Giotto, a pipeline for integrative analysis and visualization of single-cell spatial transcriptomic data

Ruben Dries[1,6], Qian Zhu[1,6], Chee-Huat Linus Eng[2], Arpan Sarkar[1,3], Feng Bao[4], Rani E George[1], Nico Pierson[2], Long Cai[2], Guo-Cheng Yuan[1,5]*

1. Department of Pediatric Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA
2. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA
3. Departments of Statistics and Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA
4. Tsinghua University, Beijing, China
5. Harvard Stem Cell Institute, Boston, Massachusetts, USA
6. These authors contribute equally.

* Corresponding Author: Email: gcyuan@jimmy.harvard.edu

Abstract

The rapid development of novel spatial transcriptomics technologies has provided new opportunities to investigate the interactions between cells and their native microenvironment. However, effective use of such technologies requires the development of innovative computational algorithms and pipelines. Here we present Giotto, a comprehensive, flexible, robust, and open-source pipeline for spatial transcriptomic data analysis and visualization. The data analysis module implements a wide range of algorithms ranging from basic tasks such as data pre-processing to innovative approaches for cell-cell interaction characterization. The data visualization module provides a user-friendly workspace that allows users to interactively visualize, explore and compare multiple layers of information. These two modules can be used iteratively for refined analysis and hypothesis development. We illustrate the functionalities of Giotto by using the recently published seqFISH+ dataset for mouse brain. Our analysis highlights the utility of Giotto for characterizing tissue spatial organization as well as for the interactive exploration of multi- layer information in spatial transcriptomic and imaging data. We find that single-cell resolution spatial information is essential for the investigation of ligand-receptor mediated cell-cell interactions. Giotto is generally applicable and can be easily integrated with external software packages for multi-omic data integration.

Introduction

Most tissues consist of multiple cell types that operate together to perform their functions. The behavior of each cell is in turn mediated by its tissue environment. With the rapid development of single-cell RNAseq (scRNAseq) technologies in the last decade, most attention has gone to unraveling the composition of cell types with each tissue. However, recent studies have also shown that identical cell types may have tissue-specific expression patterns [1,2], indicating that tissue environment plays an important role in mediating cell states. Since spatial information is lost during the process of tissue dissociation and cell isolation, the

scRNAseq technologies are intrinsically limited for studying the structural organization of a complex tissue and interactions between cells and their tissue environment.

Recently, a number of technological advances have enabled single-cell transcriptomic profiling in a spatially resolved manner [3–9] (**Fig. 1A, inset**). Applications of these technologies have revealed distinct spatial patterns that previously are only inferred through indirect means at a lower resolution [10,11]. There is an urgent need for standardized spatial analysis tools that can facilitate comprehensive exploration of the current and upcoming spatial datasets [12,13]. To fill this important gap, we present a standardized and user-friendly pipeline, called Giotto, that allows researchers to process, (re-)analyze and interactively visualize spatial datasets at the single-cell resolution. Giotto implements a range of algorithms that are unique for spatial transcriptomics analysis and provides an easy-to-use workspace for interactive data exploration. Giotto has a number of strengths, including modularization, interactive visualization, reproducibility, robustness, and flexibility. As such, the Giotto pipeline will not only serve as a convenient entry point for spatial transcriptomic data analysis but also can be used as an effective tool for developing new hypotheses.

Results

Overview of Giotto pipeline

Giotto contains two independent yet fully integrated modules **(Fig. 1B, C**). The first module (Giotto Analyzer) provides step-by-step instructions about the different steps in analyzing spatial single-cell data, whereas the second module (Giotto Viewer) provides a fast and interactive viewer of spatial single-cell data and additional annotations (such as stainings, morphology and subcellular transcript localization information). These two modules can be used either independently or iteratively.

Giotto Analyzer requires as minimal input a gene-by-cell count matrix and the spatial coordinates for the centroid position of each cell (**Fig. 1B**). At the basic level, Giotto Analyzer can be used to perform common steps similar to single-cell RNAseq analysis, such as pre-processing, feature selection, dimension reduction and unsupervised clustering; however, the main strength comes from its ability to integrate gene expression and spatial information in order to gain insights into the structural and functional organization of a tissue. To this end, Giotto Analyzer creates a spatial grid and neighborhood network connecting cells that are physically close to each other. These objects function as the basis to perform analyses that are associated with cell neighborhoods. The Giotto Viewer module is designed both to interactively explore the outputs of Giotto Analyzer and to visualize additional information such as cell morphology and transcript locations (**Fig. 1C**). Giotto Viewer provides an interactive workspace allowing users to easily explore the data in both physical and expression space and identify relationships between different data modalities. Taken together, these two modules provide an integrated pipeline for spatial data analysis and visualization.

Giotto facilitates the comprehensive analysis of single-cell spatial transcriptomic data

Giotto Analyzer is written in the popular language R. The core data structure is a simple and flexible S4 object (**Fig. 2A**). Raw and processed count matrices are represented as a base matrix in R, while other annotations and metadata is encoded by an *igraph* network or a *data.table*. The former is a powerful library to work with networks, and the latter is a simple but intuitive table format with excellent performance for large-scale operations. In total, the

Giotto object contains 6 main slots, corresponding to count matrix, gene metadata, cell metadata, dimensionality reduction outputs, nearest-neighbor (NN) network analysis outputs, and spatial grid/neighborhood network analysis outputs, respectively (**Fig. 2A**). Each step is customizable by choosing from a list of algorithms and varying parameter values. The specific setting for each run can be saved to allow for reproducible analysis. Results from multiple runs using different settings can be stored in the same object, enabling users to easily evaluate the robustness of each method.

A typical Giotto analysis starts with steps that are similar to scRNAseq data analysis, but then increasingly incorporates additional information associated with the microenvironment (**Fig. 2B**). In brief, the analysis starts by considering only the gene-by-cell count matrix, carrying out a sequence of steps including normalization and quality control of raw counts, adjustment for batch effects or technical variations, feature selection, dimensionality reduction (such as PCA, tSNE [14], and UMAP [15]), clustering (such as Louvain [16] and Leiden algorithms [17]), and marker gene detection. These steps are similar to scRNAseq analysis and described in detail in the **Methods**.

The main utility of Giotto Analysis is to systematically incorporate spatial information for characterizing the spatial tissue organization and understanding the role of cell-cell interactions in mediating cellular states, as described in the following sections. A spatial grid and neighborhood network are constructed from the corresponding cell centroid coordinates. Here a spatial grid is a coarse-grained representation of the data defined on a regular grid, where the gene expression levels of cells within each grid box are averaged. Alternatively, the neighborhood network uses a graphic representation, where neighboring cells are connected through edges with either binary or continuous weights (**Methods**). The spatial grid and neighborhood network slot allows users to store multiple versions which could be used as input for various downstream analyses.

As an example to illustrate the different steps of Giotto Analyzer, we re-analyzed the recently published seqFISH+ dataset [9], where 10,000 genes were profiled in 913 cells of the mouse somatosensory cortex, subventricular zone and choroid plexus. We restricted the main analysis for cells in the somatosensory cortex region (523 cells) due to its richness of layer-specific excitatory neurons, interneuron diversity, and distinct cell morphology. By default, Giotto Analyzer can carry out a series of steps, including pre-processing, gene selection, dimensionality reduction, and Leiden clustering on a sNN. As a result, we identified six global and distinctive clusters for which we assigned a cell type based on known marker genes[18] (**Fig. 2C, Supplementary Fig. 1A**), including excitatory neurons (*Icam*), GABAergic neurons (*Slc32a1*), and four smaller groups of spatially scattered astrocytes (*Gli3*), endothelial cells (*Cldn5*), oligodendrocytes (*Sox10*) and microglia (*Laptm5*) (**Supplementary Fig. 1B**). In addition, Giotto Analyzer provides basic functionalities to visualize both single-cell resolution heterogeneity in both expression and spatial space representations (**Fig. 2D**), whereas Giotto Viewer should be used for more sophisticated, interactive visualization and exploration. Furthermore, Giotto Analyzer provides several automatic and manual sub-clustering strategies to explore a more detailed map of cellular states. Such analysis resulted in the identification of layer-specific excitatory neurons (L2/3, L4, L5 and L6), marked by different layer-specific genes, and more specialized inhibitory neurons (*Adarb2*, *Lhx6*) [19] and oligodendrocyte-like cell types (*Tmem88b*, *Gpr17*) [20,21] (**Supplementary Fig. 1C-F**).

Giotto uncovers different layers of spatial expression variability

A key component of Giotto Analyzer is the implementation of a wide range of computational methods for spatial gene expression pattern identification. On a basic level, Giotto Analyzer can reduce the single-cell resolution data to a spatial grid through averaging (**Supplementary Fig. 2A, B**). Principal component analysis (PCA) is applied to the grid-average data and significant principal components, along with their associated genes, are identified and reported. Using the aforementioned seqFISH+ dataset as an example, we found that the first principal component (PC) separates the outer layer extremities from the other layers. This is likely due to differences in cell-type compositions as most layers correlate with *Slc17a7* expression, a marker for glutamatergic neurons, while the extremities show higher abundance of astrocytes and oligodendrocytes (**Fig. 3A, top, Fig. 2D**). In contrast, the second PC separates the outer and inner layers, which have similar cell-type compositions (**Fig. 3A, bottom, Fig. 2D**). This simple approach is effective for identifying general linear trends within the spatial transcriptomics data (**Supplementary Fig. 2C**).

In addition, Giotto Analyzer provides an alternative, more sophisticated approach to identify more complex spatial patterns using a hidden Markov random field (HMRF) model. This approach was developed in our recent study in order to identify spatial domains with coherent gene expression patterns [22]. This is particularly useful for identifying distinct nonlinear patterns with sharp boundaries. By applying HMRF to the aforementioned somatosensory cortex seqFISH+ dataset, we identified 9 HMRF domains that resembles the anatomic layer structure (**Fig. 3B**). For example, Domain D7 is similar to Layer L1, Domain D2 is similar to Layer L2/3. This layered structure is less evident from the PCA analysis.

HMRF requires spatial genes as input. To identify genes whose expression patterns are spatially coherent, Giotto implements two simple yet effective approaches that utilize the underlying neighborhood network or physical distance between cells (**Supplementary Fig. 3A,B**, **Methods**). For example, *Cux2*, *Grm2*, *Cadm4* and *Islr2* show layer specific expression patterns (**Fig. 3C**). Some genes are associated with distinct PCs. For example, *Cux2 is* positively correlated with the second principal component, *Cadm4* is negatively correlated (**Fig. 3A, right**). However, other genes have more complex spatial patterns that are not revealed by PCA. The identification of spatially coherent genes is useful for development of new spatial pattern algorithms.

Giotto facilitates the exploration of the cellular neighborhood

Cells within a tissue do not live in isolation but closely interact with each other through specific molecules and signaling pathways (**Fig. 4A**). Giotto Analyzer provides a number of tools to explore the cell neighborhood organization and to infer the effects of cell-cell interactions. More specifically, Giotto Analyzer can determine pairs of cell types that are more frequently adjacent to each other than expected by chance within the neighborhood network (**Fig. 4B, Supplementary Fig. 4A**). These pairs can be either homo-typic (i.e., from the same origin) or hetero-typic (i.e., from different origins). For example, the layer-specific excitatory neurons tend to form homo-typic interactions, whereas spatially scattered populations, such as astrocytes, inhibitory neurons and oligodendrocytes, tend to form hetero-typic pairs (**Fig. 4B, C, Supplementary Fig. 4B-D**).

Besides the formation of organized cellular structures, the cell neighborhood has also an important effect on the intrinsic gene expression patterns of each individual cell. To systematically identify the effect of cell-cell interactions on gene expression, Giotto searches for genes whose expression levels are significantly affected by neighboring cell types. Based on the cell type of neighboring cells, cells within a cell type are divided into two groups: one group contains those cells that neighbor with a specific cell type of interest, and the other group contains all cells, i.e. those that interact with a specific cell type of interest or not.. Giotto compares the gene expression patterns between these two groups of cells (**Fig. 4A**) and identifies the subset of genes that are differentially expressed (see Methods for details). For the seqFISH+ dataset, we identified 5910 genes (4466 unique genes) that were differentially expressed in a pair of neighboring cell types (33 unique cell-cell pairs) (**Fig. 4A**, red vs black example, **Supplementary Fig. 3D**). The broad impact and complexity of the neighborhood influence is further illustrated by highlighting the diversity of different neighboring cell types that have an effect on gene expression (**Fig. 4D, Supplementary 3D**). Recently, several published reports have investigated a similar problem based on single-cell RNAseq analysis [23,24]. Based on expression levels of ligand-receptor pairs in pairwise cell type clusters, these studies have provided novel insights into the role of cell-cell communication in the tissue environment. However, the inferred ligand-receptor pairs may not be expressed in neighboring cells and such pairs would likely not be directly involved in cell-cell communication. Since seqFISH+ preserves spatial information, we can analyze gene expression patterns as well as spatial proximity to more precisely identify cell-cell interactions. By analyzing the seqFISH+ dataset, we noticed that, while there is indeed a significant correlation between co-expression and cell-cell interaction (**Supplementary Fig. 4G**), the information of gene expression alone is insufficient for accurate prediction of cell-cell interactions. For example, of all test ligand-receptor combinations (n = 416), which were selected as the number one ranked cell-cell pair based on expression, only 18% is in concordance with that of the spatially informed ranking (**Fig. 4E**). To systematically integrate spatial information into predictions, Giotto Analyzer uses a permutation-based approach of the cell neighborhood network to create a null distribution, which is similar to the method in (ref. [24]) but extended for spatial information. In total, we identified 424 instances (258 unique ligand-receptor pairs among 49 different cell-cell interactions) were cell communication was dependent on the spatial context and here exemplified by showing the spatially induced changes for the top 50 ligand-receptor pairs in all cell-cell interaction combinations (**Fig. 4F, Supplementary Fig. 4H**). As an example, we found that the expression levels of the ligand Semaphorin 3F (*Sema3f*) and receptor Neuropilin 2 (*Nrp2*) pair are specifically enriched in neighboring microglia-endothelial and inhibitory neuron-astrocyte cell pairs (**Fig. 4F** green squares, **Fig.4G, H** and **Supplementary Fig. 4I,J**), suggesting this ligand-receptor pair may play a significant role in cell-cell communication in a spatially context dependent manner. Consistent with our prediction, Semaphorin and Neuropilins are known to be important in brain related processes, such as axon guidance, and are also considered potential drug targets as they regulate brain tumor progression [25,26]. As such, the single-cell resolution spatial distribution of gene expression patterns is critical for understanding the logic of cell-cell communication.

Giotto Viewer: interactive visualization and exploration of spatial transcriptomic data

Giotto Viewer is designed for interactive visualization and exploration of spatial transcriptomic data. While Giotto Analyzer also generates figure outputs, Giotto Viewer provides an additional web-based, user-friendly visualization tool which can help generate new insights that are not apparent by examining static images. Giotto Viewer takes the results

from Giotto analyzer as input, and creates an interactive workspace for displaying, navigating, and highlighting various components of the spatial transcriptomic data, such as gene expression, cell type, and spatial domain.

Giotto Viewer displays multiple panels, each showing a different property of the data. A key aspect is that it automatically links and synchronizes these panels to enhance data navigation. For example, the gene expression patterns can be viewed both in the expression space (such as tSNE or UMAP) (**Supplementary Fig. 5**, bottom panel) and the physical space (**Supplementary Fig. 5**, top panel). Cell-type annotations can be visualized next to spatial-domain results (**Fig. 5A**). Each feature can be visualized at any desired resolution via the zoom and pan functions. Additional layers of information, such as cell segmentations, antibody staining images, and subcellular transcript localization can also be loaded into Giotto Viewer and overlaid with other information (**Fig. 5, Supplementary Fig. 5**). Finally, cells of interest can be highlighted, compared across different panels, and exported to a file for refined analysis by Giotto Analyzer.

Giotto Viewer is written in Javascript, HTML, and CSS, and is run in a locally supported web-based environment. Multiple panels, each displaying a specific component, are dynamically created according to the user's needs using a JSON configuration file, which serves as a step by step guide. This library is available as an Application Programming Interface (API) for further customization and easy adoption by web developers and advanced users. We use a number of state-of-the-art Javascript technologies in its implementation, including Leaflet.js, Turf.js, jQuery, to enhance dynamic user interaction (**Methods**). For example, we use a Google Maps-like algorithm to facilitate efficient navigation of cells against the staining image background. An input image is divided into many smaller, indexed units at various zoom levels. At each viewpoint, only information contained in a specific unit is loaded, thus avoiding memory overload and enabling the efficient rendering of large images. Giotto Viewer defines three classes of panels to accommodate different data representations (see **Methods**). The information from different panel classes is linked through sharing of cell IDs and annotations across panels. This allows seamless integration of different views and facilitates synchronous updates across all panels

Comparative visualization and analysis of multiple annotations

We use the aforementioned seqFISH+ dataset as an example to demonstrate the utility of Giotto Viewer. In addition to the gene expression matrix, spatial coordinates, and cell annotations generated by Giotto Analyzer, we also loaded Nissl, DAPI, and polyA staining images, cell segmentations, and transcript subcellular information together. Next, to facilitate comparison, we created four panels corresponding to cell type and spatial domain annotations represented in the physical and expression space, respectively (**Figure 5A**). The cell positions in these panels are linked and can be synchronously updated through zoom and pan operations. This feature allows users to easily navigate different areas and examine the relationship between cell type and spatial domain annotations. As the user zooms in onto the L1-L2/3 region (**Figure 5B**), it becomes apparent that domain D7 consists of a mixture of cell types including astrocytes, microglias, and interneurons. To identify the expression clusters of these cells, a user can use the lasso tool to select cells of domain D7 from the upper-left panel (representing the physical space) and visualize the selected cells in the expression space (upper-right panel) or corresponding cell-type annotations (lower-left and lower-right panels)

(**Figure 5B**). By comparing different viewpoints, it is clear both cell type and spatial domain differences contribute to cellular heterogeneity.

To gain further insights into the difference between cell type and spatial domain annotations, we saved the selected cells to an output file. The corresponding information can be directly loaded into Giotto Analyzer for further analysis. This allows us to identify a number of additional marker genes, such as *Cacng3*, *Lmtk3*, and *Scg3* (**Figure 5C**). The seamless iteration between data analysis and visualization is a unique strength of Giotto.

Subcellular transcript localization visualization

seqFISH+ technology can detect transcripts at ~100nm resolution, thereby providing an opportunity to study patterns of subcellular transcript localization. Giotto Viewer provides a utility to visualize the exact locations of individual transcripts (**Figure 6**). To facilitate real-time exploration of the transcript localization data, which is much larger than other data components, we adopted a position-based caching of transcriptomic data (see **Methods**). From the original staining image (**Figure 6A**), the users can zoom in any specific region or select specific cells and visualize the locations of either all detected transcripts (**Figure 6B**) or selected genes of interest (**Figure 6C**). The spatial extent of all transcripts is useful for cell morphology analysis (**Figure 6B** and **Figure 6C**), whereas the localization pattern of individual genes may provide functional insights into the corresponding genes (**Figure 6C)**. For example, transcripts of *Snrnp70* and *Car10* are preferentially localized to the cell nucleus (delineated by DAPI background), while *Agap2* and *Kif5a* transcripts are distributed closer to the cell periphery (**Figure 6C**). Such differences may have functional implications.

Discussion

Single-cell analysis has entered a new phase -- from characterizing cellular heterogeneity to interpreting the role of spatial organization. To overcome the challenge for data analysis and visualization, we have developed Giotto as a standardized pipeline that can be broadly applied in conjunction with a wide range of spatial transcriptomics technologies. Since the only requirement is a count matrix and spatial coordinates, it is possible that, with minor changes, the pipeline can also be applied to spatial proteomic data. We implemented a modularized approach, where the two modules (Giotto Analyzer and Giotto Viewer) are seamless integrated but can also be run independently. Each module is designed to optimize easiness-to-use, speed, flexibility, and reproducibility. These two modules can be used iteratively to refine data exploration and generate new hypotheses.

Giotto implements a wide range of algorithms that are unique for spatial transcriptomics analysis. One example is a recently developed hidden Markov random field (HMRF) method for detecting spatial domains, which is useful for dissecting the contribution of intrinsic and environment-mediated cellular state variation [22]. In this current study, we have further developed an algorithm for detecting cell-cell interactions by integrating gene expression and spatial information and showed that incorporating spatial information is critical for the investigation of cell-cell interactions (**Figure 4**).

Giotto differs from existing spatial transcriptomics data analysis [8,22,27–29] and visualization pipelines [30–32]. To our knowledge, Giotto Analyzer is the first general-purpose pipeline for spatial transcriptomic analysis, while the other methods are designed for specific tasks, such as the identification of cell types [8,29], marker genes [27,28] or domain patterns [22]. The flexible

design of Giotto makes it an ideal platform for incepting new algorithms and integration with external pipelines. As single-cell multi-omics data become more available, such integration may greatly enhance mechanistic understanding of the cell-state variation in development and diseases.

Methods

Giotto Analyzer

Data usage and availability. seqFISH+ data from the primary somatosensory cortex and olfactory bulb [9] was used to test and build Giotto. All datasets required for Giotto Analyzer are also part of the Giotto R package. Codes and data are publicly available at http://spatial.rc.fas.harvard.edu and links therein.

Quality control, pre-processing and normalization. seqFISH+ data was first filtered using the *filterGiotto* function on the raw discrete count matrix. Using a detection threshold of 1, genes were excluded if not detected in at least 10 cells and cells were excluded if they did not contain at least 10 detectable genes. Using these criteria no genes or cells were removed. The raw count matrix was subsequently normalized using the *normalizeGiotto* function by first dividing each cell by library size and then rescaling by a factor of 2000 followed by a log2(counts+1) transformation. Next the function *addStatistics* computes general cell and gene statistics such as the total number of detected genes and counts per cell. To adjust for variation due to the former technical covariates the *adjustGiottoMatrix* was applied to the normalized data.

Feature selection. To identify informative genes for clustering the *calculateHVG* function was used. Highly genes were defined as those genes that showed high variance relative to their mean expression. Genes were further filtered by only retaining those genes that displayed a variance > 1 compared to the predicted variance, were detected in at least 5% of genes and had an average log-normalized expression > 0.5 in cells were the gene was detected.

Dimensionality reduction. The identified highly variable genes were then used to reduce dimensions by performing principal component analysis on the adjusted and normalized count matrix with the function *runPCA*. Further non-linear dimension reduction is performed with umap and t-sne on the pca-space (first 15 PCs) using *runUMAP* and *runtSNE* respectively. Significant principal components can be estimated with the *signPCA* function using a scree plot or the jackstraw method [33].

Clustering. First a shared nearest neighbor (sNN) network was constructed with *createNearestNetwork* in pca-space using the default parameters. Leiden clustering [17] is implemented as *doLeidenCluster* and was used to identify clusters at a predefined resolution. In addition, a novel iterative clustering method was developed and implemented as *iterCluster*, as described below.

Iterative clustering. This method iterates through a predefined number of rounds of feature selection, dimension reduction and clustering as previously described. After each round the most stable cluster is selected by computing the ratio of within-cluster sNNs over outside-

cluster sNNs and subsequently removed from next rounds. This approach removes distortions in expression space that arise due to the presence of distinct cell types and might hence mask more subtle transcriptomic changes in other cell types. In an iterative manner manual or automatic global sub-clustering with the implemented function *doLeidenSubCluster* can be combined with *iterCluster* to obtain a more detailed cell fate map.

Marker genes detection. We have developed a new method to identify marker genes and implemented it as *findGiniMarkers*. For each gene in each cluster, *findGiniMarkers* calculates a Gini-coefficient for both gene expression values and gene detection scores. The Gini-coefficients corresponding to gene expression and detection scores are multiplied, which are then used to rank each gene in all clusters to identify potential marker genes. We have found this approach is effective for identifying genes that are both specific and widely expressed in a particular cluster therefore used it as the default setting in Giotto. In addition, an extended version is also implemented as *findMarkers_one_vs_all,* which performs systematic pairwise comparisons between each cluster and all other clusters merged together. As an alternative, Giotto Analyzer also implements an algorithm from the scran package [34] to identify marker genes as the *findMarker* function.

Spatial grid and neighborhood network. A spatial grid is defined as a Cartesian coordinate system with defined units of width and height and is created with the function *createSpatialGrid*. The gene expression levels of cells within each grid box are averaged. *detectSpatialPatterns*, *showPattern* and *showPatternGenes* can be used to perform a PCA-based approach to detect spatial gene expression patterns and associated genes. Another representation of the spatial relationship is the neighborhood network, where each node represents a cell, and each pair of neighboring cells are connected through an edge. The number of neighbors can be defined by setting (a minimal) k and/or radial distance from the centroid position of each cell, and the edge weights can be either binary or continuous.

Spatially coherent gene detection. Two methods, implemented in the function *calculateSpatialGenes*, can be used to discover genes with a spatially coherent expression pattern. The first method uses the neighborhood network and binarized gene expression data as input. The binarization can be done by either *kmeans* clustering (k = 2) or simple thresholding (default = 10%). Genes are selected based on the association of their binarized expression level and neighborhood relationship, which is evaluated by using the Fisher exact text. The second method is similar, except that the continuous values of cell-cell distance, normalized by ranking, is used for evaluating statistical association.

HMRF. Spatial domains were identified with HMRF as previously described [22]. In brief, HMRF is a graphic-based model for spatial gene expression patterns. A HMRF enables the detection of spatial domains by systematically comparing the gene signature of each cell with its surroundings to search for coherent patterns. The original model was implemented in Python and is incorporated in Giotto by using the consecutive wrapper functions *doHMRF*, *viewHMRF* and *addHMRF* to discover, visualize and select HMRF domain annotations respectively.

Identification of interacting cell types. The edges of the neighborhood network (k = 3) described are labelled as homo- or hetero-typic, if they connect cells of identical or different annotated cell types, respectively. To determine the ratio of observed over expected frequencies between two identical or different cell types, the observed number of edges

between any two cell types was compared to a random permutation (n = 200) distribution by reshuffling the cell labels. Associated P values were calculated by observing how often the simulated values were higher or lower than the observed value for respectively increased or decreased frequencies. A wrapper for this analysis is implemented in Giotto Analyzer as *cellProximityEnrichment*.

Analysis of the effect of cell-cell interactions on gene expression. Changes in gene expression that are associated with the type of neighboring cells were identified with *getCellProximityGeneScores* which considers the average gene expression of a single gene per cell type (global average expression) and compares that with the average gene expression for only those cells that spatially interact with another cell type (spatially restricted average expression). A Wilcoxon rank sum test is performed between the global and spatially restricted average expression. To visualize and/or extract the neighbor dependent gene enrichment or cell proximity gene (CPG) scores, Giotto Analyzer implements *showCPGscores* for which we filter on the minimum number of cells (n = 5), p-value (p <= 0.05), spatial difference (>= 0.2) and log2 fold-change (>= 0.5).

To further systematically explore neighbor associated changes in gene pairs, such as known ligand-receptor combinations, Giotto Analyzer provides two complementary algorithms, The first algorithm uses the function *getGeneToGeneSelection* to combine the results obtained for individual genes from the CPG scores for all possible or a selected set of gene pairs and cell-cell interaction combinations. In a similar manner as for individual genes this dataset is filtered for both genes of the gene-pair using a more relaxed threshold (n = 5, p < 0.1, spatial difference > 0.2 and log2 fold-change > 0.3). The second algorithm is based on a statistical framework that corroborates the statistical significance of neighborhood dependent changes in gene-pairs and is implemented as *specificCellCellcommunicationScores* for a specific pair of interacting cell types or for all possible combinations of cell types as *allCellCellcommunicationsScores.* In brief, a null distribution is created, based on multiple rounds of reshuffling cell ids within the same cell types that are being considered. By comparing the observed spatial change in gene expression for a gene pair in the considered spatially interacting cell types with that seen in the permutations, we evaluate the statistical significance associated with both increased or decreased neighborhood dependent gene expression. As a comparison to the previous cell-cell communication scores or to provide an estimated guess when there is no spatial information available (e.g. scRNAseq) we also implement a version of the algorithm (*exprOnlyCellCellcommunicationScores*) that considers only gene expression. Here statistical significance is evaluated based on the null distribution of gene expression estimated by reshuffling cell-type labels. This latter approach is similar to a previous study [23]. For the focused analysis on ligand-receptor pairs, the ligand-receptor pair information was retrieved from FANTOM5 [35].

Giotto Viewer

Input files. The minimal input files for Giotto Viewer contains the gene expression matrix and the cell centroid coordinates. Such information can be either provided manually in tabular format, or directly loaded through the output files of Giotto Analyzer. If available, additional input files such as cell segmentations (ROI files), staining images (TIFF files), transcript locations (TXT files), can also be incorporated. Giotto Viewer provides tools to process such information for visualization (see next sections).

Image input and initial processing. Multistack cell staining images, such as Nissl, DAPI, and polyA, that are acquired by microscopy instruments and accompanying software packages are input in multi-channel TIFF format. Giotto Viewer provides a utility to extract and decouple the multi-channel information based on the ImageMagick library (https://imagemagick.org/). To enhance visualization and exploration, Giotto provides an option to stitch images across multiple fields of view (FOV) with gaps in between. The layout can be manually controlled by modifying a coordinate offset file specifying the relative positions of FOVs.

Giotto Viewer uses a Google Maps-like algorithm to facilitate multi-level zooming and navigation. To this end, Giotto Viewer further processes the stitched images by dividing each into equal-sized smaller images using the image tiling function in the tileup package in Ruby (https://github.com/rktjmp/tileup). This creates a set of tiled images corresponding to 6 zoom levels with 1.5X increment. The size of each tile is fixed at 256 by 256 pixels.

Cell boundary segmentation is a necessary step for assignment of each transcript to its corresponding cells. Giotto Viewer does not provide this functionality but can accept user-provided cell boundary segmentation information as input, in the form of Region-of-Interest (ROI) files, for visualization. Giotto Viewer extracts information from the ROI files by adapting a JAVA program based on the ImageJ framework. An overview of the Giotto Viewer processing steps is in **Supplementary Fig. 6**.

Implementation of interactive visualization of multi-layer spatial transcriptomic information. Giotto Viewer is implemented as a web-based application that can be installed on any Linux, Windows or MAC OS based computer. The package is written in Javascript utilizing a number of state-of-the-art toolboxes including Leaflet.js (https://leafletjs.com/), Turf.js (https://turfjs.org/), Bootstrap (https://getbootstrap.com/), and jQuery (https://jquery.com/). The Leaflet.js toolbox is used to efficiently visualize and explore multiple layers of information in the data, based on a Google Maps-like algorithm. The implementation of Giotto Viewer consists of a cell object layer, containing all the differently shaped cells, an image layer corresponding to staining images, and an annotation property that specifies the cluster membership and gene expression information.

Giotto Viewer allows users to 1) load and visualize any number of panels simultaneously, and 2) add different types of data to each panel. To permit flexibility, there are three types of panels implemented in the Giotto Viewer: PanelTsne, PanelPhysical, and PanelPhysicalSimple. PanelTsne requires cell coordinates in the state space as input. PanelPhysical lays out the cells in the physical space in the segmented cell shapes. PanelPhysicalSimple is a simplified version to PanelPhysical except that cell segmentation and staining images are not required, and instead renders cell objects as fixed size circle markers. The implementation of these 3 panels follows closely the paradigm of object-oriented design in Javascript, specified by the MDN Web Docs and ECMAScript.

Briefly, the three panel types are motivated by the fact that depending on data availability, property of cells change from dataset to dataset, so different ways of cell representation should be considered. In the presence of cell staining images, images should serve as background overlays to the data. If segmentation information is available, cells should be

represented in their true cell shape. Yet when neither staining nor segmentation is available, Giotto Viewer should represent cells as basic geometric shapes (circles) so that the viewer can still run in the absence of staining or segmentation data. We designed the panel classes with these considerations in mind. Giotto Viewer makes it easy for users to specify the number of panels, the type of each panel, and the layout configuration. Users specify such information in a JSON formatted configuration file. A script then automatically generates the HTML, CSS, JS files of the comparative viewer that is ready for exploration in a standard web browser.

To enable interactivity, panels should be linked to each other. We first define mouseover and mouseout events for each cell object. The exact specification of events depends on the type of panel, the action chosen by the user, and the context in which the action is performed. Next, we maintain equivalent cell objects across panels by creating a master look-up table to link cell IDs in different panels. This is useful to facilitate interactive data exploration and comparison, synchronous updates of zoom and view positions during data exploration. Finally, the order and dependency with which interactions are executed are enforced by constantly polling element states and proceeding each step only when states are changed. In the API, the functions addInteractions(), addTooltips() enable the easy specification of cross-panel interactions. In the JSON configuration file, interactions between panels are simply defined by the user using the “interact_X: [panel ids]” and “sync_X: [panel ids]” lines.

Giotto Viewer provides an intuitive utility to select a subset of cells of interest for visualization and further analysis. The toggle lasso utility allows a user to hand-draw an enclosed shape in any displayed panel to select cells directly. We implemented this function by modifying the Leaflet-lasso.js toolbox to add support for the selection of the dynamic polygon shaped markers. Giotto Viewer can also highlight Individual cells with summary information displayed. This is achieved by using built-in functions in the Turf.js toolbox.

Visualizing subcellular transcript localization. To visualize subcellular transcript localization information, an additional layer is created in Leaflet.js. To efficiently handle the large amount of data, we adopted the Leaflet.csvTils.js plugin (https://github.com/gherardovarando/leaflet-csvtiles) whereby only a small subset of transcripts that falls within the current viewing area, are rendered by the Leaflet engine and thereby saves the systems resources.

Figure Legends

Figure 1. Simplified diagram of the modular Giotto pipeline
A. Diagram of an organized multi-cellular tissue. Recent spatial transcriptomic technologies allow the identification of multiple transcripts at single-cell resolution.
B. Overview of the spatial transcriptomics analysis module Giotto Analyzer. This module requires a raw count matrix and x,y coordinates for the corresponding cells. It follows standard scRNA-seq processing and analysis steps to identify differentially expressed genes and cell types. In a following step a spatial grid and neighborhood network is created which is further used to incorporate the spatial information of the single-cell environment and which is used for spatial analysis. **C.** Cell coordinates, annotations, and clustering information are utilized and incorporated in the Giotto Viewer. This interactive viewer allows users to explore the link between cells' physical positions and their clustering pattern in the expression space (UMAP or tSNE). The addition of staining images and cell segmentation information is also supported.

Figure 2. Framework of the Giotto R package to analyze single-cell resolution spatial transcriptomic data.
A. Schematic representation of the Giotto S4 object showing the primary slots for accessing and storing data and analyses. **B.** Example of the basic Giotto pipeline steps with typical and spatial transcriptomics specific operations. **C.** Heatmap (left) and violinplot (right) visualization of marker genes for identified cell types in the seqFISH+ somatosensory cortex data. **D.** Co-visualization of somatosensory cortex single cells, color-coded by their cell types (as in **C)**, in both expression space (first 2 dimensions of UMAP, **top**) and physical space (**bottom**).

Figure 3. Layers of spatial gene expression patterns.
A. Spatial patterns identified through PCA analysis of grid-averaged expression values. **B.** Visualization of HMRF domains in the physical space. **C.** Examples of genes that display a spatially coherent expression pattern.

Figure 4. Cell neighborhood network analysis uncovers cell-cell interaction effects.
A. Schematic of a multi-cellular tissue with an organized cellular structure (left) and environment specific gene expression (right). **B.** Barplot depicting the ratio of observed over expected frequencies of pairwise interacting cell types. **C.** Example of an enriched cell-cell interaction in layer L5 of the mouse somatosensory cortex. **D.** Sankey plot showing the number of genes per cell type whose expression change is associated with neighboring cell types. **E.** Histogram illustrating the comparison of spatially informed rankings of ligand-receptors in cell-cell combinations with that of the highest ranked cell-cell combination based on gene expression only. **F.** Dotplot showing the permutation based enrichment (log2 fold change) and statistical scores (p-values) for the top 50 ligand-receptor pairs within spatial cell-cell interaction combinations. **G-H**. Detailed visualization of ligand-receptor expression changes associated with cell-cell interactions. Blue open dots indicate the average expression level in all cells of the cell type of interest. Red open dots represent the average expression level within cells interacting with specific cell types. Filled dots represent the combined scores from individual genes.

Figure 5. Giotto Viewer provides an interactive workspace to visualize and compare multiple cell annotations.

A. Side-by-side comparison using 4 panels created within Giotto Viewer, including 2 panels (left) representing the physical space and 2 panels (right) representing the expression space as UMAP plots. Spatial domain (top panels) and cell type (bottom panels) annotations are displayed next to each other for easy comparison. B. Zoomed-in view of L1-3 regions. Selection of cells in domain D7 is indicated in red outline and highlights domain and cell type annotation differences. As the selection is made, the corresponding UMAP panels are automatically updated with the selection. C. Expression patterns of domain marker genes identified through iterative analysis by Giotto Analyzer.

Figure 6. Visualization of transcript subcellular localization patterns.
A. Nissl staining images in the somatosensory cortex region. A region in layer L4 (indicated by the box) is selected for further analysis. **B.** Distribution of all detected transcripts in L4 delineates cell boundaries. Colors indicate different cell types. Three excitatory neurons (indicated by the yellow) are selected for further analysis **C.** Subcellular transcript localization patterns in the three highlighted cells. Each dot represents an individual transcript. Giotto Viewer can be used to visualize either all transcripts (top row) or those from selected genes (middle and bottom rows).

Supplementary Figure 1. Detailed cell-type identification.
A. Heatmap depicting Pearson correlation scores between cell type averages. **B.** Visualization of highlighted cell types in physical space. **C.** Violinplot showing marker gene expression levels for the identified cell types. **D.** Co-visualization of the identified cell types in umap (top) and physical space (bottom). **E.** Heatmap depicting the expression levels of cell-type specific marker genes for cell type clusters. **F.** Heatmap depicting Pearson correlation scores between sub-clustered cell type averages.

Supplementary Figure 2. Spatial grid applications.
A. Illustration of a spatial grid in a Cartesian coordinate system. **B.** Boxplot illustrating the average number of cells per grid. **C.** Spatial patterns identified through PCA analysis of grid-averaged expression values.

Supplementary Figure 3. Neighborhood network implementation and applications.
A. Examples of single-cell neighborhood networks based on different parameters of k (number of neighbors) and/or distance from each individual cell.

Supplementary Figure 4. Cell neighborhood analyses and visualizations.
A. Heatmap illustrating the preferential interactions between cell types. **B.** Visualization of highlighted cell types in physical space. **C-D**. Visualization of highlighted cell-cell interactions in physical space. **E.** Barplot showing the number of neighborhood associated gene expression changes in different pairwise cell-cell interactions. **F.** Barplot showing the number of genes per cell type whose spatial expression change is associated with another neighboring cell type. **G.** Density plot comparing the ranking of ligand-receptor expression scores between cell type pairs based on expression only or spatially informed cell-cell interaction expression. **H.** Heatmap depicting the row and column z-scored counts for significant ligand-receptor expression changes between depicted cell types. **I-J.** Visualization of highlighted cell-cell interactions in physical space.

Supplementary Figure 5. Overview of the Giotto Viewer processing steps.

Giotto Viewer is run locally using Python's simple HTTP server. Giotto Viewer is made up of three types of panels: PanelPhysical (top), PanelTsne (bottom), and PanelPhysicalSimple (not shown) (see **Methods** for explanations). Multiple panels can be simultaneously displayed side-by-side (two are shown). Each panel reflects a different aspect of the data, and panels must be linked to enable interactive exploration. Left hand control panels includes buttons for basic functions such as changing the staining image, switching between spatial domain and cell type annotations, querying gene expression pattern, and selecting cells directly on the map (i.e. the toggle lasso button). Cells can be selected from either panel. All panels are synchronously updated to highlight the cell selection.

Supplementary Figure 6. **Preparations steps of Giotto Viewer.**
The different components of the set up includes staining images, cell segmentation files, gene expression, and annotation files. Some of these inputs are optional. The annotation files are generated from Giotto Analyzer. The layout of the panels can be customized by the user.

References

1. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **173**, 1307 (2018).
2. Tabula Muris Consortium *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
3. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
4. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
5. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
6. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* **92**, 342–357 (2016).
7. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, (2018).
8. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
9. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
10. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
11. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
12. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, (2017).
13. Snyder, M. P. *et al.* Mapping the Human Body at Cellular Resolution -- The NIH Common Fund Human BioMolecular Atlas Program. *arXiv [q-bio.OT]* (2019).
14. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**,

2579–2605 (2008).

15. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4314
16. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).
17. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
18. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
19. Hodge, R. D. *et al.* Conserved cell types with divergent features between human and mouse cortex. *bioRxiv* 384826 (2018). doi:10.1101/384826
20. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
21. Bonfanti, E. *et al.* The role of oligodendrocyte precursor cells expressing the GPR17 receptor in brain remodeling after stroke. *Cell Death Dis.* **8**, e2871 (2017).
22. 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
23. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).
24. Kumar, M. P. *et al.* Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep.* **25**, 1458–1468.e4 (2018).
25. Karayan-Tapon, L. *et al.* Semaphorin, neuropilin and VEGF expression in glial tumours: SEMA3G, a prognostic marker? *Br. J. Cancer* **99**, 1153–1160 (2008).
26. Meyer, L. A. T., Fritz, J., Pierdant-Mancera, M. & Bagnard, D. Current drug design to target the Semaphorin/Neuropilin/Plexin complexes. *Cell Adh. Migr.* **10**, 700–708

(2016).

27. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).
28. Edsgärd, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* **15**, 339–342 (2018).
29. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
30. Schapiro, D. *et al.* histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873–876 (2017).
31. Fernández Navarro, J., Lundeberg, J. & Ståhl, P. L. ST viewer: a tool for analysis and visualization of spatial transcriptomics datasets. *Bioinformatics* **35**, 1058–1060 (2019).
32. *starfish*. (Github).
33. Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545–554 (2015).
34. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* **5**, 2122 (2016).
35. Ramilowski, J. A. *et al.* A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).

Figure 1

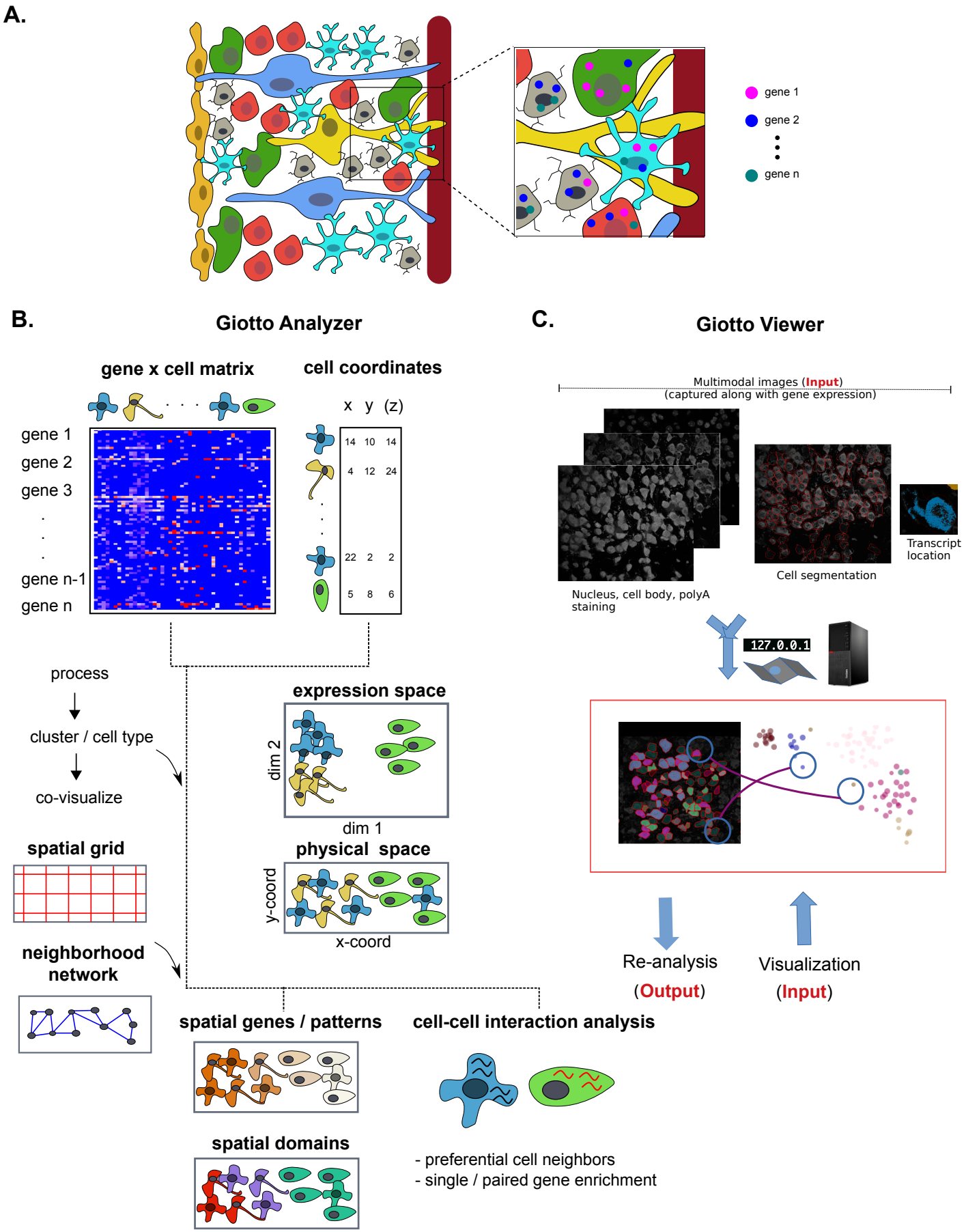
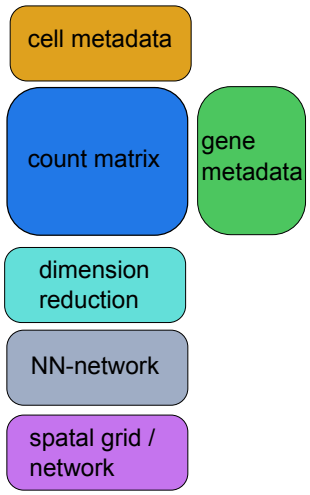
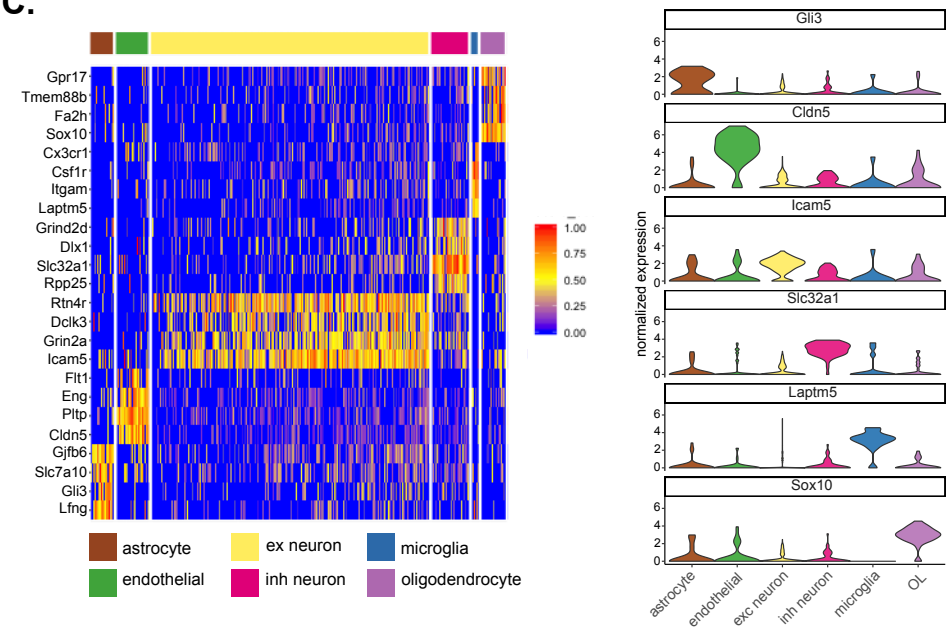

A.
gene 1
gene 2
gene n
B.
Giotto Analyzer
gene x cell matrix
cell coordinates
x y (z)
gene 1
gene 2
gene 3
gene n-1
gene n
14 10 14
4 12 24
22 2 2
5 8 6
process
cluster / cell type
co-visualize
expression space
dim 2
dim 1
spatial grid
physical space
y-coord
x-coord
neighborhood network
spatial genes / patterns
cell-cell interaction analysis
spatial domains
- preferential cell neighbors
- single / paired gene enrichment
C.
Giotto Viewer
Multimodal images (Input)
(captured along with gene expression)
Nucleus, cell body, polyA staining
Cell segmentation
Transcript location
127.0.0.1
Re-analysis
(Output)
Visualization
(Input)

Figure 2

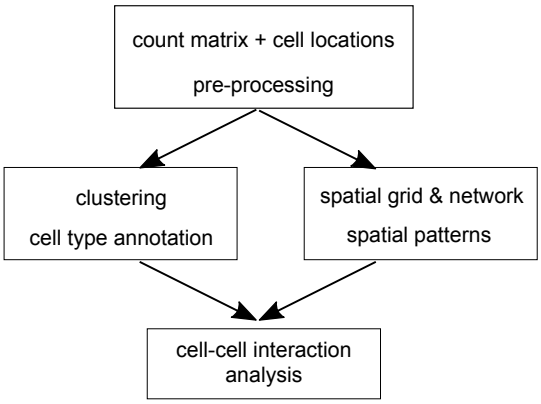
A.



C.



B.



D.

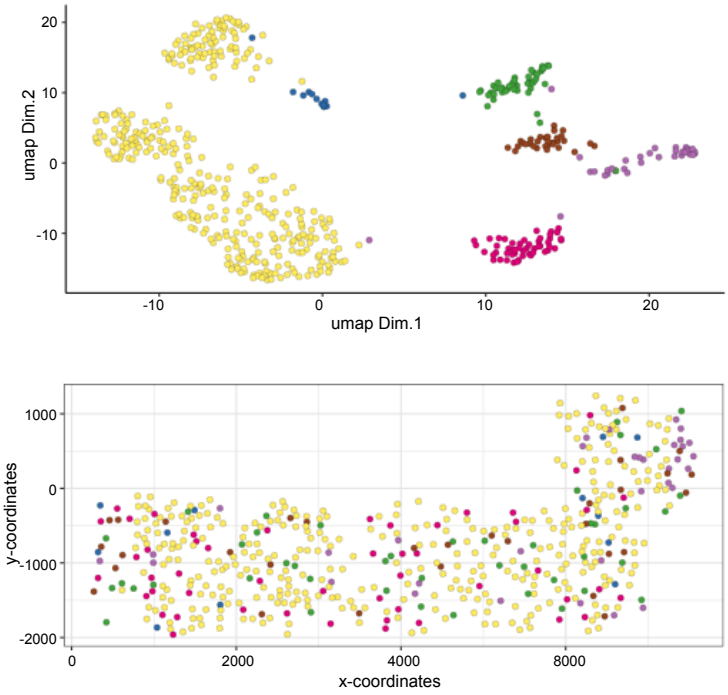
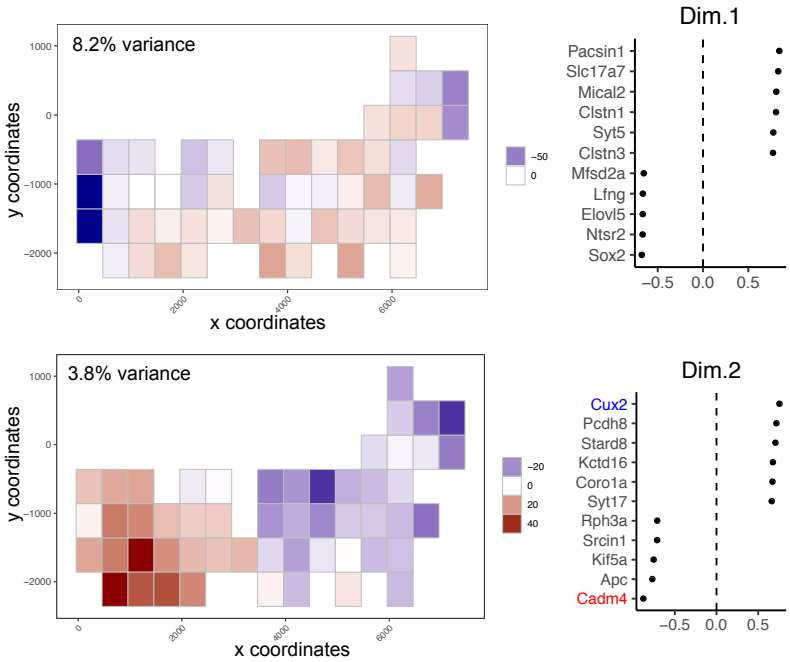
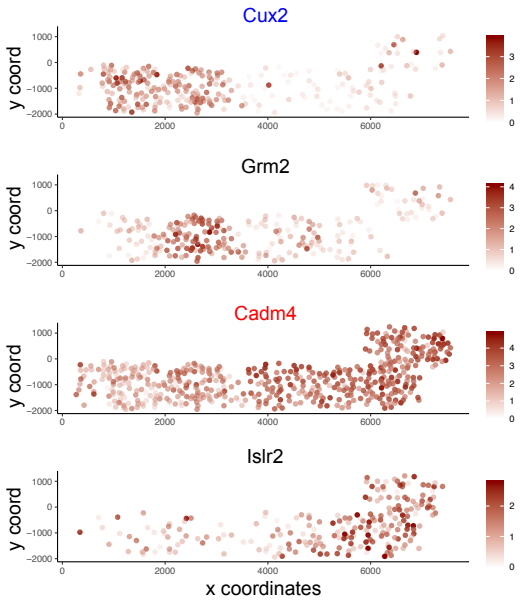


Figure 3

A.


8.2% variance
y coordinates
x coordinates
Dim.1
Pacsin1
Slc17a7
Mical2
Clstn1
Syt5
Clstn3
Mfsd2a
Lfng
Elovl5
Ntsr2
Sox2
3.8% variance
Dim.2
Cux2
Pcdh8
Stard8
Kctd16
Coro1a
Syt17
Rph3a
Srcin1
Kif5a
Apc
Cadm4

C.


Cux2
Grm2
Cadm4
Islr2
y coord
x coordinates

B.

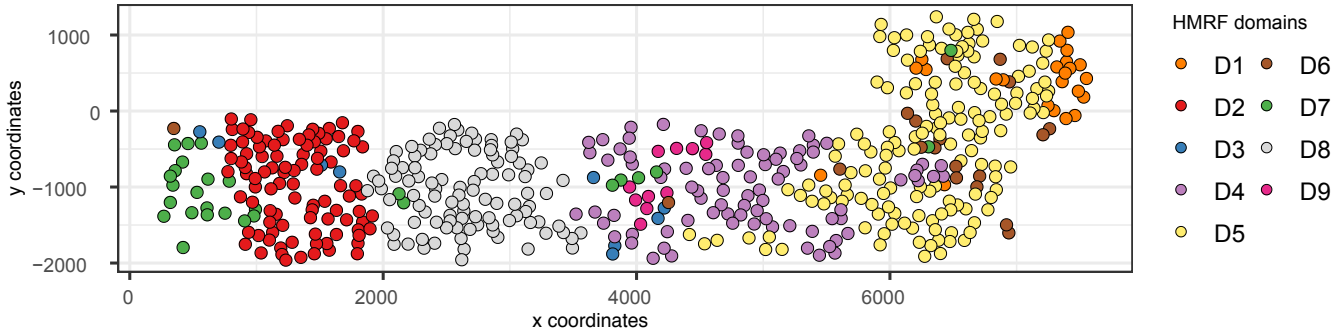

y coordinates
x coordinates
HMRF domains
D1
D2
D3
D4
D5
D6
D7
D8
D9

Figure 4

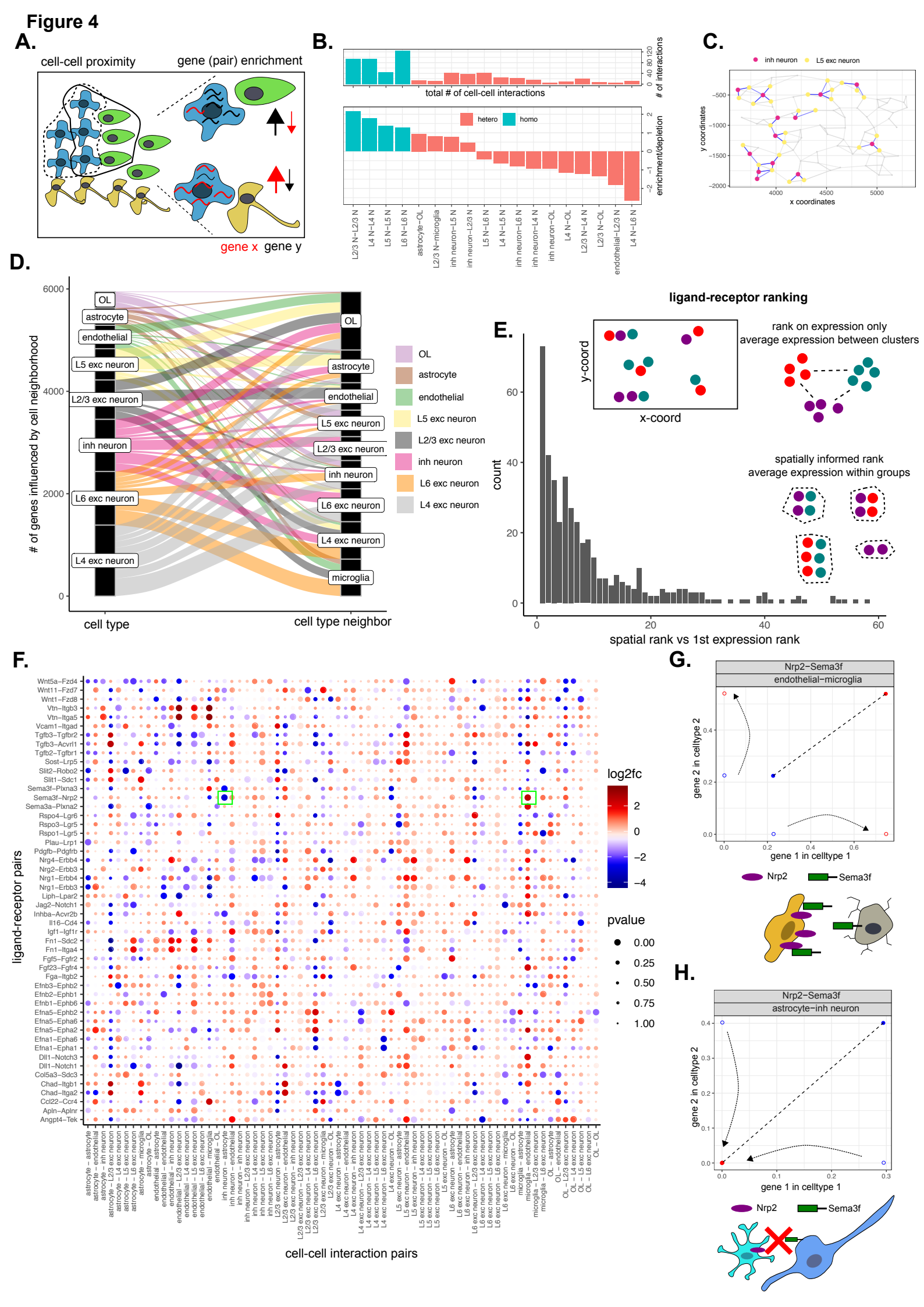

A.
cell-cell proximity
gene (pair) enrichment
gene x
gene y
B.
total # of cell-cell interactions
# of interactions
hetero
homo
enrichment/depletion
L2/3 N–L2/3 N
L4 N–L4 N
L5 N–L5 N
L6 N–L6 N
astrocyte–OL
L2/3 N–microglia
inh neuron–L5 N
inh neuron–L2/3 N
L5 N–L6 N
L4 N–L5 N
inh neuron–L6 N
inh neuron–L4 N
inh neuron–OL
L4 N–OL
L2/3 N–L4 N
L2/3 N–OL
endothelial–L2/3 N
L4 N–L6 N
C.
inh neuron
L5 exc neuron
y coordinates
x coordinates
D.
# of genes influenced by cell neighborhood
cell type
cell type neighbor
OL
astrocyte
endothelial
L5 exc neuron
L2/3 exc neuron
inh neuron
L6 exc neuron
L4 exc neuron
microglia
E.
ligand-receptor ranking
rank on expression only
average expression between clusters
spatially informed rank
average expression within groups
y-coord
x-coord
count
spatial rank vs 1st expression rank
F.
ligand-receptor pairs
cell-cell interaction pairs
log2fc
pvalue
G.
Nrp2–Sema3f
endothelial–microglia
gene 2 in celltype 2
gene 1 in celltype 1
Nrp2
Sema3f
H.
Nrp2–Sema3f
astrocyte–inh neuron
gene 2 in celltype 2
gene 1 in celltype 1
Nrp2
Sema3f

Figure 5

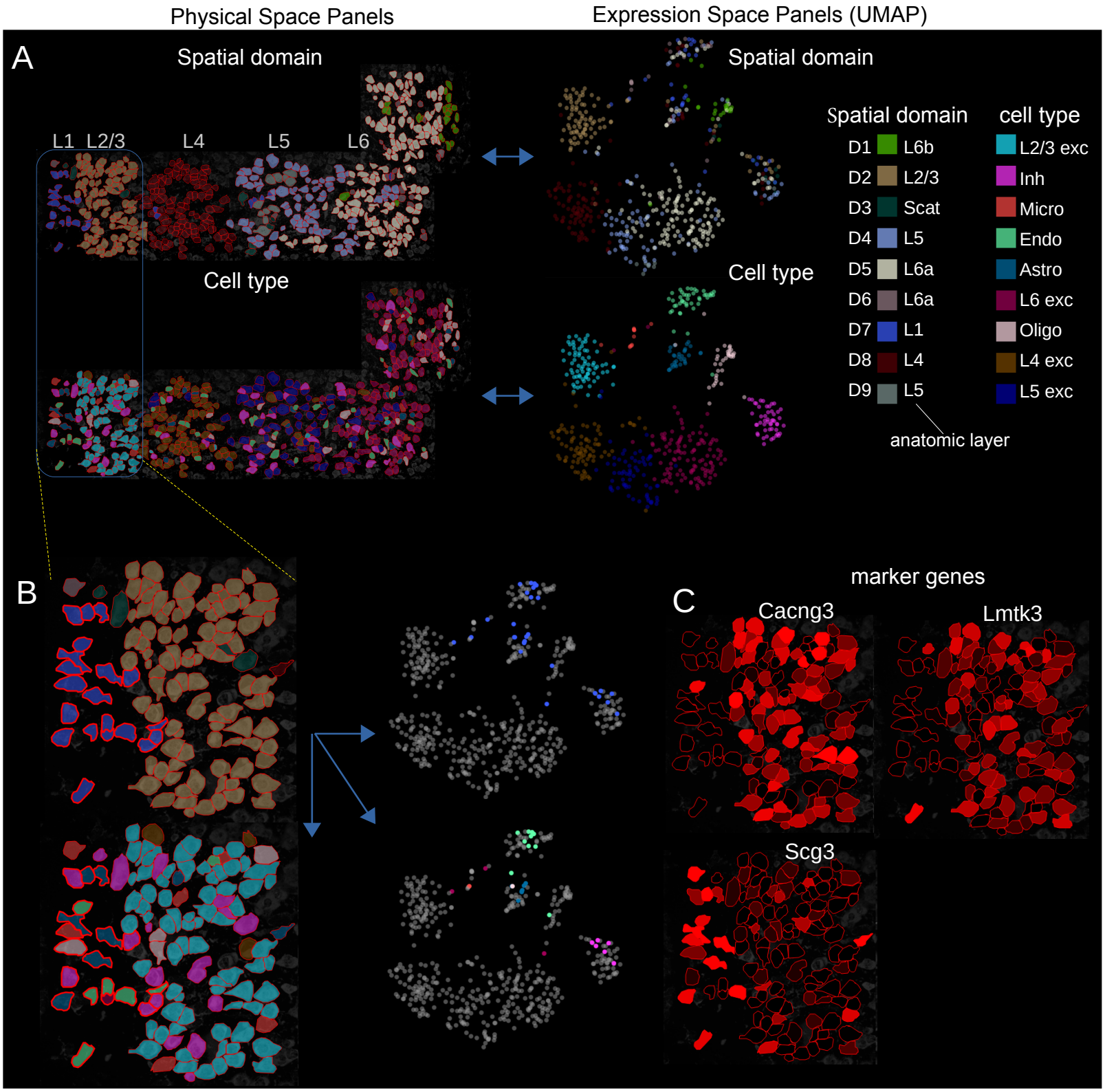
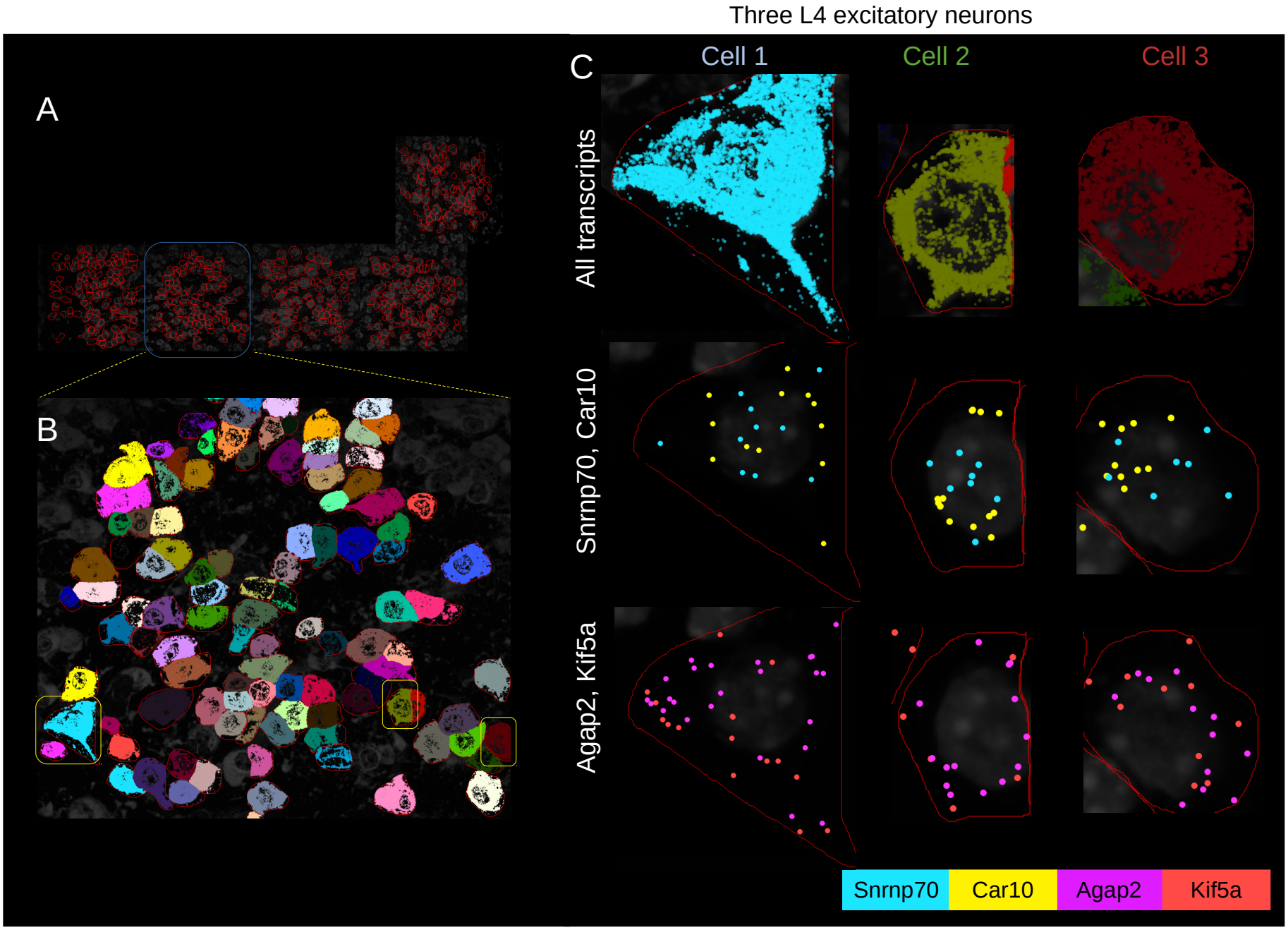
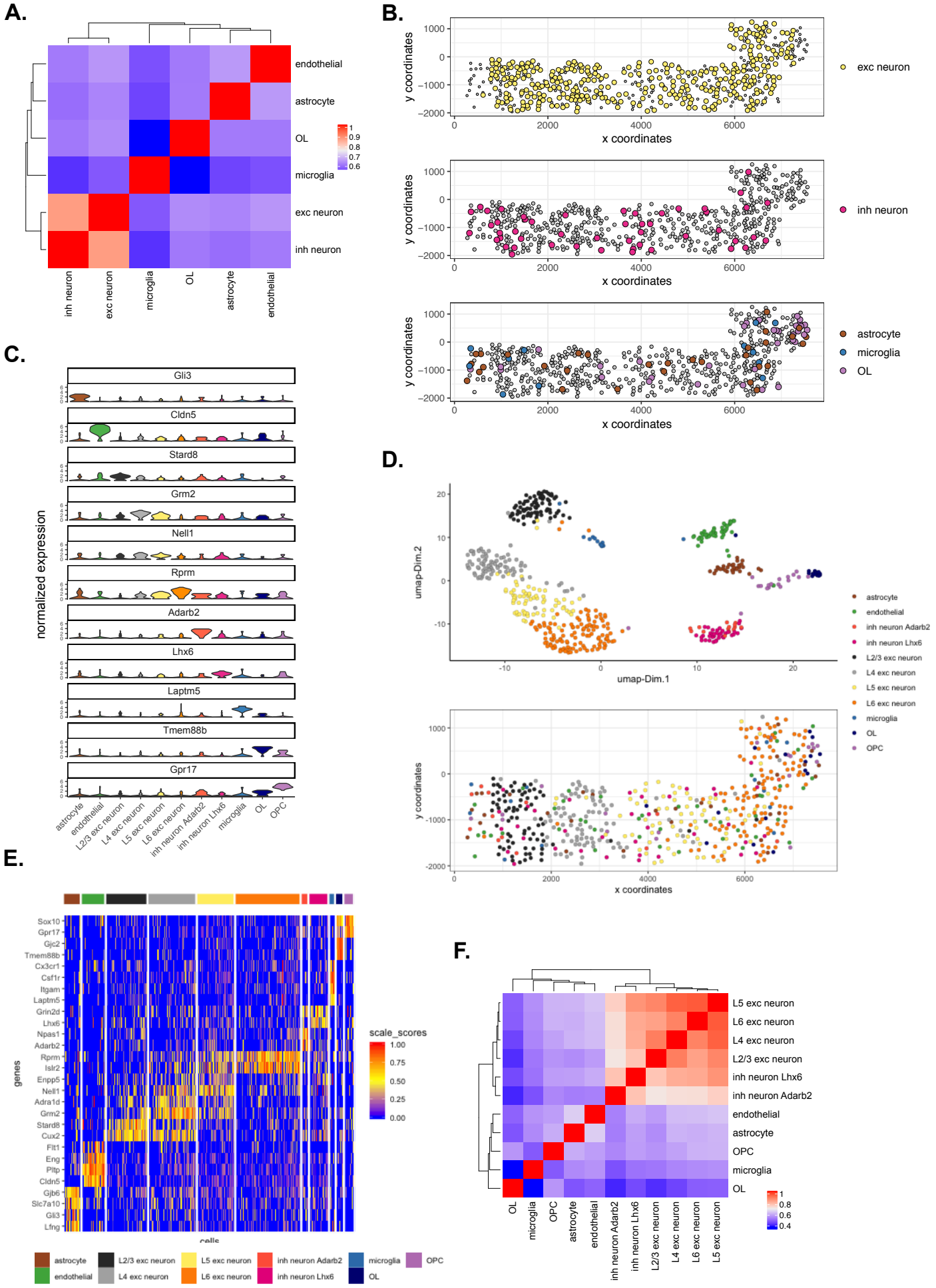

Physical Space Panels
Expression Space Panels (UMAP)
A
Spatial domain
L1 L2/3
L4
L5
L6
Cell type
Spatial domain
Cell type
Spatial domain
D1 L6b
D2 L2/3
D3 Scat
D4 L5
D5 L6a
D6 L6a
D7 L1
D8 L4
D9 L5
anatomic layer
cell type
L2/3 exc
Inh
Micro
Endo
Astro
L6 exc
Oligo
L4 exc
L5 exc
B
C
marker genes
Cacng3
Lmtk3
Scg3

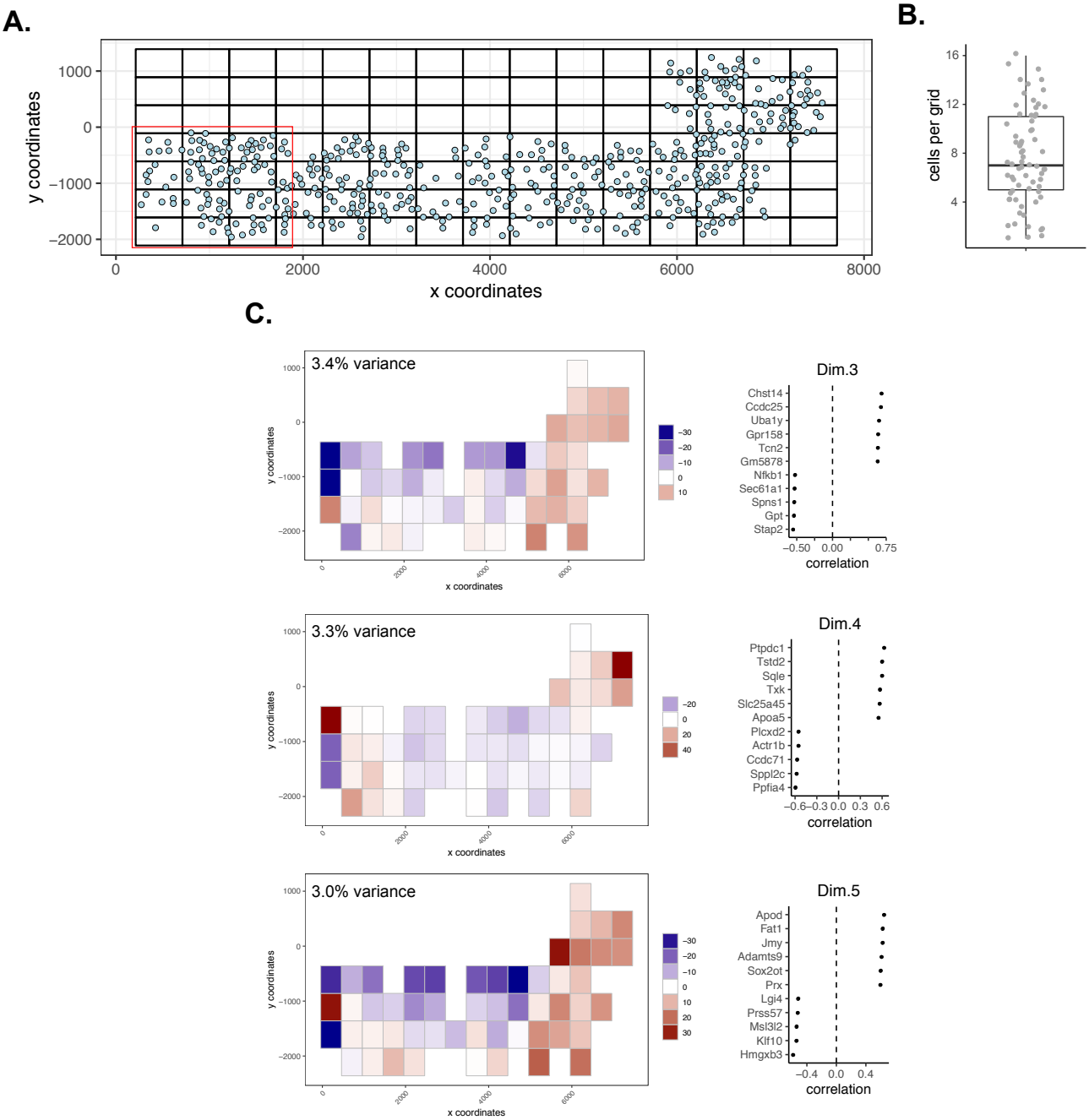
Figure 6


Three L4 excitatory neurons
A
B
C
Cell 1
Cell 2
Cell 3
All transcripts
Snrnp70, Car10
Agap2, Kif5a
Snrnp70
Car10
Agap2
Kif5a

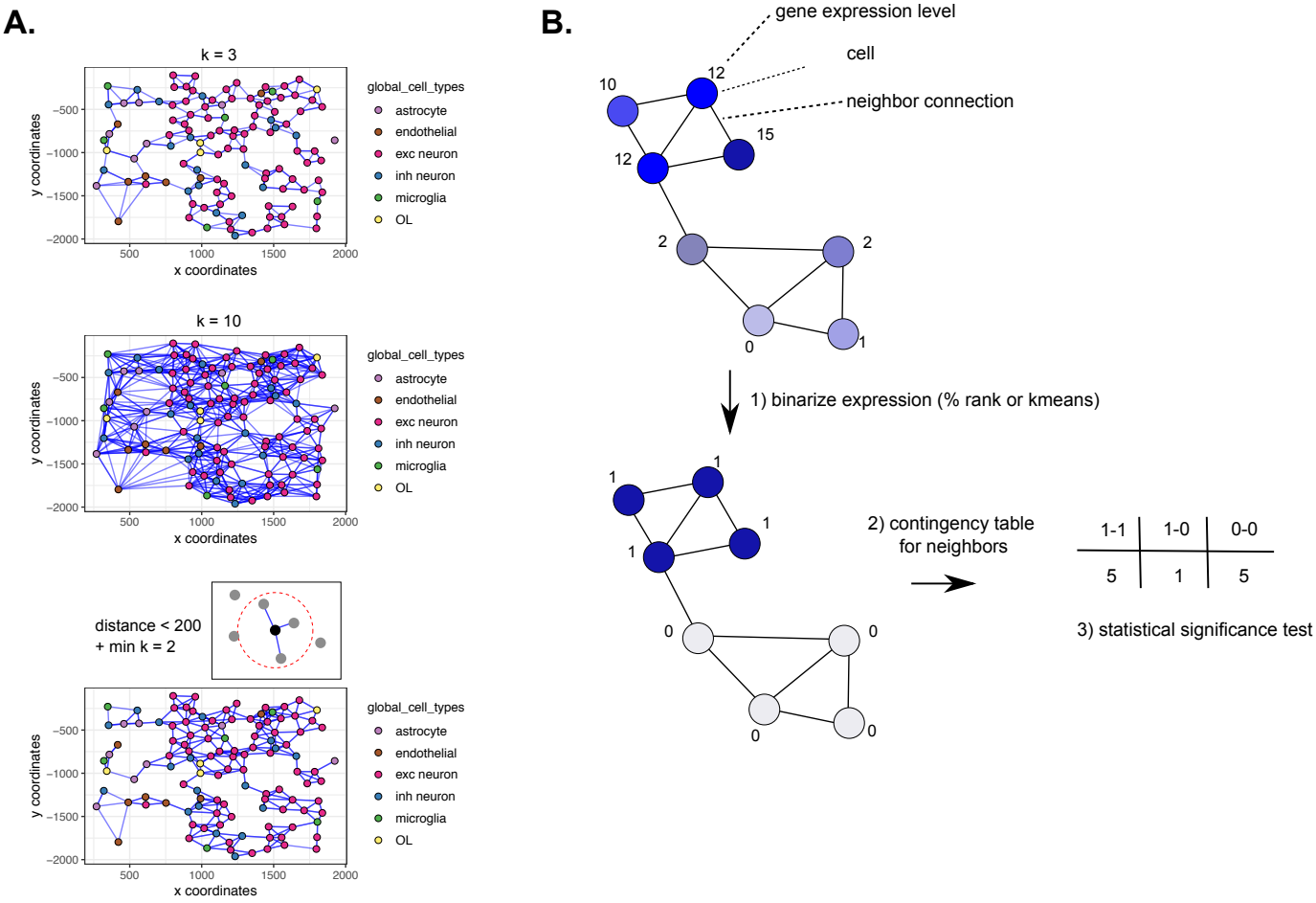
Supplementary Figure 1



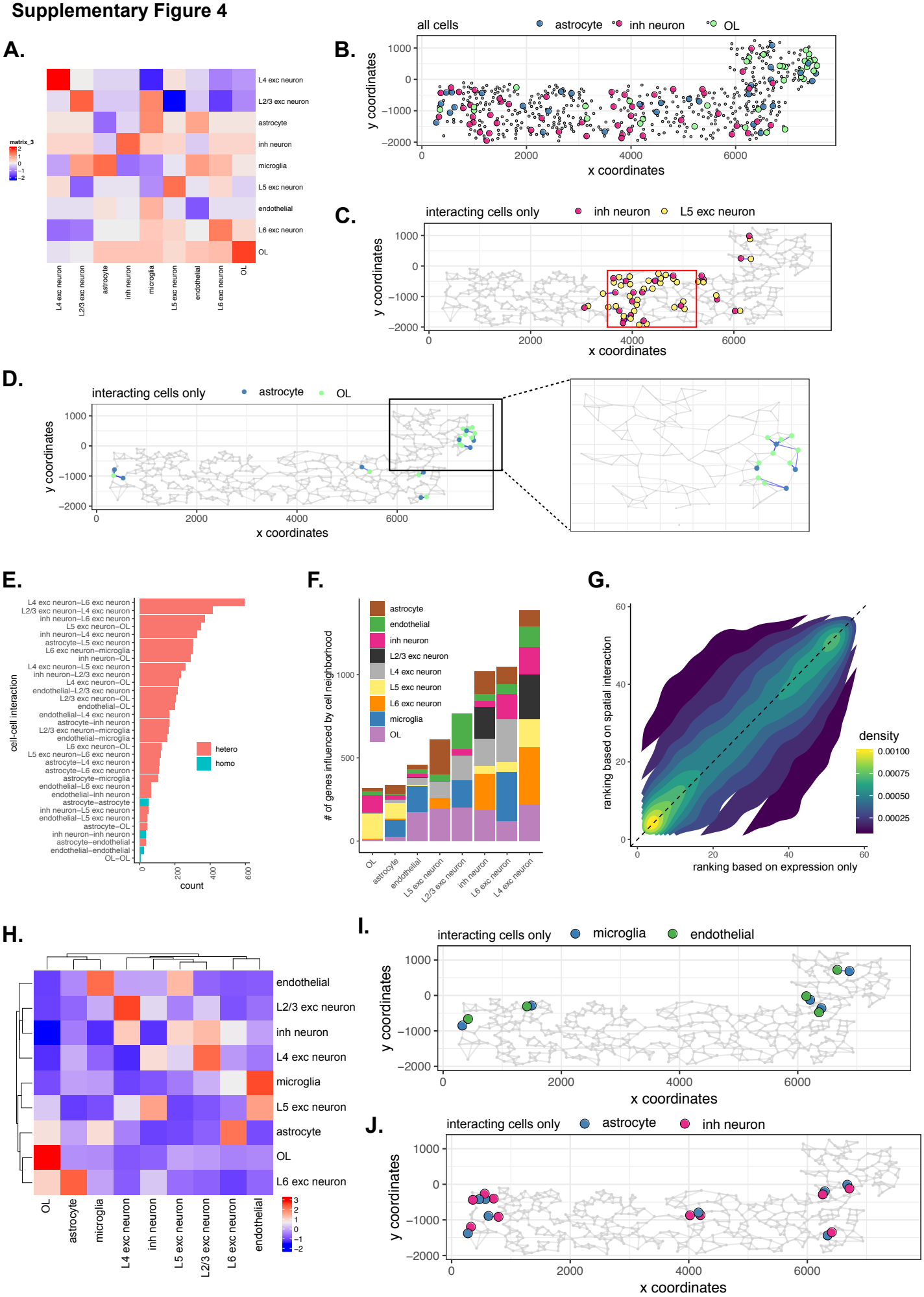
Supplementary Figure 2



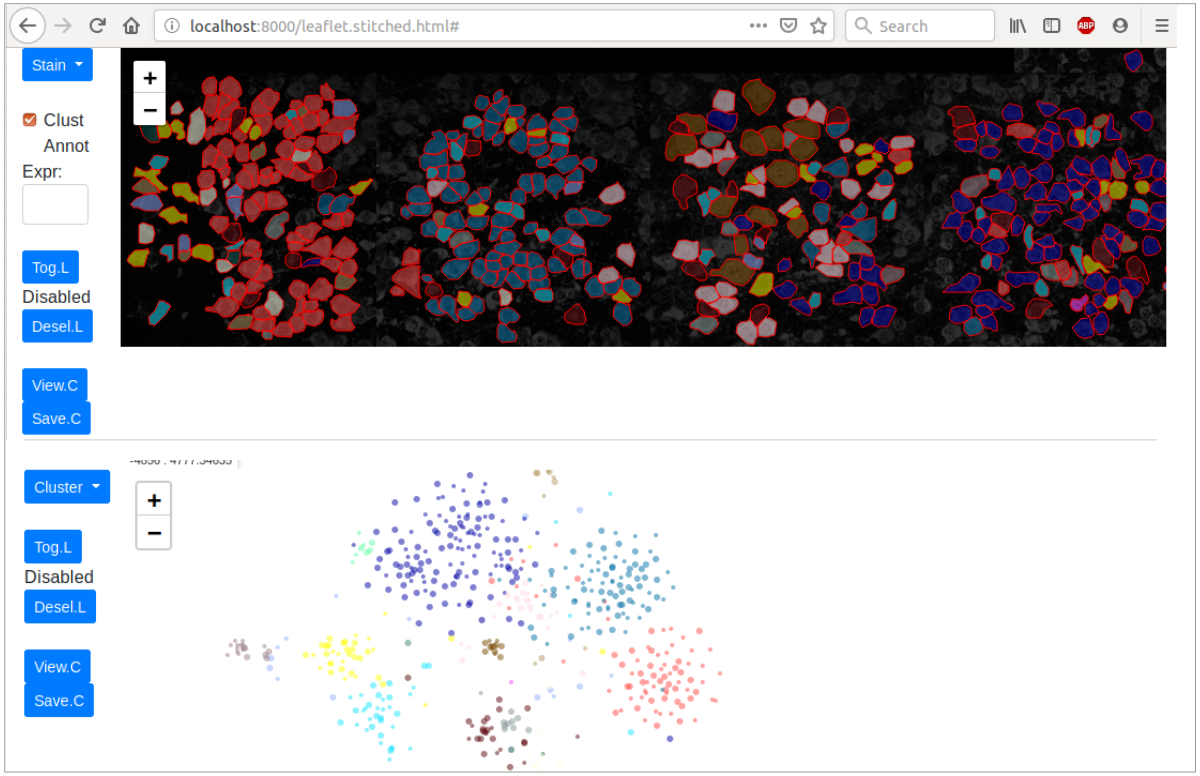
Supplementary Figure 3

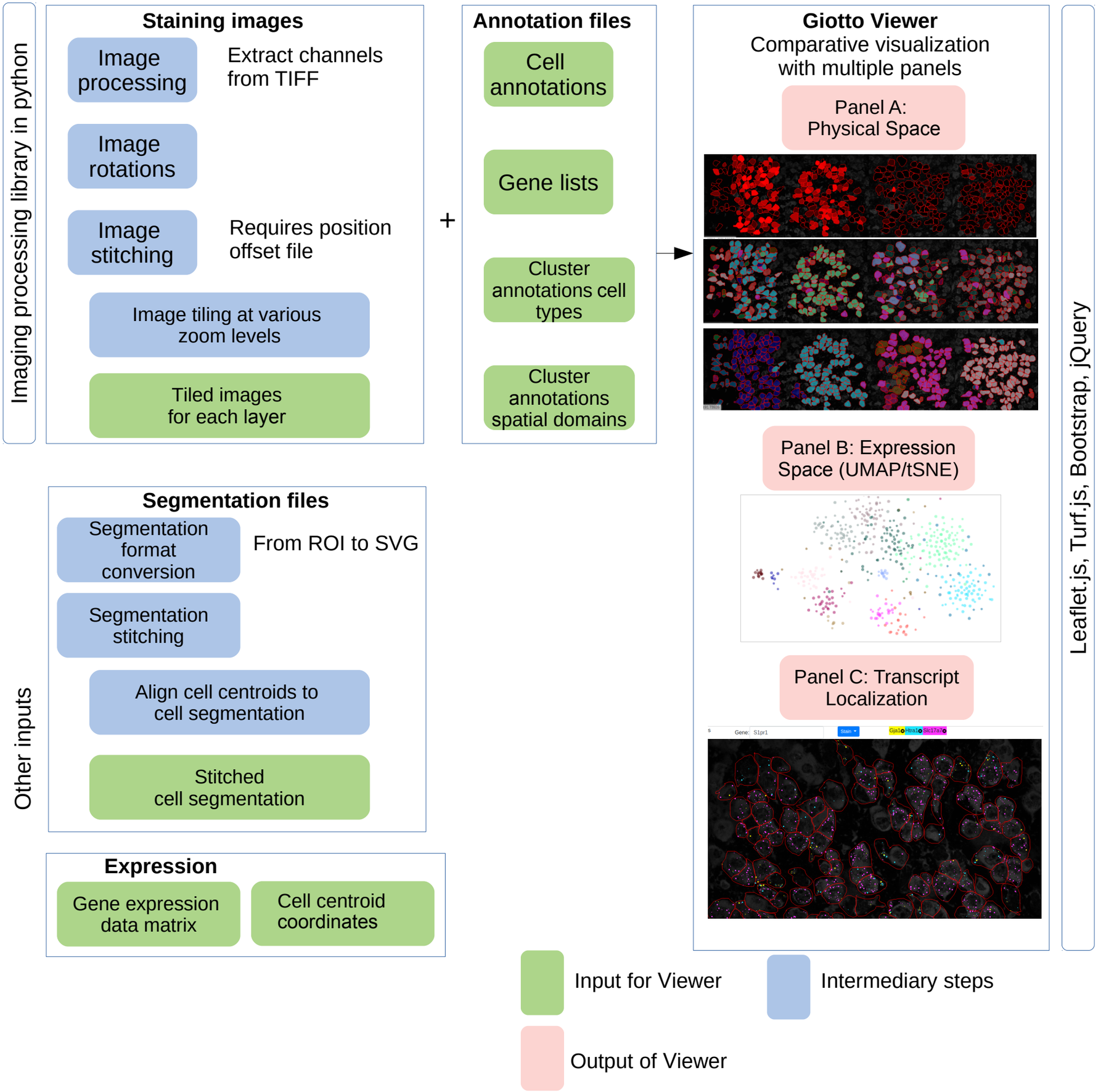


Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6

Preparations