Integrative Analysis of Spatial and Single-Cell Transcriptomics Reveals Principles of Tissue Organization and Intercellular Communication in Mouse Olfactory Bulb

Francisco Jose Grisanti Canozo¹, Zhen Zuo¹, 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

#: Corresponding Author (samee@bcm.edu)

Abstract

Intercellular communication and spatial organization of cells are two critical aspects of a tissue's function. Understanding these aspects requires integrating data from single-cell RNA-Seq (scRNA-seq) and spatial transcriptomics (ST), the two cutting edge technologies that offer complementary insights into tissue composition, architecture, and function. Integrating these data types is non-trivial since they differ widely in the number of profiled genes and often do not share marker genes for given cell-types. We developed STANN, a neural network model that overcomes these methodological challenges. Given ST and scRNA-seq data of a tissue, STANN models cell-types in the scRNA-seq dataset from the genes that are profiled by both ST and scRNA-seq. The trained STANN model then assigns cell-types to the ST dataset. We apply STANN to assign cell-types in a recent ST dataset (SeqFISH+) of mouse olfactory bulb (MOB). Our analysis of STANN's assigned cell-types revealed principles of tissue architecture and intercellular communication at unprecedented detail. We find that cell-type compositions are disproportionate in the tissue, yet their relative proportions are spatially consistent within individual morphological layers. Surprisingly, within a morphological layer, there is a high spatial variation in cell-type colocalization patterns and intercellular communication mechanisms. Our

analysis suggests that spatially localized gene regulatory networks may account for such variability in intercellular communication mechanisms.

Keywords: Single-Cell RNA-Seq, Spatial Transcriptomics, Tissue Architecture, Cell Colocalization, Intercellular Communication, Deep Neural Network

Introduction

All organs are composed of a complex mixture of cells that are further organized into tissues that make up an organ. Organ physiology and function depend on the coordinated action of cells embedded within the tissues of the organ. In pathologic conditions, the normally coordinated organ physiology is disrupted resulting in loss of tissue architecture and eventual organ failure. A mechanistic understanding of a tissue depends on its cell-types and their transcriptional properties, spatial organization, and communication¹. Methodological advances to address these mechanistic details are essential for the study of development, physiology, and disease²⁻⁸. Spatial transcriptomics (ST) and single-cell RNA-seq (scRNA-seq) are complementary technologies, each with its own strengths and weaknesses, for studying complex tissues at a single-cell resolution². While scRNA-seq can profile the transcriptional expression of thousands of genes per cell (typically 10-20k)⁹ and characterizes a broad panel of cell-types, it does not preserve the spatial context of cells within the tissue¹⁰. In order to uncover the spatially varying principles of tissue architecture and intercellular communication, alternative strategies that maintain spatial information are required¹¹. The ST technologies complement scRNA-seq by preserving individual cells' spatial location in their native tissue¹⁰. However, current ST protocols profile the transcriptional expression of only about half as many genes as scRNA-seq (typically ~1k-10k)¹⁰, an issue that can make it problematic to identify cell-types in ST datasets. In particular, when the marker genes of different cell-types are absent in an ST

dataset, it is unclear if one can assign correct types to the cells in that dataset. Errors in cell-type assignment, in turn, may lead to inaccurate biological conclusions from an ST data analysis.

To harness the potentials of both scRNA-seq and ST, we developed STANN (Spatial Transcriptomics Cell Types Assignment Using Neural Networks) -- a neural network that learns from an scRNA-seq dataset to assign cell-types to the cells of an ST dataset (Fig. 1A,B). The inputs to STANN are the transcriptional expression of genes that are shared between the two datasets and an assignment of cell-types to the scRNA-seq cells. Given this information, STANN first learns the mapping of cell-types to scRNA-seq cells (Fig. 1A). STANN next predicts the cell-type of each cell in the ST dataset given the transcriptional expression of the shared genes in that cell (Fig. 1B). We applied STANN to learn cell-type mapping from an scRNA-seq dataset of mouse olfactory bulb (MOB)¹² and assign cell-types to a recent ST dataset (sequential Fluorescence In Situ Hybridization plus; seqFISH+¹³) of MOB. Utilizing these annotated cell-types, we pursued a detailed investigation of the architecture and intercellular communication mechanisms in MOB (Fig. 1C).

Our analysis revealed that the proportions of cell-types vary across different morphological layers of MOB, but remain consistent within a given morphological layer. To investigate the spatial organization of different cell-types, we developed methods akin to colocalization analysis in microscopy images¹⁴. Interestingly, these methods revealed that colocalization patterns between a pair of cell-types vary significantly across different spatial locations of the MOB and even within a morphological layer. This variation ranges from spatially separated occurrences to high degrees of colocalization between the two cell-types. We also find a remarkable variation in intercellular communication mechanisms between the same pair of cell-types, defined by

significant co-expression of receptor-ligand pairs¹⁵, across different locations within a morphological layer. Importantly, by using different sets of receptor-ligand pairs at different spatial locations, a given pair of cell-types also activates different sets of genes at those locations. This suggests that even when cell-type identities do not change, the overall transcriptional signatures of the corresponding cells can change across the tissue giving rise to spatially localized gene regulatory networks.

In summary, the STANN model and the downstream analyses pinpointed the consistent and the variable aspects of the architecture and intercellular communication mechanisms in MOB in unprecedented detail. To our knowledge, this is the first such systematic and rigorous attempt to delineate the principles of tissue architecture and intercellular communication by harnessing the unique features of the ST and scRNA-seq technologies.

Results

STANN: a Deep Neural Network Model for the Supervised Mapping of Cell-Types from scRNA-seq to ST Data

Deep neural networks have recently shown remarkable performance in classification tasks¹⁶. The models can efficiently deal with high dimensional data and learn sufficiently complex, non-linear functions in a data-driven manner to map data points to their respective classes. With proper statistical measures, one can identify the architecture and the parameters that yield an accurate and generalizable model¹⁷. The STANN (Spatial Transcriptomics Cell Types Assignment Using Neural Networks) model in this work implements a fully connected deep neural network (Fig. 1A). The inputs to the model are gene expression values of a cell (input layer) and the outputs are the probabilities of the cell belonging to different cell-types (output layer). Each intermediate layer comprises neurons that linearly combine inputs from the neurons

of the previous layer and propagate the result through a nonlinear activation function to the next layer (Methods). The linear weights in each neuron are optimized using the backpropagation algorithm¹⁸.

We trained STANN on 15 cell-types with 11744 cells and 9031 genes. In order to identify an accurate and generalizable model, STANN performed a hyper-parameter search for the optimal number of intermediate layers, numbers of neurons in each layer, and the activation function for neuron's in each layer. This identified an optimal model with two hidden layers (with 160 and 145 neurons) and the hyperbolic tangent function as the nonlinear function. In order to test the generalizability of STANN on unseen data, we performed a 10-fold cross validation. On each fold, the data is split into 80% training and 20% testing. Over the 10 folds, the model showed an average accuracy of $99.5 \pm 0.1\%$ on the training data and $95.4\% \pm 0.5\%$ on the separately held-out test data.

STANN also achieved remarkable accuracy for individual cell-types, even for cell-types with small numbers of cells (Fig. 2A). We attribute this capability of STANN to its use of a class-balance aware loss function, that penalizes under or over-represented cell-types. Other current methods do not use this approach¹⁹; hence it is likely that those methods will be biased to the majority classes, underfit rare cell-types, and their reported performances are likely inflated (Methods).

To compare STANN with other available tools, we noted that the majority of these methods rely on marker genes to assign cell-types in the ST dataset²⁰⁻²². However, the MOB scRNA-seq and SeqFISH+ datasets share only 33% of the marker genes (Supplementary Table S1), which makes it unlikely that marker gene based methods will be effective in this case. We therefore used the widely used Seurat method²³, which first aligns the two datasets in low dimensions

before utilizing the marker genes. STANN performed better than Seurat by a margin of 4.9% in our benchmarking runs (Fig. 2B, Methods). We also compared STANN against a scPred²⁴, another robust method that combines unbiased feature selection from a reduced-dimension space with support vector machines. STANN performed better than scPred by 10.9% in our benchmarking runs (Fig. 2B, Methods). Briefly, we compared Seurat and scPred against STANN over 50 independent runs. In each run, we simulated an ST dataset by taking a random subsample of ~9k genes (equal to the number of shared genes between scRNA-seq and ST datasets) from the original scRNA-seq data and tested each method on this subsampled data. We found that STANN outperformed Seurat and scPred in all 50 runs, with accuracies of 99.2% $\pm$ 0.21% vs. Seurat's 94.4% $\pm$ 4.51% and scPred's 88% $\pm$ 0.64% (Fig. 2B). STANN's performance was consistently high in rare cell types (<30 cells). For example, for monocytes (Mono), macrophages (MF), oligodendrocyte progenitor cells (OPC), and mural cells, STANN showed improvements of 0.4 – 17% over Seurat and scPred (Fig 2C).

STANN Provides a Spatially Resolved Annotation of Cell-Types in the Mouse Olfactory Bulb

The mouse olfactory bulb (MOB) SeqFISH+ dataset¹³ comprises six fields of view (FOV) from a slice of adult MOB (Fig. 3A). Each FOV is approximately 0.2 mm X 0.2 mm in size and captures the expression of 10,000 genes from 2050 cells¹³ (Fig. 3A). The MOB is structured into several layers, known as the MOB morphological layers. In a coronal image of MOB, these layers appear in the following order from the center to the outside: granule cell layer (GCL), internal plexiform layer (IPL), mitral cell layer (MCL), external plexiform layer (EPL), and glomerular layer (GL). The FOVs in the SeqFISH+ data span all of these morphological layers¹³ (Fig. 3A; Supplementary Table S2). This dataset has 9031 genes in common with our previously

published scRNA-seq data¹² of MOB. In the rest of this manuscript, we refer to these genes as the “shared” genes. Importantly, the shared genes include only 33% of the marker genes for MOB cell-types (Supplementary Table S1) – suggesting that marker gene based methods^{20–22} would not be ideal for predicting cell-types in this SeqFISH+ data.

As discussed above, we trained STANN on our scRNA-seq dataset¹² and utilized the trained model to assign cell-types to the SeqFISH+ data. From the original scRNA-seq dataset¹², 14 major cell-types were found in the WT condition (Supplementary Table S3; Fig. S1). Consistent with previous studies^{4,25}, STANN showed an abundance of olfactory ensheathing cells in FOVs originating from the GL (Fig. 3B). STANN also predicted neuronal granule cells as the most abundant cell-type in these six FOVs. This is expected since the majority of the FOVs are located within the GCL or map to an extension of it (Fig. 4A). Overall, STANN revealed a high imbalance in the proportions of cell-types in SeqFISH+ data (all FOVs considered together; Fig. 4B). This likely reflects MOB biology and is not an artifact of STANN, since the MOB scRNA-seq data has a similar imbalance of cell-type proportions (Fig. 4A). However, STANN found subtle differences between the two datasets in terms of the relative abundance of cell-types. For example, the most abundant cell-type across the six FOVs of SeqFISH+ data is neuronal granule cells, which is the second most abundant cell-type in the scRNA-seq data. Similarly, the most abundant cell-type in scRNA-seq, olfactory ensheathing cells, is the third-most abundant cell-type in SeqFISH+. These variations are expected since the SeqFISH+ FOVs constitute a small sample of the scRNA-seq data. Importantly, these variations imply that STANN did not overfit the scRNA-seq data.

STANN also highlighted the importance of utilizing models to learn cell-types from shared genes. In particular, the SeqFISH+ data UMAP (Uniform Manifold Approximation and Projection;

Fig. 3C) built from the shared genes did not converge to a good separation of neuron clusters, implying that the shared genes do not capture sufficient variation to have a straightforward classification of those cell-types. However, as STANN shows, it may be possible for nonlinear functions to integrate information from the shared genes and accurately classify all cell-types.

STANN Predictions Reveal Consistency in Cell-Type Composition but Variability in Colocalization Patterns across the Tissue

STANN predictions revealed that the proportion of cell-types are consistent within a given morphological layer of MOB, but may vary between two different morphological layers. Interestingly, the patterns of colocalization between a pair of cell-types can vary substantially even within a given morphological layer.

We define the cell-type composition of an FOV as the proportion of different cell-types in it. Within a morphological layer, STANN revealed that the cell-type compositions of each FOV are very similar (based on both Chi-square test and relative entropy; Methods; Supplementary Table S4). For example, the three FOVs located within the GCL have nearly identical cell-type compositions (FOVs 2, 3, and 4 in Fig. 5A; Supplementary Table S4). However, FOVs differ in cell-type compositions when they span different morphological layers. For example, FOV 1 (spanning GCL and IPL) and FOV 5 (spanning GCL, IPL, EPL and MPL) have significantly different cell-type compositions (Fig. S2, Supplementary Table S4). The same applies for the FOVs 0 (spanning GL and EPL) and 4 (spanning GCL).

We next quantified colocalization between any pair of cell-types in a given FOV. For each cell-type predicted by STANN in an FOV, we calculated a multivariate kernel density²⁶ map from its locations within the FOV (Fig. 5C; Methods). We computed the Pearson correlation coefficient between the probabilities of observing the two cell-types across all spatial

coordinates in that FOV. This colocalization analysis, built upon previous studies of colocalized fluorescence probes²⁷, identified cell-type pairs with significant colocalization (strong positive correlation) or cell-type pairs that were spatially separate (strong negative correlation) in each FOV (Fig. 5B,C). Interestingly, we found a high variation in colocalization patterns between every pair of cell-types across the FOVs (Fig. S3). The three FOVs located within the GCL (FOV 2, 3, and 4) demonstrate this point. The most conservative prediction will be that FOVs that share the same morphological layer or are adjacent, will be more similar than FOVs located in different morphological layers or at a distance. While this conservative prediction holds for 36% of cell-type pairs in FOVs 2, 3, and 4, we found that 64% of the cell-type pairs in these FOVs show highly variable colocalization patterns, ranging from spatial separation to significant colocalization across the FOVs (Figure 5B).

Intercellular communication is critical for tissue function^{28–33} and is also associated with disease states^{21,34,35}. Identifying the communicating pairs of cell-types and their communication mechanisms (in terms of receptors and ligands) is currently a major focus of scRNA-seq studies. Recent studies suggest that the communications between certain cell-types may be enriched within specific parts of a tissue (commonly termed microenvironments or niches^{34,36}). However, to what extent and how intercellular communication varies across a tissue is still an open question¹⁵. It is also unclear if the communication between two cell-types changes with their abundance or colocalization in different parts of a tissue.

We applied CellPhoneDB¹⁵, an intercellular communication analysis tool, on STANN's predicted cell-types in each FOV of the SeqFISH+ data (Methods). For every pair of cell-types in an FOV, CellPhoneDB identifies the receptors and ligands that are significantly and specifically enriched between those cell-types. However, CellPhoneDB does not take the spatial location of cell-types

into account and could incorrectly infer communication between two spatially separated cell-types, e.g., when they each communicate with a third cell-type but not with each other (Fig. 6A). To make our analysis robust against this type of incorrect inferences, we have applied an additional filtering on the intercellular communications between spatially separated cell-types (Methods). This eliminated 63.07% of the communications inferred by CellPhoneDB. Although we were over-conservative in this step, the possibility of such a large fraction of inferences being incorrect underscores the importance of studying intercellular communication with spatial information.

This analysis discovered widespread intercellular communication in MOB while capturing bona fide receptors and ligands. For example, we found endothelial and microglial cells communicating through the Type 3 TGF- β receptor and the Type 1 TGF- β ligand in FOV 4 (Fig. 6B). Previous studies have experimentally validated this TGF family receptor-ligand pair in MOB^{13,37}. These two cell-types are also colocalized in this FOV (Pearson correlation coefficient for colocalization = 0.71) (Fig. 6C), further supporting the possibility of their communication. Similarly, we found neuronal transition cells and olfactory ensheathing cells to communicate through the BMP pathway ligand Bmp7 and the receptor Slamf1 in FOV 6, although the two cell-types are spatially separated in this FOV (Pearson correlation coefficient for colocalization = -0.42). The BMP pathway is classically known to mediate long-range intercellular communication^{38,39}. As expected overall, we found the most abundant receptors and ligands are from the well-known signaling pathways such as Fibroblast growth factor (FGF), Notch, Bone morphogenetic protein (BMP), Transforming growth factor beta (TGF- β), and Wnt^{40–43}.

The analysis also revealed the details of spatial variation in intercellular communication across the FOVs. On average, a given cell-type pair communicates in 4 FOVs (range: 1-7). We observe

this variation even when we focus on FOVs from the same morphological layer. For FOVs 2, 3, and 4, which are from the GCL layer, we see that 25% of cell-type pairs communicate in one FOV, 39% in two and 36% in all three FOVs. Some intercellular communications are specific to an FOV or to a particular morphological layer. For example, olfactory ensheathing cells to astrocytes cells communicate only in FOVs 2, 3, and 4 located specifically in the GCL layer. Similarly, myelinating-oligodendrocyte cells and neuronal transitional cells communicate only in FOV 0 that covers the GL, EPL layers. On the other hand, certain pairs of cell-types communicate in nearly every FOV and their communications appear to be a common aspect of MOB. These examples include the intercellular communication between astrocytes and endothelial cells, and neuronal granule cells and endothelial cells.

Interestingly, we found that a cell-type pair often uses different sets of receptors and ligands in different FOVs. On average, the number of receptors and ligands used by two cell-types in different FOVs can change by 3.5-fold (Fig. 6D; ratio of the maximum and the minimum number of receptors-ligands used across different FOVs; range: 0-19), with minimal overlap between the sets of receptors and ligands across FOVs (Median similarity between cell-type pairs in receptor-ligand usage: 3%, range: 0-42%). We observed this variation even across FOVs where the two cell-types colocalize. For example, endothelial and neuron transitional cells colocalize in FOVs 3 and 5 (Pearson Correlation Coefficient > 0.5), but only 7% of their 17 ligand-receptor pairs are common in these FOVs. We found that a higher co-localization between cell-types shows a positive trend with a higher number of ligands and receptors (Fig. 6E), although there are instances when cell-types co-localize but do not communicate.

The above analyses implied spatially variable usage of receptors and ligands for intercellular communication. We next asked if the gene regulatory network (GRN) of a receptor or ligand,

i.e., the genes in regulatory relationships with it, in a given cell-type may change across FOVs.

This concept of spatially localized GRNs has been noted in the literature⁴⁴. However, to our knowledge, the existence and role of spatially localized GRNs in mediating intercellular communication have never been discussed.

We applied SCENIC⁴⁵, a GRN construction algorithm, on STANN's predicted cell-types in each FOV (Methods). We found that the GRN of nearly every receptor and ligand in every cell-type changes across FOVs, suggesting that spatially localized GRNs are very common in MOB. For example, endothelial cells use the TGF β -1 ligand to communicate with microglia cells in FOVs 6 and 4 (Fig. S4), yet only 8.3% of genes in the GRN of TGF β -1 in endothelial cells are shared in these two FOVs. On average, only 20.38% (Fig. 6F; range 0.93%-48.46%) genes in the GRN of a receptor or ligand in a cell-type were common across FOVs.

Overall, our analyses in this section revealed widespread spatial variation in intercellular communications. A pair of cell-types often interacts in a region-specific manner and may communicate through utilizing very different sets of receptors and ligands. These mechanistic variations are associated with spatially varying gene regulatory networks, pointing to a new level of heterogeneity in single-cell data where cells of the same type show subtle changes in GRNs of receptors and ligands depending on their spatial location.

Methods

Cell type annotation in scRNA-seq data

We collected the annotated cell-types from our previous scRNA-seq dataset of mouse olfactory bulb (MOB)¹². Following the common steps for cell-type annotation in scRNA-seq data^{23,46}, we first ranked genes based on the variance-to-mean ratio of their expression values. Then, using

principal component analysis, we computed a low dimensional manifold that best explains the variance in the data, and further reduced the dimensionality into two dimensions using uniform manifold approximation and projection (UMAP). We then clustered cells in this two-dimensional space using the Louvain community detection algorithm. Finally, based on marker genes detected using the Wilcoxon rank-sum test, we annotated the clusters (Fig. S1).

Data normalization

For the scRNA-seq data, Scanpy toolkit 1.3⁴⁷ was used to perform standard pre-processing, including normalization of library-size with 10,000 reads per cell, log transformation, regression of effects of total counts per cell and the percentage of mitochondrial genes expressed. We finally scaled the data to unit variance (clipping values exceeding standard deviation greater than 10).

Following the original SeqFISH+ manuscript, we performed a quantile normalization on the normalized count matrix of each gene in the scRNA-seq and seqFISH+ datasets to correct for cross-platform differences^{13,48,49}. Finally, we applied Min-Max scaling to scale the data between 0 and 1. This final step was used to ensure efficient updates in the gradient descent optimization of our neural network model.

STANN architecture and training

We implemented a multi-layer perceptron model and searched for its optimal architecture (using random initializations in terms of the number of hidden layers, the number of nodes in hidden layers, and the activation functions) using the TensorFlow framework. This approach identified a model with two hidden layers (with 165 and 145 nodes, hyperbolic tangent activation function, and a learning rate of 0.01) as the optimal choice for our preliminary model on the MOB

scRNA-seq data. The input layer of this model has size equal to the number of genes that have been profiled under both scRNA-seq and sc-ST. The output layer is a softmax layer with size equal to the number of cell-types detected above from the scRNA-seq data.

The softmax output layer is defined as:

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^k \exp(x_j)}$$

Where all the x_i values are the elements of the output of the fully connected layer immediately preceding it, k is the number of classes and $\sum_{j=1}^k \exp(x_j)$ is the normalization term, which ensures that all output will sum to one, and thus constitute a probability distribution.

The hyperbolic tangent function is defined as:

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$$

Where $\tanh(x)$ is the hyperbolic tangent function evaluated at x .

We used categorical cross-entropy (CE) as the loss function to train the model. CE is defined as:

$$\text{CE} = - \sum_{i=1}^N \sum_{j=1}^k t_{i,j} \cdot \log(p_{i,j})$$

Where N is number of observations (data points), k is the number of classes, $t_{i,j}$ is the binary indicator whether class j is the correct classification for observation i . Finally, $p_{i,j}$ is the predicted probability that observation i belongs to class j .

However single-cell datasets are usually imbalanced, so we used an alternative loss function that penalizes under or over-represented cell-types based on their number of cells in the dataset, defined as:

$$\text{Weighted CE} = - \sum_{i=1}^N \sum_{j=1}^k w_j \cdot t_{i,j} \cdot \log(p_{i,j})$$

Where w_j is class weight balance defined as:

$$w_j = \frac{N}{k \cdot N_j}$$

Where N_j is the number of observations in class j . In our analysis, this loss function remedied the class-imbalance issue and was our key to achieving superior performance for cell-type assignment¹⁹. Most of the current methods are not sensitive to class-imbalances. Hence it is likely that those methods will be biased to the majority classes, underfit rare cell-types, and their reported performances will likely be inflated.

STANN Implementation

STANN is implemented in Python 3 using Keras and its TensorFlow⁵⁰ back end. The model can be trained in a local computer within minutes for a dataset of ~10K cells and ~10K genes. Training on CPU or GPU is supported using Keras and TensorFlow. The hyperparameter optimization is implemented using keras-tuner.

STANN comparison against alternative methods

We compared Seurat²³ and scPred²⁴ against our model over 50 runs. In each run, we subsampled 2500 cells randomly out of the main scRNA-seq dataset. From this subsampled

dataset, we then eliminated all the genes that are not shared between the original scRNA-seq and SeqFISH+ datasets; but we retained the cell-types as computed from the original scRNA-seq dataset. This way, we could treat the subsampled dataset as a simulated ST data where we already know the true cell-types. We evaluated each method on this subsampled dataset through 10-fold cross-validation. We found that Seurat achieved a mean estimated accuracy = $94.4 \pm 4.51\%$, and scPred a mean estimated accuracy = $88.26 \pm 0.64\%$, compared to our method with a mean estimated accuracy = $99.2 \pm 0.21\%$ (Fig. 2B).

Colocalization analysis

The method is based on principles of co-localization analysis in microscopy images²⁷. Briefly, we first use kernel density estimation to estimate the density of each cell-type in each FOV. Kernel density estimation is a method to estimate the density of a random variable from a finite sample size²⁶. For any pair of cell-types, we then used the Pearson's Correlation Coefficient (PCC)²⁷ between their kernel densities in an FOV as a measure of their colocalization in that FOV. Details of the kernel density estimation step is as follows.

The kernel density estimator is defined as²⁶.

$$\hat{f}(\mathbf{x}; \mathbf{H}) = n^{-1} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$$

where $K_{\mathbf{H}}$ is a kernel function that depends on $\mathbf{H}$, the symmetric and positive-definite bandwidth matrix. $\mathbf{x}$ is an estimation point and $\mathbf{X}_i$ is an observation point. We use the scaled and translated normal kernel function:

$$K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) = (2\pi)^{-1/2} |\mathbf{H}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{X}_i)^{\top} \mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)\right\}$$

which is a normal density centred at $\mathbf{X}_i$ and with variance matrix $\mathbf{H}$. Normal Kernel is the most widely used multivariate kernel function. We used the kernel density estimator from the R package `ks`⁵¹. This package estimates $\mathbf{H}$ through cross-validation.

We used Pearson's Correlation Coefficient (PCC)²⁷ to measure the similarity between the kernel densities of two cell-types in an FOV. It is the covariance normalized by the product of their standard deviations:

$$\text{PCC} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

where x_i represents the probability of observing cell-type x at the i^{th} location in that FOV, as estimated in the kernel density for x . Similarly, y_i represents the probability of observing cell-type y at the i^{th} location in that FOV, as estimated in the kernel density for y . Here n is the number of discrete locations in this FOV. We did not apply PCC on the overall kernel densities, we selected only those locations where probabilities are greater or equal to the 2-quantiles of the kernel densities.

Ligand-Receptor analysis

With STANN's predictions of cell-types in each FOV of the SeqFISH+ data, we identified every pair of communicating cell-types and the receptor-ligands that they use. We used the tool CellPhoneDB¹⁵ for this analysis. CellPhoneDB uses a curated database of receptor-ligand pairs and constrains that a receptor or ligand should be expressed in at least 10% cells of a cell-type to be considered for this analysis. Based on an empirical null distribution, the tool then identifies receptors-ligands that are significantly and specifically enriched within a pair of cell-types.

Eliminating potential false-positive predictions of long-range intercellular communication

To flag potential false-positive predictions of intercellular communication between spatially separated cell-types, we checked if the predicted ligand-receptor pairs are used for communication with other co-localized cell-types. In particular, we perform the following check for each spatially separated but communicating cell-type pair A-B ($PCC < 0$). Let the communication between A and B is mediated by the ligand L and the receptor R. We eliminate this ligand-receptor pair from consideration if A is co-localized with cell-type C and communicates with C using L, and if B is co-localized with cell-type D and communicates with D using R. We outline this scheme in Fig. 6A. We reported the remaining ligand-receptors between A-B as potential mediators of long distance communication between A-B (Supplementary Table S5).

Gene Regulatory Networks

We used pySCENIC⁴⁵ to compute gene regulatory networks for each cell-type within each region of the tissue. Based on these networks, we identified the genes having regulatory relationships with the receptors and ligands mediating intercellular communication.

Jaccard Similarity coefficient

We used the jaccard similarity coefficient (also known as Jaccard Index) to gauge similarity between two finite sample sets A and B , e.g., between sets of receptors and ligands across FOVs. The Jaccard index is defined as the size of the intersection ($N(A \cap B)$) divided by the size of the union ($N(A \cup B)$) of the sample sets, it's expressed as:

$$\text{Jaccard similarity}(A, B) = N(A \cap B) / N(A \cup B)$$

Statistical comparison of cell-type compositions

To compare cell-type compositions between a pair of FOVs, we used the chi-square test of independence on the counts of different cell-types in those FOVs. We used a p-value cutoff of 0.01 after Bonferroni correction (Supplementary Table S4). We further corroborated these results using relative entropies as follows. Let x and y denote the compositions of cell-types in the two FOVs. We then computed the KL-divergences of x given y and also of y given x . Let p and q denote these two relative entropy terms. We checked both the arithmetic and the harmonic means of p and q and confirmed that the results are in agreement with our chi-square test, i.e., we confirmed that the cases where we could not reject the null hypothesis are indeed the cases with the lowest mean relative entropies.

All analyses were performed using Scipy v1.5.2.

Mouse Models

Wild-type mice C57BL/6J of post-natal day P40 (male) was used for the olfactory bulb SeqFISH+ experiments¹³. The olfactory bulb scRNA-seq data¹² was generated from wild-type male and female C57BL/6NJ mice at 14 weeks of age.

Discussion

Mammalian brains are complex ecosystems of several millions to hundreds of billions of cells^{52,53}. Brain cells exhibit a high degree of spatial organization^{54,55} and intercellular communication², which are disrupted in nearly every brain disease. A critical step toward understanding physiology and disease of the brain is to chart the spatial organization of cells and their communications in this tissue.

By integrating scRNA-seq (single-cell transcriptomics) with ST (spatial transcriptomics) data, and performing a series of systematic analyses, we have charted the spatial organization of cell-types and their communications in mouse olfactory bulb (MOB). Our analyses revealed both consistent and variable aspects of cell organization and intercellular communications in the MOB. Interestingly, we found more aspects to be spatially variable than consistent. On the one hand, as we had expected, we found that the proportions of cell-types vary across different morphological layers of MOB, but remain consistent within a given morphological layer. On the other hand, we found remarkable variations in their colocalization patterns and intercellular communication mechanisms throughout MOB. The analysis points to a widespread use of spatially localized gene regulatory networks in the MOB. While the motivation of our study was indeed to discover spatially variable principles of tissue architecture and function, without studying more samples, it is impossible to state which of these variable aspects are characteristic of MOB. For example, we have observed spatially localized gene regulatory networks; however, we cannot comment whether we would observe the same network at the same MOB location across every sample.

The key to this work's success is a neural network model, STANN, that learns to assign the cell-type to a cell by modeling scRNA-seq data. The trained STANN model can assign cell-types to every cell in an ST data. The two data sources do not profile the same set of genes, so STANN uses only the common genes to both data sources. STANN was highly accurate in this modeling task, but it would be impossible if we had relied only on the well-known marker genes of the MOB cell-types. The ST dataset in this study¹³ shares only 33% of the marker genes with the scRNA-seq dataset¹², readily implying that the marker gene based methods^{20–22} for integrating different single-cell datasets would not suffice in this case. STANN filled in this gap, and even performed better than the other state of the art methods, such as Seurat²³ and

scPred²⁴, that do not rely on marker genes. We attribute STANN's success primarily to its ability to identify complex nonlinear functions that map the expression values of shared genes in a cell to its cell-type. This highlights that to integrate two single-cell transcriptomic datasets, their shared genes do not necessarily have to include all the marker genes. One can solve this task as long as the shared genes capture sufficient information and the model is sufficiently complex (see below). Another important aspect boosting STANN's performance was its use of a class imbalance aware loss function. This helped STANN to accurately model rare cell-types, which is very common in single-cell studies. In our comparisons, the widely popular method Seurat lost much of its performance due to rare cell-types.

Although STANN performed very well in integrating the scRNA-seq and ST datasets, it is not guaranteed that such a solution will always exist. From a model design perspective, it is important to determine the cases when it is possible to integrate a given scRNA-seq and an ST dataset. Answering this question also enables us to answer a broader question: which genes should we select in an ST assay if the technology limits the number of genes?

We note that, given a choice of the model (e.g., a support vector machine or a neural network), we can establish a necessary condition on the shared genes that should hold for the model to accurately learn cell-types. Intuitively, for any choice of model, the shared genes need to capture the minimum variance necessary to distinguish the cell-types. Starting with the set of genes in the scRNA-seq dataset ranked according to their variability, one can then train cell-type mapping models (such as STANN) using an increasingly larger number of the most variable genes. The performance of the models in this process should never decrease since each model will use more variance in gene expression than its previous models. Let R denote the number of genes used in the first such model that learns cell-type mapping with a

satisfactory performance and the performance “saturates” for all models that use more genes. Then, for the model to be able to correctly learn cell-types using the shared genes, it is a necessary condition that the shared genes have rank $\leq R$. This necessary condition also enables us to find a trade-off between performance in cell-type learning and the number of genes to be profiled in an ST assay. For example, when we do not want to use as many genes as necessary for the model to saturate in performance, we can pick up a sub-optimal performance threshold and decide to profile only the corresponding genes in our ST assay. We tested this condition on the current dataset. Interestingly, we find that only about 5000 genes are necessary to accurately model cell-types in this dataset (Fig. S5). These 5000 genes include only 30.59% of the marker genes, further underscoring the point that with a proper model, we do not need to rely entirely on marker genes for cell-type learning.

The quality and availability of scRNA-seq and ST datasets are increasing at a fast pace, urging for methods to integrate these datasets accurately and without depending on expert-derived prior information, e.g., information on marker genes. The STANN model and the follow-up analyses were an initial step toward this goal. We anticipate that our observations will fuel the development of more sophisticated methods to analyze colocalization patterns and intercellular communication over long and short distances. We anticipate seeing more applications of these tools and methods on datasets collected from developmental stages and diseased tissues. Finally, our work has paved the way toward comparative spatial transcriptomics between healthy and diseased samples. We anticipate future studies will characterize the spatially aberrant transcriptomics of diseased samples enabling the design of targeted therapies.

Data Availability

The datasets analyzed in this study are available with the respective manuscripts and their supplementary information^{12,13}. The accession number for the raw single-cell RNA sequencing data is GEO: GSE121891¹². We collected the spatial transcriptomics SeqFISH+ data from the authors' site at <https://github.com/CaiGroup/seqFISH-PLUS>

Software Availability

<https://github.com/sameelab/STANN>

Acknowledgments

This work was supported by grants from the National Institutes of Health (HL127717, HL130804, HL118761 (J.F.M.); Vivian L. Smith Foundation (J.F.M.), State of Texas funding (J.F.M.), Fondation LeDucq Transatlantic Networks of Excellence in Cardiovascular Research (14CVD01) 'Defining the genomic topology of atrial fibrillation' (J.F.M.)

Conflicts: JFM is a founder and owns shares in Yap Therapeutics.

References

1. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, (2018).
2. Lein, E., Borm, L. E. & Linnarsson, S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64–69 (2017).
3. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
4. Ortiz, C. *et al.* Molecular atlas of the adult mouse brain. *Sci Adv* **6**, eabb3446 (2020).

5. Adler, M., Korem Kohanim, Y., Tendler, A., Mayo, A. & Alon, U. Continuum of Gene-Expression Profiles Provides Spatial Division of Labor within a Differentiated Cell Type. *Cell Syst* **8**, 43–52.e5 (2019).
6. Gillies, R. J., Brown, J. S., Anderson, A. R. A. & Gatenby, R. A. Eco-evolutionary causes and consequences of temporal changes in intratumoural blood flow. *Nat. Rev. Cancer* **18**, 576–585 (2018).
7. Junttila, M. R. & de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501**, 346–354 (2013).
8. Maley, C. C. *et al.* Classifying the evolutionary and ecological features of neoplasms. *Nat. Rev. Cancer* **17**, 605–619 (2017).
9. Vieth, B., Parekh, S., Ziegenhain, C., Enard, W. & Hellmann, I. A systematic evaluation of single cell RNA-seq analysis pipelines. *Nat. Commun.* **10**, 4667 (2019).
10. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).
11. Mayr, U., Serra, D. & Liberali, P. Exploring single cells in space and time during tissue development, homeostasis and regeneration. *Development* **146**, (2019).
12. Tepe, B. *et al.* Single-Cell RNA-Seq of Mouse Olfactory Bulb Reveals Cellular Heterogeneity and Activity-Dependent Molecular Census of Adult-Born Neurons. *Cell Rep.* **25**, 2689–2703.e3 (2018).
13. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
14. McDonald, J. H. & Dunn, K. W. Statistical tests for measures of colocalization in biological microscopy. *J. Microsc.* **252**, 295–302 (2013).
15. 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).
16. Angermueller, C., Pärnamaa, T., Parts, L. & Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **12**, 878 (2016).
17. Yu, T. & Zhu, H. Hyper-Parameter Optimization: A Review of Algorithms and Applications. *arXiv [cs.LG]* (2020).
18. Kelley, H. J. Gradient Theory of Optimal Flight Paths. *ARS Journal* **30**, 947–954 (1960).
19. Ma, F. & Pellegrini, M. ACTINN: automated identification of cell types in single cell RNA sequencing. *Bioinformatics* **36**, 533–538 (2020).
20. Zhang, A. W. *et al.* Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat. Methods* **16**, 1007–1015 (2019).
21. Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
22. Qian, X. *et al.* Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* **17**, 101–106 (2020).
23. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
24. Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q. & Powell, J. E. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019).
25. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018).
26. Chacón, J. E. & Duong, T. Multivariate Kernel Smoothing and its Applications. (2018)

doi:10.1201/9780429485572.

27. Manders, E. M. M., Verbeek, F. J. & Aten, J. A. Measurement of co-localization of objects in dual-colour confocal images. *Journal of Microscopy* vol. 169 375–382 (1993).
28. Cohen, M. *et al.* Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. *Cell* **175**, 1031–1044.e18 (2018).
29. Xiong, X. *et al.* Landscape of Intercellular Crosstalk in Healthy and NASH Liver Revealed by Single-Cell Secretome Gene Analysis. *Mol. Cell* **75**, 644–660.e5 (2019).
30. Wang, L. *et al.* Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. *Nat. Cell Biol.* **22**, 108–119 (2020).
31. Skelly, D. A. *et al.* Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Rep.* **22**, 600–610 (2018).
32. Ximerakis, M. *et al.* Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22**, 1696–1708 (2019).
33. Li, H. *et al.* Crosstalk Between Liver Macrophages and Surrounding Cells in Nonalcoholic Steatohepatitis. *Front. Immunol.* **11**, 1169 (2020).
34. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
35. Ji, A. L. *et al.* Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* **182**, 497–514.e22 (2020).
36. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).
37. Yousef, H. *et al.* Systemic attenuation of the TGF- β pathway by a single drug simultaneously rejuvenates hippocampal neurogenesis and myogenesis in the same old

- mammal. *Oncotarget* **6**, 11959–11978 (2015).
38. Ramel, M.-C. & Hill, C. S. Spatial regulation of BMP activity. *FEBS Lett.* **586**, 1929–1941 (2012).
 39. Zakin, L. & De Robertis, E. M. Extracellular regulation of BMP signaling. *Curr. Biol.* **20**, R89–92 (2010).
 40. Marei, H. E. S. *et al.* Gene Expression Profile of Adult Human Olfactory Bulb and Embryonic Neural Stem Cell Suggests Distinct Signaling Pathways and Epigenetic Control. *PLoS ONE* vol. 7 e33542 (2012).
 41. Bialas, A. R. & Stevens, B. TGF- β signaling regulates neuronal C1q expression and developmental synaptic refinement. *Nat. Neurosci.* **16**, 1773–1782 (2013).
 42. Maier, E., Nord, H., von Hofsten, J. & Gunhaga, L. A balance of BMP and notch activity regulates neurogenesis and olfactory nerve formation. *PLoS One* **6**, e17379 (2011).
 43. Hébert, J. M., Lin, M., Partanen, J., Rossant, J. & McConnell, S. K. FGF signaling through FGFR1 is required for olfactory bulb morphogenesis. *Development* **130**, 1101–1111 (2003).
 44. Yang, Y., Fang, Q. & Shen, H.-B. Predicting gene regulatory interactions based on spatial gene expression data and deep learning. *PLoS Comput. Biol.* **15**, e1007324 (2019).
 45. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
 46. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
 47. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
 48. Zhu, Q., Shah, S., Dries, R., Cai, L. & Yuan, G.-C. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization

- data. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4260.
49. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
50. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv [cs.DC]* (2016).
51. Duong, T. ks: Kernel density estimation for bivariate data. (2020).
52. Herculano-Houzel, S. The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* **3**, 31 (2009).
53. Herculano-Houzel, S., Catania, K., Manger, P. R. & Kaas, J. H. Mammalian Brains Are Made of These: A Dataset of the Numbers and Densities of Neuronal and Nonneuronal Cells in the Brain of Glires, Primates, Scandentia, Eulipotyphlans, Afrotherians and Artiodactyls, and Their Relationship with Body Mass. *Brain Behav. Evol.* **86**, 145–163 (2015).
54. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
55. Defelipe, J. The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front. Neuroanat.* **5**, 29 (2011).

Figure Legends

Figure 1. A) Learning stage of STANN, model training on annotated scRNA-seq data. B) ST Cell-type mapping with trained STANN model. C) Downstream analyses on cell-type annotated ST, spatial density maps of locations of cell-types, co-localization scores, intercellular communication and gene regulatory networks.

Figure 2. A) STANN's 10-fold accuracy scores across cell-types on separately held-out data, mean accuracy across all classes is 95.3%. Astro: Astrocytes, EC: Endothelial Cells, OECs: Olfactory Ensheathing Cells, MicroG : Microglia Cells, Neuron_Inmature: Neuron Immature Cells, Neuron_Mature: Neuron Mature Cells, Neuron_GC: Neuron Granule Cells, MyOligo: Myelinating-Oligodendrocyte Cells, Mes: Mesenchymal Cells, Mural: Mural Cells, Mono: Monocytes Cells, Neuron_MTC: Neuron M/T cells, Mf: Macrophages, OPC: Oligodendrocyte Progenitor Cells, Neuron_PGC: Periglomerular Cells. Boxplots showing overall accuracy (B) and cell type specific accuracies for STANN, Seurat's CCA and scPred (C). 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 3. A) A schematic showing the locations of the six fields of views (FOVs) in the MOB SeqFISH+ data. B) Spatial distribution of predictions in FOV six, located in the glomerular layer and external plexiform layer (cf. Fig 4F in Eng. et al¹³). C) model predictions mapped to the uniform manifold approximation and projection (UMAP) embedding of SeqFISH+ data.

Figure 4. MOB cell-type compositions according to STANN's predictions (A) and the scRNA-seq data (B).

Figure 5. a) Composition of cell-types across different FOVs in the Granule Cell Layer (GCL). b) PCC scores of FOVs that mapped the GCL c) Kernel density maps of Neuronal Transition cells vs Olfactory Ensheathing cells in two different FOVs located within the GCL layer, showing completely opposite co-localization patterns. STANN Predictions Reveal Spatially Localized Inter-cellular Communication Networks

Figure 6. A) Outline of the scheme to identify potential false-positive long-range ligand-receptor communications. B) Variability in expression of TGF β ligand in microglia cells and type III TGF β receptor in endothelial cells. Vertical axis represents a fraction of the cell-type population that expresses the ligand/receptor and the horizontal axis represents FOV (Field of View). C) Kernel density maps of endothelial cells and microglia, and their co-localization pattern in FOV 4 (located within the GCL layer). D) Ratio between the maximum and the minimum numbers of ligand-receptor pairs utilized by each pair of cell-types across FOVs. E) Number of ligand-receptor observations across FOVs and cell-type pairs vs Pearson Correlation Coefficient (PCC) F) Shared genes in receptor or ligand GRNs across FOVs and cell-types pairs.

Supplementary Figure Legends

Figure S1. A) Cellular composition of mouse olfactory bulb 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 average expression level of that gene in each cell type cluster. C) Individual single-cell transcriptomes were colored according to cluster/cell identity.

Figure S2 A) Composition of cell-types across different FOVs and morphological layers

Figure S3 A) Colocalization between cell-types pairs in SeqFISH+ data, quantified as Pearson correlation coefficient (PCC) across different fields of views (FOVs). 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 S4. TGF beta ligand upstream regulators in FOV 6 and FOV 4.

Figure S5. A) Metric scores of different models ran with varying numbers of genes/features.

Supplementary Table Titles

Supplementary Table S1. Marker genes from Tepe et al, 2018 and their intersection with SeqFISH+

Supplementary Table S2. Morphological layer annotation of each Field of Views (FOVs) (also see Figure 3)

Supplementary Table S3. Acronyms of each cell-type referred in this study.

Supplementary Table S4. Statistical comparison of cell-type compositions results

Supplementary Table S5. Ligand-receptors potentially mediating long-range intercellular communication.

Figure 1

Figure 2

A) STANN's 10-fold accuracy scores across cell-types on separately held-out data

Figure 3

A) Field of views schematics

B) STANN predicted cell-types in FOV 6

C) SeqFISH+ -- UMAP

Figure 4

Figure 5

A) Cell-type composition across FOV's in the Granule Cell layer

B) Pearson correlation coefficient across FOV's in the Granule Cell layer

C) Co-localized and mutually exclusive spatial density estimates

Figure 6**A) Long-distance L-R interactions potential false-positives** **B) Fractions of microglia cells expressing TGBR3 and endothelial cells expressing TGB1** **C) Spatial density estimate of microglia and endothelial cells**
FOV 4 -- Co-localized **D) Cell-type pairs ratio of LR obs across FOVs** **F) L-R observations across FOVs and cell-type pairs vs PCC** **E) Shared genes (%) in given receptor or ligand across FOVs and cell-type pairs in GRNs**