Cell-Type Modeling in Mouse Brain Spatial Transcriptomics Data Elucidates Spatial Variation of Cellular Colocalization and Intercellular Communication

Francisco Jose Grisanti Canozo[1,2], Zhen Zuo[1], James F. Martin[1,2], Md. Abul Hassan Samee[1#]

1: Baylor College of Medicine, Houston, TX 77030, USA

2: Texas Heart Institute, Houston, TX 77030, USA

#: Lead Contact

E-mail address of corresponding author: samee@bcm.edu

Figure S1

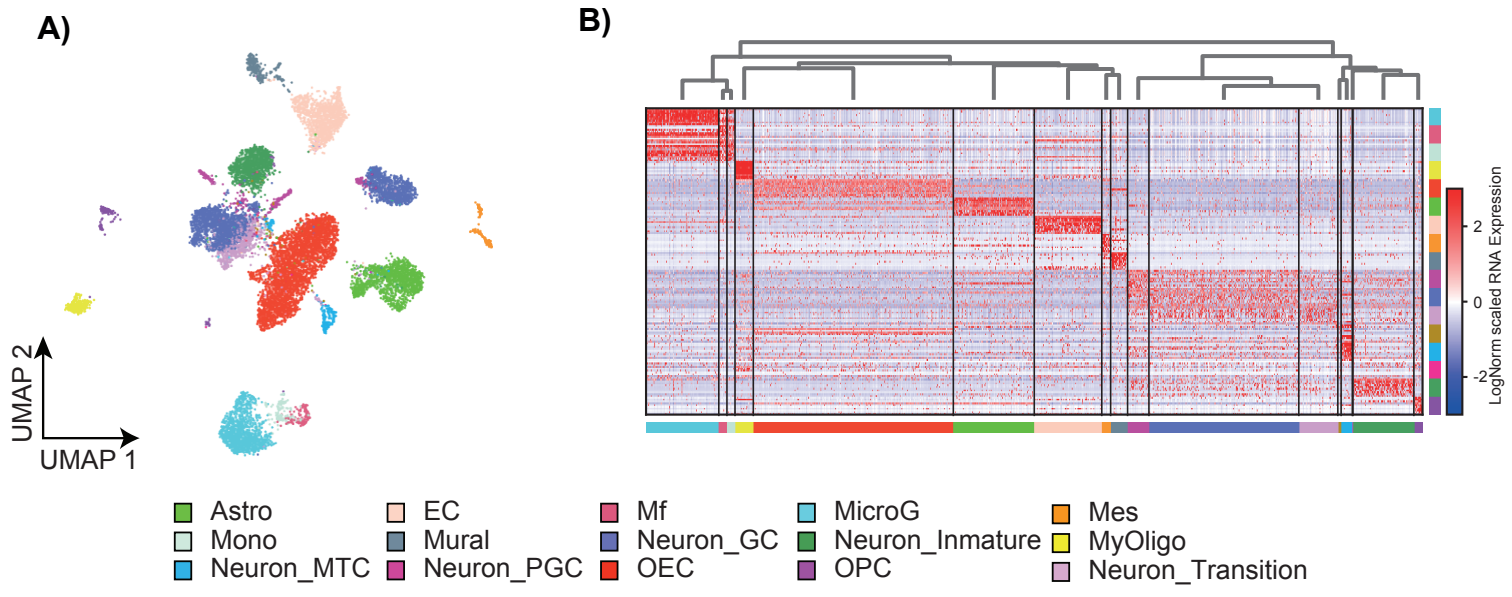

Figure S1. Cell-type clusters and cluster enriched genes in the mouse olfactory bulb scRNA-seq (Tepe et al., 2018) data, Related to Figure 2 and Figure 3. A) The data visualized using uniform manifold approximation and projection (UMAP). B) Heatmap illustrating the genes most highly enriched in each cluster, with each column representing a gene, and each row representing the log-normalized scaled RNA expression level of that gene in each cell-type cluster. Individual single-cell transcriptomes were colored according to cluster (cell-type) identity. Color bar represents in linear scale the log-normalized scaled RNA expression values.

Figure S2

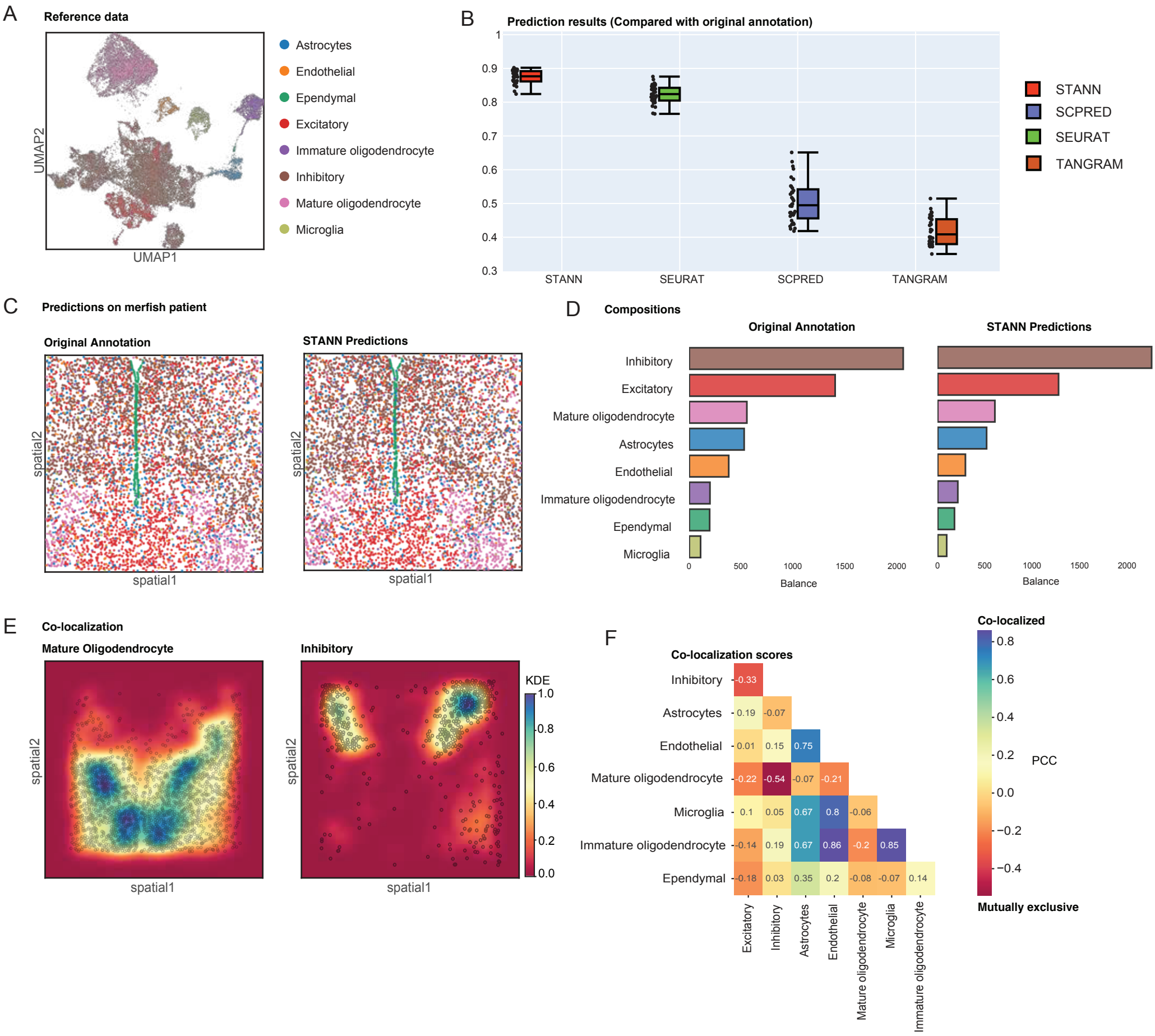

Figure S2. MERFISH Benchmarking, Related to Figure 2. A) UMAP of the reference scRNA-seq data (Moffitt et al., 2018). B) Prediction accuracies compared with original annotation. C) Spatial plot of STANN predictions and MERFISH original annotations for an example slide from Moffitt et al.'s data. The same slide is used for panels D-F. D) Cell-type compositions predicted by STANN and as annotated by Moffitt et al. E). Kernel density maps of mature oligodendrocytes cells vs. inhibitory cells showing completely opposite co-colocalization patterns. Color bar represents in linear scale the kernel density estimates (KDE) values. F) Heatmap of Pearson Correlation Coefficients (PCC) scores of cellular colocalization. Color bar represents in linear scale the PCC.

Figure S3

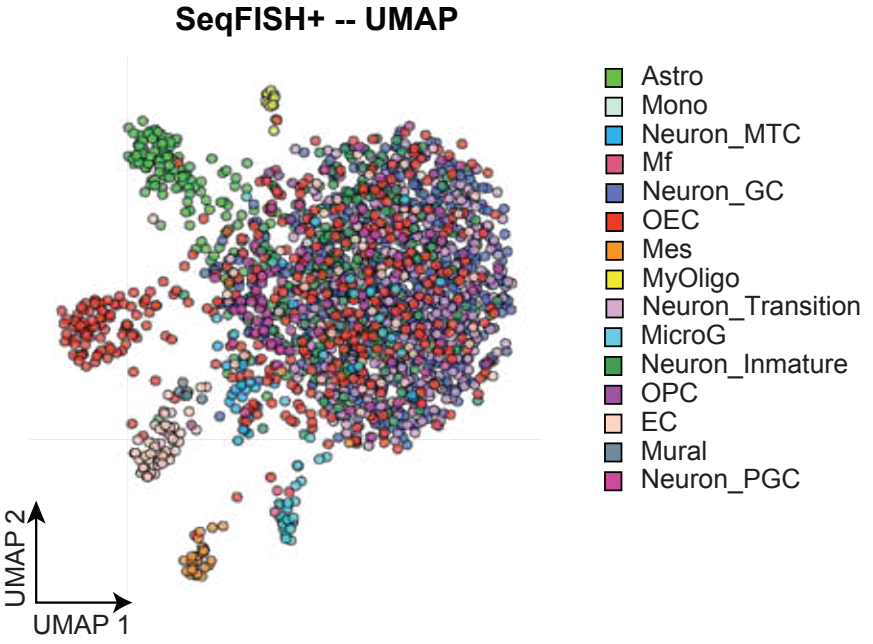

Figure S3. UMAP embedding of seqFISH+ data (Eng et al., 2019) with cells annotated using STANN, Related to Figure 3.

Figure S4

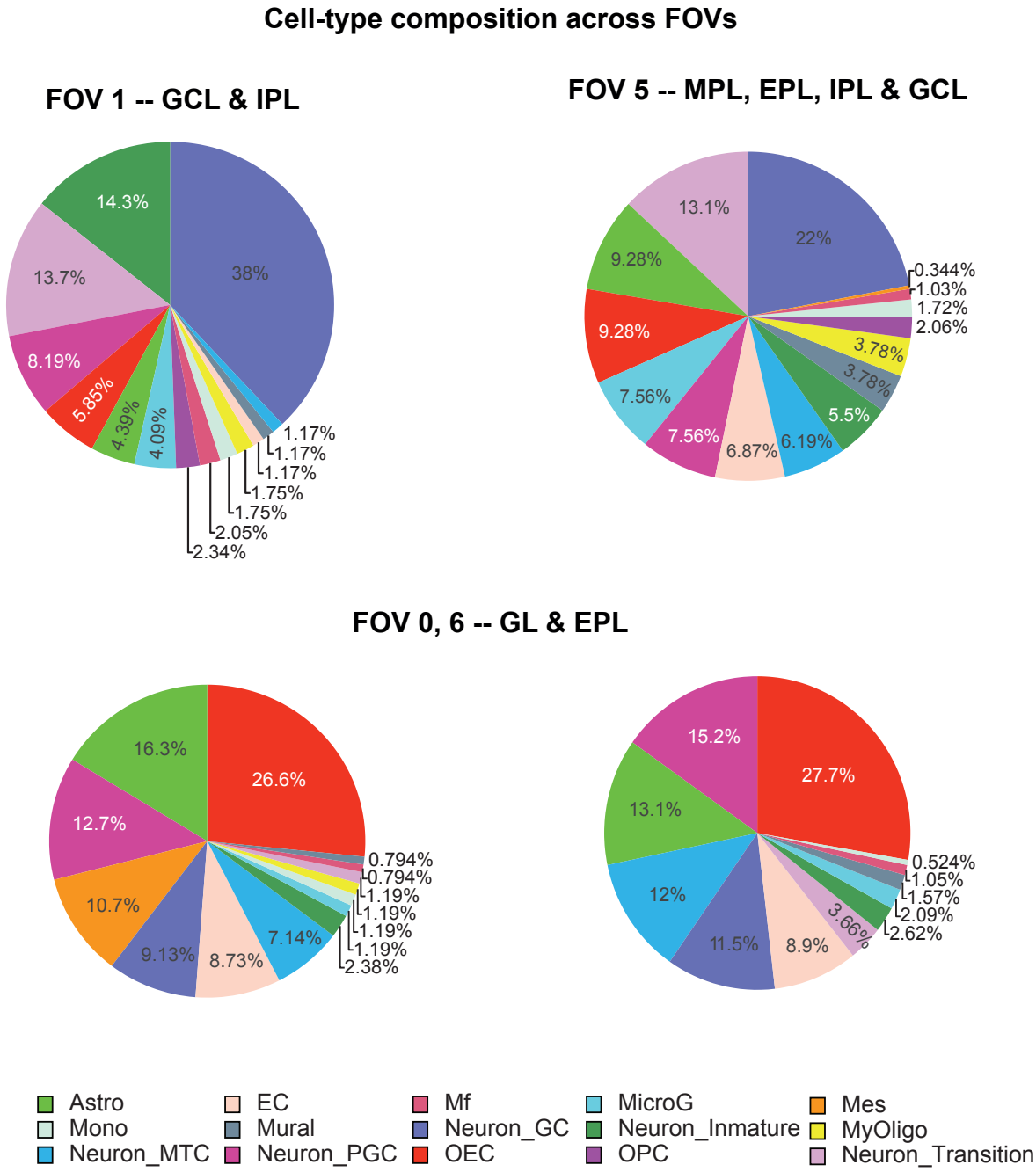

Figure S4. Composition of cell-types across different FOVs and morphological layers in MOB seqFISH+ data, Related to Figure 4.

Figure S5

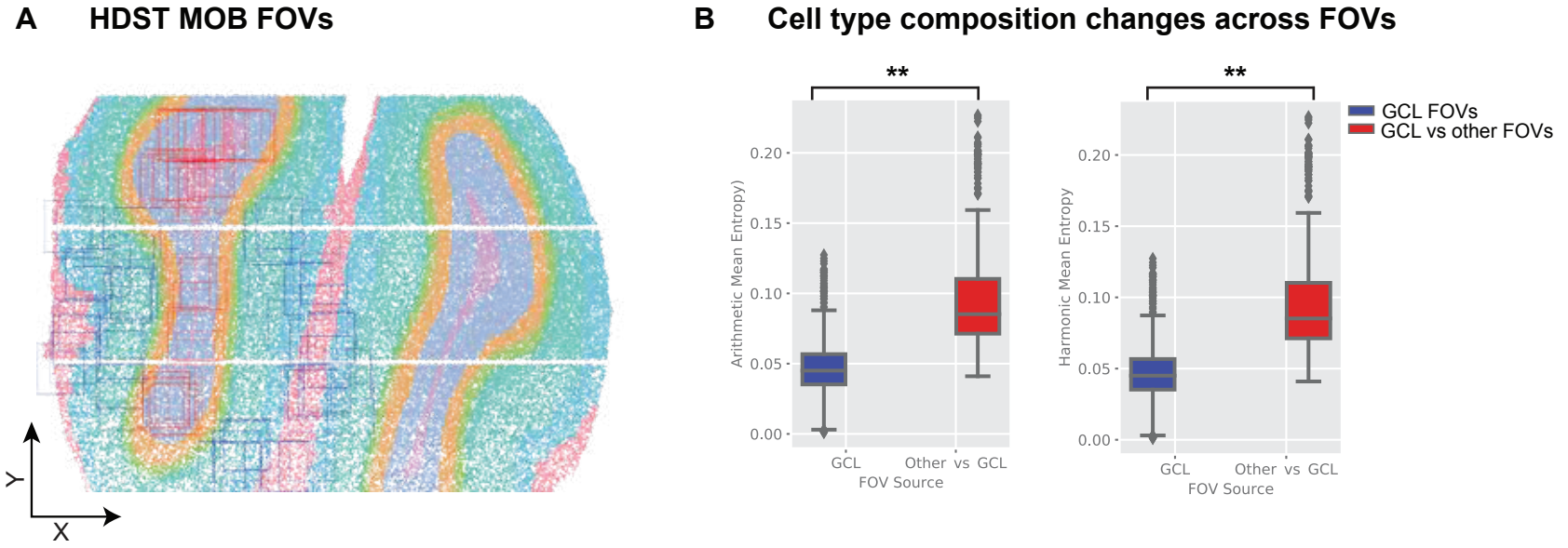

Figure S5. Changes in cell-type compositions in MOB HDST data, Related to Figure 4. A) Visualization of HDST MOB Granule Cell Layer data. We randomly selected 100 regions, each region of size 0.2 mm X 0.2 mm (the same size as an FOV in seqFISH+ data). Red regions are located within the GCL and blue regions are located in other morphological regions. B) Cell-type composition changes in terms of entropy, with pairwise comparisons of FOVs within the GCL or the GCL vs other morphological regions. Chi-square test (** FDR < 0.05)

Figure S6

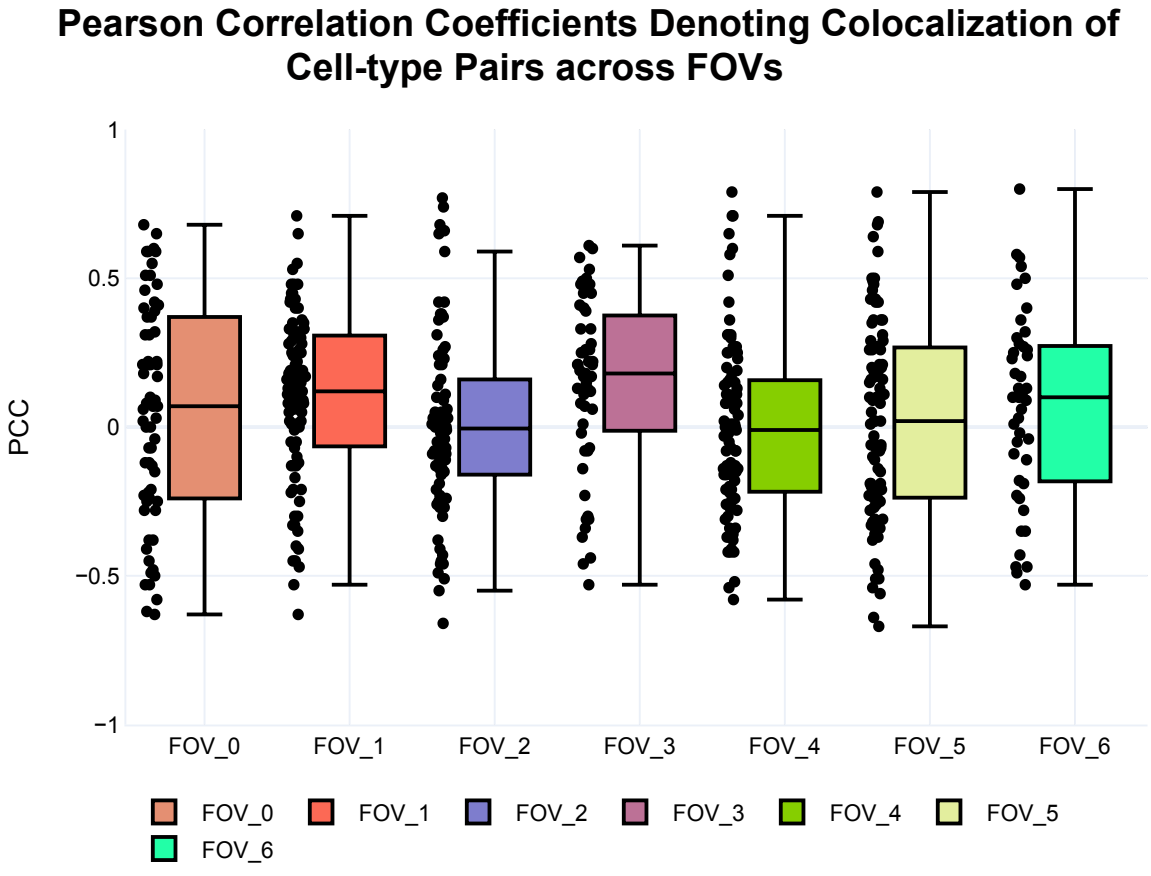

Figure S6. Colocalization between cell-type pairs in seqFISH+ data across different FOVs, Related to Figure 4. Shown are the Pearson Correlation Coefficients between the kernel densities of each cell-type pair. Horizontal lines in the boxplots mean the following: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

Figure S7

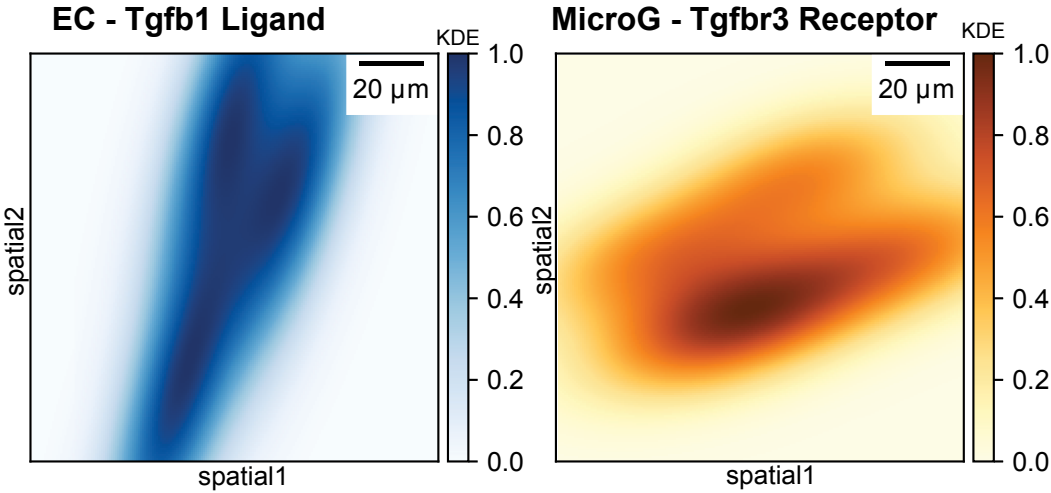

Figure S7. Kernel density estimates of endothelial cells expressing TGF beta ligand and microglias expressing TGF beta receptor in FOV 2, Related to Figure 5. Shown is the data from FOV 4. Color bar represents in linear scale the kernel density estimates (KDE) values.

Figure S8

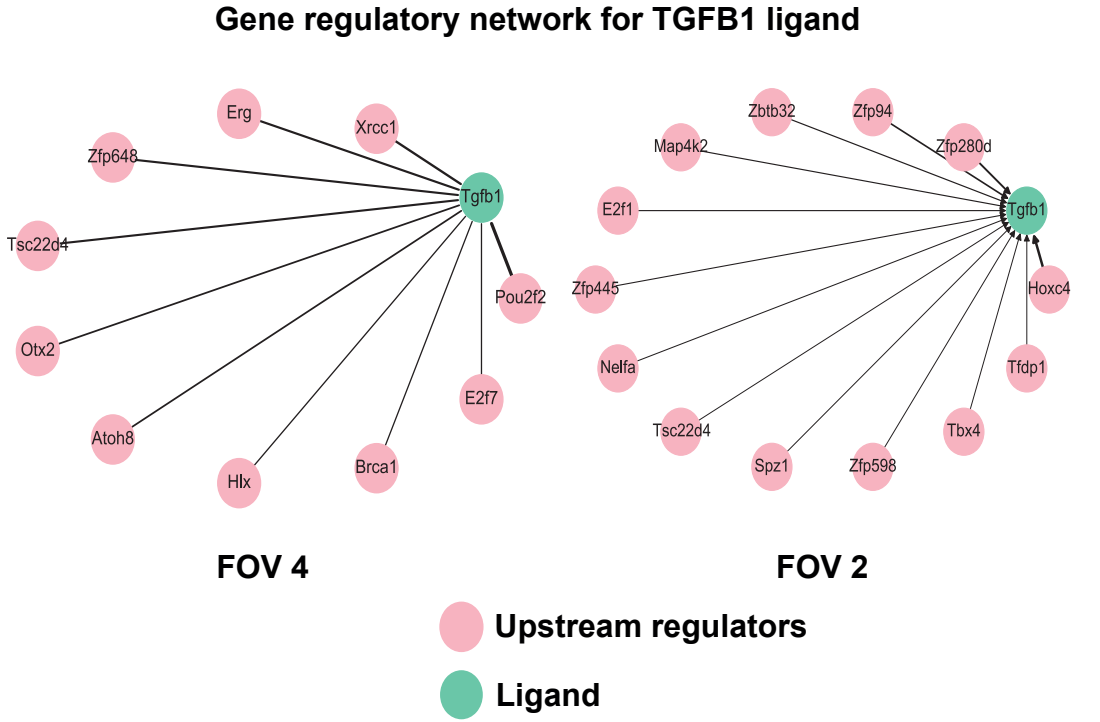

Figure S8. Regulators of the TGF beta ligand in FOV 4 and FOV 2, Related to Figure 5.

Figure S9

Up-stream regulators of marker genes across FOVs

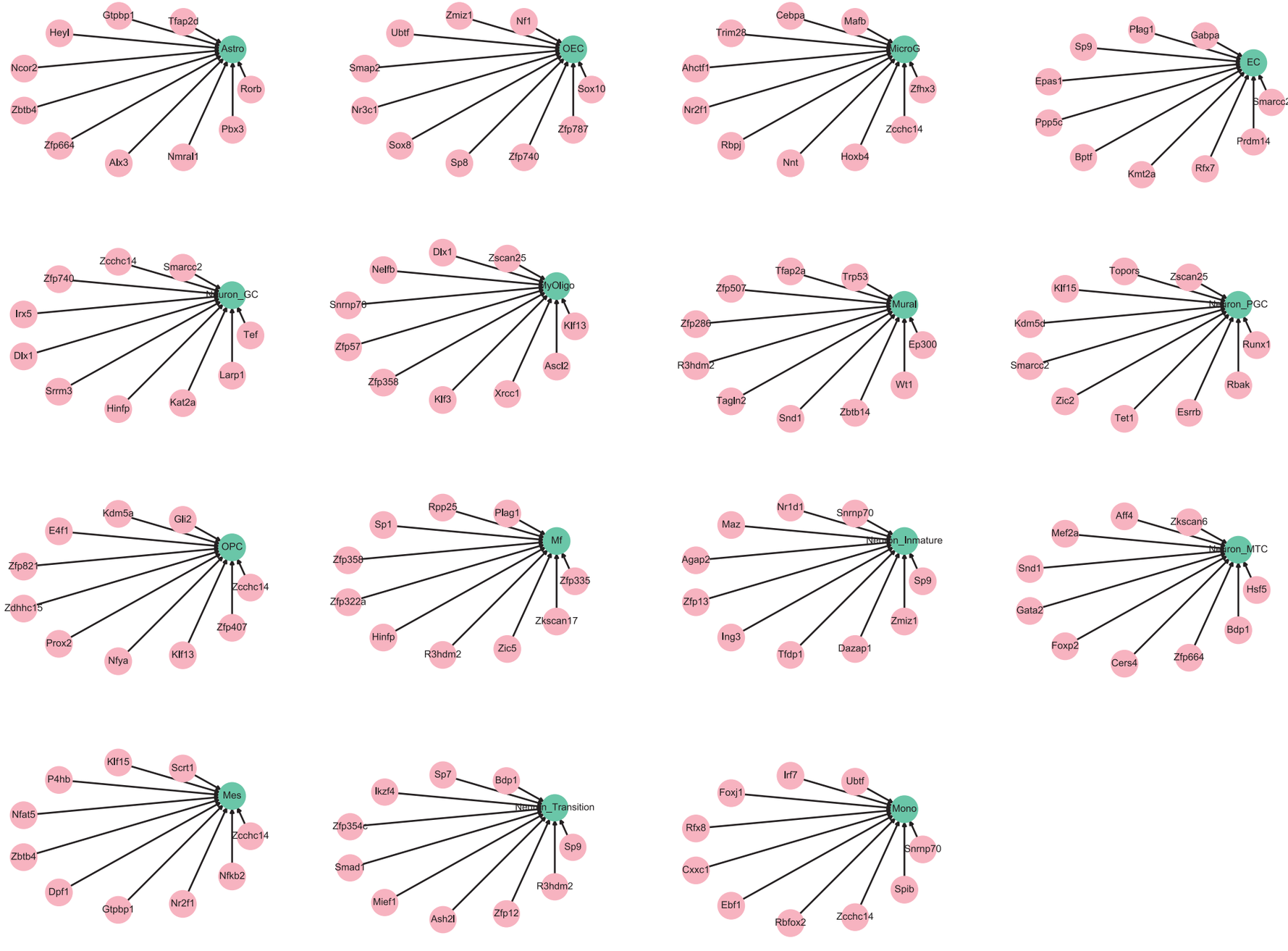

Figure S9. Genes that appear consistently across all FOVs as the upstream regulators of different cell-types' marker genes, Related to Figure 6. The networks were derived using the same pipeline as we used to show spatially localized ligands and receptors refine cell-subtypes, but by using marker genes across FOVs.

Figure S10

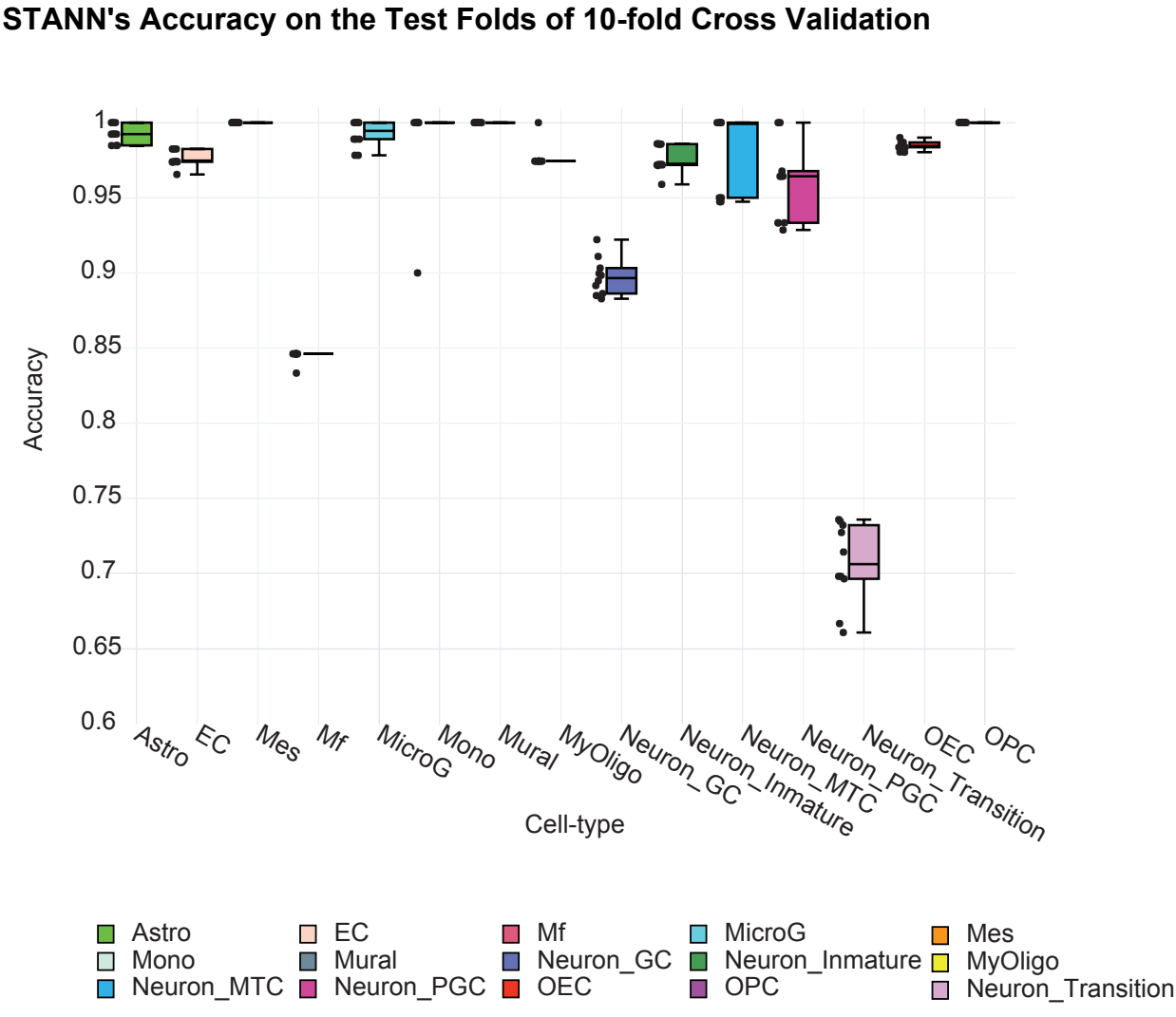

Figure S10. STANN's accuracy across cell-types for the test folds in 10-fold cross validation, Related to Figure 2 and STAR Methods. Mean accuracy across all cell-types is 95.15 +/- 0.325%.

Figure S11

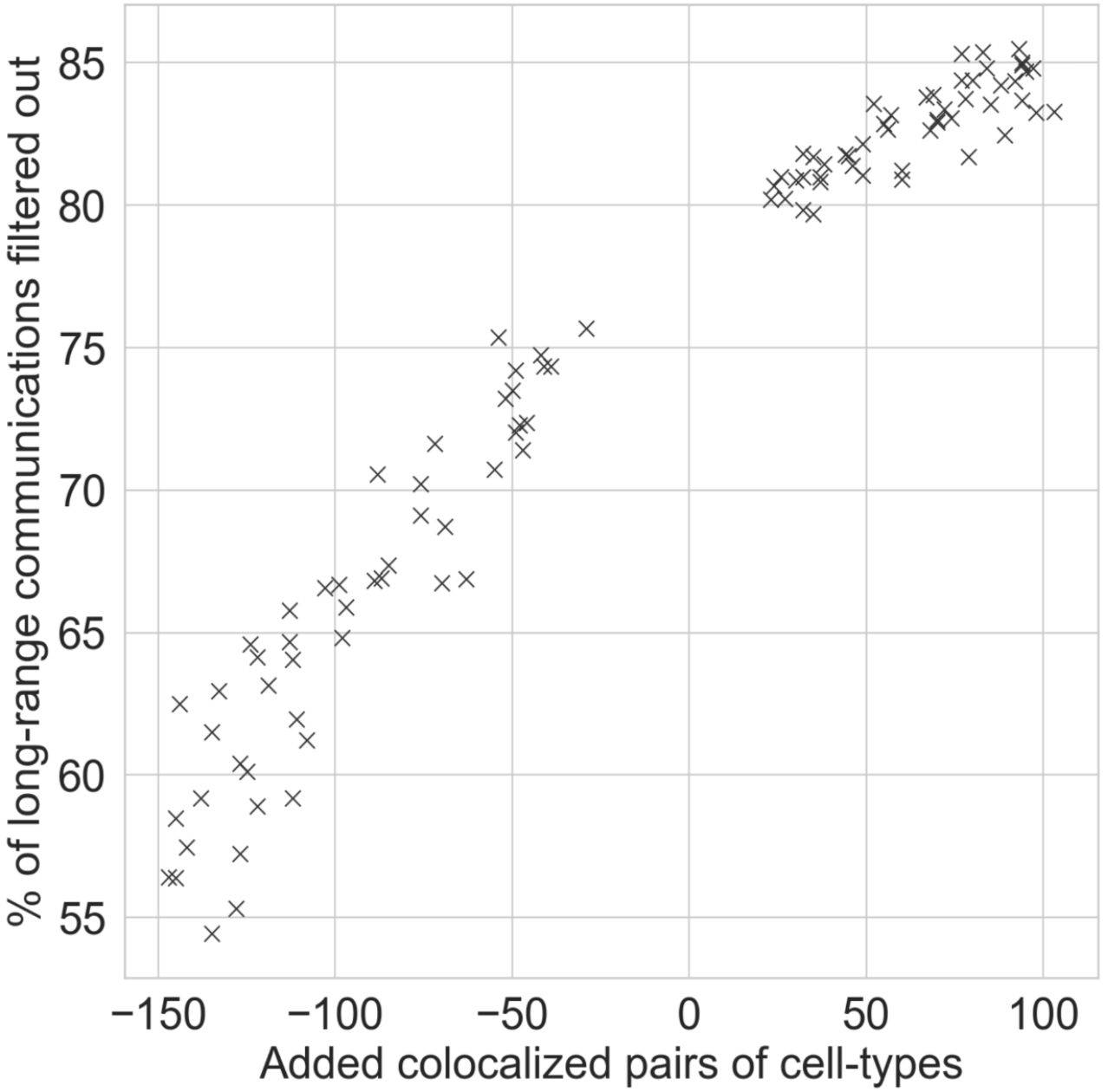

Figure S11. Simulation of synthetic false-positive predictions of long-range intercellular communication, Related to Figure 5. The Y-axis represents the percentage of long-range communicators filtered out and the X-axis represents the added colocalization pairs of cell-types. Negative values in the X-axis denote removal of colocalization (increasing spatial separation).