Check for updates

OPEN

Modeling intercellular communication in tissues using spatial graphs of cells

David S. Fischer [1,2,4], Anna C. Schaar[1,3,4] and Fabian J. Theis [1,2,3] ✉

Models of intercellular communication in tissues are based on molecular profiles of dissociated cells, are limited to receptor–ligand signaling and ignore spatial proximity in situ. We present node-centric expression modeling, a method based on graph neural networks that estimates the effects of niche composition on gene expression in an unbiased manner from spatial molecular profiling data. We recover signatures of molecular processes known to underlie cell communication.

Cells interact on multiple length-scales through direct contact of surface-bound receptors and ligands, tight junctions and mechanical effects, and through indirect mechanisms, including soluble factors. Usually, these communication events cannot be directly observed but are critical to understand emergent phenomena in tissue niches[1]. Molecular signatures of sender and receiver cell types are used to infer latent cell communication events in a tissue through co-occurrence of ligand and receptor expression across putatively communicating cell types[2,3] and through gene expression signatures in the receiving cell[2,4]. Here we propose node-centric expression models (NCEM) to improve cell communication inference through the use of spatial graphs of cells to constrain axes of cellular communication. We infer cell communication from image-structured molecular profiling assays of RNA or proteins with subcellular resolution (Fig. 1a). We defined an NCEM as a graph neural network that predicts a cell's observed gene expression vector from its cell type label and its niche[5] (Fig. 1b and Methods). Cell–cell dependencies may be caused by diverse molecular mechanisms not limited to ligand–receptor-based communication. Therefore, we consider the effects of niche composition on all genes in an unbiased manner. Previous mathematical models of cell–cell interactions in spatial data differed in at least one out of the following central design choices that constitute an NCEM (Supplementary Table 1): they did not represent statistical dependencies of gene expression[6,7], did not model cell communication events[6–8], did not work on targeted protocols with limited ligand and receptor gene capture[8–10] or relied on leave-one-gene hold-outs[4,9], which can result in false discoveries of dependencies (Extended Data Fig. 1 and Methods).

We demonstrate cell communication inference with NCEMs on six datasets measured with MERFISH[11,12], CODEX[13], MIBI-TOF[14], MELC[15] and chip cytometry[16] (Extended Data Fig. 2 and Methods). On average, intracell-type variance accounted for 40.6% of the total variance (Supplementary Fig. 1 and Methods). We defined the screened neighborhood sizes such that they cover the range of average node degrees of the given dataset (Extended Data Fig. 2b). We fit a linear model of gene expression based on a niche represented as interaction terms between the receiver cell type and the presence of each (sender) cell type in the neighborhood (Methods). Linear NCEMs were most predictive on an intermediate length scale of 69 μm across the six datasets (Fig. 1c), showing that cell–cell dependencies appear on length scales characteristic of molecular mechanisms of cell communication. NCEMs outperformed nonspatial baseline models consistently by an average ΔR^2 (Online Methods) of 0.016 (Fig. 1c). As expected, the ΔR^2 is small compared to the baseline model R^2 that characterizes between-cell-type variance (0.39–0.79) because cell type identity accounts for a large fraction of variance in single-cell gene expression assays[17]. The inferred length scales were robust to data downsampling (Extended Data Fig. 3a), out-of-domain data from an unseen genetic knockout condition (Extended Data Fig. 3b–e), to simulated segmentation errors (Extended Data Fig. 3f,g) and to removal of the interaction terms from the linear model (Supplementary Fig. 2). The spatial effect on model performance varies between target cell types, suggesting that cell-type-specific molecular mediators of cell–cell dependency are captured (Supplementary Fig. 3).

NCEMs can be extended to spot transcriptomics if within-cell-type variation can be recovered from spot transcriptomics datasets in deconvolution analyses[18,19]. Here NCEMs model the expression variation within cell types across spots as a function of the inferred spot composition (Fig. 1d and Methods). We considered data from lymph nodes[18,19] (Extended Data Fig. 4a,b) for which a deconvolution was previously demonstrated with cell2location[19]. Linear NCEMs were substantially better at predicting gene expression states of cell types in particular spots than a nonspatial baseline model, both globally (Fig. 1e) and for each cell type (Fig. 1f). The inferred couplings were stable to moderate subsampling of the transcriptomics spots in the training data (Extended Data Fig. 4d). We found putatively interacting ligand–receptor pairs for almost all type couplings in CellPhoneDB[3] and NicheNet[2] analyses of matched single-cell RNA sequencing (scRNA-seq) data, thus demonstrating the need for a quantitative description of statistical couplings in niches (Extended Data Fig. 4e–g). We also identified spatial dependencies between entire niches when modeling spot graphs (Supplementary Fig. 4).

Next we interpreted the spatial dependencies in the MERFISH brain data. We found that L2/3 intratelencephalic (IT) cells depend on the presence of Sncg expressing cells, vascular leptomeningeal cells, and L4/5 cells in their niche (Extended Data Fig. 5 and Supplementary Fig. 5). These associations are reproduced by CellPhoneDB (Supplementary Fig. 6a,b). The L2/3 IT cell subclusters are spatially localized in distinct areas of the primary motor cortex[12]. Indeed, the relative performance of NCEM is spatially structured (Extended Data Fig. 5f and Supplementary Fig. 5e). We quantified these dependencies between cell types as 'cell type couplings', the number of significant gene-wise coefficients of the cell type pair in an NCEM fit (Extended Data Fig. 5g and Methods).

[1]Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany. [2]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. [3]Department of Mathematics, Technical University of Munich, Garching bei München, Germany. [4]These authors contributed equally: David S. Fischer, Anna C. Schaar. ✉e-mail: fabian.theis@helmholtz-muenchen.de
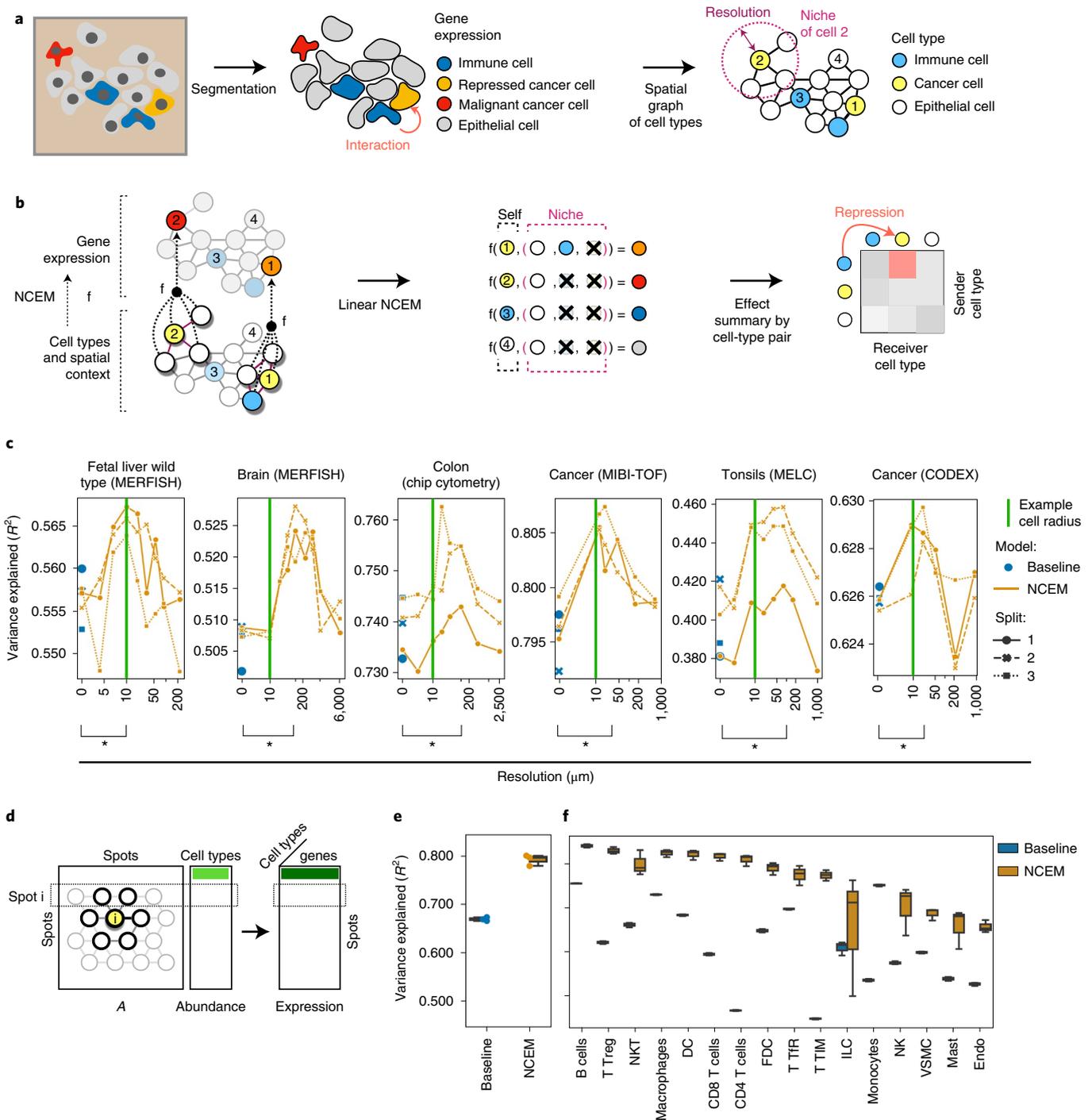
Fig. 1 | Node-centric expression models capture statistical dependencies between cells in space. a, Spatial graphs of cells are based on segmentation masks of cells in spatial molecular profiling data. Resolution is the radius of neighborhood used to define a niche. Numbers label cells and are used in Fig. 1b. **b**, NCEMs describe the gene expression observation of a cell as a function (f) of its spatial neighborhood (niche). **c**, Linear models capture neighborhood dependencies in spatially resolved single-cell data. Shown are the R^2 values between predicted and observed expression vectors on held-out test cells by resolution for six datasets. Green line, 10 μm; baseline (blue points with cross-validation split indicated as shape), a nonspatial linear model; bracket (*), significant difference in paired t-test. **d**, Variation in deconvoluted expression vectors over spots for a given cell type can be attributed to spot composition with a linear NCEM. A, spot adjacency matrix. **e,f**, NCEM performance on deconvoluted data. Shown are the R^2 values between predicted and observed gene expression vectors for held-out test spots of a linear NCEM in comparison to a baseline model that does not use the spot composition information. The performance is shown across the entire test set (**e**) and split by cell type (**f**) ($n = 3$ cross-validation splits). For each box in (**e,f**), the centerline defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 × IQR and outliers are given as points beyond the minimum or maximum whisker.
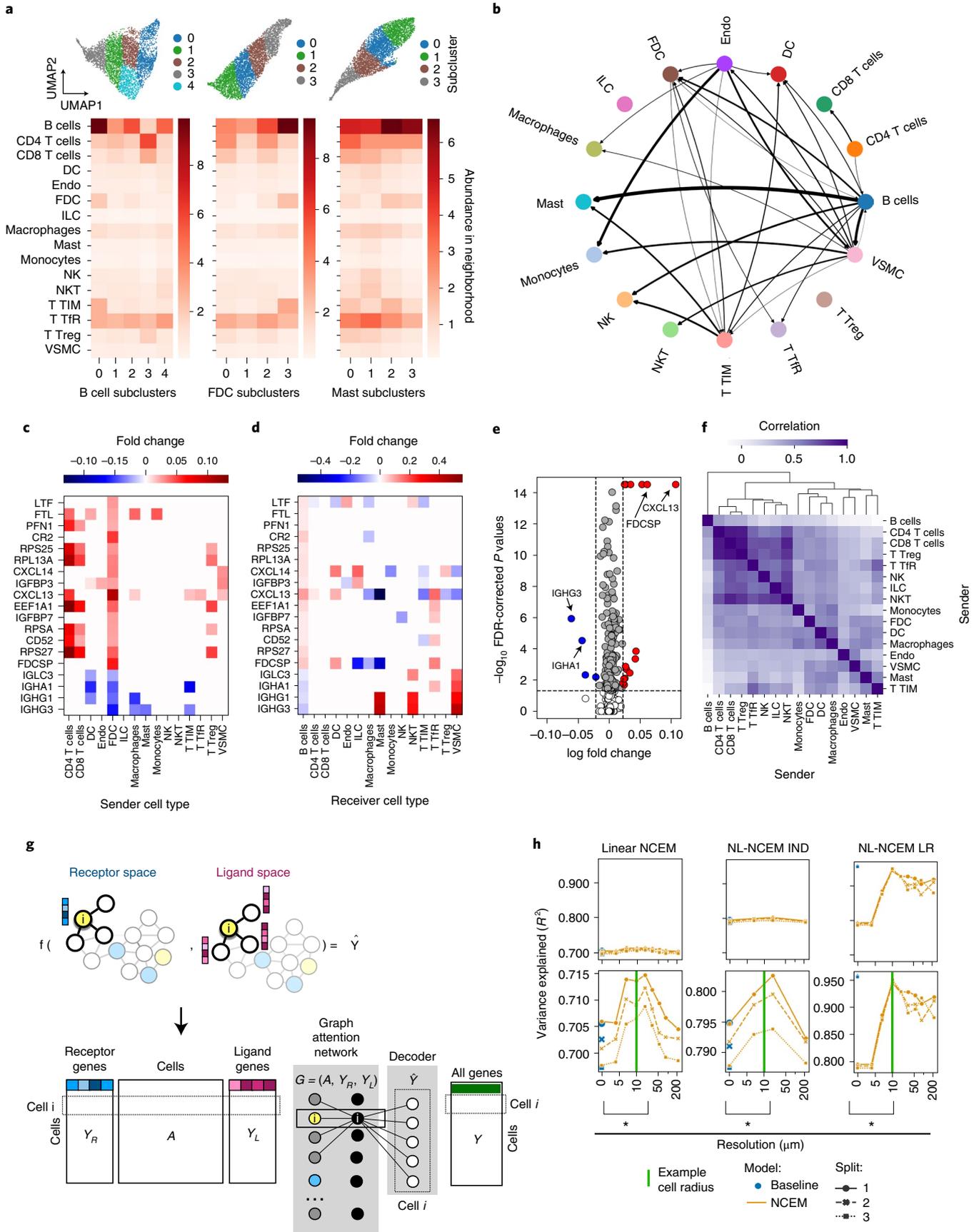
Fig. 2 | Node-centric expression models identify interacting cell types. a, Top: UMAPs with subclusterings of spot-wise gene expression of B cells, FDCs and mast cells from the Visium lymph node dataset. Bottom: the average abundance of cell types per neighborhood after deconvolution of the subclusters. **b**, Type coupling analysis with edge width proportional to the L1 norm of the vector of fold changes of differentially expressed genes for each pair of sender and receiver cell types. Only edges with at least 200 genes are shown. **c,d**, Sender effect (**c**) and receiver effect (**d**) analysis of the FDC–B cell signaling axis. Shown is the estimated fold change that the sender cell type on the x axis induces in the gene on the y axis in receiving B cell (**c**) and conversely for receiver cell types given sender FDCs (**d**). **e**, Volcano plot of differentially expressed genes of B cells in the neighborhood of FDC. **f**, Sender similarity analysis based on a correlation of the coefficient vectors of each sender type with respect to B cell receivers. **g**, A graph kernel to specifically model receptor activity of a cell given the observed ligand expression in its niche extends NCEM to ligand–receptor modeling. **h**, R^2 values between predicted and observed expression vectors for held-out test cells of linear (NCEM), nonlinear (NL-NCEM IND) and ligand–receptor-kernel NCEMs (NL-NCEM LR) by resolution on imputed MERFISH fetal liver data. Green line, 10 μm; baseline (blue points with cross-validation split indicated as shape), a nonspatial linear model; bracket (*), significant difference in paired t-test.

We discovered a dependency of CD8 T cells on multiple other cell types in the chip cytometry colon data (Extended Data Fig. 6) and a dependency of CD8 T cells on proximity to the tumor–immune boundary[14] in colorectal cancer (Extended Data Fig. 7).

Similarly, we interpreted NCEM fits on the deconvoluted Visium lymph node data. We identified a bidirectional dependency of B cells and follicular dendritic cells (FDCs) that is indicative of positive feedback between both cell types in germinal center organization[20] (Fig. 2a,b and Extended Data Fig. 4c). Similarly, we found evidence for a dependency of mast cells on B cells (Fig. 2b) and a mast cell subcluster associated with niches enriched in B cells (Fig. 2a). The FDC subcluster associated with niches enriched in B cells (cluster 3) showed increased expression of *Cxcl13*, a key chemokine for germinal centers[20] (Extended Data Fig. 4c), supporting the association of these couplings with germinal centers. We further dissected these couplings based on the gene-wise effects of all senders on one receiver type ('receiver effect analysis', Fig. 2c) and of one sender on all receivers ('sender effect analysis', Fig. 2d), which contextualizes differential expression results of the FDC–B cell axis (Fig. 2e and Supplementary Data 1). Multiple T cell clusters had a similar effect on B cells in a 'sender similarity analysis' (Fig. 2f), in which we correlated the coefficient vectors of sender cell types that correspond to B cell receivers, which demonstrates conservation of cell type identity in the sender profile.

In contrast to linear NCEMs, nonlinear NCEMs can account for weighted or higher-order interactions between cell types (Extended Data Fig. 8a). As for linear NCEMs, we found resolution-dependent prediction performance profiles in nonlinear NCEMs (Extended Data Fig. 8b and Methods) and a dependency between L4/5 IT and L2/3 IT cells (Extended Data Fig. 8c and Methods). Notably, the nonlinear models did not outperform linear models in gene expression prediction, which suggests that the spatial dependencies in the given datasets are well described by linear models (Extended Data Fig. 8b). Next, we considered a conditional variational autoencoder (CVAE) version of an NCEM to model cell-intrinsic latent states. We conditioned the distribution over node expression states on a graph embedding of the niche and the cell type (Extended Data Fig. 9a). Even though CVAE–NCEM attained much higher predictive performance in reconstruction tasks (Extended Data Fig. 9b and Supplementary Fig. 7a), these models did not consistently capture spatial dependencies because niche states were represented in latent variables (Extended Data Fig. 9c–f and Supplementary Fig. 7b–e).

The interpretation of spatial dependencies inferred on targeted spatial molecular profiling assays is constrained by the limited capture of ligand–receptor pairs. We imputed the cell-wise gene expression in the MERFISH fetal liver data using corresponding scRNA-seq data[21] (Extended Data Fig. 10). We designed a graph kernel of cell-wise receptor activity to directly model ligand–receptor interactions between neighboring cells (Fig. 2g). The peak predictive performance of the ligand–receptor nonlinear NCEM was much higher compared to the nonlinear NCEM (R^2 of 0.799 and 0.947), demonstrating the increased complexity of the input

compared to categorical cell type labels. We observed differential receptor signaling as differential latent neuron activation of *Kit*[11] in sinusoidal endothelial cells (SECs) depending on their proximity to arterial endothelial cells (AECs) (Supplementary Data 2).

NCEMs are linear and nonlinear graph neural networks and CVAEs that model cell communication events in spatial omics assays (Supplementary Table 2). We identified statistical dependencies between cells on physiologically relevant length scales and interpreted fits based on model parameters. The statistical identifiability of cell type couplings may improve with increased capture of niche heterogeneity, through the inclusion of three-dimensional data, by increasing the number of cells measured and by increasing the variation in the training data through perturbations of niche structure. Uncertainty in segmentation of cells or nuclei can be improved on the level of the spatial measurement[22] or may be addressed in model extensions. We found that linear NCEMs perform well in the presented tasks and are promising candidates for cell communication inference. The complexity of the graph neural network used in the NCEM defines the complexity of the motifs of cell communication that can be discovered and may be expanded given more complex datasets. CVAE–NCEMs may be used to model cell-intrinsic variation together with niche effects. Similarly, the graph kernels tailored to ligand–receptor signaling presented here provide constrained latent variables that explain extrinsic variation and could be used together with variables for cell-intrinsic variation. Nonlinear NCEMs learn a cellular representation within the spatial graph[23], and we demonstrated that these representations can model niches and may be exploited for unsupervised analysis of tissue structures.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-022-01467-z.

References

1. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nat. Biotechnol.* **40**, 308–318 (2022).
2. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
3. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
4. Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J. & Stegle, O. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell Rep.* **29**, 202–211 (2019).

5. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. Proceedings of the 5th International Conference on Learning Representations (2017).

6. Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).

7. Palla, G. et al. Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).

8. Yuan, Y. & Bar-Joseph, Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol.* **21**, 300 (2020).

9. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).

10. Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711 (2011).

11. Lu, Y. et al. Spatial transcriptome profiling by MERFISH reveals fetal liver hematopoietic stem cell niche architecture. *Cell Discov.* **7**, 47 (2021).

12. Zhang, M. et al. Molecular, spatial and projection diversity of neurons in primary motor cortex revealed by in situ single-cell transcriptomics. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2020.06.04.105700v1.abstract (2020).

13. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **183**, 838 (2020).

14. Hartmann, F. J. et al. Single-cell metabolic profiling of human cytotoxic T cells. *Nat. Biotechnol.* **39**, 186–197 (2021).

15. Pascual-Reguant, A. et al. Multiplexed histology analyses for the phenotypic and spatial characterization of human innate lymphoid cells. *Nat. Commun.* **12**, 1737 (2021).

16. Jarosch, S. et al. Multiplexed imaging and automated signal quantification in formalin-fixed paraffin-embedded tissues by ChipCytometry. *Cell Rep. Methods* **1**, 100104 (2021).

17. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

18. Lopez, R. et al. DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat Biotechnol.* **40**, 1360–1369 (2022).

19. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).

20. Ansel, K. M. et al. A chemokine-driven positive feedback loop organizes lymphoid follicles. *Nature* **406**, 309–314 (2000).

21. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).

22. Lohoff, T. et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* **40**, 74–85 (2022).

23. Hetzel, L., Fischer, D. S., Günnemann, S. & Theis, F. J. Graph representation learning for single cell biology. *Curr. Opin. Syst. Biol.* **28**, 100347 (2021).

Methods

Data. *Fetal liver (MERFISH).* Lu et al.[11] measured wild type (WT) and Tet[2−/−] fetal livers with MERFISH in 140 (WT) and 195 (Tet[2−/−]) images across four WT fetal livers at E14.5 and two Tet[2−/−] knockout fetal livers at E14.5, with 132 genes observed in 40,864 (WT) and 54,970 (Tet[2−/−]) cells. We used cell types as originally annotated by Lu et al.: AEC, erythroid cell, erythroid progenitor, hepatocyte, megakaryocyte, macrophage, myeloid and SEC. We removed cells with an unknown label from the dataset. We scaled model outputs by the node size in the respective output layer of each model class to mitigate count noise (Supplementary Fig. 8f).

Brain (MERFISH). Zhang et al.[12] measured mouse primary motor cortex with MERFISH in 64 images across two mice, with 254 genes observed in 284,098 cells. We used the cell types originally annotated by Zhang et al.: astrocytes, endothelial, L2/3 IT neurons, L4/5 IT, L5/6 near-projecting neurons, L5 IT, L5 pyramidal tract neurons, L6 cortico-thalamic projection neurons, L6 IT, L6 IT Car3, L6b, Lamp5, microglia, oligodendrocyte precursor cells, oligodendrocytes, perivascular macrophages, pericytes, parvalbumin, smooth muscle cells, Sncg, somatostatin (Sst), Sst Chodl, vascular leptomeningeal cells, vasoactive intestinal polypeptide, and other cells, where L identifies the layer (L1–L6) of the distinctive laminar structure based on cytoarchitectural features (Extended Data Fig. 2a). Parvalbumin, Sst, vasoactive intestinal polypeptide, Sncg and Lamp5 define five subclasses of GABAergic cells. We removed cells labeled as 'other' from the dataset. We used an identifier for the respective mouse as domain information.

Colon (chip cytometry). Jarosch et al.[16] measured an inflamed colon with chip cytometry in two images from one patient, with 19 genes observed in 11,321 cells. We used the cell types originally annotated by Jarosch et al.: B cells, CD4 T cells, CD8 T cells, GATA3[+] epithelial, Ki67 high epithelial, Ki67 low epithelial, lamina propria cells, macrophages, monocytes, PD-L1[+] cells, intraepithelial lymphocytes, muscular cells and other lymphocytes (Extended Data Fig. 2a). We coarsened the cell type annotation by combining Ki67 high epithelial and Ki67 low epithelial to a joined annotation of Ki67 epithelial. We log-transformed the gene expression values for use in the analyses presented here to mitigate count noise (Supplementary Fig. 8b).

Cancer (MIBI-TOF). Hartmann et al.[14] measured colorectal carcinoma and healthy adjacent tissue with MIBI-TOF in 58 images across four individuals, with 36 genes observed in 63,747 cells. We used the cell types originally annotated by Hartmann et al.: endothelial, epithelial, fibroblast, CD11c myeloid, CD68 myeloid, CD4 T cells, CD8 T cells and other immune cells (Extended Data Fig. 2a). The cohort in this dataset includes two patients with colorectal carcinoma and two healthy controls. We scaled the model outputs by cell-wise size factors.

Tonsils (MELC). Pascual-Reguant et al.[15] measured tonsils from patients undergoing tonsillectomy with multiepitope ligand cartography (MELC), an immunohistochemistry approach, in one image across one patient, with 51 genes observed in 9,512 cells. We used the cell types originally annotated by Pascual-Reguant et al.: B cells, endothelial cells, innate lymphoid cell (ILC), monocytes/macrophages/dendritic cells (DC), natural killer (NK) cells, plasma cells, T cytotoxic cells, T helper cells (Extended Data Fig. 2a). We removed cells labeled as 'other' from the dataset.

Cancer (CODEX). Schürch et al.[13] measured advanced-stage colorectal cancer with CODEX in 140 images across 35 patients, with 57 genes observed in 272,266 cells. We used the cell types originally annotated by Schürch et al.: B cells, CD11b[+] monocytes, CD11c[+] dendritic cells, CD11b[+] CD68[+] macrophages, CD163[+] macrophages, CD68[+] macrophages, CD68[+] macrophages GzmB[+], CD68[+] CD163[+] macrophages, CD3[+] T cells, CD4[+] T cells, CD4[+] T cells CD45RO[+], CD4[+] T cells GATA3[+], CD8[+] T cells, NK cells, T regs, adipocytes, dirt, granulocytes, immune cells, immune cells/vasculature, lymphatics, nerves, plasma cells, smooth muscle, stroma, tumor cells, tumor cells/immune cells and vasculature (Extended Data Fig. 2a). We removed cells with an annotation of dirt or an undefined label from the dataset. We merged the macrophage subclusters (CD11b[+] CD68[+] macrophages, CD163[+] macrophages, CD68[+] macrophages GzmB[+] and CD68[+] CD163[+] macrophages) and the CD4[+] T cell subclusters (CD4[+] T cells, CD4[+] T cells CD45RO[+] and CD4[+] T cells GATA3[+]). We scaled model outputs by the node size in the respective output layer of each model class to mitigate count noise (Supplementary Fig. 8e).

Lymph node (Visium). We performed deconvolution with cell2location[19] on a 10x Visium lymph node dataset based on a scRNA-seq dataset from the same tissue as previously described[19]. The cell type labels used for deconvolution were B cells, DC, endothelial, follicular dendritic cells, ILC, macrophage, mast, monocytes, NK, natural killer T cells (NKT), CD4 T cells, CD8 T cells, T cells (TIM3), T follicular regulatory cells _fr_, T regulatory cells _reg_) and vascular smooth muscle cells.

Dataset partitions. We randomly selected 10% of all nodes across all images and patients as the test set. From the remaining nodes, 10% of all nodes are selected as the validation set.

MERFISH-scRNA-seq integration. We integrated scRNA-seq with MERFISH data to impute the full gene expression in the spatial graph of cells measured in MERFISH. We performed this integration between the MERFISH fetal liver (WT) data and scRNA-seq of E14.5 whole fetal liver cells sequenced by 10x Genomics platform[11], which is available as GSE172127 on GEO. The scRNA-seq dataset contains 9,448 cells across 28,692 features. We performed quality control and removed cells with fewer than 500 detected genes, genes expressed in less than three cells and cells with more than 10% of the transcripts coming from mitochondrial genes from the dataset. We applied Tangram[21] to generate a spatially resolved representation of the scRNA-seq fetal liver dataset[21]. We used 131 out of 132 genes from the MERFISH fetal liver (WT) data, which were both present in the MERFISH and the scRNA-seq dataset, to perform the mapping.

Variance decomposition into inter- and intra-cell-type variance. The variance of a single-cell resolved dataset can be decomposed into intercell-type variance, intracell-type variance and gene variance. The gene variance is independent of cell type definitions and can therefore be considered separately from intra- and inter-cell-type variance:

$$\sum_i^N \sum_j^J \left(y_{i,j} - \bar{y}\right)^2$$

$$= \underbrace{\sum_i^N \sum_j^J \left(y_{i,j} - \bar{y}_{k(i),j}\right)^2}_{\text{intracell—type variance}} + \underbrace{\sum_i^N \sum_j^J \left(\bar{y}_{k(i),j} - \bar{y}_j\right)^2}_{\text{intercell—type variance}} + \underbrace{\sum_i^N \sum_j^J \left(\bar{y}_j - \bar{y}\right)^2}_{\text{gene variance}}$$

(1)

where $y_{i,j}$ is the gene expression of cell i out of N and gene j out of J, $\bar{x}_{k,j}$ is the mean expression of each gene j in each cell type k, $k(i)$ is the cell type of cell i, $\bar{y}_j$ is the mean expression of each gene j and $\bar{y}$ is the mean of the dataset.

Simulations. *Segmentation errors.* We simulated segmentation errors in which the segment boundary between two neighboring cells is misplaced (Extended Data Fig. 3f,g). We selected a fraction of cells (10% or 50%) in the MERFISH fetal liver data at random, selecting one neighbor at random for each selected cell, and transferred a fraction of the total molecular abundance vector of the selected cell to its neighbor.

Spatial dependencies. We simulated single-cell resolved spatial data from scratch by using the cell graph from the chip cytometry colon data (Extended Data Fig. 2) and the simulated node-wise gene expression vectors. We modeled cell types using the number of genes originally defined in the dataset and drew a mean expression value for each gene from a uniform distribution between 0 and 10. We considered two scenarios of dependencies between cells: (1) a dataset without spatial dependencies in which all cells are drawn from one cell type and are independent and identically distributed and (2) a dataset with spatial dependencies in which cells belong to either one of two cell types, where we introduced dependencies on the presence of the respective other cell type in the neighborhood in 50% of the genes, with strong effect sizes drawn from a uniform distribution between 4 and 6. We fitted NCEMs, Misty[9] and SVCA[4] on both simulated datasets (Supplementary Fig. 2). The simulated images contained over 4,500 cells per image. To reduce the runtime for SVCA for these samples we cropped both images in the lower-right corner to create images with approximately 850 cells each.

Models. The inputs to NCEMs are (1) a gene expression matrix $Y \in R^{N \times J}$ where N is the number of cells and J is the number of genes, (2) a matrix of observed cell types $X^1 \in R^{N \times L}$ where L is the number of unique cell type labels and (3) a matrix specifying the batch assignments $X^c \in R^{N \times C}$ of C distinct batches or domains, such as images or patients. We denote the adjacency matrix of connected cells as $A \in R^{N \times N}$, which is calculated based on the spatial proximity of cells per image. For linear models and models with an indicator aggregator, we used a binary adjacency matrix $A_{ij} = 1$ if $d(x_a, x_b) \leq \delta_{\max}$ where $d(\cdot, \cdot)$ describes the euclidean distance between nodes $a, b \in N$ in space and $\delta_{\max}$ is the neighborhood size (resolution), and $A_{ij} = 0$ otherwise. For models using graph convolutions, we normalized A such that all rows sum to one: D^{-1}_A where D is the diagonal node degree matrix. The output of NCEMs is $\hat{Y} \in R^{N \times J}$, a reconstruction of the gene expression matrix. In selected datasets, we applied size factor scaling to the network output Y using the size factors $sf_i = \frac{\sum_j^J y_{ij}}{\frac{1}{N} \sum_{i'}^N \sum_{j'}^J y_{i'j'}}$. The global data handling per dataset is reported in Supplementary Table 3, model hyperparameters for linear models are reported in Supplementary Table 4, and the parameters for nonlinear and CVAE models in Supplementary Table 5.

Loss functions. We use a Gaussian log-likelihood, ll, loss as an optimization objective for linear and nonlinear models with

$$ll(y) = \frac{1}{N*J} \sum_i^N \sum_j^J \left(-\log\left(\sqrt{2\pi}\sigma_j\right) - 0.5\frac{(y_{ij} - \hat{y}_{ij})^2}{\sigma_j^2}\right)$$ over cells i and genes j, where σ_j is the predicted standard deviation of a gene (Supplementary Fig. 8). The loss function of CVAE model is the negative data log-likelihood in addition to the

Kullback–Leibler divergence between the variation posterior $q_\phi(z)$ and the prior $p(z)$ on the latent variables: $\text{ll}_{\text{CVAE}} = -\text{ll}(y) + D_{\text{KL}}\left(q_\phi(z)||p(z)\right)$.

Optimization. We ran grid searches to find the optimal set of hyperparameters for each dataset where the batch size is the number of images per dataset. We selected the number of nodes evaluated per image per batch to improve convergence. We trained all models with the Adam optimizer: linear models with 0.05, the remaining models with multiple learning rates of {0.5, 0.05, 0.005}. Additionally, we used a learning scheduler on the validation loss with a patience of 20 epochs, which reduces the learning rate by a factor of 0.5, so $\text{lr}_{\text{new}} = \text{lr} * 0.5$ and early stopping with a patience of 100 epochs. The exact description of all grid searches in code are supplied in the benchmarking repository (Code Availability). We trained linear models for hypothesis testing using ordinary least squares estimators on the full dataset.

Linear NCEM. The linear nonspatial baseline model infers a reconstruction $\hat{Y}$ from a node's cell type and respective domain information via $\hat{Y} = X^D \beta$, where X^D is the design matrix and $\beta \in R^{(L+C)\times J}$ are the parameters learned by the model. The design matrix of nonspatial baseline models is given by $X^D = (X^1, X^c) \in R^{N\times(L+C)}$. The spatial counterpart model, the linear NCEM, has access to an additional spatial sender–receiver interaction matrix. First, we computed the binary sender cell presence in the neighborhood of each cell $X^S = 1_{(AX^l>0)} \in \{0,1\}^{N\times L}$, where $1_{()}$ represents an indicator function. To generate a matrix representation of sender–receiver cell interactions, we compute the interaction terms between the cell type of the index cell and the presence of each cell type in its neighborhood as the outer product between X^l and X^S. The resulting interaction matrix is $X^{TS} \in \{0,1\}^{N\times L^2}$, and the design matrix for the linear model with interaction terms is given by $X^D = (X^1, X^{TS}, X^c) \in R^{N\times(L+L^2+C)}$. This design matrix can be related to a graph neural network: X^l and X^c are node-wise condition vectors that can be appended to a local graph embedding centered on an index cell, and X^{TS} is equivalent to an outer product of the one-hot-encoded representation of an index cell with the projection obtained from a single-layer graph neural network that embeds one-encoded cell type feature vectors with a feature-wise max pooling operator across the neighborhood without the index cell. This projection is a cell-type-dimensional indicator for the presence of each cell type in the neighborhood. Linear NCEMs perform parameter inference on $\hat{Y} = X^D \beta$ where $\beta \in R^{(L+L^2+C)\times J}$. We also considered an NCEM without interaction terms which does not have receiver-specific sender effects but only global sender effects, which account for the presence of senders in the niche via $X^D = (X^1, X^S, X^c) \in R^{N\times(L+C)}$. We evaluated significance of coefficients corresponding to the interaction matrix X^{TS} with a Wald test.

Linear NCEM for deconvoluted spot transcriptomics. The baseline model is the same as for the standard linear NCEM. The corresponding NCEM treats the spot as a neighborhood and uses the deconvoluted cell type abundances per spot $X^F \in R^{(N*L)\times L}$ as a vector-shaped neighborhood summary, replacing a kernel on a graph. Note that $(N*L)$ is the number of spots times the number of cell types: this model treats every type- and spot-wise gene expression vector, a result of the deconvolution, as an observation. The overall design matrix of the linear model includes the interaction between the target cell type and the spot composition, and spot-wise covariates: $X^D \in R^{(N*L)\times(L+L^2+C)}$. Note that here, the spot composition is the same for all L gene expression prediction problems per spot. As for the linear NCEM, we again fit a linear model to this design matrix to predict deconvoluted gene expression. One can define a corresponding nonlinear model that uses the deconvoluted cell type abundances per spot $X^F \in R^{N\times L}$ as a vector-shaped node feature space. Note that N is the number of spots in this nonlinear model. These feature vectors can be connected based on spot proximity in a graph embedding of spots $f_{\text{enc}} : q_\phi(z_s|g(A, X^1)_s)$. The cell-type-wise gene expression decoder for spot s and cell type k is then $f_{\text{dec}} : p_\theta(Y_{sk}|z_s, X^k, X_s^c)$, where X^k is a one-hot embedding of the cell type k.

Nonlinear NCEM. NCEMs include nonlinear models that encode the neighborhood through a graph neural network (NL-NCEM) and decode expression vectors. The corresponding nonspatial baseline model is a nonlinear model (NL) that predicts expression from cell type and graph-level predictors. A local graph embedding is given by $f_{\text{enc}} : q_\phi(z_i|X_i^l, g(A, X^1)_i, X_i^c)$, which encodes the cell type labels X^l, some graph-level predictors X^c and the local graph embedding $g(A, X^l)$, based on the adjacency matrix A, into a latent state z. The latent state of cell i is input to a fully connected layer stack given by $f_{\text{dec}} : p_\theta(Y_i|z_i, X_i^l, X_i^c)$. If one uses an indicator embedding function as described in the section Linear NCEM and all hidden layers are removed from the NL-NCEM, a single linear transformation of the input remains, which is equivalent to the linear NCEM. Alternatively, $g(A, X^l)$ can be a graph embedding learned by a graph-convolutional network (GCN)[5]. A one-layer GCN is given by $g(A, X^l) = \text{softmax}(\text{ReLU}(\bar{A}X^lW))$, where $W \in R^{L\times H}$ is a weight matrix, H is the dimension of the learned node representation and $\bar{A}$ is the normalized adjacency matrix.

Ligand–receptor NCEM (NL-NCEM-LR). Here we consider a specific NL-NCEM with a tailored graph kernel. This graph kernel embeds each cell into a receptor dimensional latent space z based on the receptor gene expression on the index cell, ligand gene expression on neighboring cells, and the adjacency matrix A, which encodes the set of neighbors MN of cell i: $f_{\text{enc}} : z_{ik} = g(A_i, Y_{i,r(k)}, Y_{:,l(k)}) = \sum_{m}^{M} f_R(Y_{i,r(k)}) * f_L(Y_{m,r(k)})$. Here, $r(k)$ and $l(k)$ encode the gene index of receptor and ligand that correspond to ligand–receptor pair k. The latent state and graph-level predictors X^c are input to a fully connected layer stack that decodes gene expression $f_{\text{dec}} : p_\theta(Y_i|z_i, X^c)$. The corresponding nonspatial baseline is a nonlinear model (NL) that receives receptor expression of the index cell as bottleneck activation and has the same decoder. This baseline model is not nested in the nonlinear ligand–receptor NCEM (NL-NCEM-LR) but models a baseline which imputes all genes' expression based on ligand gene expression within the cell.

Conditional variational autoencoder NCEM. A conditional variational autoencoder NCEM (CVAE–NCEM) learns a distribution over node states Y based on a node-wise latent space z. The nonspatial CVAE null model contains the cell type and graph-level predictors as a condition in the variational posterior and the likelihood model. In CVAE–NCEM, the conditions are the cell type labels X^l, some graph-level predictors X^c and the local graph embedding $g(A, X^l)$. The encoder is given by $f_{\text{enc}} : q_\phi(z_i|Y_i, X_i^l, g(A, X^1)_i, X_i^c)$ and the decoder is defined by $f_{\text{dec}} : p_\theta(Y_i|z_i, X_i^l, g(A, X^1)_i, X_i^c)$. A CVAE–NCEM for a full dataset depends on both the niche and the type of the cell itself. This setting presents the challenge of encountering a nonidentifiability between variance attributed to latent variables, cell type conditions and neighborhood context. In this study, we consider the CVAE–NCEM trained on the molecular vectors of a single target cell type as a function of the full neighborhood context to remove the nonidentifiability with respect to cell type variation and focus on the nonidentifiability between latent variables and neighborhoods.

Model evaluation. We evaluated model performance using the coefficient of determination: $R_i^2 = 1 - \frac{\sum_j^J (y_{ij} - \hat{y}_{ij})^2}{\sum_j^J (y_{ij} - \bar{y}_{ij})^2}$ for cells i and over genes j. We selected the best performing models based on R^2 on a validation dataset and showed this metric evaluated on test data in the manuscript. The performance of CVAEs is additionally assessed in style transfer tasks. In style transfer, the gene expression state and neighborhood of a reference node a from the source domain is encoded to estimate the latent states of this node. This latent representation is then decoded to the target domain of cell b, which implies conditioning the decoding on the target neighborhood:

$$z_a \sim q_\phi\left(\mathbf{z}|\mathbf{Y_a}, \mathbf{X_a^1}, \mathbf{g(A, X^1)_a}, \mathbf{X_a^c}\right) \qquad (2)$$

$$\hat{Y}_b = p_\theta\left(\mathbf{z_a}, \mathbf{X_b^1}, g(A, X^1)_b, \mathbf{X_b^c}\right) \qquad (3)$$

where a, b are cell indices, q_ϕ is the amortized variational posterior and p_θ is the decoder network. See also Conditional variational autoencoder NCEM for details on the notation.

Unsupervised analysis. We used uniform manifold approximation and projection (UMAP) to embed the cells in two dimensions for visualization of high-dimensional data.

We computed the UMAP of B cell, FDC and mast cell substates (Fig. 2a) based on 50 principal components (PCs) and $k = 500$. We computed the UMAP of the scRNA-seq reference dataset of lymph nodes (Extended Data Fig. 4b) based on 50 PCs with $k = 100$. We computed the UMAP of the MERFISH brain data[12] matrix (Extended Data Fig. 5a) based on the first 35 PCs and the k-nearest neighbor graph with $k = 10$. We computed the UMAP of L2/3 IT neurons in slice 153 (Extended Data Fig. 5a) and slice 162 (Supplementary Fig. 5a) of the MERFISH brain dataset based on the first 40 PCs with $k = 40$ and performed Louvain community detection using Scanpy[17] to define stable L2/3 IT substates. We computed UMAPs of CD8 T cells in area 1 in the chip cytometry dataset (Extended Data Fig. 6) based on the gene expression matrix directly and $k = 22$, and UMAPs of CD8 T cells in image 1, 5, 8 and 16 of the MIBI-TOF cancer dataset (Extended Data Fig. 7) based on the gene expression directly and $k = 60$. We performed Louvain community detection of the latent space in CVAE and CVAE–NCEM IND models (Extended Data Fig. 9d,e and Supplementary Fig. 7c,d) based on the latent space using $k = 80$ for the MERFISH brain dataset and $k = 250$ for the chip cytometry colon dataset.

We performed cluster enrichment with Fisher's exact test. Each contingency table is composed of two categorical variables. The first variable describes the binary assignment of cells to one L2/3 IT subcluster. The second variable describes the presence of a source cell type in their neighborhood. We performed Benjamini and Hochberg false discovery rate correction (FDR) of cluster enrichment P values. A similar approach was used for the cluster enrichment analysis of CD8 T cells in the chip cytometry colon and the MIBI-TOF cancer datasets.

Type coupling analysis. We performed type coupling analysis, sender effect and receiver effect analysis based on a Wald test on the parameters estimates of linear NCEM obtained on the full dataset as ordinary least squares estimates. We performed FDR-correction of the resulting P values using the Benjamini-Hochberg correction method. The coupling measure between sender and receiver cells is the L1-norm of coefficients of significant coefficients that correspond to the specific receiver–sender interaction term in the linear model, or the number of differentially expressed genes. The sender effect and receiver effect analysis consists of the set of coefficients and their significance for a particular sender and receiver, respectively. The sender similarity analysis is a hierarchical clustering of the Pearson product-moment correlation coefficients of coefficient vectors of sender cell types for one defined receiver cell type.

Differential receptor activity in NL-NCEM-LR. We used a t-test to obtain a ranking for highly differential receptor signaling in SEC depending on the presence of neighboring AEC. We used the neighborhood size that corresponded to the best performing resolution of the NL-NCEM-LR model.

Subsampling robustness analysis. We randomly subsampled the spatial transcriptomics spots from the Visium lymph node data to 5%, 25%, 50% and 75% of all spots across three cross validations. We deconvoluted the resulting subsampled slide with cell2location and used this inferred spot composition as input to the NCEM type coupling analysis for spot-transcriptomic data. In order to assess the robustness with respect to identified putative dependencies, we computed the R^2 between the inferred coefficient vectors over genes for each cell type pair between the fit to the complete data and the fit to the subsampled slide.

CellPhoneDB and NicheNet. We inferred putatively communicating ligand–receptor pairs in lymph nodes using CellPhoneDB as implemented in squidpy[7] on scRNA-seq data[24] with $n = 53,275$ cells on the 10,000 most variable genes. We quantified sender–receiver interactions as the number of significant ligand–receptor pairs at an FDR-corrected P value of 0.05. Additionally, we considered the presence of nonzero expression of cognate ligand–receptor pairs (Extended Data Fig. 4e). We performed the CellPhoneDB analysis shown in Supplementary Fig. 6 based on $n = 1,000$ permutations. Additionally, we used randomly subsampled data for the analysis of MERFISH brain[12] 10% with $n = 27,655$, MIBI TOF cancer[14] 40% with $n = 25,498$ and CODEX cancer[13] 10% with $n = 25,186$.

We defined the 5,000 most variable genes per receiver cell type as target genes in a NicheNet analysis. For the following cell types, we limited the number of highly variable genes to the number given in brackets depending on the respective intracell-type heterogeneity: DC (500), endothelial (1,500), erythrocyte (250), HSC (1,000), macrophages (4,000), mast (1,000), monoctyes (2,000), myeloid (2,000), neutrophil (400), stromal cells (1,500) and T T_{reg} (3,000). We defined all remaining genes as background genes for NicheNet. We selected the top-100-ranked ligands from NicheNet and thresholded the putative ligands to be expressed in at least 5% of all sender cells.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The MERFISH fetal liver[11], MERFISH brain[12], MIBI TOF cancer[14], MELC tonsils[15], CODEX cancer[13], chip cytometry colon[16] and Visium lymph node[19] datasets are publicly available (Methods).

Code availability

All models described here are implemented in a Python package available at https://github.com/theislab/ncem. All benchmarking and analysis codes are provided at https://github.com/theislab/ncem_benchmarks. Tutorials for model usage are available from https://github.com/theislab/ncem_tutorials.

References

24. Tabula Sapiens Consortium et al. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).

Author contributions

D.S.F. and F.J.T. conceived the project. D.S.F. and A.C.S. performed the analysis and wrote the code. D.S.F., A.C.S. and F.J.T. wrote the manuscript.

Funding

Competing interests

F.J.T. consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd and Omniscope Ltd, and has ownership interest in Dermagnostix GmbH and Cellarity. The remaining authors declare no competing interests.

Additional information

Extended data are available for this paper at https://doi.org/10.1038/s41587-022-01467-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-022-01467-z.

Correspondence and requests for materials should be addressed to Fabian J. Theis.

Peer review information *Nature Biotechnology* thanks Qing Nie, Tommaso Biancalani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
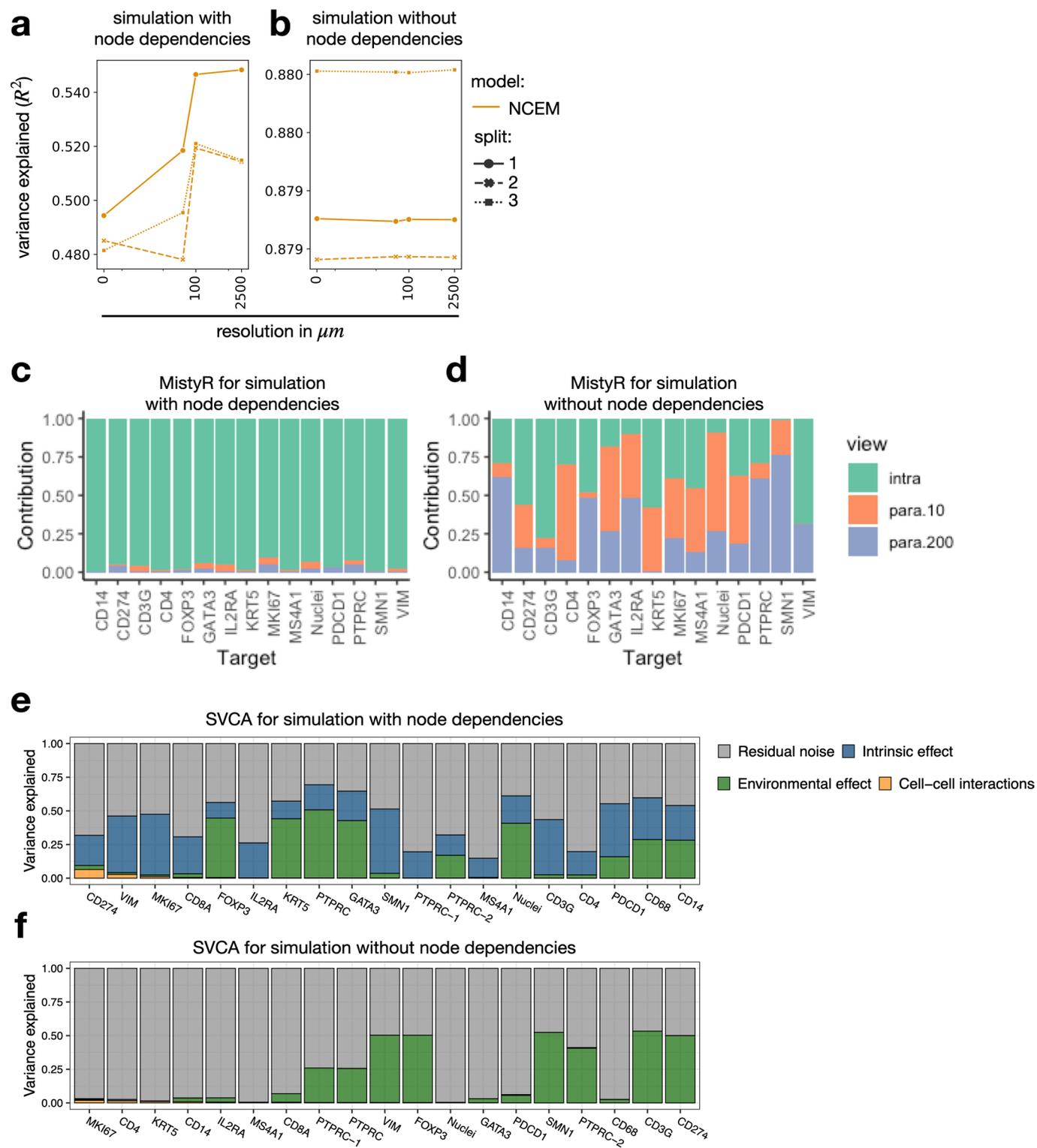
Extended Data Fig. 1 | Benchmarking cell–cell dependency inference on simulated data. Shown are fits of NCEMs **(a, b)**, MistyR **(c, d)**, and SVCA **(e, f)** on simulated data with simulated dependencies between cells **(a, c, e)** and without simulated dependencies between cells **(b, d, f)**. The simulation is described in the Methods.
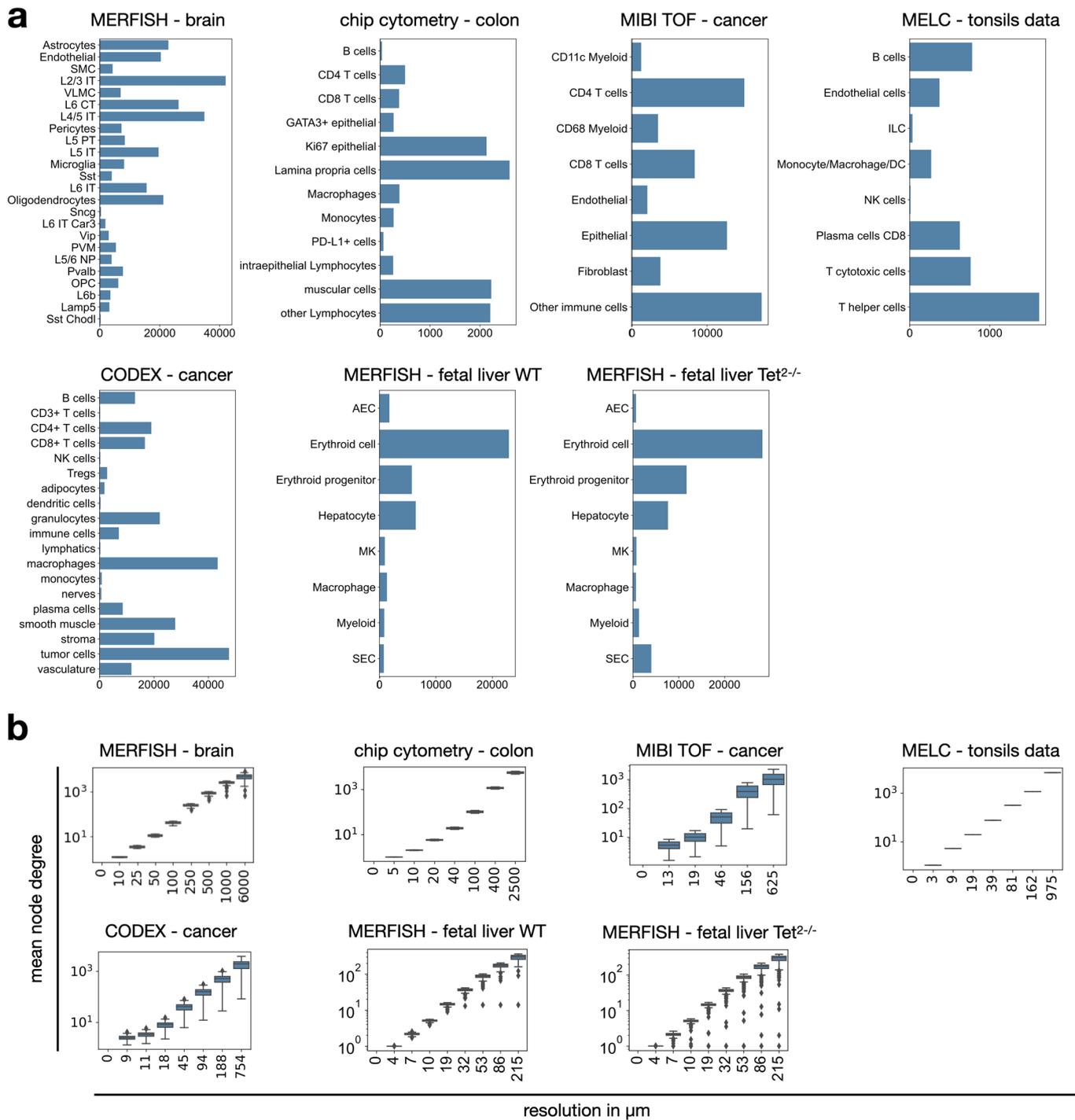
Extended Data Fig. 2 | Cell-type centric summary statistics of the considered datasets. (a) Cell-type frequencies by dataset. Shown is a barplot with the number of cells in each cell-type for MERFISH – brain data, chip cytometry – colon data, MIBI TOF – cancer data, MELC – tonsils data, CODEX – cancer data, MERFISH - fetal liver wild type and MERFISH - fetal liver *Tet2−/−*. **(b)** Mean node degree (number of neighbors) by resolution in μm and dataset for MERFISH – brain data (n = 284,098 cells), chip cytometry – colon data (n = 11.321 cells), MIBI TOF – cancer data (n = 63,747 cells), MELC – tonsils data (n = 9,512 cells), CODEX – cancer data (n = 272,266 cells), MERFISH - fetal liver wild type (n = 40,864 cells) and MERFISH - fetal liver *Tet2−/−* (n = 54,970 cells). For each box in **(b)**, the centerline defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker.
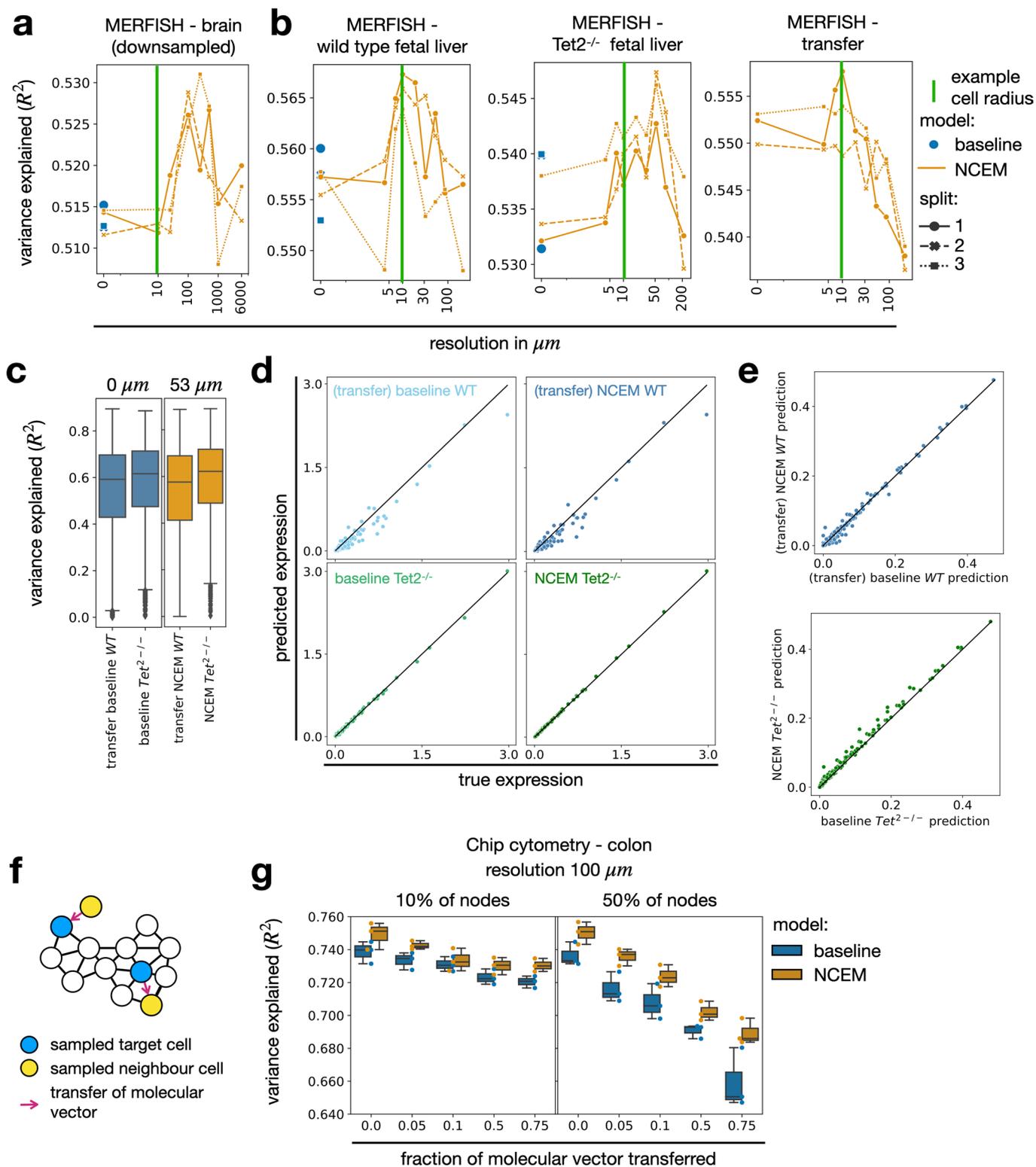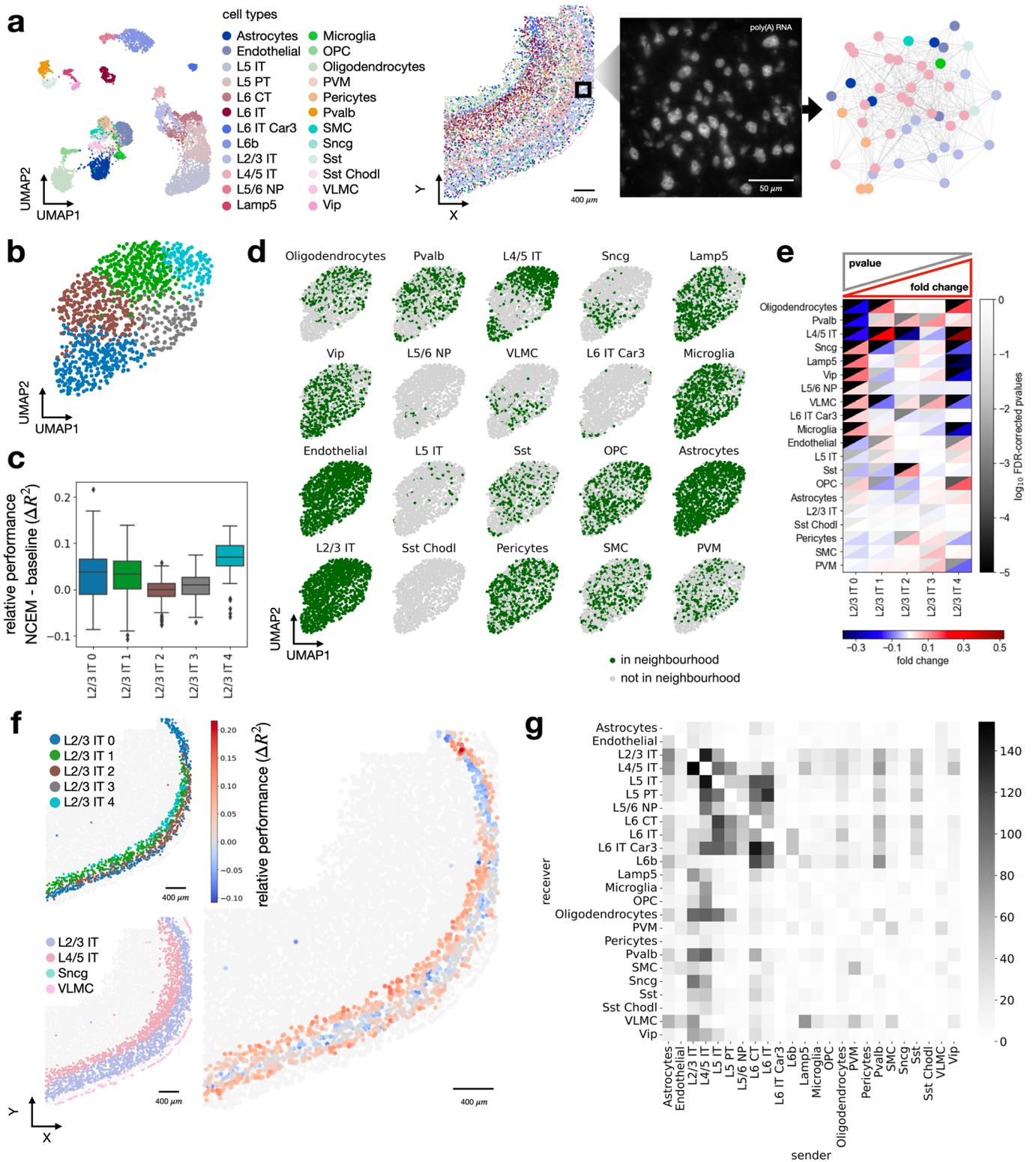
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | NCEM robustness with respect to data perturbation. (a) Robustness of length scales of spatial dependencies to data down-sampling. Shown are the R^2 values between predicted expression vectors and observed expression vectors for held-out test cells of linear models by resolution in μm with cross validation indicated as point shape and line style, showing the relative performance of NCEM model and baseline model. The downsampling was performed on the full set of images. **(b)** Out-of-domain generalization of NCEM fits across genotypes. We translated a linear NCEM fit on wild type data to *Tet2-/-* knockout data and found a similar spatial dependency structure as in a linear NCEM fit on the knockout data alone. **(c)** Comparison of baseline and optimal resolution linear NCEM fits between the transferred wild type model and the knockout fit on knockout test data, based on R2 values between predicted expression vectors and observed expression vectors (n = 4,740 cells). **(d)** Comparison of true and predicted gene-wise mean expression for different models on knockout evaluation data. **(e)** Comparison of R2 values attained by NCEM and baseline model in transfer and in within-domain prediction task. **(f)** Concept of simulation of segmentation errors. Cells from a measured spatial graph are sampled at random and a fraction of their molecular counts is transferred to neighboring cells, simulating a misplacement of a segmentation boundary between both cells (Methods). **(g)** Robustness in terms of segmentation errors for baseline and NCEM linear model on the chip cytometry - colon dataset for 10% and 50% of all nodes in the dataset and different strengths of augmentation (n = 3 cross-validation splits). For each box, the centerline defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker.
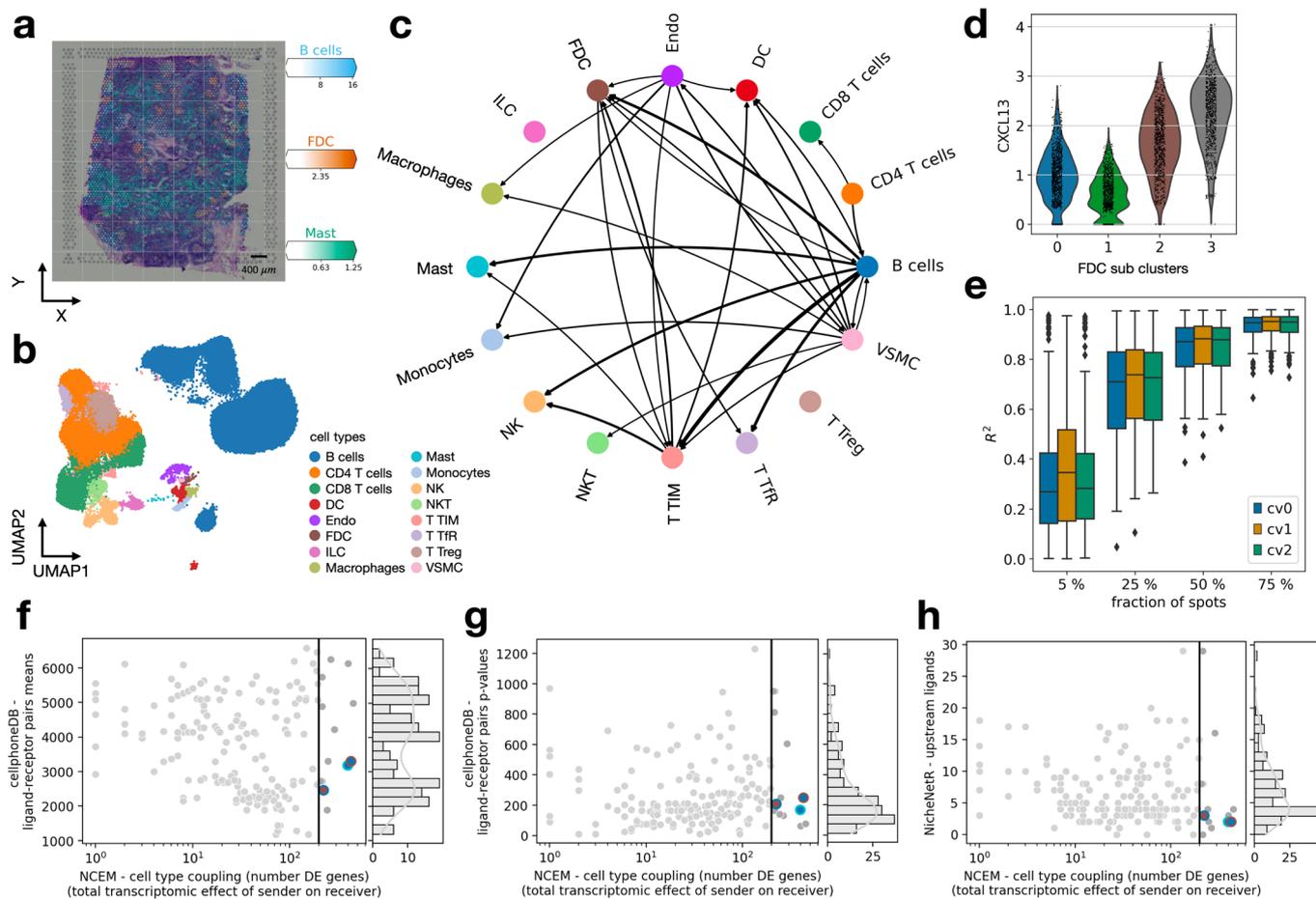
Extended Data Fig. 4 | Robustness of cell communication inference on deconvoluted spot transcriptomics data. (a) 10x Visium slide of a lymph node with the spot-wise abundance of B cells, follicular dendritic cells (FDC), and mast cells inferred with cell2location superimposed. **(b)** UMAP of cells in a matched scRNA-seq data set of human lymph nodes, spleen and tonsils with cell types superimposed. **(c)** Type coupling heatmap of the Visium – lymph node dataset, with edge width proportional to the number of differentially expressed genes at a false-discovery-rate-corrected p-value threshold of 0.05 for each pair of sender and receiver cell types. Only edges with at least 200 genes are shown. **(d)** Violin plot of *Cxcl13* expression per cell for FDC subclusters in the Visium - lymph node data. **(e)** Robustness of type coupling analysis in a Visium slide on human lymph nodes. Shown are R^2 between inferred cell type coupling vectors of randomly subsampled spots and the complete data for subsampling ratios of 5 %, 25 %, 50% and 75 % in three cross validations (n = 256 type couplings in each boxplot). For each box, the centerline defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker. **(f-h)** Correlation of measures of cell communication events between pairs of cell types compared with type coupling scores from NCEM on the tabular sapiens lymph node dataset. Shown are CellphoneDB permutation test results with the number of ligand-receptor pairs with positive mean expression **(f)**, number of ligand-receptor pairs with a FDR-corrected p-value below a threshold of 0.05 **(g)** and the number of ligands associated with a pair of cell types as identified by NicheNet **(h)** (Methods). Each point is one pair of cell types. The vertical line indicates the threshold for showing edges in Fig. 2b.

a

cell types
- Astrocytes
- Endothelial
- L5 IT
- L5 PT
- L6 CT
- L6 IT
- L6 IT Car3
- L6b
- L2/3 IT
- L4/5 IT
- L5/6 NP
- Lamp5
- Microglia
- OPC
- Oligodendrocytes
- PVM
- Pericytes
- Pvalb
- SMC
- Sncg
- Sst
- Sst Chodl
- VLMC
- Vip

poly(A) RNA

50 μm

400 μm

b

c

d

Oligodendrocytes   Pvalb   L4/5 IT   Sncg   Lamp5

Vip   L5/6 NP   VLMC   L6 IT Car3   Microglia

Endothelial   L5 IT   Sst   OPC   Astrocytes

L2/3 IT   Sst Chodl   Pericytes   SMC   PVM

- in neighbourhood
- not in neighbourhood

e

pvalue

fold change

log$_{10}$ FDR-corrected pvalues

fold change

f

- L2/3 IT 0
- L2/3 IT 1
- L2/3 IT 2
- L2/3 IT 3
- L2/3 IT 4

relative performance (ΔR^2)

- L2/3 IT
- L4/5 IT
- Sncg
- VLMC

400 μm

g

receiver

sender

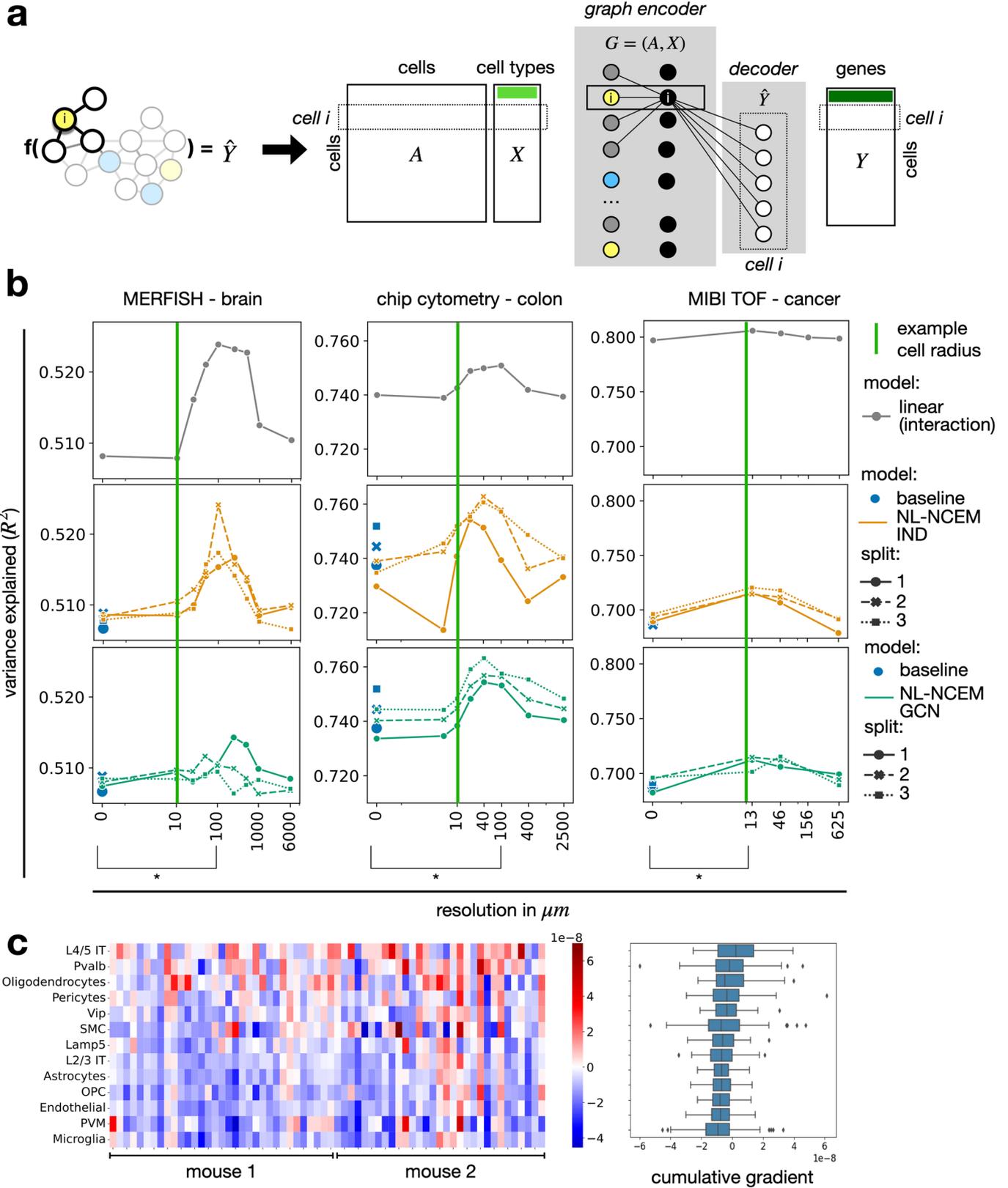**Extended Data Fig. 5 | See next page for caption.**

Extended Data Fig. 5 | Cell heterogeneity can be attributed to niche composition. (a) Spatial cell type distribution in the mouse brain. Shown are a UMAP of molecular embedding of all cells in slide 153 (n = 7439 cells) with the cell type superimposed, followed by slice 153 of mouse brain in the MERFISH – brain dataset with the spatial allocation of all cell types superimposed, field of view number 486 of the same slice with poly(A) RNA channel superimposed at central z-plane (z = 4.5 μm), and the spatial proximity graph of the same field of view with a resolution of 100 μm. **(b)** UMAPs of molecular embedding of L2/3 IT cells with molecular subclustering superimposed (colors as in b)**. (c)** Distribution of cell-wise difference of R^2 between NCEM and non-spatial baseline model by molecular subcluster (L2/3 IT 0: n = 316, L2/3 IT 1: n = 314, L2/3 IT 2: n = 313, L2/3 IT 3: n = 133, L2/3 IT 4: n = 128). The centerline of the boxplots defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker. **(d)** UMAPs of molecular embedding of all L2/3 IT cells in an example image (n = 1204 cells) showing if a given cell-type is present in the neighborhood. The underlying neighborhoods were defined at the optimal resolution defined in Fig. 1d (100 μm). **(e)** Heatmap of fold change and false-discovery rate corrected p-values of cluster enrichment of binary neighborhood labels, where fold changes are the ratio between the relative neighboring source cell-type frequencies per subtype cluster and the overall source cell-type frequency in the image. **(f)** Model performance on L2/3 IT cells in space on slice 153 of mouse brain in the MERFISH – brain dataset with L2/3 IT sub-states (first panel), L2/3 IT, L4/5 IT, Sncg, and VLMC (second panel) and the difference of R2 between the NCEM at resolution of 100 μm and the best nonspatial baseline model (third panel) superimposed. **(g)** Type coupling analysis of MERFISH – brain data, showing the number of differentially expressed genes at a false-discovery-rate-corrected p-value threshold of 0.05 for each pair of sender and receiver cell types.
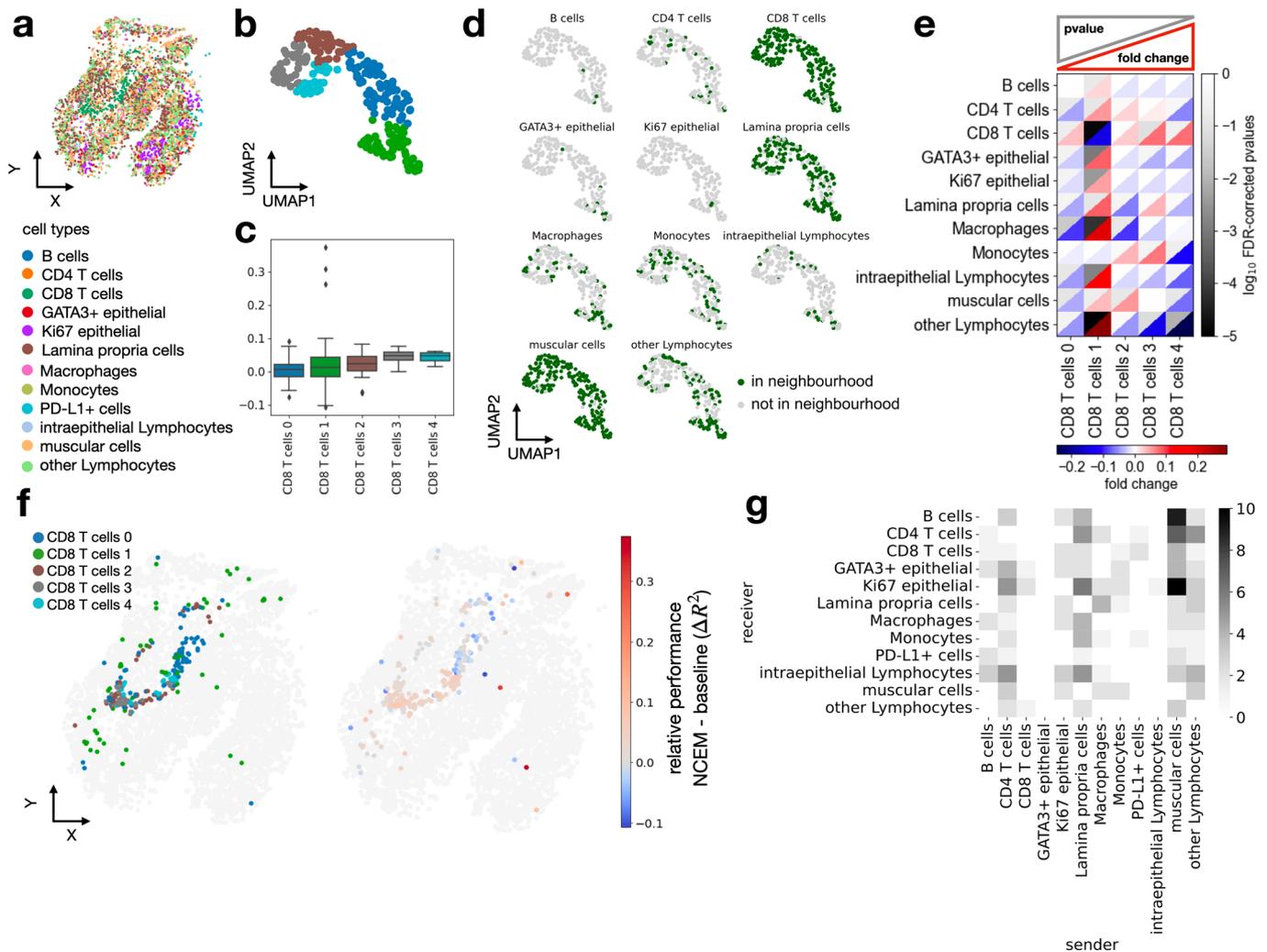
Extended Data Fig. 6 | Attributing cell heterogeneity to niche composition in inflamed colon. (a) Area 1 of chip cytometry – colon dataset with cell-types superimposed. **(b)** UMAPs of molecular embedding of CD8 T cells only with molecular subclustering superimposed (colors as in c)**. (c)** Distribution of cell-wise difference of R^2 between spatial model non nonspatial baseline model by molecular sub-cluster (CD8 T cells 0: n = 74, CD8 T cells 1: n = 58, CD8 T cells 2: n = 41, CD8 T cells 3: n = 37, CD8 T cells 4: n = 24). The centerline of the boxplots defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker. **(d)** UMAPs of molecular embedding of all CD8 T cells in area 1 (n = 234 cells) showing if a given cell-type is present in the neighborhood. The underlying neighborhoods were defined at the best performing resolution identified in Fig. 1c (40 μm). **(e)** Heatmap of fold change and false-discovery rate corrected p-values of cluster enrichment of binary neighborhood labels, where fold changes are the ratio between the relative neighboring source cell-type frequencies per subtype cluster and the overall source cell-type frequency in the image. **(f)** Area 1 of colon in the chip cytometry – colon dataset with CD8 T cell sub-states (left) and the difference of R^2 between the NCEM interaction model at resolution of 40 μm and the best nonspatial baseline model (right). **(g)** Type coupling analysis, showing the number of differentially expressed genes at a false-discovery-rate-corrected p-value threshold of 0.05 for each pair of sender and receiver cell types.
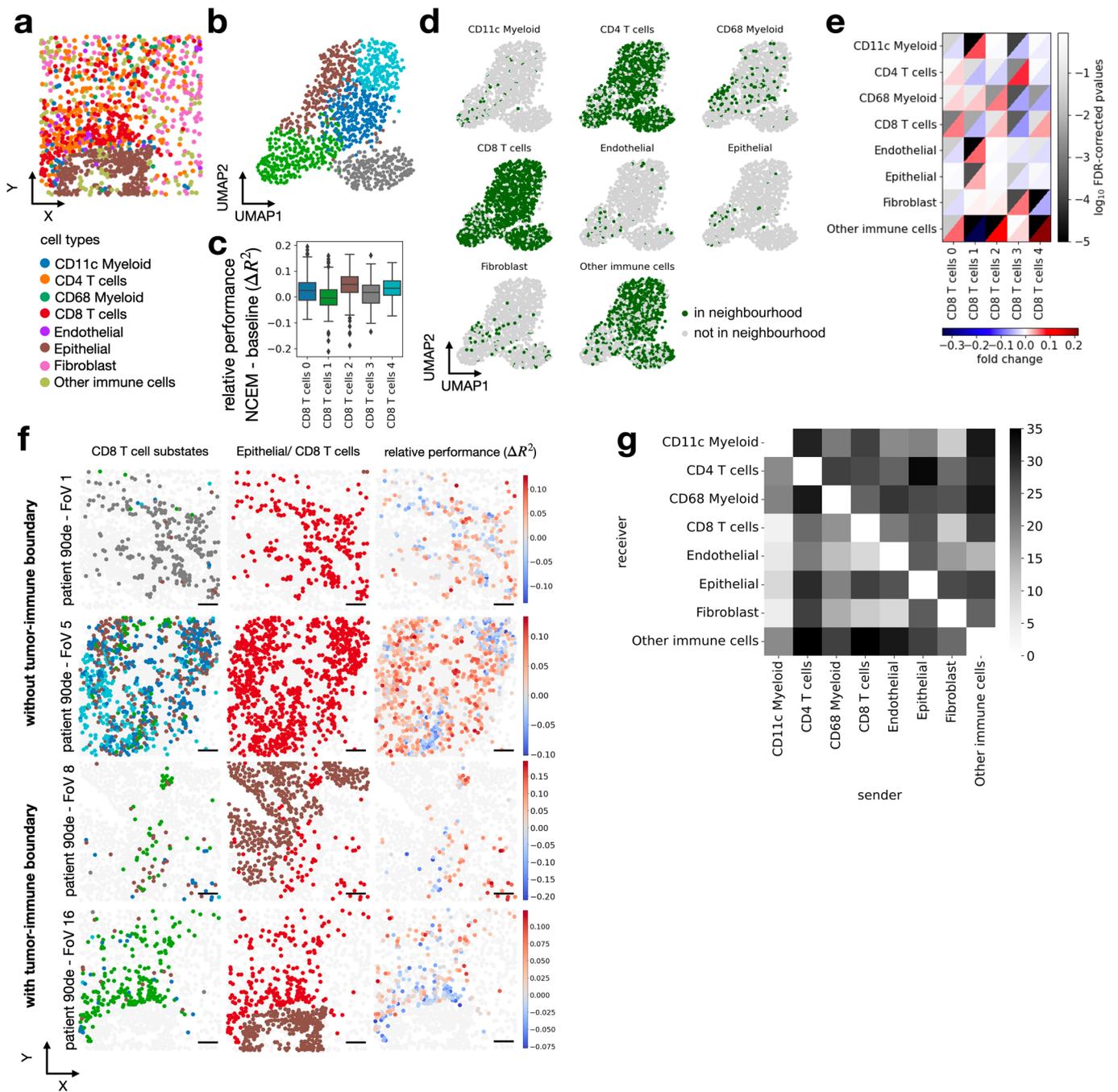
Extended Data Fig. 7 | Attributing cell heterogeneity to niche composition in colorectal cancer. (a) Field of view 16 of MIBI TOF – cancer dataset with the spatial allocation of all cell-types superimposed. **(b)** UMAPs of molecular embedding of CD8 T cells only with molecular sub-clustering superimposed (colors as in c). **(c)** Distribution of cell-wise difference of R^2 between spatial model non-spatial baseline model by molecular sub-cluster (CD8 T cells 0: n=304, CD8 T cells 1: n=293, CD8 T cells 2: n=278, CD8 T cells 3: n=247, CD8 T cells 4: n=207). The centerline of the boxplots defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker. **(d)** UMAPs of molecular embedding of all CD8 T cells in area 1 (n=1,329 cells) showing if a given cell type is present in the neighborhood. The underlying neighborhoods were defined at the optimal resolution identified in Fig. 1d (13 μm). **(e)** Heatmap of fold change and false-discovery rate corrected p-values of cluster enrichment of binary neighborhood labels, where fold changes are the ratio between the relative neighboring source cell-type frequencies per subtype cluster and the overall source cell-type frequency in the image. **(f)** Field of view 1, 5, 8 and 16 of colon in the MIBI TOF – cancer dataset with CD8 T cell sub-states, cell type assignments to epithelial and T cells, and the difference of R^2 between the NCEM interaction model at a resolution of 13 μm and the best nonspatial baseline model (scale bar 50 μm). **(g)** Type coupling analysis, showing the number of differentially expressed genes at a false-discovery-rate-corrected p-value threshold of 0.05 for each pair of sender and receiver cell types.

a

b

c

Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Nonlinear models of spatial dependencies of expression states. (**a**) A node-supervised model in which the label is the expression vector of a cell and the input consists of categorical cell type assignments and a spatial proximity graph. This model can also be viewed as a nonlinear regression model: a local graph embedding of each cell is reconstructed to a cell-wise expression state. The forward pass for a cell i is shown. (**b**) Inferred nonlinear spatial dependencies. Shown are the R^2 values for held-out test data of nonlinear models by resolution in µm with cross validation indicated as point shape and line style and comparatively mean performance of linear model in Fig. 1d. *Linear (interaction) (gray line)*: linear model with interaction effects; *NL*: nonlinear model; *IND*: the graph kernel is an indicator function across cell types in the neighborhood (yellow lines); *GCN*: the graph kernel is a graph convolution, a linear embedding of the cell types in the neighborhood (teal lines); *split (point shapes)*: cross-validation split; bracket (*): significant difference in paired t-test between baseline model and best spatial model with (MERFISH – brain dataset $p_{IND} = 0.033$, chip cytometry – colon dataset $p_{GCN} = 0.026$, MIBI-TOF – cancer dataset $p_{IND} = 0.005$ and $p_{GCN} = 0.036$). (**c**) Heatmap of cumulative gradients (saliency) of gene expression prediction of L2/3 IT with respect to the input cells, aggregated by the sender cell type clusters, on test data. Shown is a cumulative gradient matrix of L2/3 IT predictions by source cell type and image (n = 64 images). The cumulative absolute gradients are derived from the absolute gradients tensor across each cell's molecular vector prediction with respect to the cells in the neighborhood (source cells) per image, by taking a sum across the molecular output features and by taking a sum across source cells of the same type. We aggregated these saliency maps per sender-receiver cell type pair as $SALS \in R^{L*L}$, where L is the number of distinct cell types in the model. Non-normalized saliencies will show a pattern similar to the contact frequency matrix as cell types with frequent connections will skew the learned importance of cell connections. Therefore, we normalized the saliencies by the absolute number n_{ab} of occurrences of each cell type pair: $SALS^{norm}_{ab} = \frac{1}{n_{ab}} SALS_{ab}$. For each box in (**c**), the centerline defines the median, the width of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker.
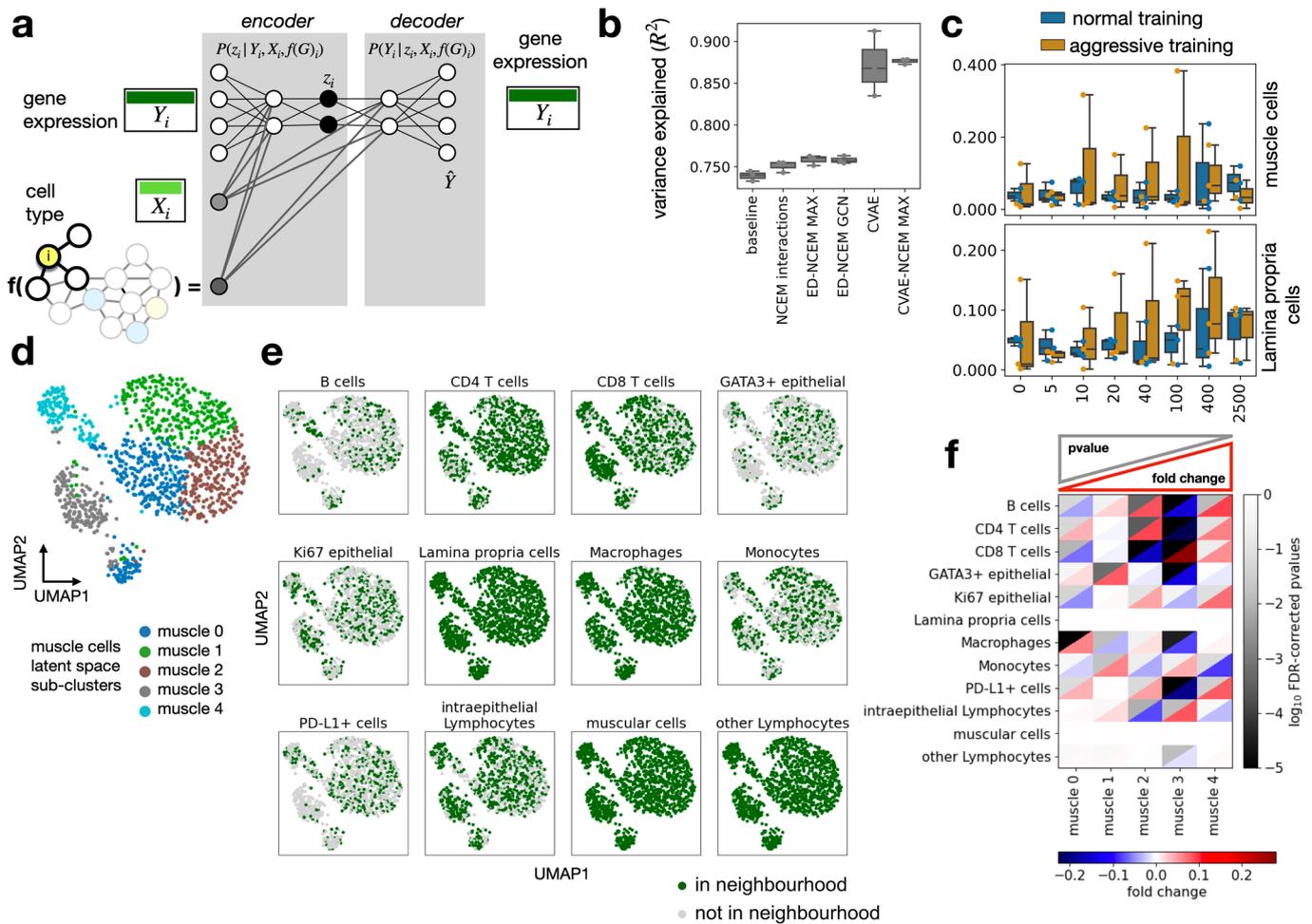
Extended Data Fig. 9 | Modeling intrinsic and extrinsic variation in deep latent variable models. (a) A node generative network (CVAE–NCEM) is a conditional variational autoencoder in which the condition is not a constant but a graph embedding, which is also learned. The forward pass for a cell i through the model is shown. **(b)** Latent variable models improve reconstructive performance. Shown are the R^2 values of held-out test data based on the forward pass model evaluation from chip cytometry – colon data for linear models, encoder–decoder models, and variational autoencoders for both NCEM and nonspatial models (n=3 cross-validation splits). *baseline*: a nonspatial linear model of gene expression per cell-type; *NCEM interactions*: linear model with interaction effects; *NL*: nonlinear model; *IND*: the graph kernel is an indicator function across cell-types in the neighborhood; *GCN*: the graph kernel is a graph convolution, a linear embedding of the cell-types in the neighborhood. **(c)** Neighborhood transfer performance of NCEM and nonspatial models. Shown is the R^2 over cells in the test set for models trained on predicting muscular cells and Lamina propria cells for both CVAE and CVAE–NCEMs trained on neighborhoods with different radii with optimization algorithm as color (n=3 cross-validation splits). *Plain*: normal CVAE training; *aggressive*: aggressive encoder training. For each box in **(b, c)**, the centerline defines the median, the height of the box is given by the interquartile range (IQR), the whiskers are given by 1.5 * IQR and outliers are given as points beyond the minimum or maximum whisker. **(d–f)** Latent variables of CVAE–NCEM are confounded with neighborhood conditions. **(d)** UMAP of molecular embedding in the CVAE–NCEM IND latent space of muscular cells in an example image (n = 1,149 cells) with molecular sub-clustering superimposed (muscle 0: n = 315, muscle 1: n = 287, muscle 2: n = 238, muscle 3: n = 183, muscle 4: n = 126). **(e)** UMAPs of molecular embedding in the CVAE–NCEM IND latent space of all muscle cells in the same image with superimposed binary label of presence of a given cell-type, as defined in the title, in the neighborhood. The underlying neighborhoods were defined at a resolution of 100 µm. **(f)** Heatmap of fold change and false-discovery corrected p-values of cluster enrichment of binary neighborhood labels, where fold changes are the ratio between the relative neighboring source cell-type frequencies per subtype cluster and the overall source cell-type frequency in the image.
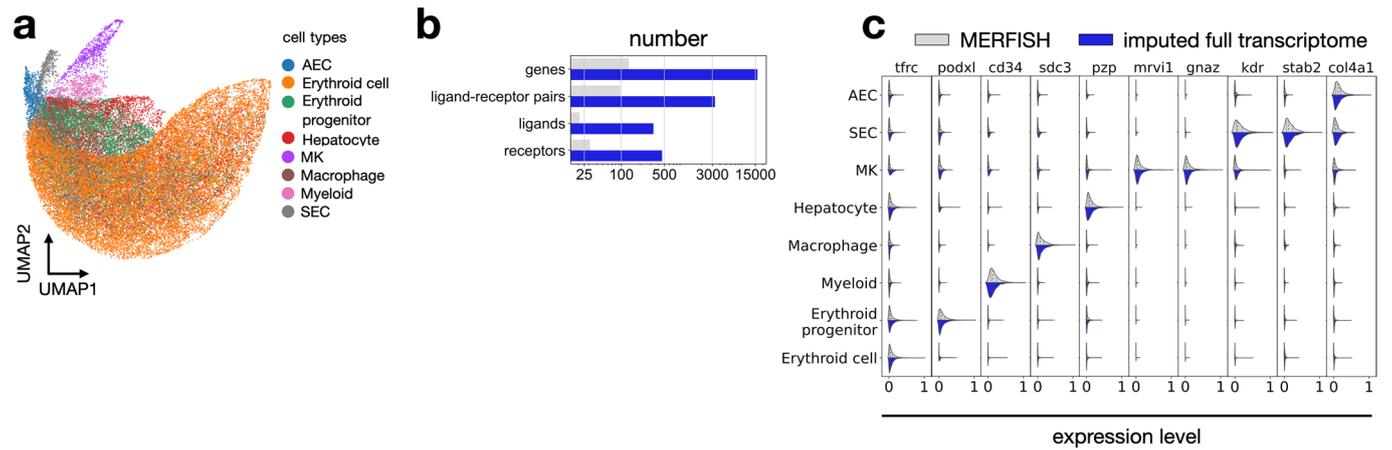
Extended Data Fig. 10 | Modeling ligand–receptor signaling with NCEM. (a) UMAP of cells in MERFISH – fetal liver data. **(b)** Imputation of MERFISH data with scRNA-seq increases the number of genes that can be modeled with NCEMs, including receptor genes, ligand genes, and ligand–receptor pairs. **(c)** Distribution of selected marker genes that are both observed in scRNA-seq and in MERFISH over cell types.

nature research

Corresponding author(s): Fabian J. Theis

Last updated by author(s): Jul 29, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NA
Data analysis	python==3.8, scanpy>=1.72, tensorflow>=2.5.0, squidpy>=1.1.2, patsy>=0.5.1, scipy>=1.6.3, matplotlib>=3.4.2, louvain>=0.7.0, diffxpy>=0.7.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The MERFISH – fetal liver, MERFISH – brain, MIBI TOF – cancer, MELC – tonsils, CODEX – cancer, chip cytometry – colon, and the Visium – lymph node datasets are publicly available (Online Methods).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No samples were collected.
Data exclusions	No data was excluded from the public data sets beyond what is described in the Online Methods.
Replication	We performed our analyses on 5 public data sets.
Randomization	No randomization was performed as this was not relevant to this study.
Blinding	No blinding was performed as this was not relevant to this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging