

Supplementary Figures for

Highly sensitive spatial transcriptomics using FISHnCHIPs

of multiple co-expressed genes

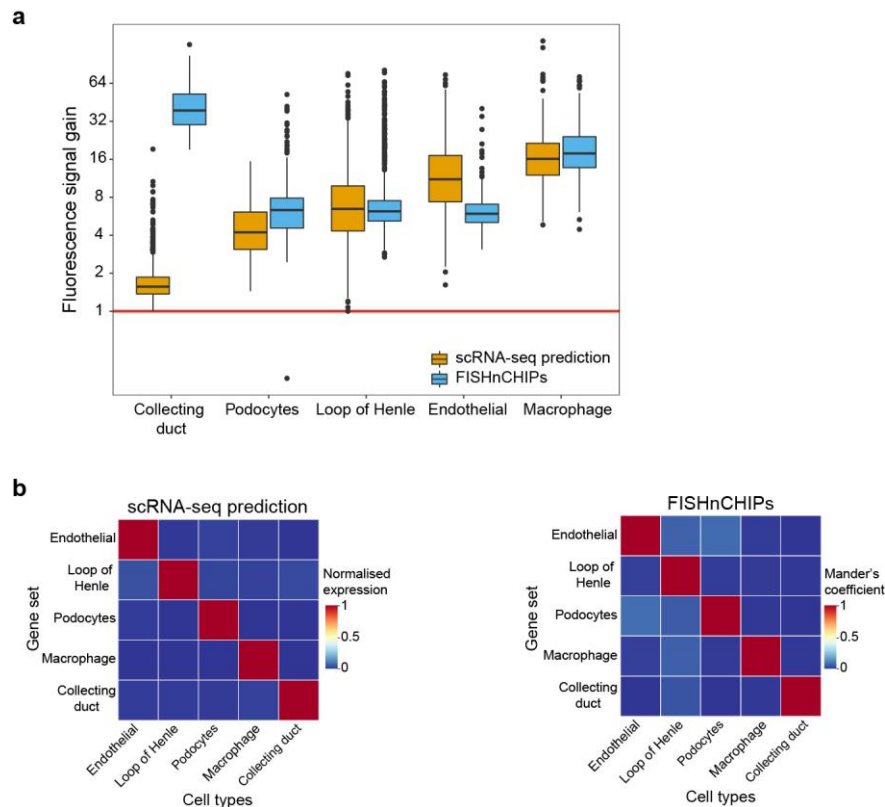
Xinrui Zhou^{1,2}, Wan Yi Seow^{1,2}, Norbert Ha¹, Teh How Cheng¹, Lingfan Jiang¹, Jeeranan Boonruangkan¹, Jolene Jie Lin Goh¹, Shyam Prabhakar¹, Nigel Chou^{1*}, Kok Hao Chen^{1*}

¹ Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR); 60 Biopolis Street, Singapore 138672, Singapore

² These authors contributed equally to this work

*Corresponding author. Email: nigel_chou@gis.a-star.edu.sg, chenkh@gis.a-star.edu.sg

Supplementary Fig. 1

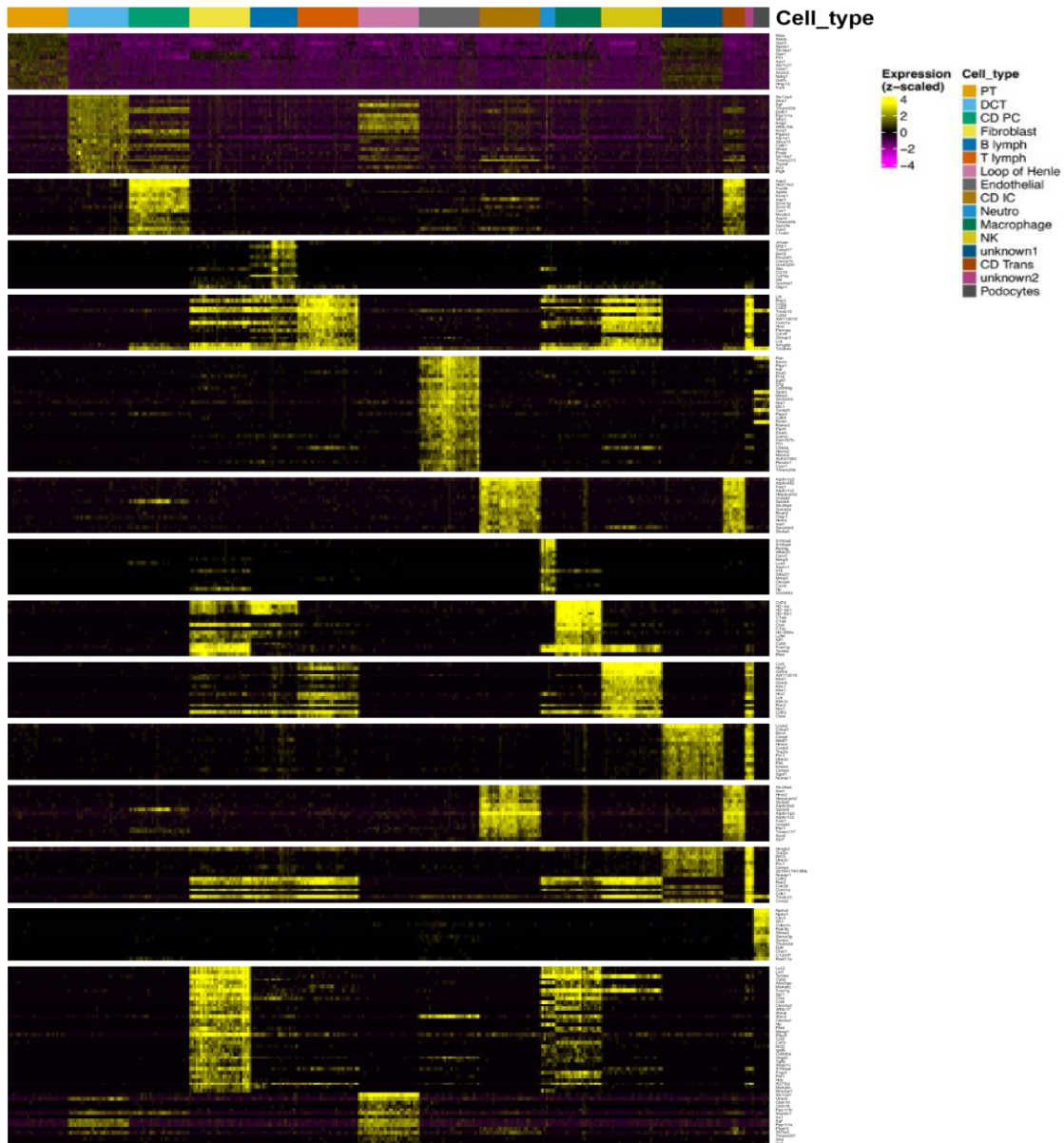


Supplementary Fig. 1 Quantification of FISHnCHIPs (cell-centric) signal and specificity in the mouse kidney (related to Fig. 2).

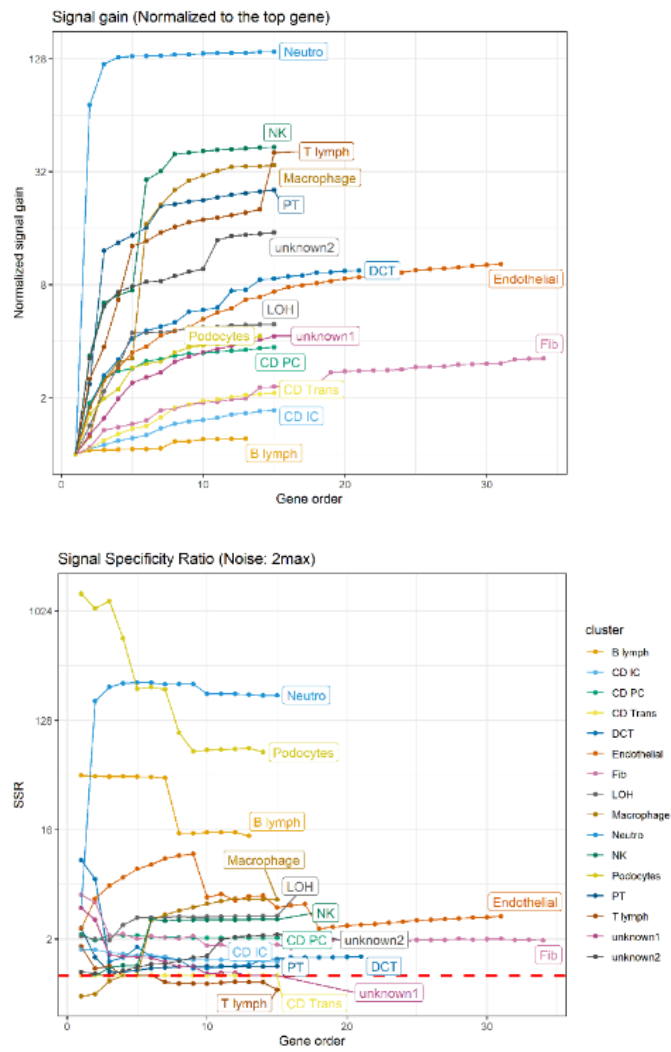
(a) Box plots of the ratio of mean fluorescence intensity per cell of FISHnCHIPs to smFISH (blue), and scRNA-seq predictions: the ratio of counts for 14-23 genes to the top DE gene (yellow). Number of cells (FISHnCHIPs): collecting duct: 146, podocytes: 461, loop of Henle: 727, endothelial: 400, macrophage: 341. Number of cells (scRNA-seq): collecting duct: 1,825, podocytes: 77, loop of Henle: 1,496, endothelial: 701, macrophage: 216. The box plots show the median (center line), the first and third quartiles (box limits), and 1.5× the interquartile range (whiskers). Red line indicates where the fluorescence signal gain is 1. **(b)** Predicted signal crosstalk (left heatmap): Normalized mean scRNA-seq counts for FISHnCHIPs genes across the 5 cell types. Number of cells analyzed is the same as in J. Measured FISHnCHIPs crosstalk (right heatmap): Mander's overlap coefficient across the 5 cell-type channels shown in Fig. 2. Source data are provided as a Source Data file.

Supplementary Fig. 2

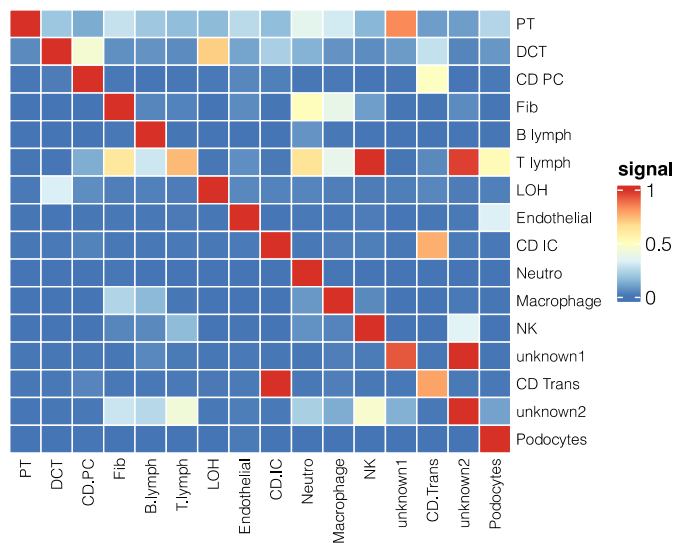
a



b



c



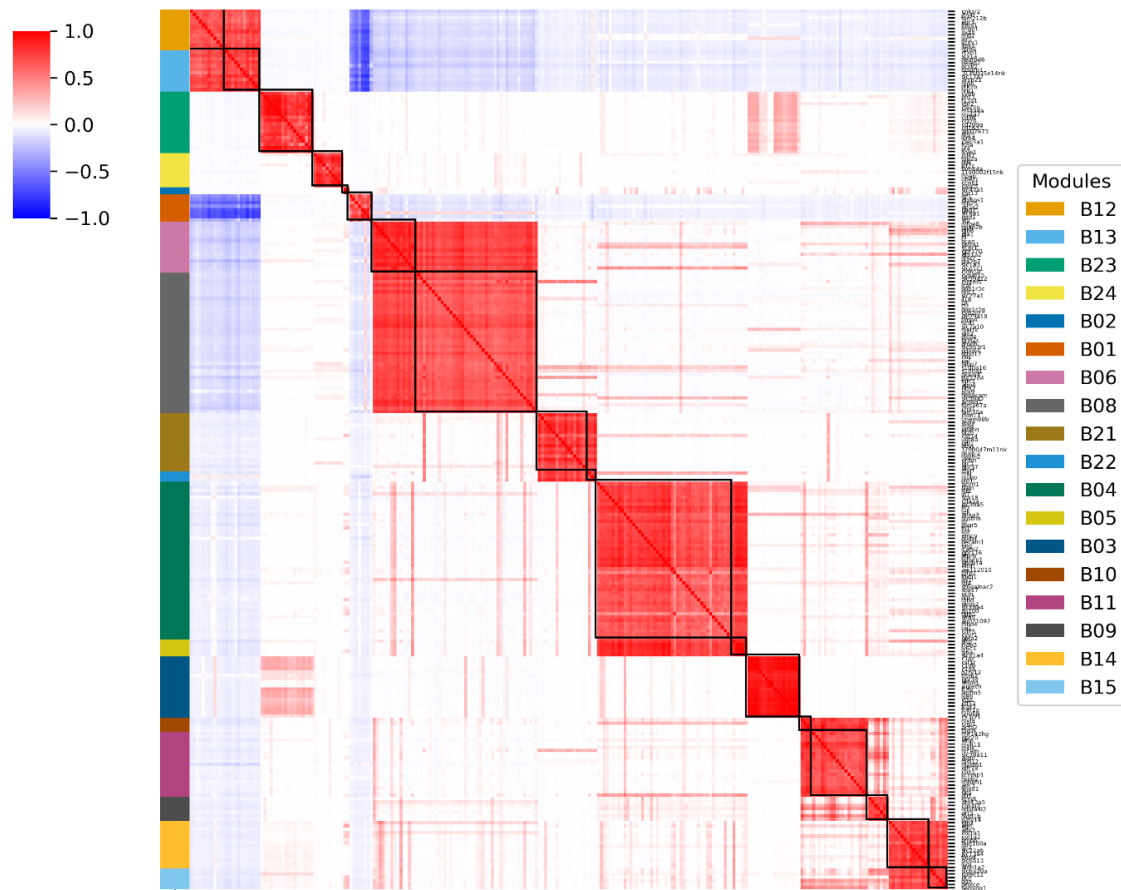
Supplementary Fig. 2 Computational prediction of FISHnCHIPs (cell-centric) signal gain and specificity.

(a) scRNA-seq²⁷ gene expression heatmap of a FISHnCHIPs gene panel targeting all the previously annotated mouse kidney cell types, sampling a maximum of 300 cells per cluster.

(b) Predicted Signal Gain (SG) and Signal Specificity Ratio (SSR) as a function of the number of FISHnCHIPs genes. We defined SG as the ratio of the sum of counts for FISHnCHIPs genes to that of the top DE gene, and SSR as the ratio of the sum of counts for FISHnCHIPs genes in the target cell type to that in the most likely off-target cell type. When SSR approaches unity, the fluorescence intensity for the cell type of interest will be equal to an off-target cell type, rendering them indistinguishable. (c) Predicted signal crosstalk: Heatmap of the normalized mean scRNA-seq counts of the FISHnCHIPs gene panel across all kidney cell types. Source data are provided as a Source Data file.

Supplementary Fig. 3

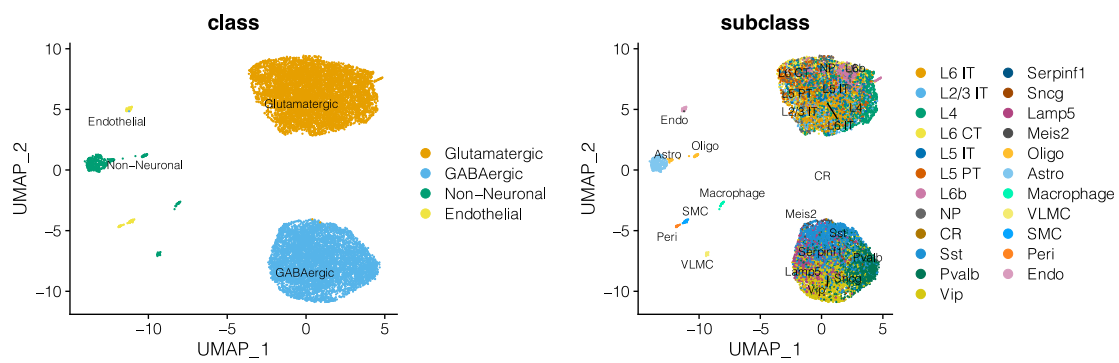
a



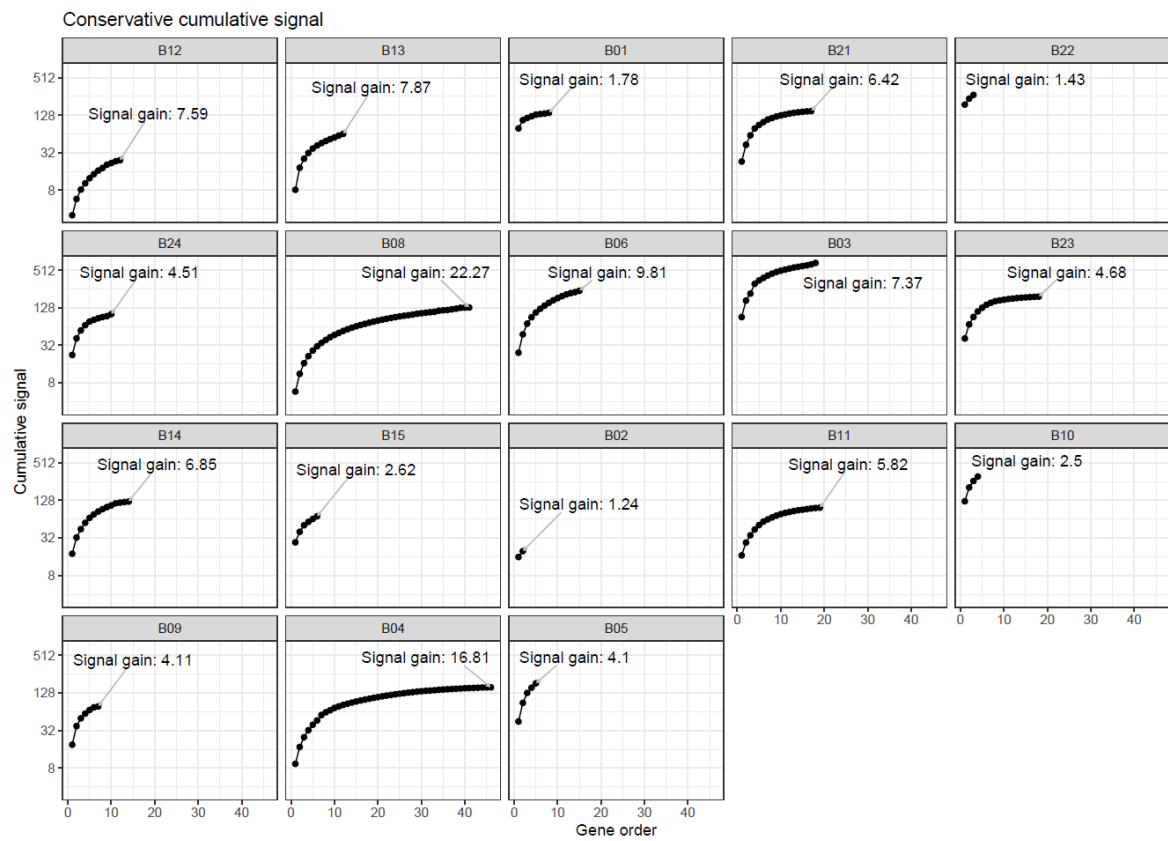
b



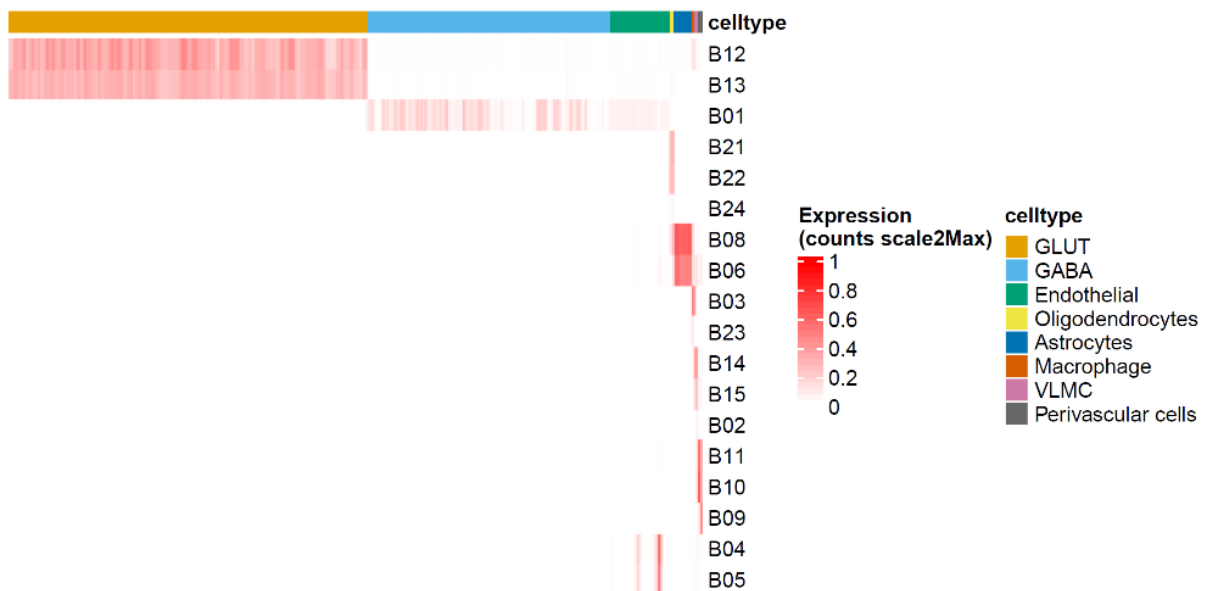
c



d



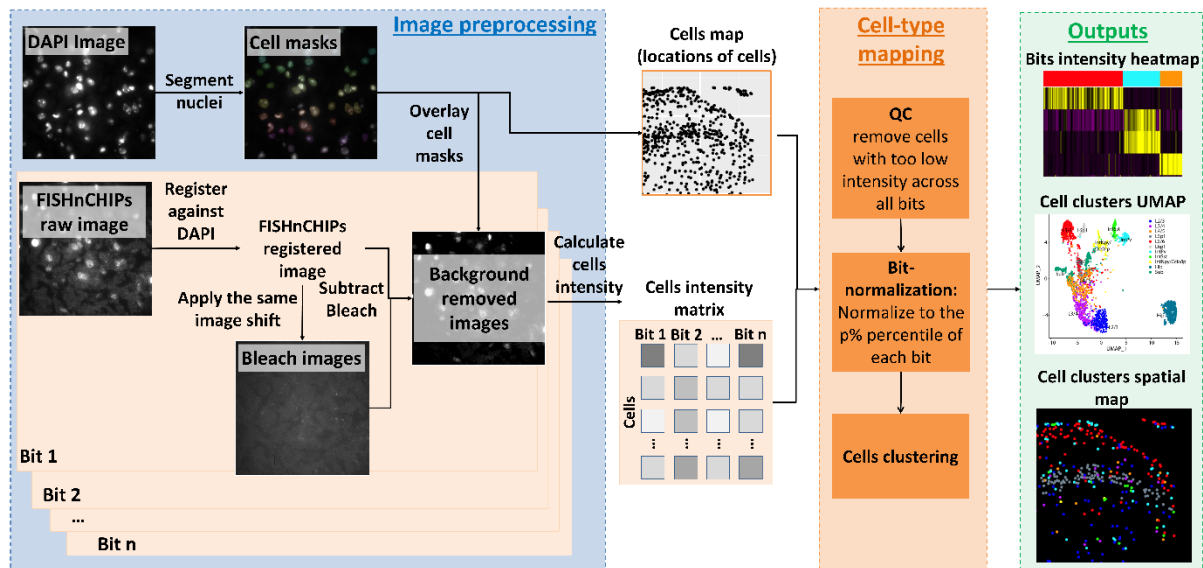
e



Supplementary Fig. 3 Gene module based FISHnCHIPs library design for the mouse cortex library (imaged in Fig. 3).

(a) scRNA-seq²⁹ gene-gene correlation heatmap (with gene names) for the 255 feature genes in the mouse cortex library (imaged in Fig. 3). We computed their pair-wise Pearson's correlation coefficient and clustered the correlation matrix using the Leiden algorithm. The gene partitions were further sub-clustered using hierarchical clustering into 18 modules. (b) scRNA-seq gene expression heatmap for the 255 genes. (c) UMAP representation of the predicted clusters from scRNA-seq simulated module-cell (meta-gene) expression, indicated by the labels provided by the scRNA-seq reference dataset. ~8 cell types are clearly separated with the selected features. (d) Predicted conservative Signal Gain (cumulative), which is defined as the ratio of the panel signal to the highest gene signal, as a function of the number of genes. (e) Predicted module-cell expression heatmap. We grouped the cluster labels into the 8 resolvable cell types (see Supplementary Table 5). Source data are provided as a Source Data file.

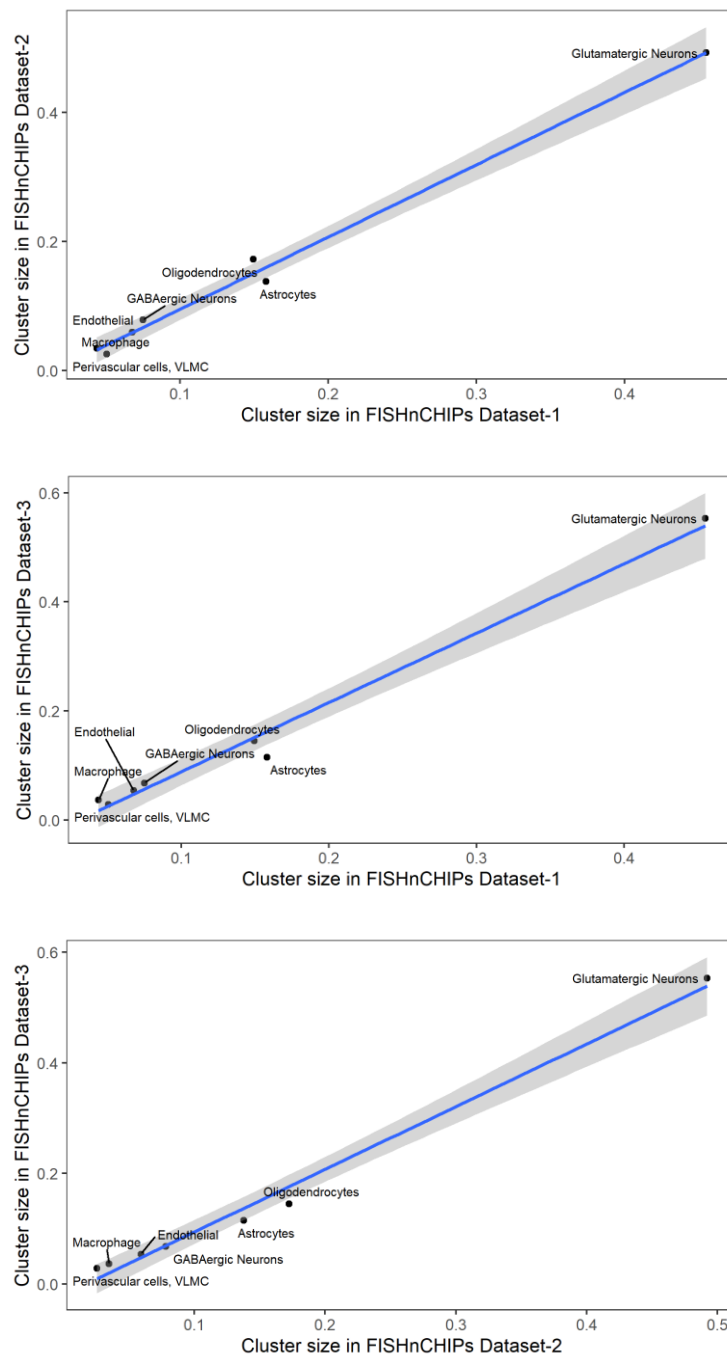
Supplementary Fig. 4



Supplementary Fig. 4 Overview of the FISHnCHIPs image processing and data analysis workflow.

Stepwise data processing (see methods, as well as supplementary software): Inputs are DAPI, FISHnCHIPs, and background (after 55% formamide wash) images. 1) Preprocessing steps include segmentation based on DAPI images to generate cell masks; 2) Registration and background subtraction of FISHnCHIPs images; 3) Cell masks are used to generate cell intensity matrix with a list of cell centroids. 4) Clustering of the cell intensity matrix. 5) Outputs can be visualized in a heatmap, UMAP, or spatial map. Outputs can also be subjected to further analyses, such as classifications of spatial patterns and analysis of cell-cell interactions.

Supplementary Fig. 5

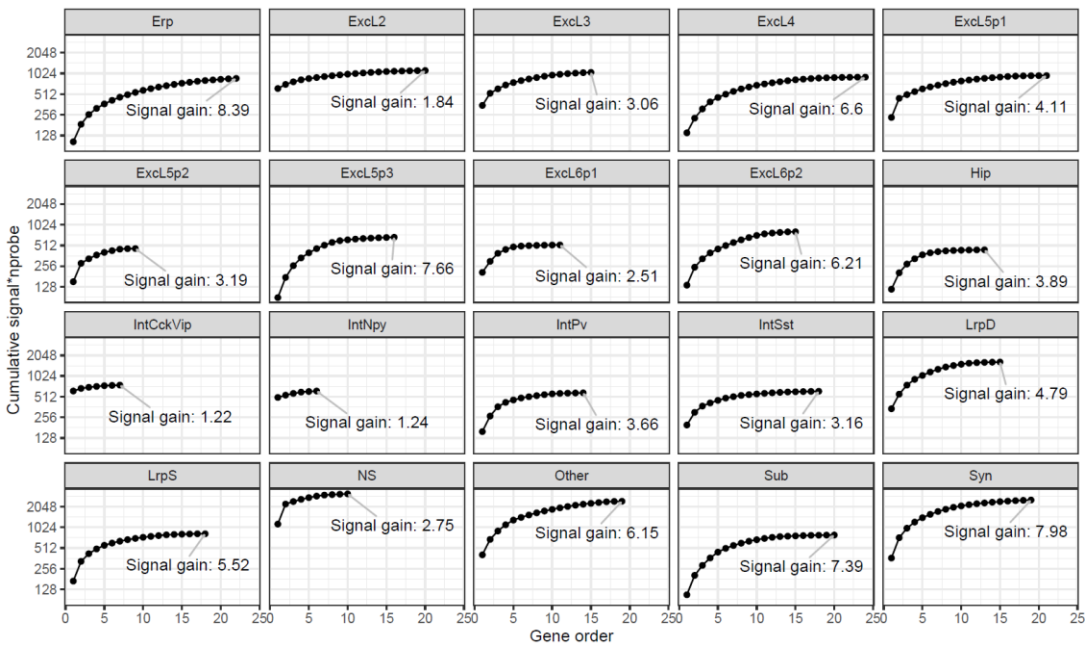


Supplementary Fig. 5 Reproducibility of the mouse brain FISHnCHiPs data (related to Fig. 3) among technical triplicates.

Scatter plots of cell type abundances between technical replicates. Source data are provided as a Source Data file.

Supplementary Fig. 6

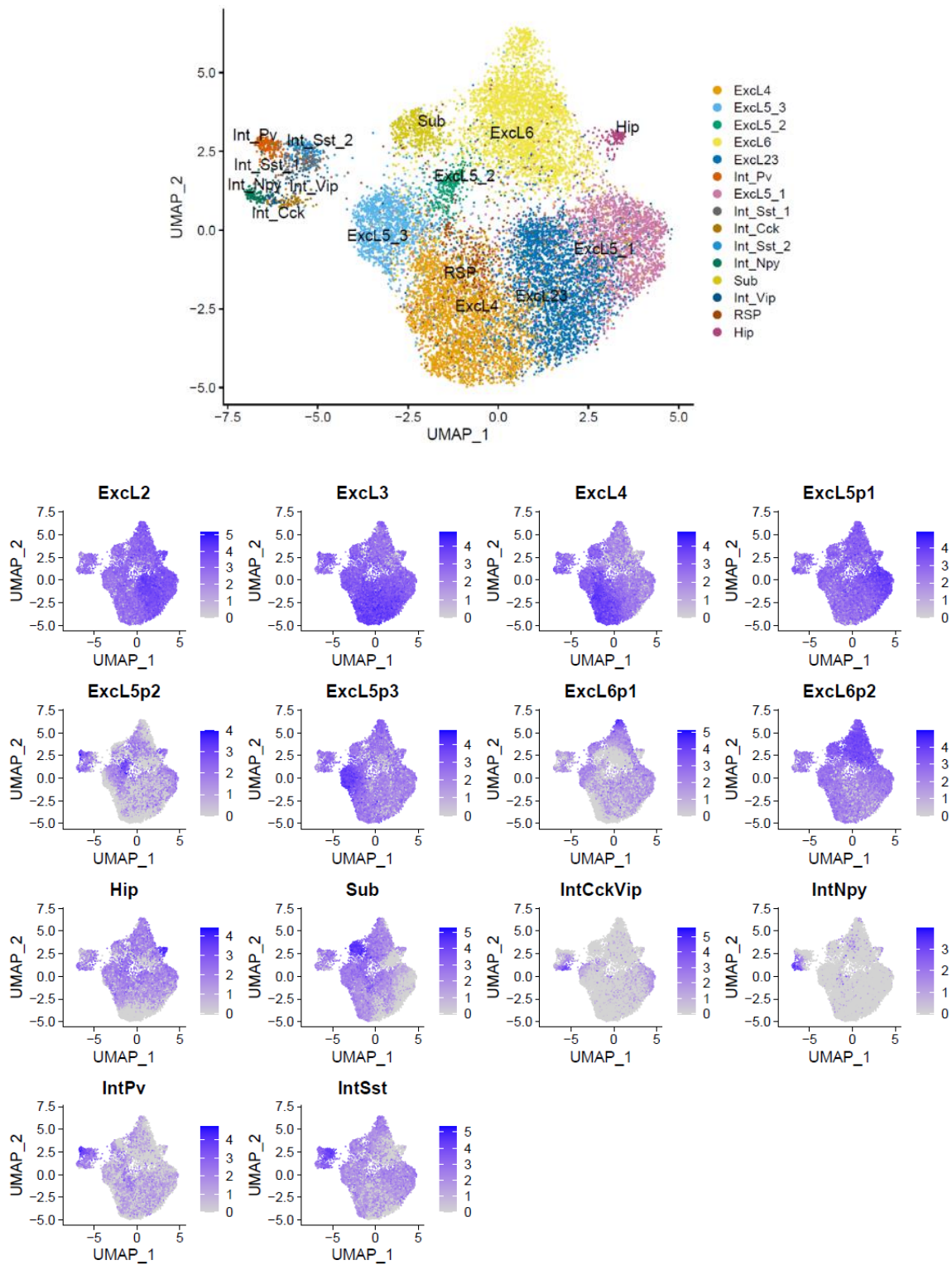
a



b



c



d

Supplementary Fig. 6 Evaluation of the gene expression programs based FISHnCHiPs library for the mouse visual cortex (imaged in Fig. 4) using the scRNA-seq reference dataset. (a) Predicted conservative Signal Gain (cumulative), which is defined as the ratio of

the panel signal to the highest gene signal, as a function of the number of genes. **(b)**

Predicted Signal Specificity: scRNA-seq expression heatmap for the 20 programs. **(c)** UMAP representation of the predicted clusters from scRNA-seq simulated program-cell (meta-gene) expression, indicated by the labels provided by the scRNA-seq reference dataset⁶⁵. **(d)**

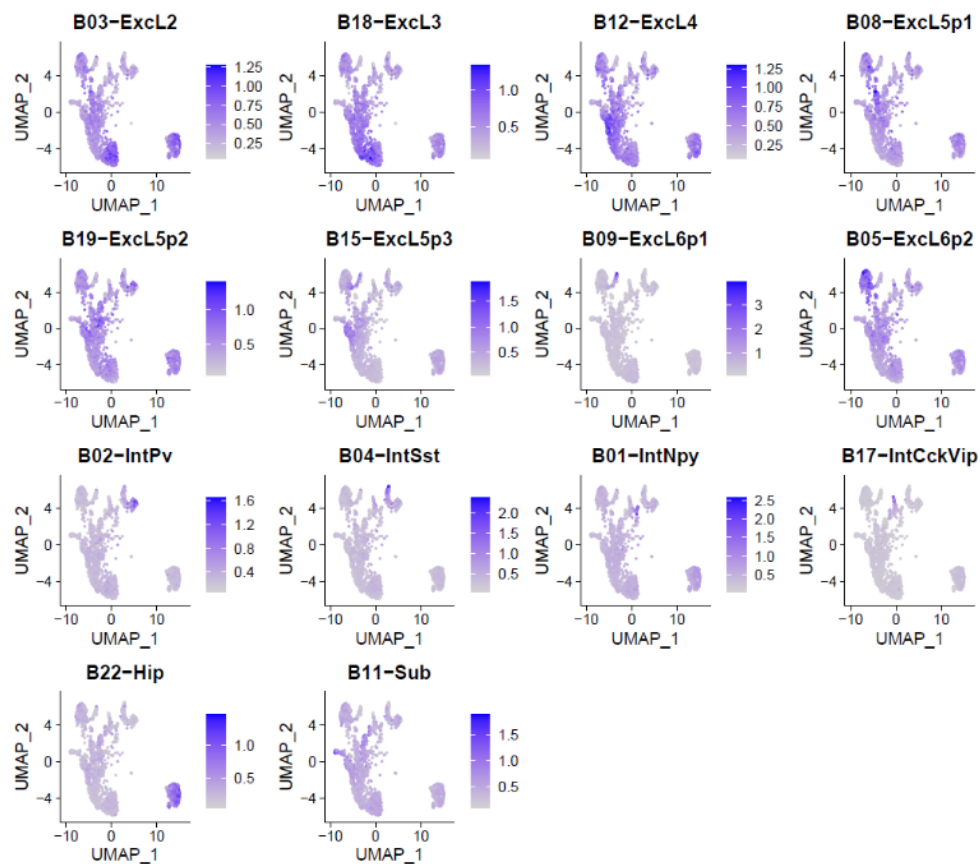
Predicted scRNA-seq feature plots of the 14 identify programs. Source data are provided as a Source Data file.

Supplementary Fig. 7

a



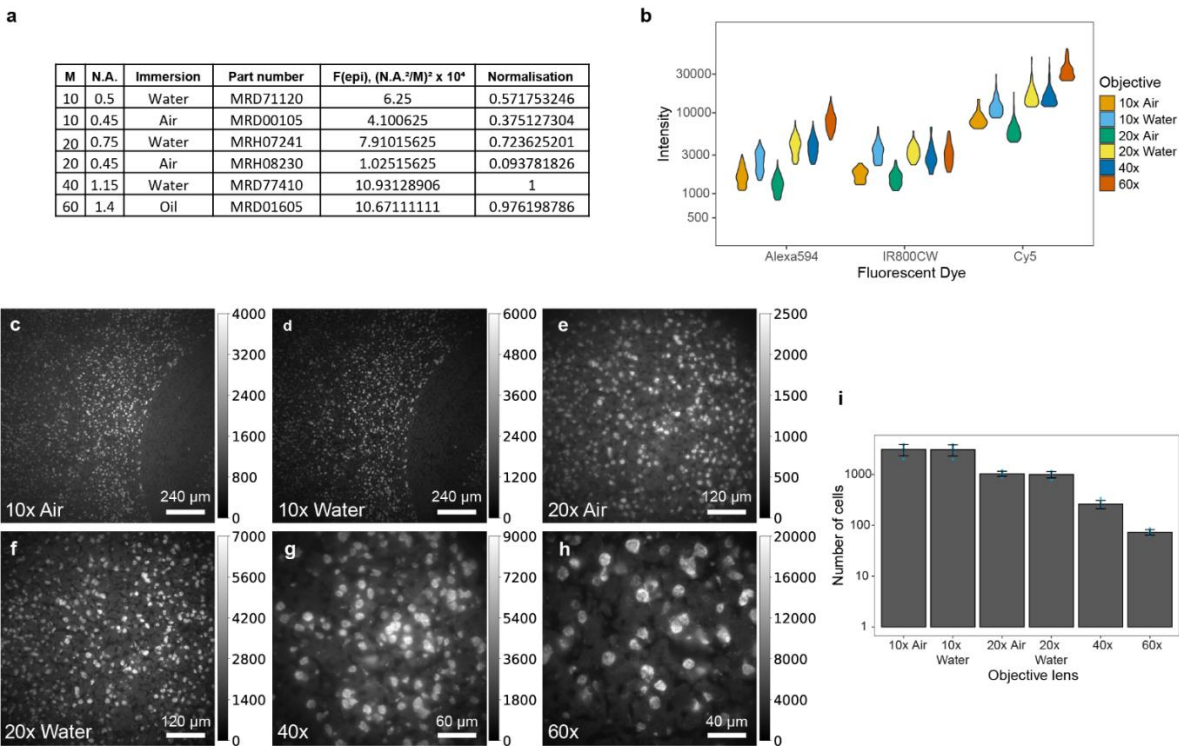
b



Supplementary Fig. 7 Gradients of gene expression along the cortical depth of the mouse visual cortex as imaged by FISHnCHIPs (related to Fig. 4).

(a) FISHnCHIPs expression heatmap of the cell-by-program-intensity matrix (cells are ordered by their distance to the outer edge). (b) FISHnCHIPs feature plots of the 14 identity programs.

Supplementary Fig. 8

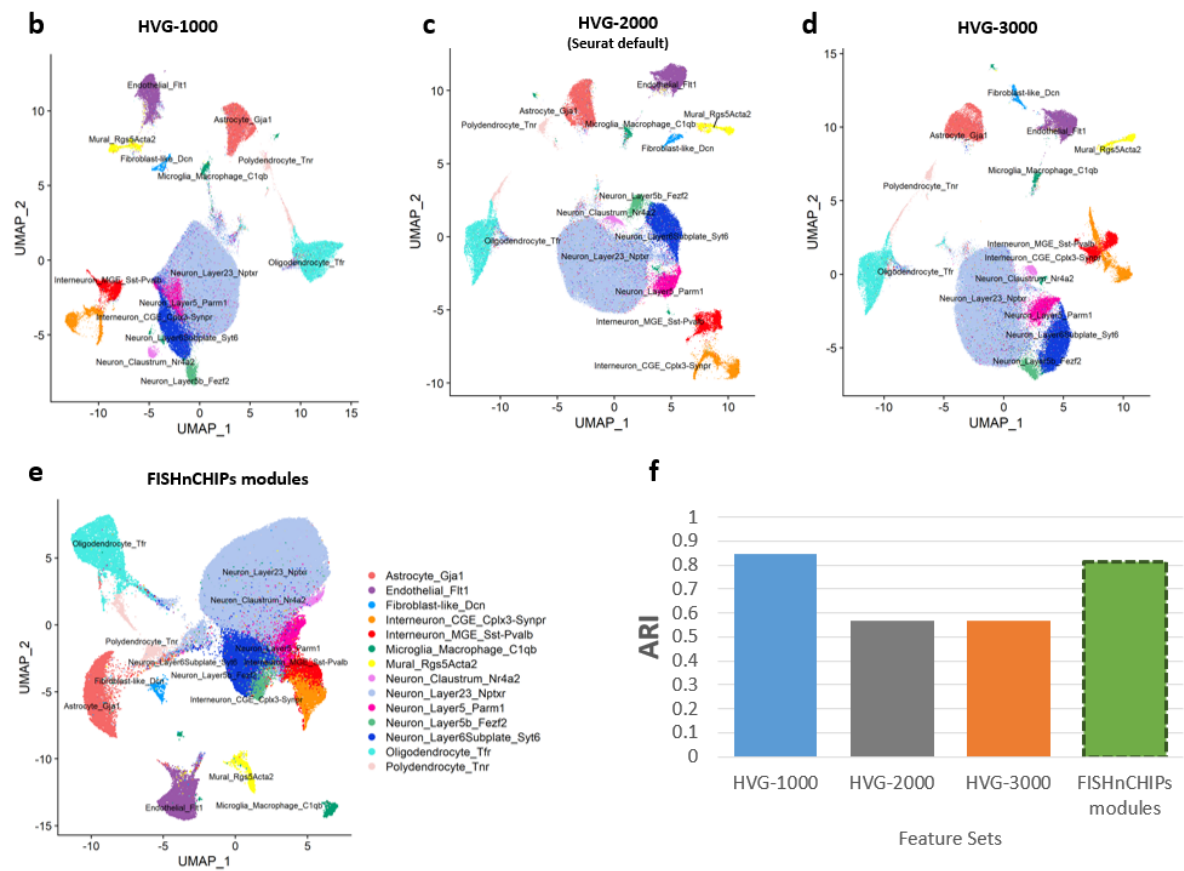
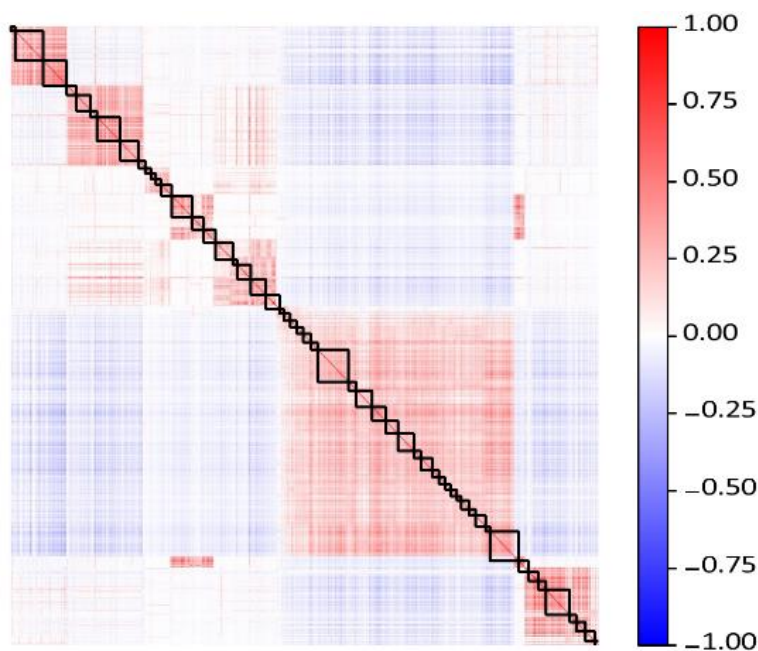


Supplementary Fig. 8 FISHnCHIPs of the mouse brain imaged under lower magnification.

(a) Table of magnification (M), numerical aperture (N.A.), and predicted light gathering power under epi-illumination configuration, $F(\text{epi})^{66}$, for 6 different objective lenses. (b) Measured mean fluorescence intensity per cell for Alexa594, Cy5, and IR800CW, for 6 different objective lens. (c-h) Example unprocessed FISHnCHIPs images (one Field of View, FOV) of the mouse cortex under 6 different objective lenses. Cells labelled with FISHnCHIPs were detected above the background level across all three-color channels, even at the 10x magnification. (i) The number of cells detected per FOV ($n = 5$ FOVs, examined over 1 experiment, error bars indicate the standard deviation). Because of the wider field of view, the number of cells imaged was $>\sim 40$ fold greater when using the 10x versus 60x objective lenses. Average number of cells: 10x air: 3130, 10x water: 3088, 20x air: 1003, 20x water: 1041, 40x: 261, 60x: 73. We chose the 10x water objectives for Fig. 5 data acquisition. Source data are provided as a Source Data file.

Supplementary Fig. 9

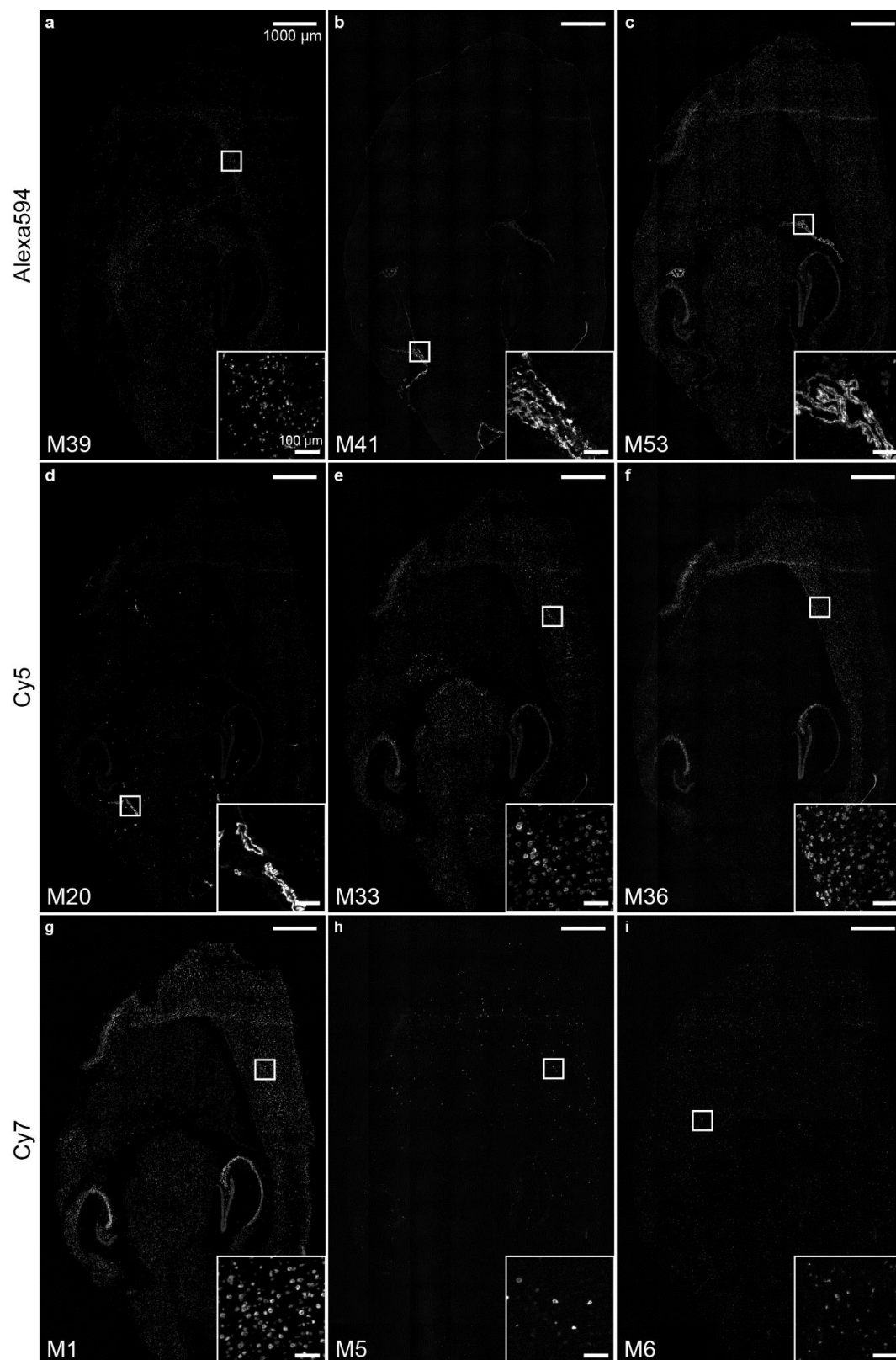
a



Supplementary Fig. 9 scRNA-seq gene-gene correlation analyses of the FISHnCHIPs genes targeting 53 gene modules in the mouse cortex (imaged in Fig. 5).

(a) scRNA-seq gene-gene correlation heatmap for the 674 feature genes in the mouse cortex library (imaged in Fig. 5). We computed their pair-wise Pearson's correlation coefficient and clustered the correlation matrix using the Leiden algorithm. The gene partitions were further sub-clustered using hierarchical clustering into 53 gene modules. (b-e) UMAP representation for cells in the scRNA-seq dataset predicted from different feature sets, namely b) 1,000 highly variable genes; c) 2,000 highly variable genes (Seurat default); d) 3,000 highly variable genes; e) 53 modules presented in Fig.5; (f) the ARI of clustering cells at a resolution of 0.1 using b-e as features against the labels from the scRNA-seq dataset as ground truth.

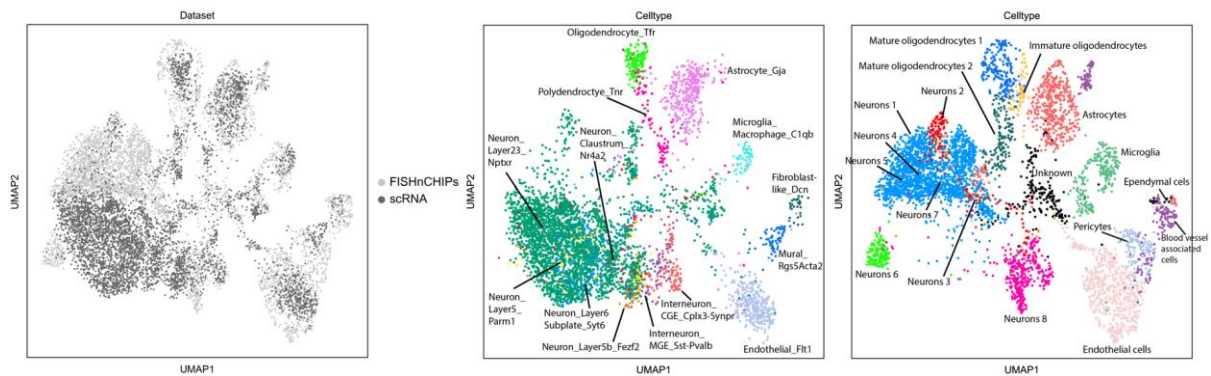
Supplementary Fig. 10



Supplementary Fig. 10 Example normalized images from the 53-modules FISHnCHIPs

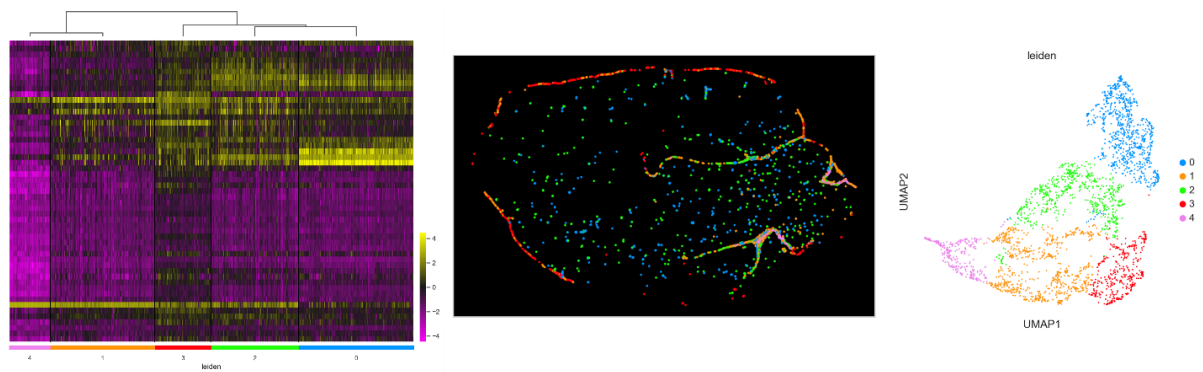
profiling of gene module 39, gene module 41, gene module 53 using Alexa 594 (**a - c**), gene module 20, gene module 33, gene module 36 using Cy5 (**d - f**), and gene module 1, gene module 5, and gene module 6 using IRDye 800CW (**g - i**) under the 10x objective lens. Scale bar, 1000 μm . Inserts: zoomed in region of the white box. Scale bar, 100 μm .

Supplementary Fig. 11



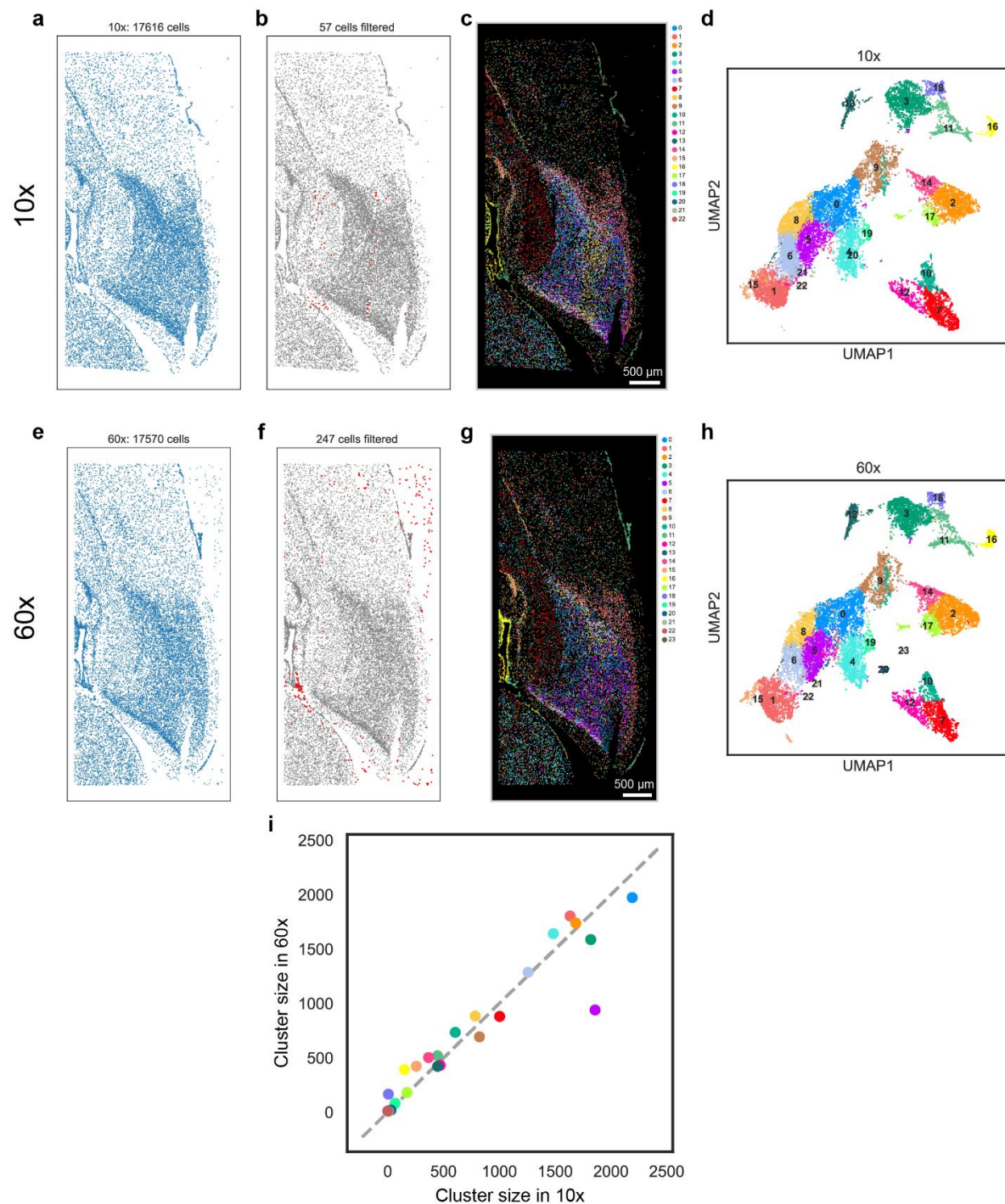
Supplementary Fig. 11 FISHnCHIPs and scRNA-seq data integration with Harmony (a) UMAP representation for frontal cortex cells from Harmony integration of the scRNA-seq reference and FISHnCHIPs data. **(b)** scRNA-seq cells with cell type labels provided by Saunders *et. al.* ³⁵. **(c)** FISHnCHIPs cells and labels.

Supplementary Fig. 12



Supplementary Fig. 12 Sub-clustering analyses of the 53-module FISHnCHiPs data revealed subtypes of blood vessel associated and inhibitory cell subtypes with distinct spatial patterns. FISHnCHiPs expression heatmap, spatial map, and UMAP of the subtypes of blood vessel associated cells.

Supplementary Fig. 13



Supplementary Fig. 13 Comparison of the 10x and 60x datasets on adjacent tissue sections. (a, e) Plot of all the segmented cells in the 10x and 60x datasets respectively. (b, f) Cells with low expression across all modules (in red) are filtered out at the first quality control stage. (c, g) Spatial map of cells after Leiden clustering. Scale bar, 500 μ m. (d, h)

UMAP representation of the clustering. (i) Scatter plot of number of cells in each cluster detected by 60x versus 10x. Dash line represents the $x = y$ line. Source data are provided as a Source Data file.

Supplementary Fig. 14

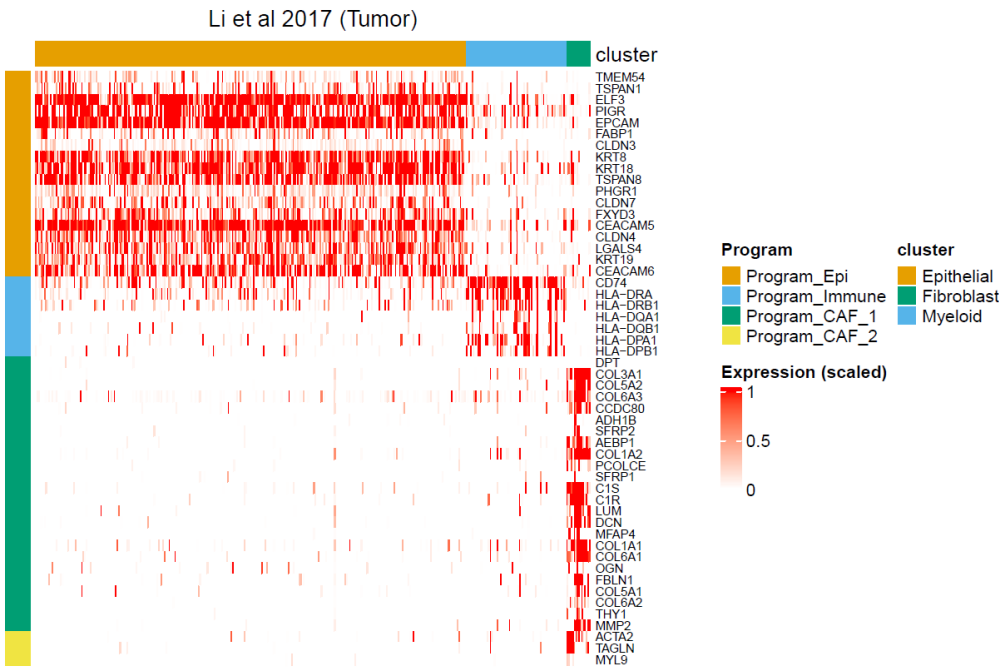
Figure	Number of cells from nuclei segmentation	Targeted cell types	Number of cells that passed quality control	Fraction of nuclei imaged that passed quality control
2	11,146	Podocytes, Loop of Henle, Endothelial, Collecting Duct, Macrophage	2,075	18.62%
3	8,578	Targeted sampling of neurons ²⁹	6,180	72.04%
4	5,552	NMF on neurons only ³¹	2,794	50.32%
5	56,352	All/untargeted ³⁵	54,834	97.31%

Supplementary Fig. 14 Comparison of the fraction of cells that passed QC for each library.

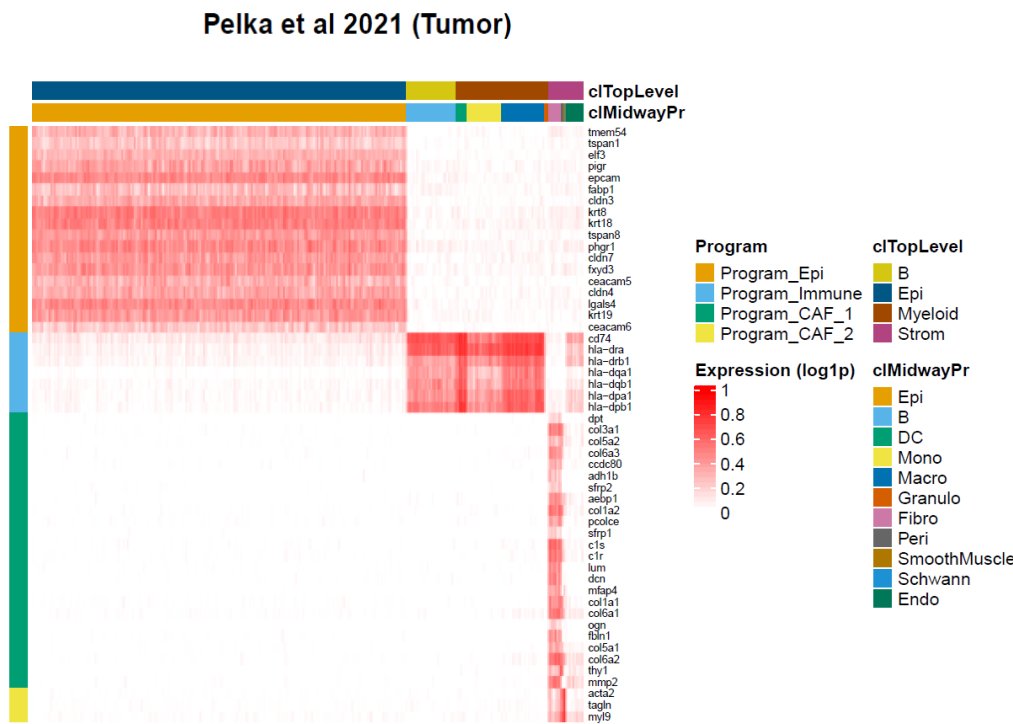
In the Fig.5 experiment, where we attempted to measure all the cell types, a very high fraction of cells (97.31%) passed QC. In the Fig.4 experiment, where we only targeted the neuronal cells, 50.32% of all cells (determined through DAPI) were quantified as neurons. This number agrees with the Fig.3 data of the cortical region, which also contained ~50% neurons.

Supplementary Fig. 15

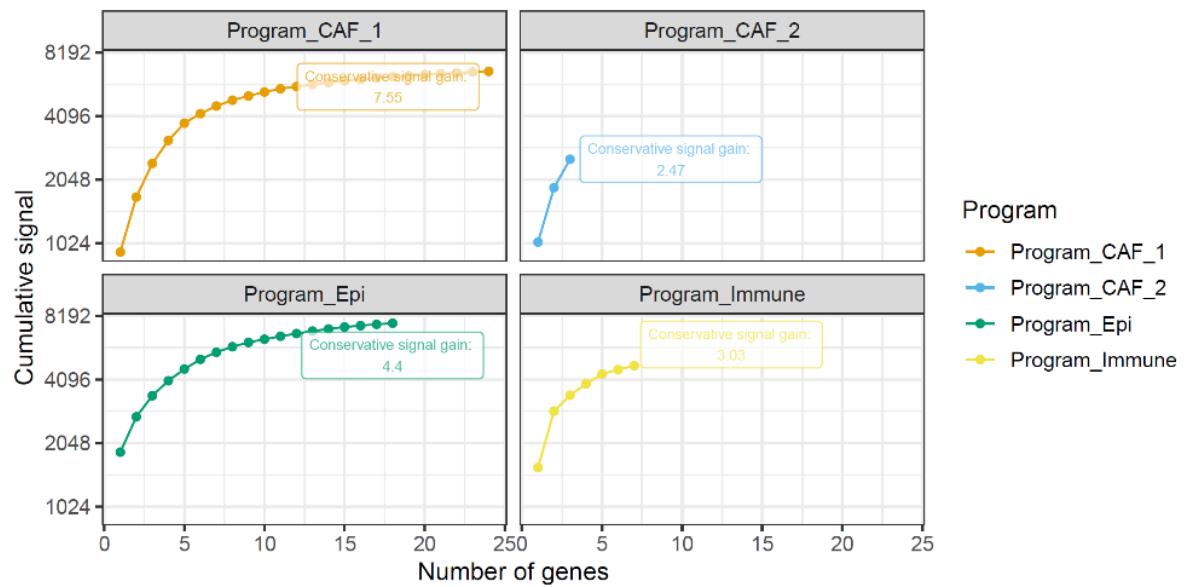
a



b



c



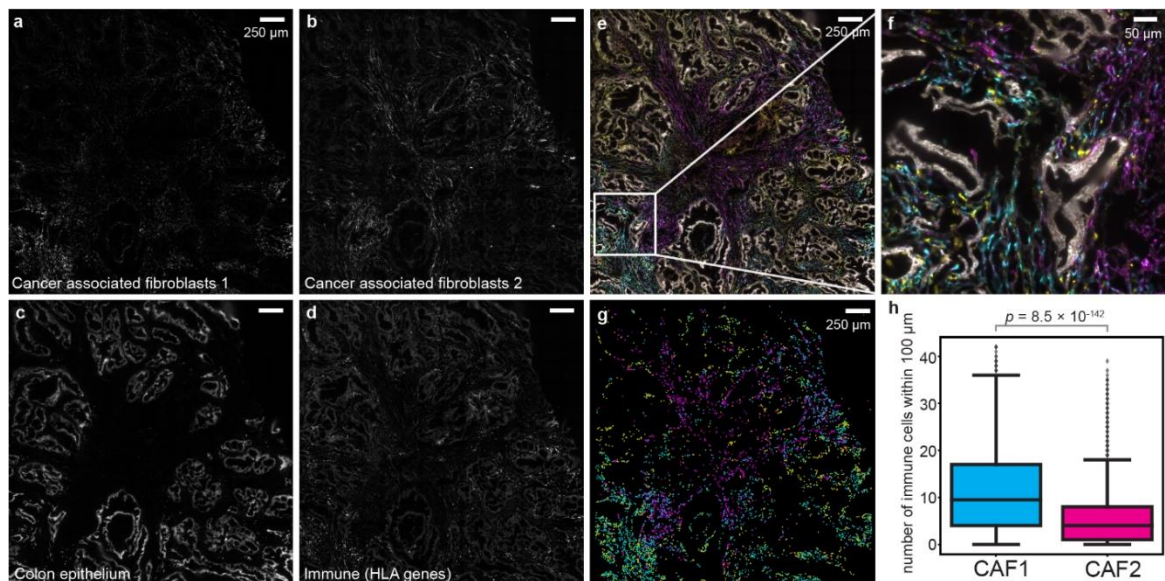
Supplementary Fig. 15 FISHnCHIPs panel for imaging cancer associated fibroblasts (CAFs) subtypes in human colorectal cancer (CRC) tissue (related to Fig. 6).

(a) scRNA-seq gene expression heatmap of the human CRC FISHnCHIPs panel³⁸. (b)

scRNA-seq gene expression heatmap of the human CRC FISHnCHIPs panel⁵³. (c)

Predicted conservative signal gain for the human CRC FISHnCHIPs panel. Source data are provided as a Source Data file.

Supplementary Fig. 16

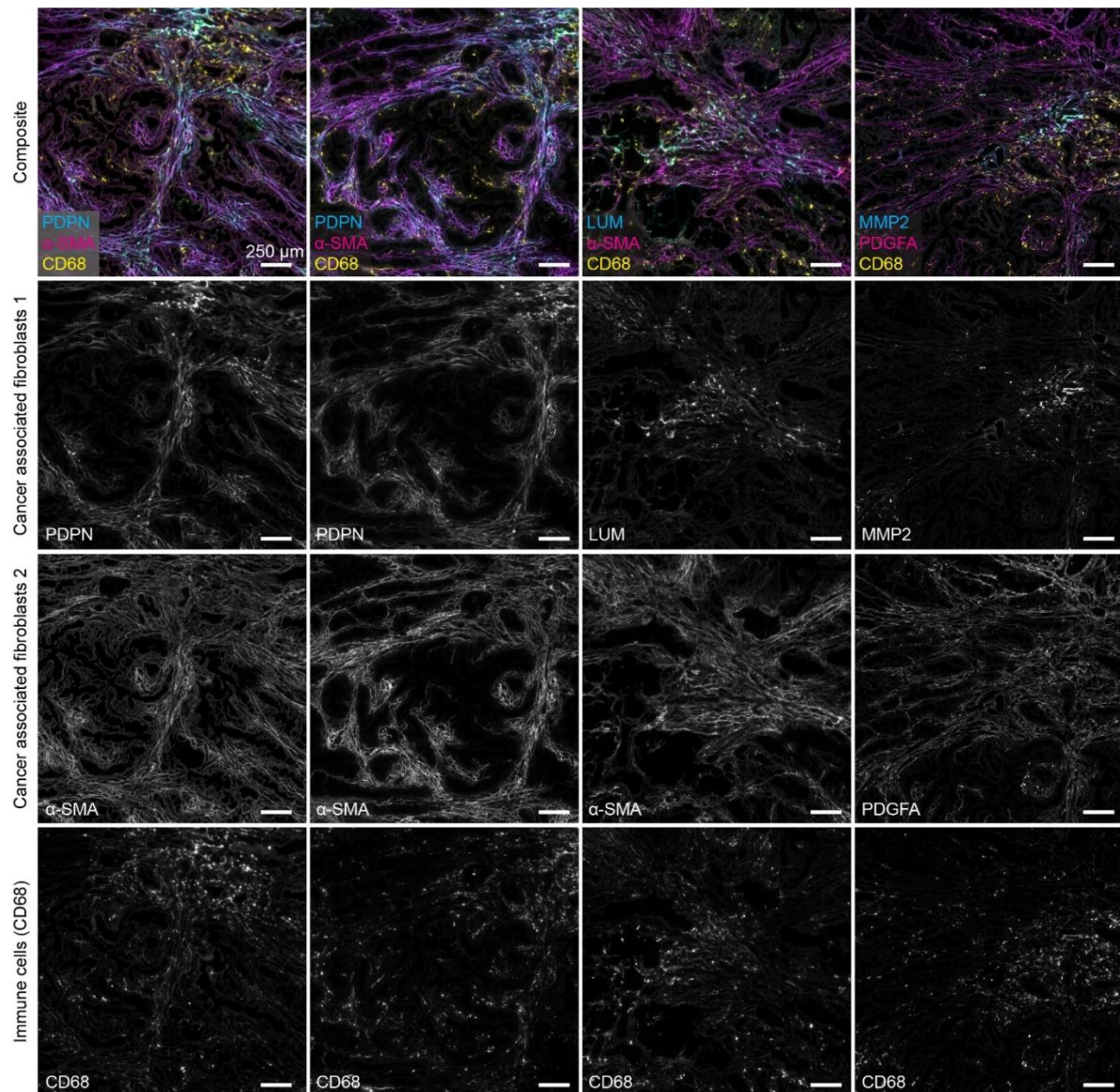


Supplementary Fig. 16 Technical replicate of FISHnCHIPS on human CRC tissue.

(a) FISHnCHIPS image of CAF-1 subtype. Scale bar, 250 μm. (b) FISHnCHIPS image of CAF-2 subtype. Scale bar, 250 μm. (c) FISHnCHIPS image of colon epithelium. Scale bar, 250 μm. (d) FISHnCHIPS image of immune (HLA genes). Scale bar, 250 μm. (e) Composite FISHnCHIPS image. Scale bar, 250 μm. (f) Zoom-in of the white box in E. Scale bar, 50 μm. (g) Box plots of the number of immune cells within 100 μm radius of CAF-1 (cyan) and CAF-2 (purple) cells. Immune cells were found 0.51-fold less frequently in the vicinity of CAF-2 than CAF-1. Number of cells, n: CAF-1: 2,548, CAF-2: 2,199, examined over 1 experiment. The box plots show the median (centre line), the first and third quartiles (box limits), and 1.5× the interquartile range (whiskers). $p = 8.5 \times 10^{-142}$, 2-sided Mann-Whitney U test.

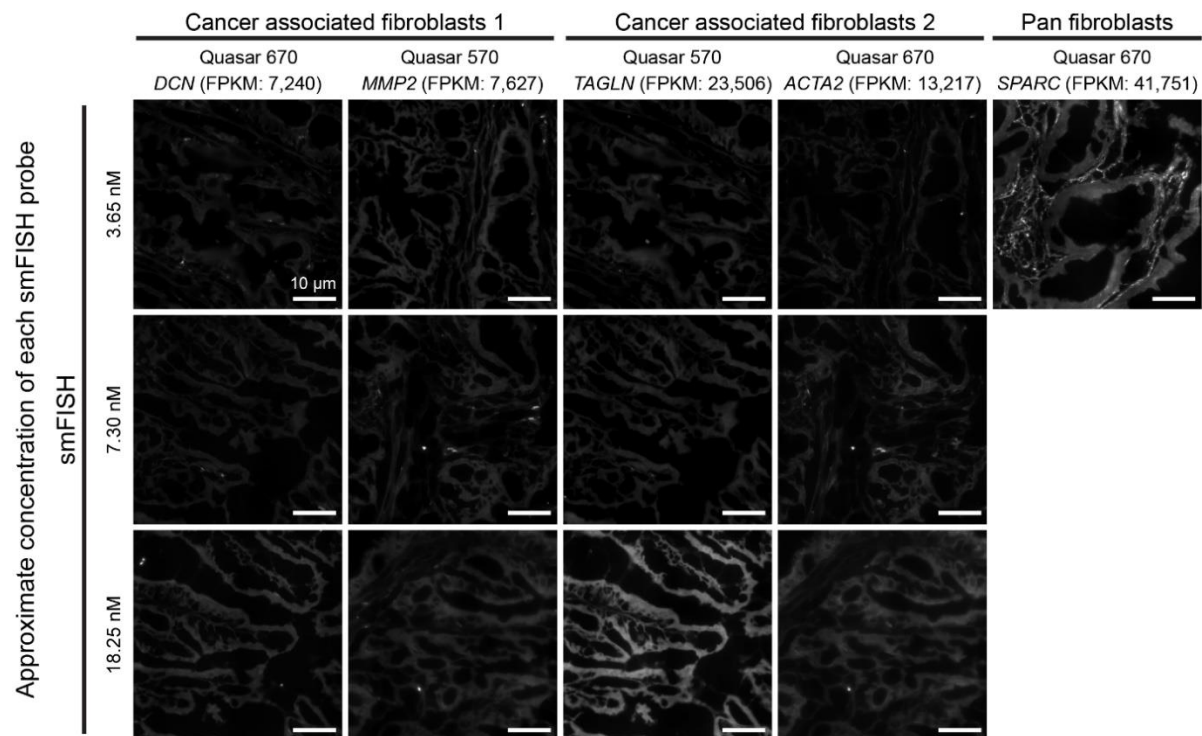
Source data are provided as a Source Data file.

Supplementary Fig. 17



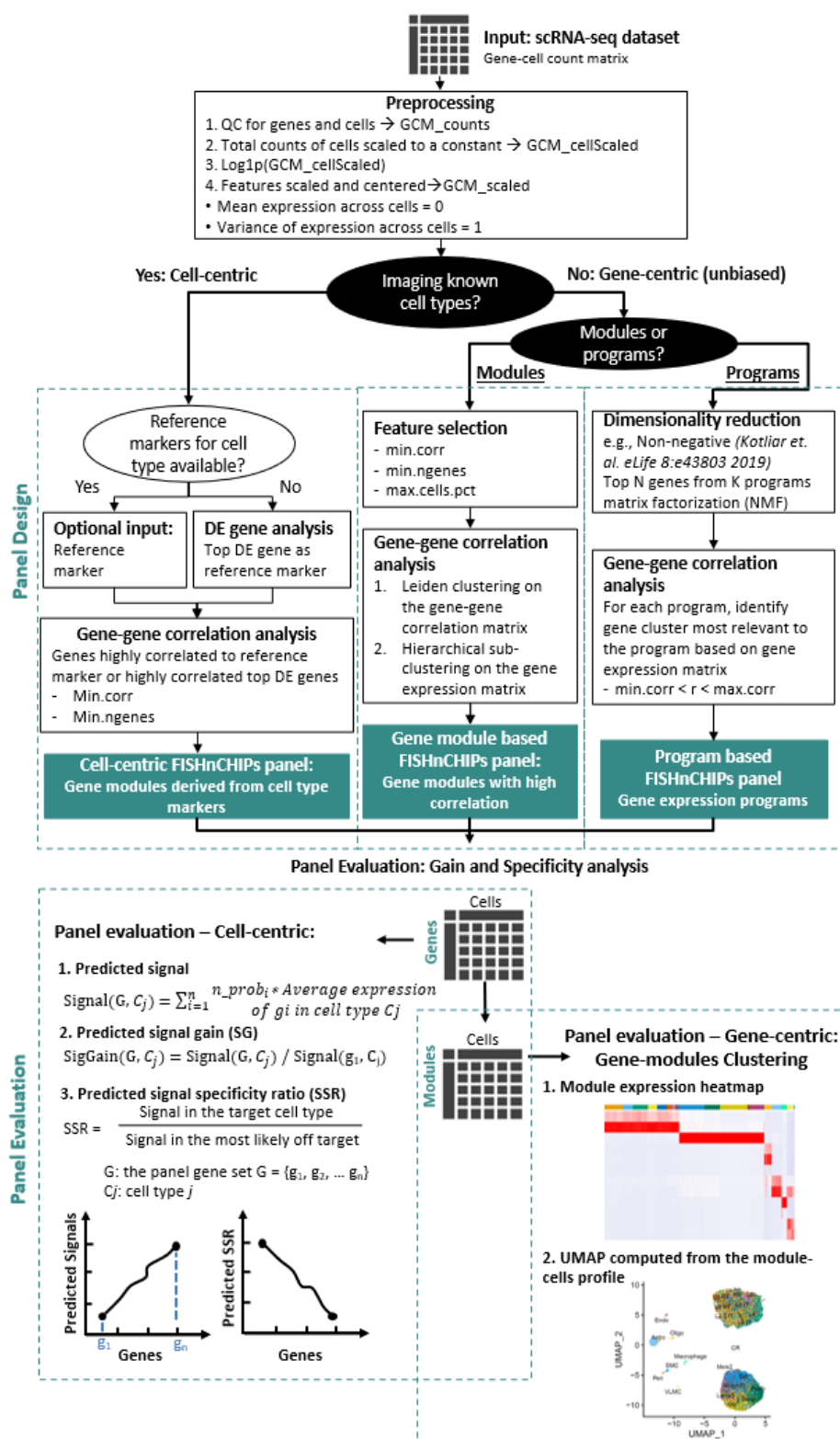
Supplementary Fig. 17 Three-color immunofluorescence staining of the immune marker CD68, CAF-1 markers PDPN, LUM and PDGFA, and CAF-2 markers αSMA and MMP2 on four slices of frozen human colorectal cancer tissue. All images are contrasted at 1 to 99.9 percentiles of the maximum intensity of each channel. Scale bar, 250 μm.

Supplementary Fig. 18



Supplementary Fig. 18. Two-color smFISH staining of the CAF-1 markers *DCN* and *MMP2*, and CAF-2 markers *ACTA2* and *TAGLN* at different concentrations on frozen human colorectal cancer tissue. *DCN* and *TAGLN* are stained together while *MMP2* and *ACTA2* are stained together on the same sample. *SPARC* smFISH staining for pan fibroblast is included as a positive control. Scale bar, 10 μm.

Supplementary Fig. 19



Supplementary Fig.19 FISHnCHiPs panel design software workflow. This flowchart summarizes the FISHnCHiPs panel design process.