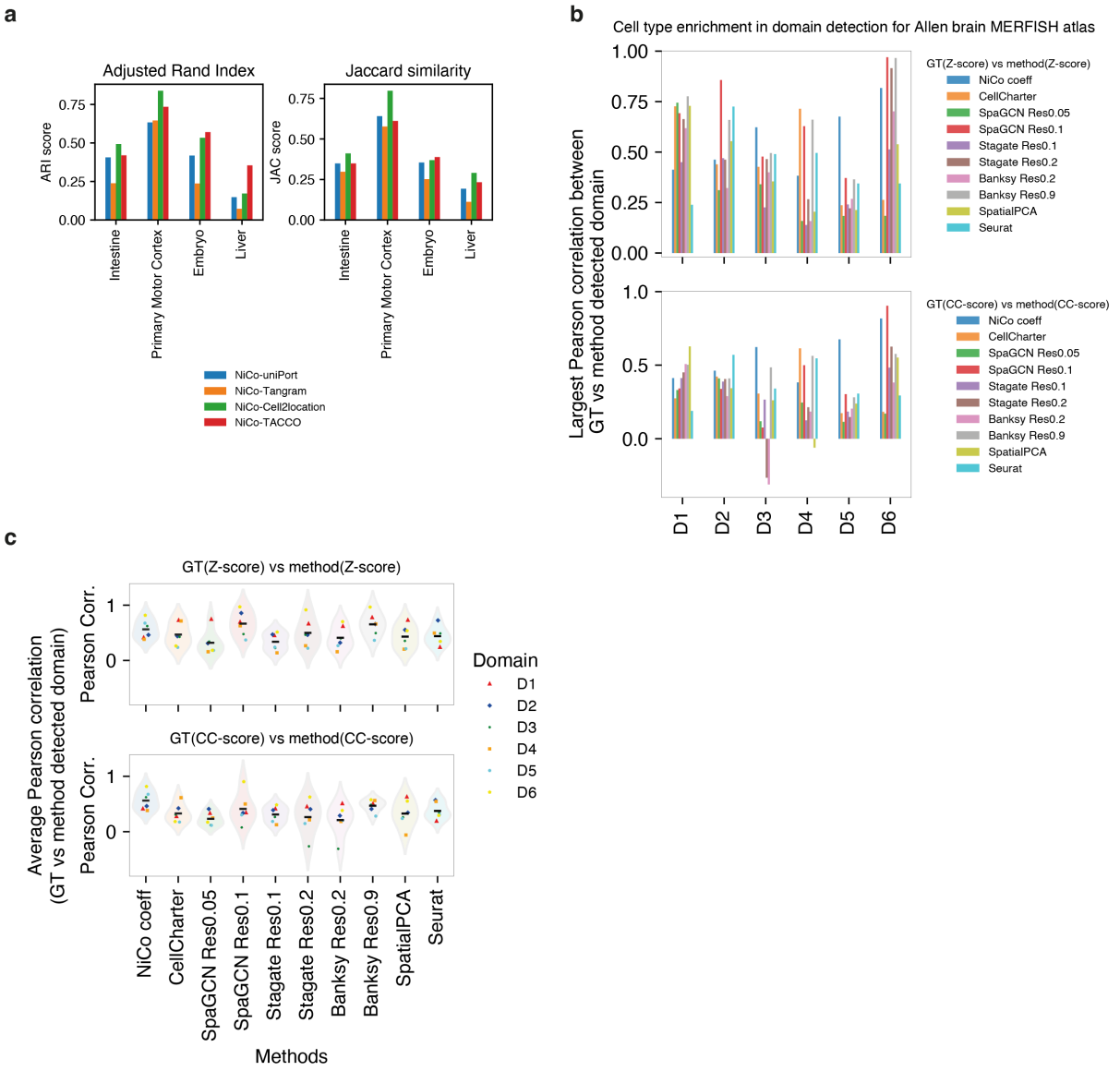
# Supplementary Information

## NiCo Identifies Extrinsic Drivers of Cell State Modulation by Niche Covariation Analysis

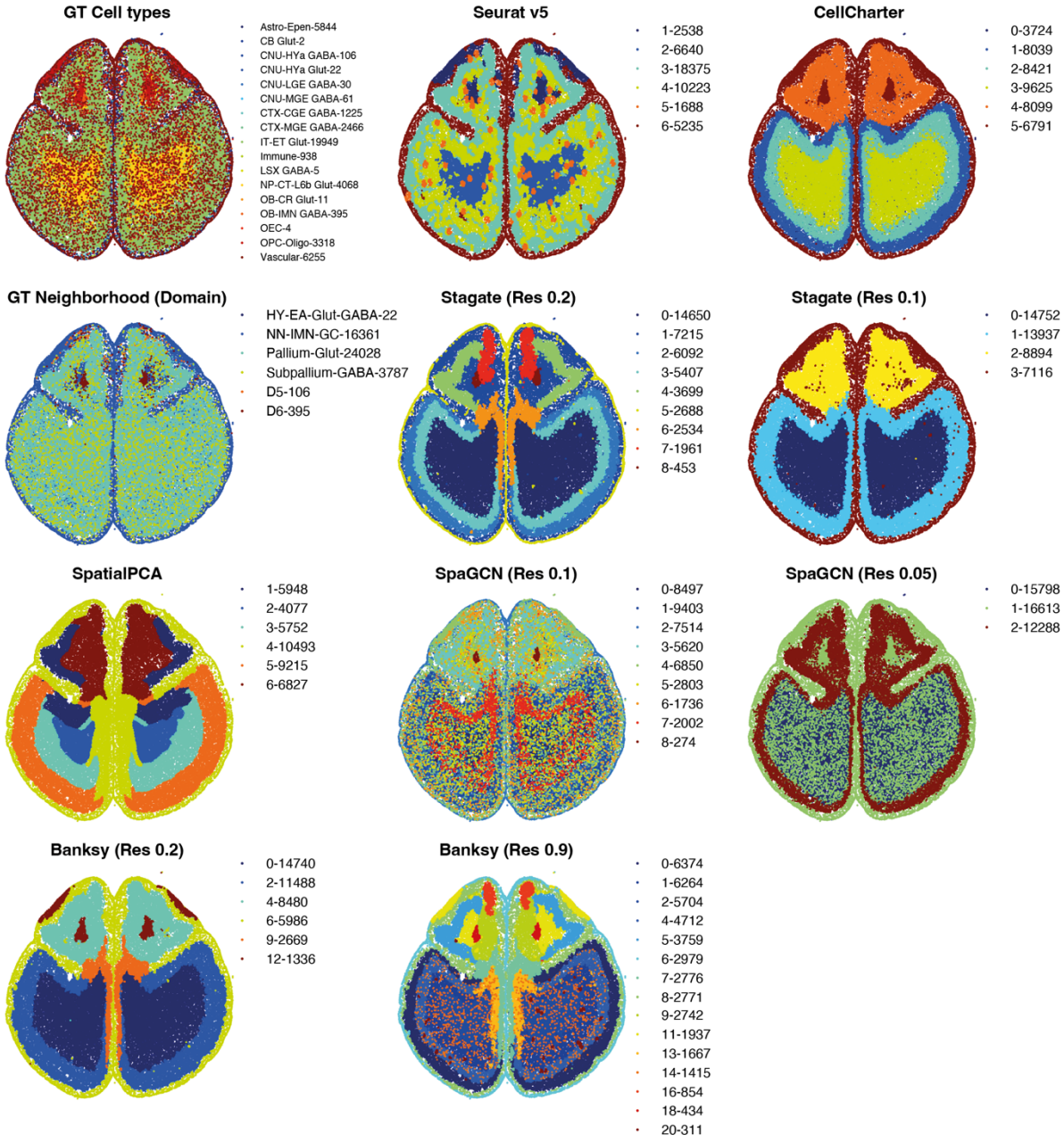
Ankit Agrawal, Stefan Thomann, Sukanya Basu, Dominic Grün

**Supplementary Figures 1-11**

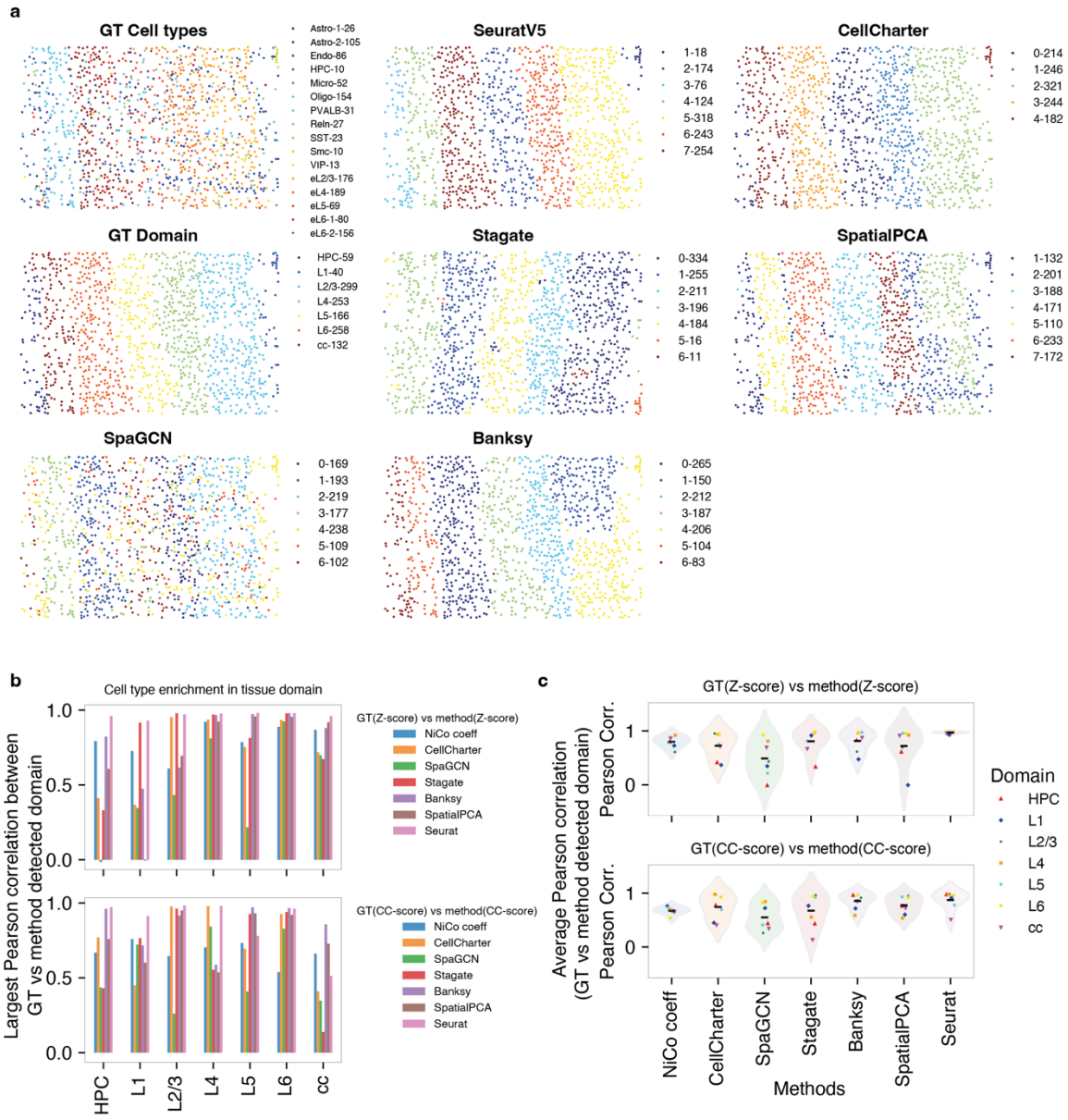
**Supplementary Table 1**



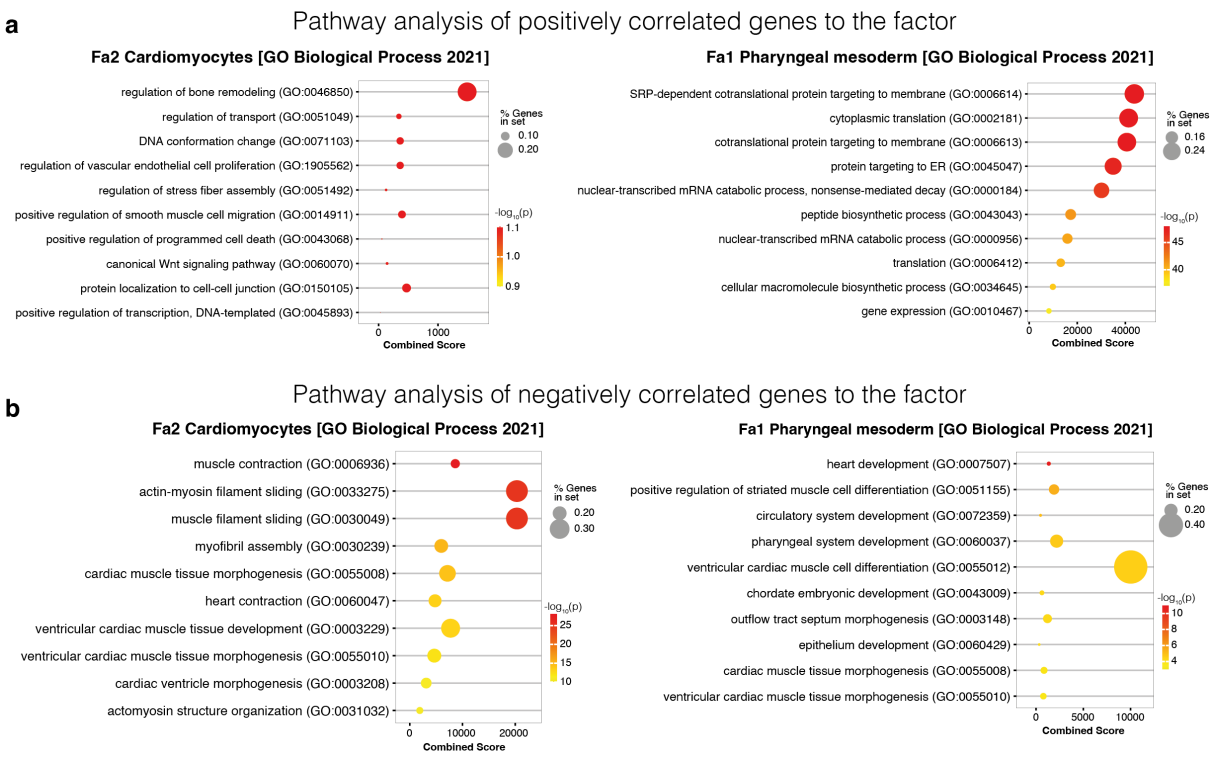

**Supplementary Figure 1| Benchmarking of tissue domain detection on Allen brain MERFISH atlas data and cell type annotation**. **a**, Comparison of NiCo annotations with uniPort, Tangram, TACCO, and cell2location for published datasets (mouse intestine and primary motor cortex MERFISH data, mouse embryo seqFISH data, and liver MERSCOPE data). Consistency of annotations between two methods was evaluated by Jaccard similarity (JAC) and adjusted rand index (ARI). Source data are provided as Source Data files (see Data Availability). (**b**, **c**), Cell type enrichment in ground truths (GT) domains was computed as enrichment Z-score (top) or CellCharter (CC) enrichment score (bottom). The Pearson correlation of this score is shown for the predicted domain with the highest correlation to the ground truth domain. For NiCo, the cell type with the highest correlation of its regression coefficients to the Z-score/CC-score was selected. Data are represented as barplots across domains (b) or violin plots across methods (c). The domain names are D1 (HY-EA-Glut-GABA), D2 (NN-IMN-GC), D3 (Pallium-Glut), D4 (Subpallium-GABA), D5 (D4; D1), D6 (D4; D2). Black line in (b), average correlation.



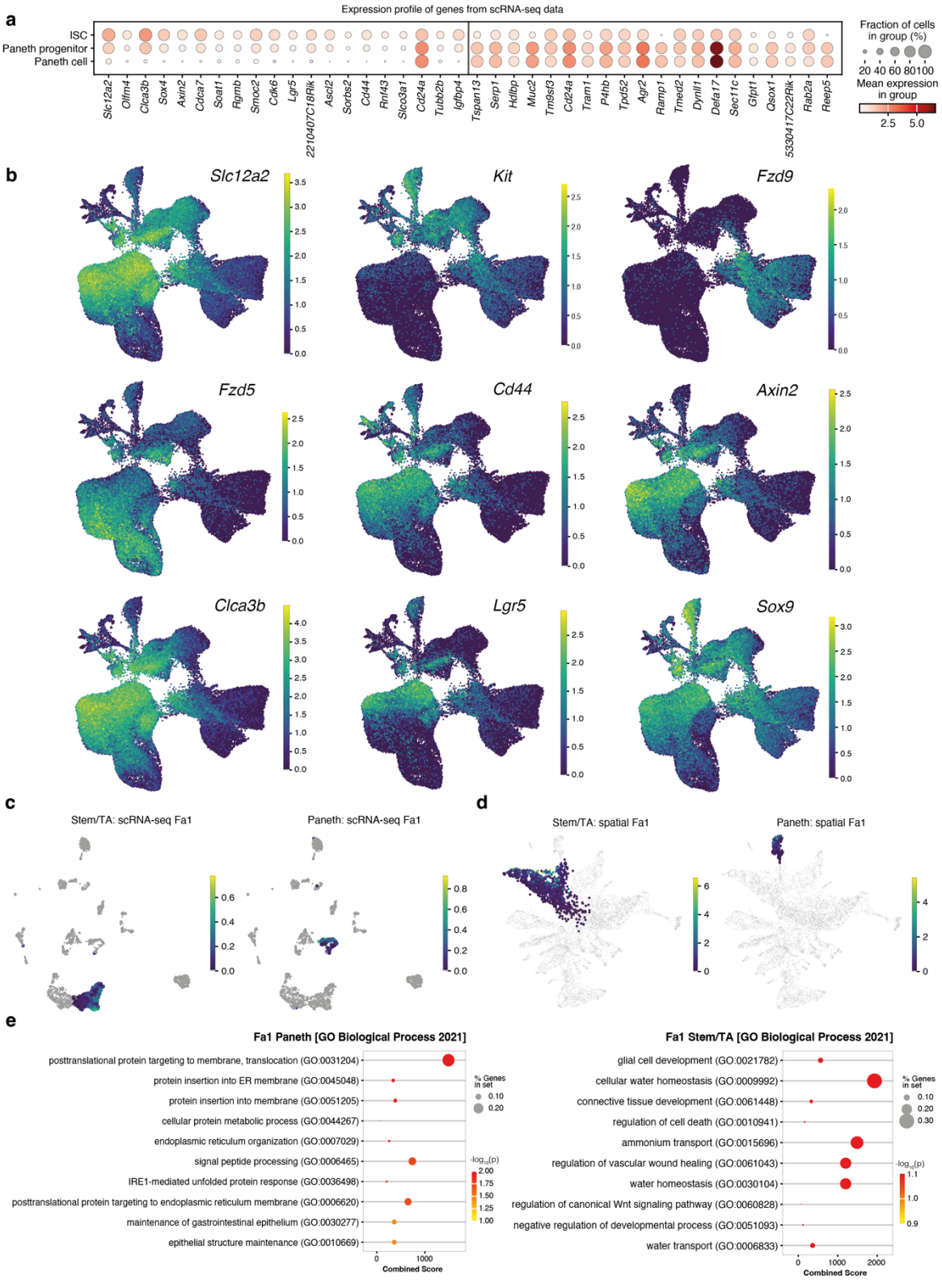

**Supplementary Figure 2| Tissue domain detection in Allen Brain MERFISH Atlas data by different methods**. The spatial maps highlight topological domains on a brain section according to the authors' ground truth (GT) annotation, and predictions obtained by tailored niche detection methods as indicated. Some methods were run with different resolution parameters as indicated to match the number of GT domains (Methods).




**Supplementary Figure 3| Tissue domain detection on STARmap visual cortex data. a**, Predicted niche domains are highlighted in spatial maps of the STARmap visual cortex data for GT annotation, Seurat V5, CellCharter, Stagate, SpatialPCA, SpaGCN, and Banksy. Domain names are defined as L1, L2/3, L4, L5, L6, HPC (hippocampus) and cc (corpus callosum). (**b**, **c**), Cell type enrichment in GT domains was computed as enrichment Z-score (top) or CellCharter (CC) enrichment score (bottom). The Pearson correlation of this score is shown for the predicted domain with the highest correlation to the ground truth domain. For NiCo, the cell type with the highest correlation of its regression coefficients to the Z-score/CC-score was selected. Data are represented as barplots across domains (b) or violin plots across methods (c). Black line in (c), average correlation. Source data are provided as Source Data files (see Data Availability).

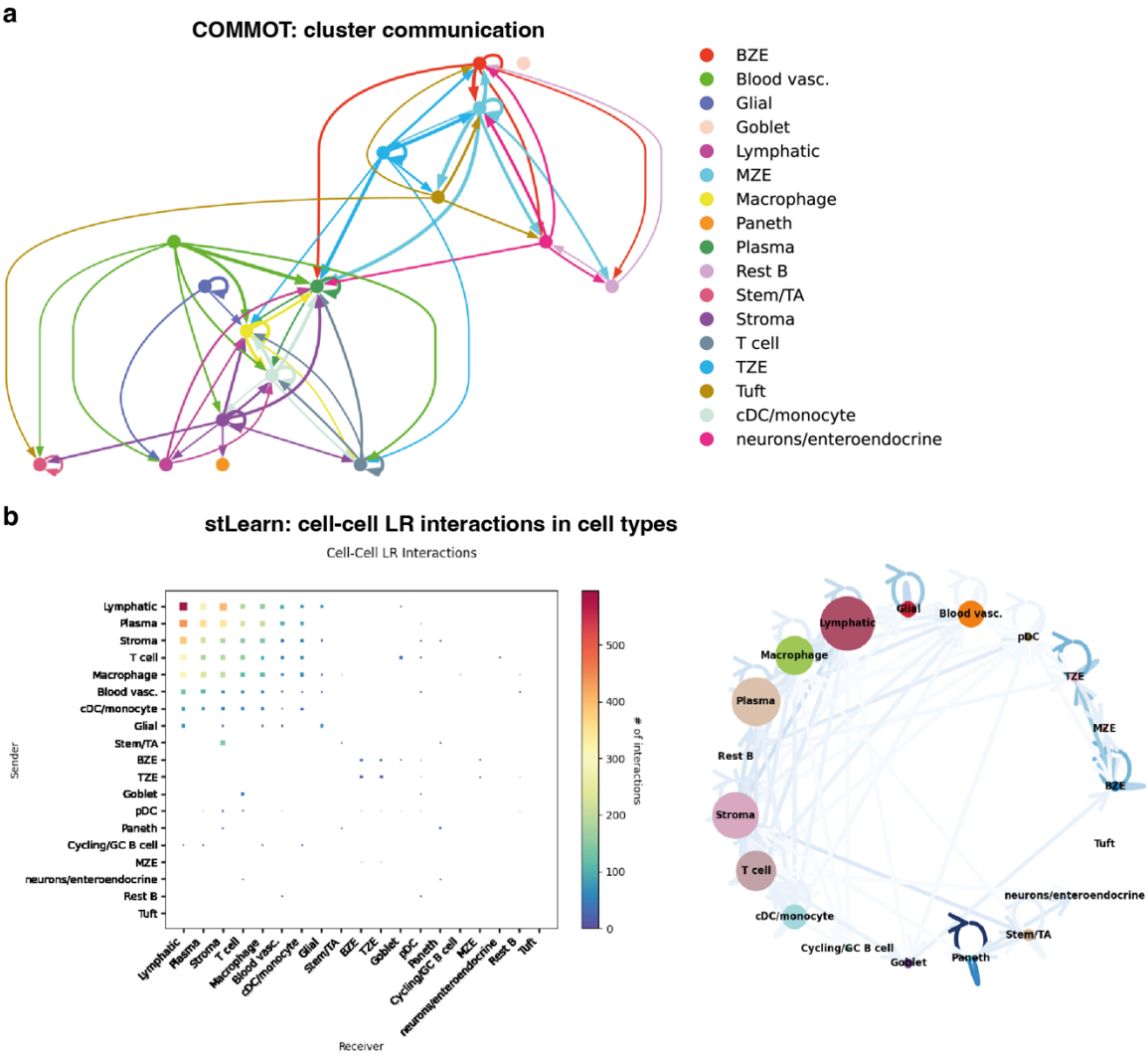



**Supplementary Figure 4| Pathway analysis of factors associated with cardiomyocytes and pharyngeal mesoderm in mouse embryonic data. a,** The top 50 positively correlated genes, identified by Spearman correlation with cardiomyocyte Fa2 (left) and pharyngeal mesoderm Fa1 (right) were analyzed for Gene Ontology (GO) Biological Process enrichment. **b,** The top 50 negatively correlated genes, identified by Spearman correlation with cardiomyocyte Fa2 (left) and pharyngeal mesoderm Fa1 (right) were analyzed for Gene Ontology (GO) Biological Process enrichment. See Methods for details.

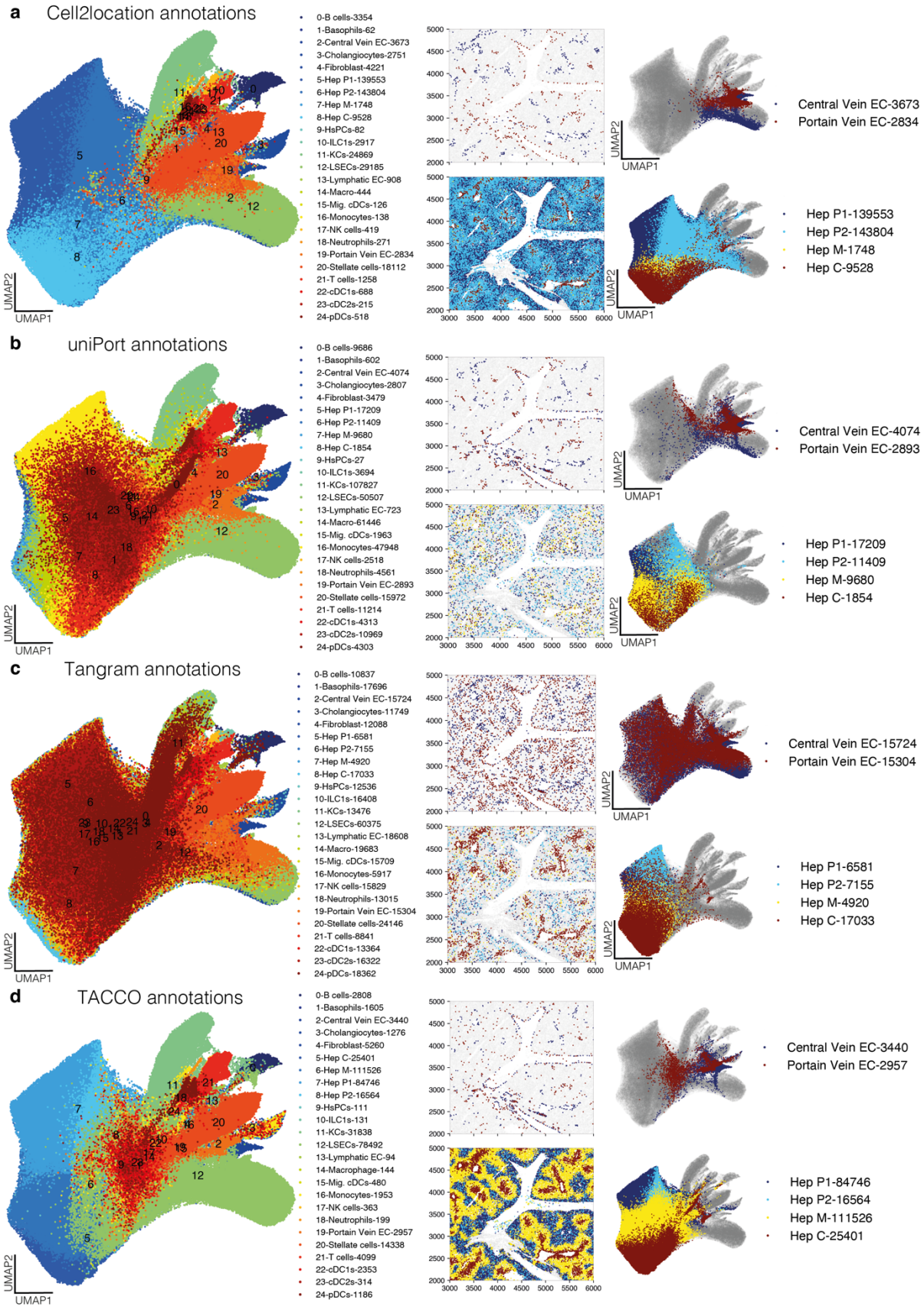



**Supplementary Figure 5| Pathway analysis and expression of genes associated with stem/TA and Paneth cell covariation in mouse intestine. a,** The top 20 positively correlated genes to Paneth Fa1 (left) and stem/TA Fa1 (right) are displayed as dot plot highlighting the fraction of cells in the respective population expressing a gene (dot size) and the mean expression level (dot color). **b,** Selected genes upregulated in intestinal stem cells and Paneth cell progenitors are highlighted in a UMAP representation of mouse intestinal scRNA-seq data

from Böttcher et al.[53]. **(c-d),** Stem/TA Fa1 and Paneth Fa1 are highlighted in the UMAP representation of the scRNA-seq reference data [54] (c) and the spatial data [31] (d). **e,** The top 50 positively correlated genes, identified by Spearman correlation with Paneth Fa2 (left) and Stem/TA Fa1 (right) were analyzed for Gene Ontology (GO) Biological Process enrichment. See Methods for details.

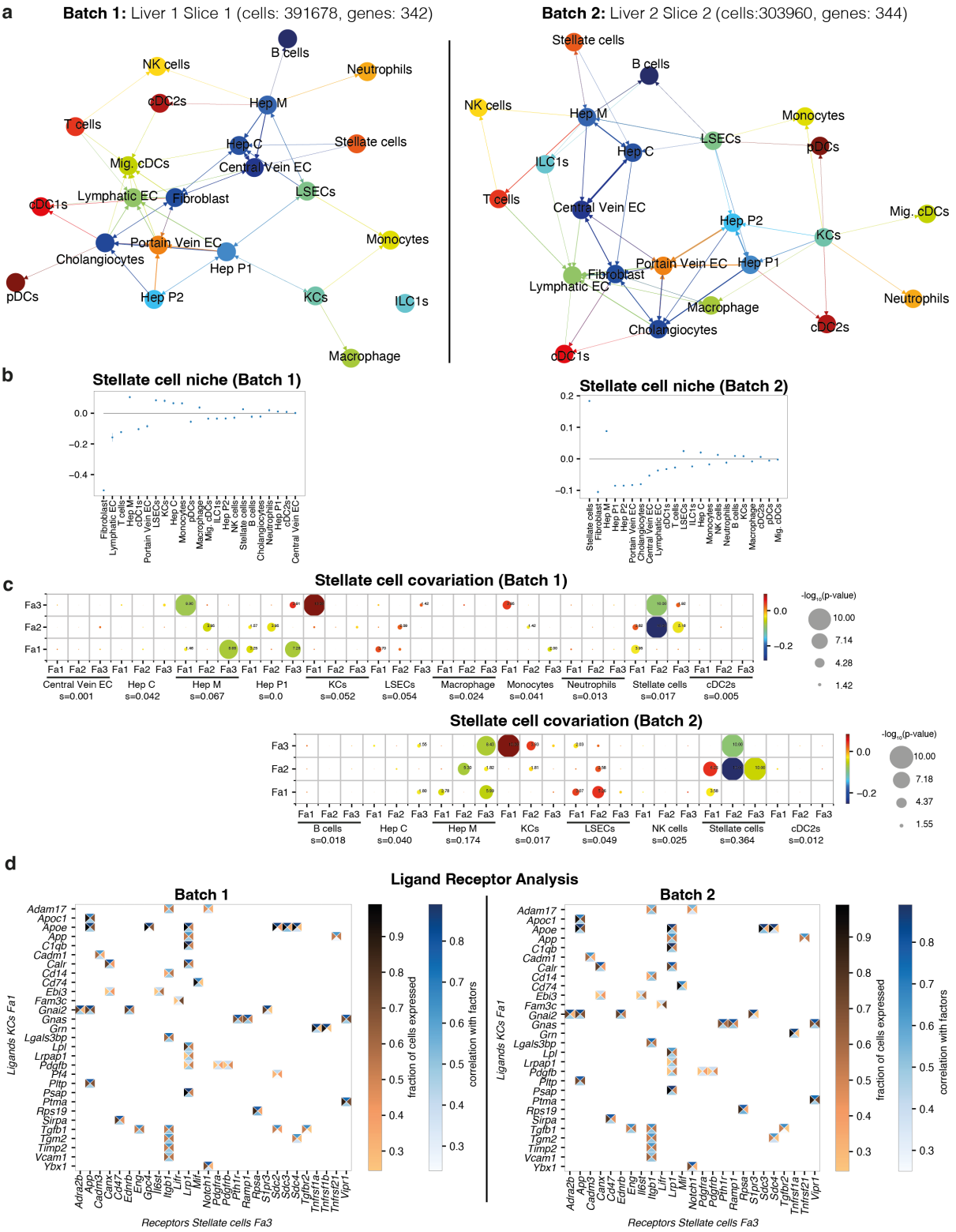



**Supplementary Figure 6| Benchmarking on cell-cell interaction analysis in the intestinal dataset. a**, The cluster communication graph for pathways (total-total) from COMMOT[21]. **b**, Cell type-cell type LR interactions from stLearn[57]. **c**, Contact-dependent genes between stem/TA and Paneth cells from CellNeighborEx[58].



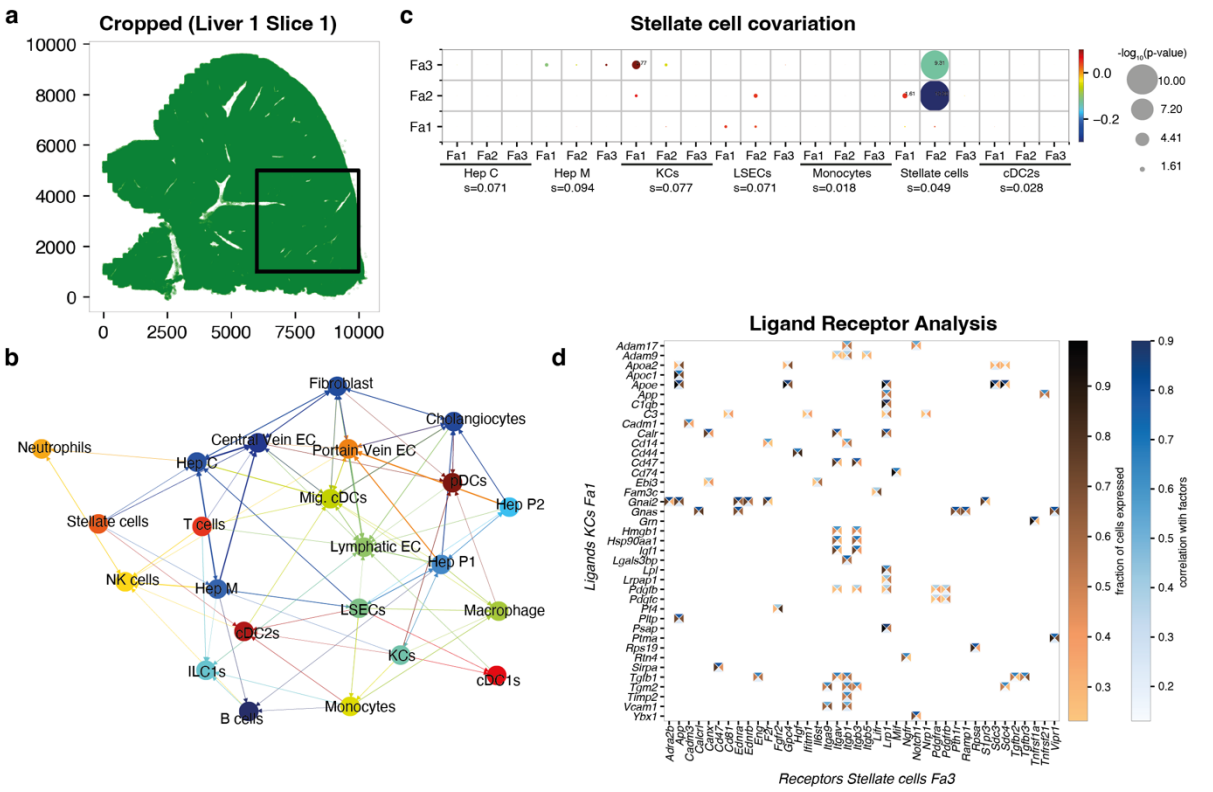

**Supplementary Figure 7| Cell type annotation of spatial mouse liver data by alternative methods. (a-d),** Left: UMAP representation based on gene expression of the spatial data. Cell type annotations are highlighted. The legend indicates the number of cells for each cell type.

Middle: Select cell types are highlighted for a representative tissue area. Top, central vein and portal vein endothelial cells (EC). Bottom, hepatocytes of different zones, i.e., central (Hep C), mid-zonal (Hep M) and portal (Hep P1/P2). Right: Cell types from the middle panel are highlighted in the expression UMAP representation. Results are shown for cell2location (a), uniPort (b), Tangram (c) and TACCO (d).

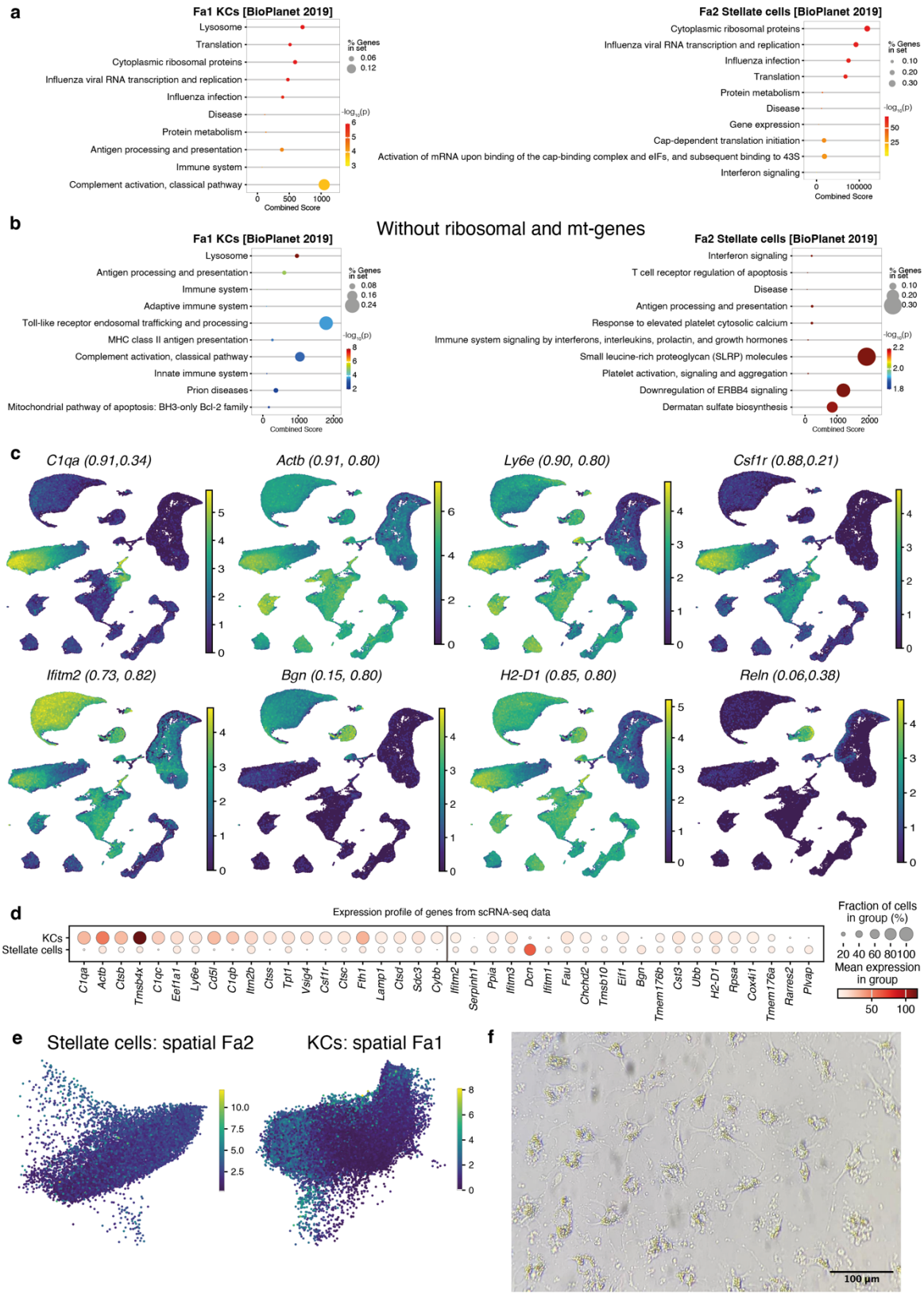



**Supplementary Figure 8| NiCo analysis on two batches of the liver dataset.** NiCo was run on experimental batches from the Vizgen liver dataset, i.e., slice 1 of liver 1 (batch 1) and slice 2 of liver 2 (batch 2). **a**, Cell type interaction map (interaction threshold c= 0.1) for batch 1 (left) and batch 2 (right). **b**, The logistic regression coefficients of the stellate cell niche for batch 1 (left) and batch 2 (right). Error bars, see Methods. **c**, Covariation between stellate cell factors (y-axis) and co-localized neighborhood cell type factors (x-axis) for batch 1 (top) and

batch 2 (bottom). Circle size scales linearly with $-\log_{10}$(p-value), and circle color indicates ridge regression coefficients. S denotes the normalized niche coefficient score. The multivariate regression p-value is derived from two-tailed t-statistics. **d**, Ligand-receptor pairs correlated with stellate cell Fa3 and KC Fa1 for batch 1 (left) and batch 2 (right). See Methods. The rectangle's north and south faces represent ligand and receptor correlation to the factors, while west and east faces represent the proportion of ligand and receptor expressing cells.

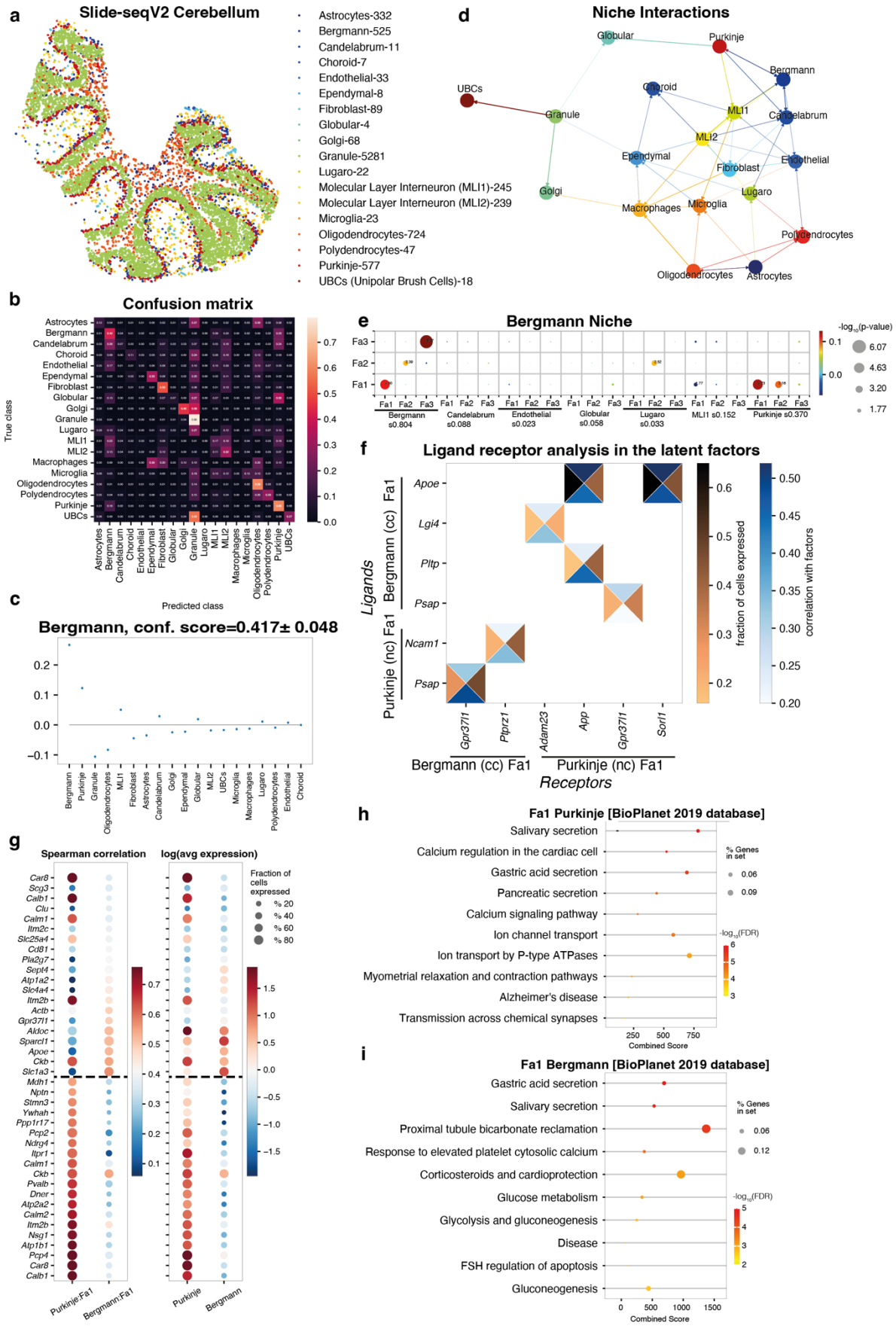



**Supplementary Figure 9| NiCo analysis on cropped region of liver data. a,** Spatial map of slice 1 from the Vizgen liver dataset analyzed in Figure 5. The cropped region analyzed with NiCo is indicated by a black box and contains a subset of 88,772 cells (22.7%). **b**, Cell-type niche interaction map (interaction threshold c=0.1). **c**, Covariation between stellate factors (y-axis) and co-localized neighborhood cell type factors (x-axis). Circle size scales linearly with $-\log_{10}$(p-value), and circle color indicates ridge regression coefficients. S denotes the normalized niche coefficient score. The multivariate regression p-value is derived from two-tailed t-statistics. **d**, Ligand-receptor pairs correlated with stellate cell Fa3 and KC Fa1. See Methods. The rectangle's north and south faces represent ligand and receptor correlation to the factors, while west and east faces represent the proportion of ligand and receptor expressing cells.




**Supplementary Figure 10| Pathway analysis and expression of genes associated with stellate and Kupffer cell covariation in mouse liver.** **a,** The top 50 positively correlated genes, identified by Spearman correlation with KC Fa1 (left) and stellate cell Fa2 (right) were analyzed for the enrichment of gene sets from the "Bio Planet 2019" database. See Methods for details. **b,** Same as (a) but after removing genes encoding mitochondrial genes, and genes

encoding ribosomal subunits. **c,** UMAP representation of scRNA-seq/snRNA-seq/CITE-seq reference data [63] highlighting normalized expression of select genes correlating to stellate cell Fa2 (top) or KC Fa1. The Pearson correlation coefficients between the gene expression and the respective factor values (KC Fa1, stellate cell Fa2) are indicated in parentheses. **d,** The top 20 positively correlated genes to KC Fa1(left) and stellate cell Fa2 (right) are displayed as dot plot highlighting the fraction of cells in the respective population expressing a gene (dot size) and the mean expression level (dot color). **e,** UMAP representation of spatial data for stellate cells (left) and KCs (right) highlighting stellate cell Fa2 and KC Fa1, respectively. Both factors exhibit expression gradients in the respective population. **f,** Brightfield image of hepatic stellate cell *in vitro* culture 48 hours after isolation.


a
Slide-seqV2 Cerebellum
Astrocytes-332
Bergmann-525
Candelabrum-11
Choroid-7
Endothelial-33
Ependymal-8
Fibroblast-89
Globular-4
Golgi-68
Granule-5281
Lugaro-22
Molecular Layer Interneuron (MLI1)-245
Molecular Layer Interneuron (MLI2)-239
Microglia-23
Oligodendrocytes-724
Polydendrocytes-47
Purkinje-577
UBCs (Unipolar Brush Cells)-18
d
Niche Interactions
Globular
Purkinje
Bergmann
UBCs
Choroid
Granule
MLI1
Candelabrum
MLI2
Ependymal
Endothelial
Fibroblast
Golgi
Lugaro
Macrophages
Microglia
Polydendrocytes
Oligodendrocytes
Astrocytes
b
Confusion matrix
True class
Predicted class
c
Bergmann, conf. score=0.417± 0.048
e
Bergmann Niche
-log10(p-value)
Bergmann s0.804
Candelabrum s0.088
Endothelial s0.023
Globular s0.058
Lugaro s0.033
MLI1 s0.152
Purkinje s0.370
f
Ligand receptor analysis in the latent factors
Ligands
Bergmann (cc) Fa1
Purkinje (nc) Fa1
Apoe
Lgi4
Pltp
Psap
Ncam1
Psap
Gpr37l1
Ptprz1
Adam23
App
Gpr37l1
Sorl1
Receptors
fraction of cells expressed
correlation with factors
g
Spearman correlation
log(avg expression)
Fraction of cells expressed
Purkinje:Fa1
Bergmann:Fa1
Purkinje
Bergmann
h
Fa1 Purkinje [BioPlanet 2019 database]
Salivary secretion
Calcium regulation in the cardiac cell
Gastric acid secretion
Pancreatic secretion
Calcium signaling pathway
Ion channel transport
Ion transport by P-type ATPases
Myometrial relaxation and contraction pathways
Alzheimer's disease
Transmission across chemical synapses
% Genes in set
-log10(FDR)
Combined Score
i
Fa1 Bergmann [BioPlanet 2019 database]
Gastric acid secretion
Salivary secretion
Proximal tubule bicarbonate reclamation
Response to elevated platelet cytosolic calcium
Corticosteroids and cardioprotection
Glucose metabolism
Glycolysis and gluconeogenesis
Disease
FSH regulation of apoptosis
Gluconeogenesis
% Genes in set
-log10(FDR)
Combined Score

**Supplementary Figure 11| NiCo application on Slide-seqV2 cerebellum data**. **a**, The spatial map of cell types in the cerebellum as annotated in ref. [74]. **b**, NiCo confusion matrix. **c**, The logistic regression coefficients of the Bergmann glia niche. Error bars, see Methods. **d**, The cell type niche interaction map (interaction threshold c=0.08). **e**, Covariation between Bergmann glia factors (y-axis) and co-localized neighborhood cell type factors (x-axis). Circle size scales linearly with $-\log_{10}$(p-value), and circle color indicates ridge regression coefficients. S denotes the normalized niche coefficient score. The multivariate regression p-value is derived from two-tailed t-statistics. **f**, Ligand-receptor pairs correlated with Bergmann Fa1 and Purkinje Fa1. See Methods. The rectangle's north and south faces represent ligand and receptor correlation to the factors, while west and east faces represent the proportion of ligand and receptor expressing cells. cc, central cell; nc, niche cell. **g**, Spearman correlations (left) and average expression (right) for the top 20 positively correlated genes for Purkinje Fa1 and Bergmann Fa1. Dashed line demarcates the genes associated with each cell types. **h**, The top 50 positively correlated genes, identified by Spearman correlation with Purkinje Fa1 were analyzed for enriched pathways from the BioPlanet 2019 database. **i**, The top 50 positively correlated genes, identified by Spearman correlation with Bergmann Fa1 were analyzed for enriched pathways from the BioPlanet 2019 database.

| Primers qPCR(m) | Genbank ID | Synthesis direction (5'-3'): Sequences |
|---|---|---|
| Acta2 (m) | NM_007392 | Forward: CTATTCAGGCTGTGCTGTCCCTC<br>Reverse: CACGTTGTGAGTCACACCATCTC |
| Col1a1 (m) | NM_007742 | Forward: GTCTGGTTTGGAGAGAGCATGAC<br>Reverse: CGCAGGAAGGTCAGCTGGATAG |
| Gapdh (m) | NM_008084 | Forward: TGTCCGTCGTGGATCTGAC<br>Reverse: CCTGCTTCACCACCTTCTTG |
| Hprt (m) | NM_013556 | Forward: TCCTCCTCAGACCGCTTTT<br>Reverse: CCTGGTTCATCATCGCTAATC |
| Lox (m) | NM_010728 | Forward: CTATGGGTACCACAGGCGCTTTG<br>Reverse: CCGCATAGGTGTCATAACATCCAG |
| Pdgfrb (m) | NM_001146268 | Forward: GAGCTCAGTGAGAGGAAGCGTATC<br>Reverse: GAACAGGTCCTCGGAGTCCATAG |
| Ppia (m) | NM_008907 | Forward: GCATACAGGTCCTGGCATCT<br>Reverse: AGCTGTCCACAGTCGGAAAT |
| Reln (m) | NM_011261 | Forward: CAGCCGAAGGACTTCACACAAG<br>Reverse: GCAGTAGAACTCCTTTCATCCGAGC |

**Supplementary Table 1| List of primers used for qPCR experiments.**