH, Simirenko L, Henriquez C, et al. 2008. A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. *Cell* **133:** 364–374. doi:10.1016/j.cell.2008.01.053

Friedman J, Hastie T, Tibshirani R. 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9:** 432–441. doi:10.1093/biostatistics/kxm045

Gyllborg D, Langseth CM, Qian X, Choi E, Salas SM, Hilscher MM, Lein ES, Nilsson M. 2020. Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res* **48:** e112. doi:10.1093/nar/gkaa792

Hayakawa K, Pham LDD, Som AT, Lee BJ, Guo S, Lo EH, Arai K. 2011. Vascular endothelial growth factor regulates the migration of oligodendrocyte precursor cells. *J Neurosci* **31:** 10666–10670. doi:10.1523/JNEUROSCI.1944-11.2011

Hiratsuka D, Kurganov E, Furube E, Morita M, Miyata S. 2019. VEGF-and PDGF-dependent proliferation of oligodendrocyte progenitor cells in the medulla oblongata after LPC-induced focal demyelination. *J Neuroimmunol* **332:** 176–186. doi:10.1016/j.jneuroim.2019.04.016

Huber W, Von Heydebreck A, Sültmann H, Poustka A, Vingron M. 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18:** S96–S104. doi:10.1093/bioinformatics/18.suppl_1.S96

Ip YT, Park RE, Kosman D, Yazdanbakhsh K, Levine M. 1992. *dorsal–twist* interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev* **6:** 1518–1530. doi:10.1101/gad.6.8.1518

Jun JI, Lau LF. 2011. Taking aim at the extracellular matrix: CCN proteins as emerging therapeutic targets. *Nat Rev Drug Discov* **10:** 945–963. doi:10.1038/nrd3599

Kosch R, Limm K, Staiger AM, Kurz NS, Seifert N, Oláh B, Solbrig S, Ziepert M, Chteinberg E, Spang R, et al. 2023. Priomics: integration of high-throughput proteomic data with complementary omics layers using mixed graphical modeling with group priors. bioRxiv doi:10.1101/2023.11.10.566517

Lee JD, Hastie TJ. 2015. Learning the structure of mixed graphical models. *J Comput Graph Stat* **24:** 230–253. doi:10.1080/10618600.2014.900500

Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SSF, Li C, Amamoto R, et al. 2014. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343:** 1360–1363. doi:10.1126/science.1250212

Li D, Ding J, Bar-Joseph Z. 2021. Identifying signaling genes in spatial single-cell expression data. *Bioinformatics* **37:** 968–975. doi:10.1093/bioinformatics/btaa769

Licht JD, Grossel MJ, Figge J, Hansen UM. 1990. *Drosophila* Krüppel protein is a transcriptional represser. *Nature* **346:** 76–79. doi:10.1038/346076a0

Liu H, Lafferty J, Wasserman L. 2009. The nonparanormal: semiparametric estimation of high dimensional undirected graphs. *J Mach Learn Res* **10:** 2295–2328.

Love S. 2006. Demyelinating diseases. *J Clin Pathol* **59:** 1151–1159. doi:10.1136/jcp.2005.031195

Maathuis MH, Kalisch M, Bühlmann P. 2009. Estimating high-dimensional intervention effects from observational data. *Ann Stat* **37:** 3133–3164. doi:10.1214/09-AOS685

Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G. 2012. Wisdom of crowds for robust gene network inference. *Nat Methods* **9:** 796–804. doi:10.1038/nmeth.2016

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. 2006. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7:** S7. doi:10.1186/1471-2105-7-S1-S7

Marx V. 2019. A dream of single-cell proteomics. *Nat Methods* **16:** 809–812. doi:10.1038/s41592-019-0540-6

Meinshausen N, Bühlmann P. 2006. High-dimensional graphs and variable selection with the lasso. *Ann Stat* **34:** 1436–1462. doi:10.1214/009053606000000281

Miyoshi G, Hjerling-Leffler J, Karayannis T, Sousa VH, Butt SJ, Battiste J, Johnson JE, Machold RP, Fishell G. 2010. Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. *J Neurosci* **30:** 1582–1594. doi:10.1523/JNEUROSCI.4515-09.2010

Mouton-Liger F, Thomas S, Rattenbach R, Magnol L, Larigaldie V, Ledru A, Herault Y, Verney C, Créau N. 2011. PCP4 (PEP19) overexpression induces premature neuronal differentiation associated with Ca^{2+}/calmodulin-dependent kinase II-δ activation in mouse models of Down syndrome. *J Comp Neurol* **519:** 2779–2802. doi:10.1002/cne.22651

Nesterov YE. 1983. A method for solving the convex programming problem with convergence rate O(1/k2). *Dokl Akad Nauk SSSR* **269:** 543–547.

Parikh N, Boyd S. 2014. Proximal algorithms. *Found Trends Optim* **1:** 127–239. doi:10.1561/2400000003

Pfeiffer SE, Warrington AE, Bansal R. 1993. The oligodendrocyte and its many cellular processes. *Trends Cell Biol* **3:** 191–197. doi:10.1016/0962-8924(93)90213-K

Proescholdt MA, Jacobson S, Tresser N, Oldfield EH, Merrill MJ. 2002. Vascular endothelial growth factor is expressed in multiple sclerosis plaques and can induce inflammatory lesions in experimental allergic encephalomyelitis rats. *J Neuropathol Exp Neurol* **61:** 914–925. doi:10.1093/jnen/61.10.914

Readhead C, Takasashi N, Shine HD, Saavedra R, Sidman R, Hood L. 1990. Role of myelin basic protein in the formation of central nervous system myelin. *Ann N Y Acad Sci* **605:** 280–285. doi:10.1111/j.1749-6632.1990.tb42401.x

Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. 2019. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363:** 1463–1467. doi:10.1126/science.aaw1219

Schäfer J, Strimmer K. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* **4:** Article 32. doi:10.2202/1544-6115.1175

Schrod S, Lück N, Lohmayer R, Solbrig S, Wipfler T, Shutta KH, Guebila MB, Schäfer A, Beißbarth T, Zacharias HU, et al. 2024. SpaCeNet: Spatial Cellular Networks from omics data. In *International Conference on Research in Computational Molecular Biology*, Boston, pp. 344–347. Springer.

Shah S, Lubeck E, Zhou W, Cai L. 2016. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92:** 342–357. doi:10.1016/j.neuron.2016.10.001

Shutta KH, De Vito R, Scholtens DM, Balasubramanian R. 2022. Gaussian graphical models with applications to omics analyses. *Stat Med* **41:** 5150–5187. doi:10.1002/sim.9546

Shutta KH, Weighill D, Burkholz R, Guebila M, DeMeo DL, Zacharias HU, Quackenbush J, Altenbuchinger M. 2023. DRAGON: determining regulatory associations using graphical models on multi-omic networks. *Nucleic Acids Res* **51:** e15. doi:10.1093/nar/gkac1157

Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353:** 78–82. doi:10.1126/science.aaf2403

Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, Arlotta P, Macosko EZ, Chen F. 2021. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqv2. *Nat Biotechnol* **39:** 313–319. doi:10.1038/s41587-020-0739-1

Sutskever I, Martens J, Dahl G, Hinton G. 2013. On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning*, Atlanta, pp. 1139–1147.

Svensson V, Teichmann SA, Stegle O. 2018. SpatialDE: identification of spatially variable genes. *Nat Methods* **15:** 343–346. doi:10.1038/nmeth.4636

Tanevski J, Flores ROR, Gabor A, Schapiro D, Saez-Rodriguez J. 2022. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biol* **23:** 97. doi:10.1186/s13059-022-02663-5

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *JR Stat Soc: Ser B (Methodol)* **58:** 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Vaquié A, Sauvain A, Duman M, Nocera G, Egger B, Meyenhofer F, Falquet L, Bartesaghi L, Chrast R, Lamy CM, et al. 2019. Injured axons instruct Schwann cells to build constricting actin spheres to accelerate axonal disintegration. *Cell Rep* **27:** 3152–3166. doi:10.1016/j.celrep.2019.05.060

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press

Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, et al. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361:** eaat5691. doi:10.1126/science.aat5691

Wittko-Schneider IM, Schneider FT, Plate KH. 2013. Brain homeostasis: VEGF receptor 1 and 2—two unequal brothers in mind. *Cell Mol Life Sci* **70:** 1705–1725. doi:10.1007/s00018-013-1279-3

Wu Q, Luo X. 2022. Estimating heterogeneous gene regulatory networks from zero-inflated single-cell expression data. *Ann Appl Stat* **16:** 2183–2200. doi:10.1214/21-AOAS1582

Wu Z, Zhou C, Yuan Q, Zhang D, Xie J, Zou S. 2021. CTGF facilitates cell-cell communication in chondrocytes via PI3K/Akt signalling pathway. *Cell Prolif* **54:** e13001. doi:10.1111/cpr.13001

Xu C, Jin X, Wei S, Wang P, Luo M, Xu Z, Yang W, Cai Y, Xiao L, Lin X, et al. 2022. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res* **50:** e131. doi:10.1093/nar/gkac901

Yuan M, Lin Y. 2006. Model selection and estimation in regression with grouped variables. *J R Stat Soc: Ser B (Stat Methodol)* **68:** 49–67. doi:10.1111/j.1467-9868.2005.00532.x

Zhang R, Ren Z, Chen W. 2018. SILGGM: an extensive R package for efficient statistical inference in large-scale gene networks. *PLoS Comput Biol* **14:** e1006369. doi:10.1371/journal.pcbi.1006369

Received February 15, 2024; accepted in revised form August 27, 2024.

Downloaded from genome.cshlp.org on February 26, 2026 - Published by Cold Spring Harbor Laboratory Press



Spatial Cellular Networks from omics data with SpaCeNet

Stefan Schrod, Niklas Lück, Robert Lohmayer, et al.

Genome Res. 2024 34: 1371-1383 originally published online September 4, 2024
Access the most recent version at doi:10.1101/gr.279125.124

Supplemental Material http://genome.cshlp.org/content/suppl/2024/10/03/gr.279125.124.DC1

References This article cites 60 articles, 10 of which can be accessed free at:
http://genome.cshlp.org/content/34/9/1371.full.html#ref-list-1

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

© 2024 Schrod et al.; Published by Cold Spring Harbor Laboratory Press