Segmentation-free inference of cell types from *in situ* transcriptomics data

Author names

Jeongbin Park[1,2,9], Wonyl Choi[3,9], Sebastian Tiesmeyer[1], Brian Long[6], Lars E. Borm[4], Emma Garren[6], Thuc Nghi Nguyen[6], Simone Codeluppi[4,5], Matthias Schlesner[7], Bosiljka Tasic[6], Roland Eils[1,8,10,11,*] & Naveed Ishaque[1,10,*]

Affiliations

[1]Digital Health Center, Berlin Institute of Health (BIH) and Charité Universitätsmedizin, Berlin, Germany;

[2]Faculty of Biosciences, Heidelberg University, Heidelberg, Germany;

[3]Department of Computer Science, Boston University, Boston, the United States of America;

[4]Division of molecular neurobiology, Department of medical biochemistry and biophysics, Karolinska Institutet, Stockholm, Sweden;

[5]Science for life laboratory, Stockholm, Sweden;

[6]Allen Institute for Brain Science, Seattle, WA, USA;

[7]Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany;

[8]Health Data Science Unit, Heidelberg University Hospital, Heidelberg, Germany;

Author List Footnotes

[9] These authors contributed equally to this work.

[10] Senior authors.

[11] Lead Contact.

Contact information

*Correspondence: Roland Eils (roland.eils@charite.de) and Naveed Ishaque

(naveed.ishaque@charite.de)

Summary

Multiplexed fluorescence *in situ* hybridization techniques have enabled cell class or type identification by mRNA quantification *in situ*. However, inaccurate cell segmentation can result in incomplete cell-type and tissue characterization. Here, we present a robust segmentation-free computational framework, applicable to a variety of *in situ* transcriptomics platforms, called Spot-based Spatial cell-type Analysis by Multidimensional mRNA density estimation (SSAM). SSAM assumes that spatial distribution of mRNAs relates to organization of higher complexity structures (e.g. cells or tissue layers) and performs de novo cell-type and tissue domain identification. Optionally, SSAM can also integrate prior knowledge of cell types. We apply SSAM to three mouse brain tissue images: the somatosensory cortex imaged by osmFISH, the hypothalamic preoptic region by MERFISH, and the visual cortex by multiplexed smFISH. SSAM outperforms segmentation-based results, demonstrating that segmentation of cells is not required for inferring cell-type signatures, cell-type organization or tissue domains.

Keywords

In situ transcriptomics, spatial cell-type calling, segmentation-free, multiplexed FISH, SSAM, osmFISH, mFISH, multiplexed smFISH, MERFISH, spatially resolved RNA profiling

Introduction

The underlying transcriptional and spatial heterogeneity of cells gives rise to the plethora of phenotypes observed in cell types, tissues, organs and organisms. It is important to investigate this heterogeneity to better understand the basis of health and disease. Recent technological advances (Svensson et al., 2018a) have seen the profound adoption of single cell sequencing to unravel transcriptional heterogeneity in healthy and diseased tissue, and have subsequently given rise to international consortia such as the Human Cell Atlas (HCA) (Regev et al., 2017). Such efforts would not be possible without the various computational frameworks supporting analysis of single-cell sequencing data (Luecken and Theis, 2019).

Pairing this transcriptional heterogeneity with spatial heterogeneity of cells is a critical factor in understanding cell identity in the context of the tissue, for example, revealing the transcriptional basis of invasive cancer regions (Salmén et al., 2018) and highlighting rich diversity of neuronal subtype expression and localization (Moffitt et al., 2018). Recently developed multiplexed fluorescence in-situ hybridization (mFISH) (Chen et al., 2015; Codeluppi et al., 2018; Lubeck et al., 2014) and *in situ* mRNA tissue sequencing (Ke et al., 2013; Lee et al., 2015; Maniatis et al., 2019; Ståhl et al., 2016; Vickovic et al., 2019a; Wang et al., 2018) techniques have enabled the simultaneous measurement of multiple mRNAs in a spatial context. Application of cell segmentation algorithms to images obtained by these techniques identifies cells, and allows classification of classes or types of cells together with their locations (Hodneland et al., 2013; Jiang et al., 2019; Kong et al., 2015; Salvi et al., 2019). The rapid increase in establishment of *in situ* transcriptomics platforms inspired the inception of the SpaceTx Consortium (Perkel, 2019), which aims to systematically evaluate these platforms and protocols.

Segmentation-based approaches usually rely on additional signals or landmarks obtained by staining nuclei (Shah et al., 2016), cell membrane (Halpern et al., 2017; Kishi et al., 2019;

Lignell et al., 2017), or total poly-A RNA (Codeluppi et al., 2018; Moffitt et al., 2018). Accurate cell segmentation, however, is difficult to achieve due to tightly apposed or overlapping cells, uneven cell borders, varying cell and nuclear shapes, signal intensity variation, probe fluorescence emission efficiency variation, and tiling artifacts (Thomas and John, 2017). The underlying problem is that the cellular structures one would want to segment are much smaller than the resolution of a diffraction-limited microscope. Therefore, there is a need for robust segmentation-independent methods for identification of cell-type signatures, cell-type organization, and tissue domains from multidimensional mRNA expression data in complex tissues. These methods could be used for datasets lacking landmarks or to validate segmentation-based approaches and identify associated artifacts.

Here we introduce a novel computational framework named Spot-based Spatial cell-type Analysis by Multidimensional mRNA density estimation (SSAM), a multi-platform segmentation-free computational framework for identifying cell-type signatures and reconstructing cell-type and tissue domain maps from both 2D- and 3D-spatially resolved *in situ* transcriptomics data. We apply SSAM to three mouse brain tissue images obtained by different techniques: the somatosensory cortex by osmFISH, the hypothalamic preoptic region by MERFISH, and the visual cortex by multiplexed smFISH. We demonstrate the performance of SSAM in identifying 1) cell types *in situ*, 2) spatial distribution of cell types, 3) spatial relationships between cell types, and 4) tissue domains (e.g., cortical layers) based on the local composition of cell types without having even segmented a single cell.

Results

The SSAM computational framework

SSAM consists of 4 major steps (**Figure 1**), namely 1) cellular mRNA density estimation and selection of representative gene expression profiles; 2) computation of cell-type signatures; 3) generation of a cell-type map, and 4) identification of tissue domains. In the first step,

SSAM estimates intra-cellular mRNA densities and then selects representative cell-wise gene expression profiles (**Figure 1A**). SSAM models the density of intracellular mRNAs as the probability density representing the existence of observed mRNA molecules. The mRNA density model is computed by applying Kernel Density Estimation (KDE) (Parzen, 1962; Rosenblatt, 1956) with a Gaussian kernel whose dispersion pattern models average cell size. Here, we measure the dispersion pattern as the full width tenth maximum (FWTM) of the Gaussian distribution. This approach models the distribution of mRNAs in the cell body, while also preserving the shapes of cells, successfully recovering the mRNA density over the tissue (**Figure S1A**). Further, it will also recover the shape of the distributions of subcellular localizing mRNAs.

SSAM then projects the mRNA density values estimated via the Gaussian KDE onto a square lattice, which represents coordinates in the tissue. The spacing between adjacent points of the lattice is set to have an order of magnitude below the average cell diameter (1 μm in our examples), to ensure both high resolution and feasible computation time. Next, the mRNA densities estimated per gene are stacked to produce a vector field over the lattice, which will be called the gene expression vector field hereafter. The gene expression vector field is analogous to a 2D/3D image where each pixel/voxel encodes the estimated gene expression of the unit area.

Next, the vectors representing likely cell locations are selected based on their total gene expression. Recalling that most mRNAs would be found inside the cell body, the total mRNA densities of a unit area reflects the probability of the unit area locating inside a cell body. Moreover, since the smoothing effect of the Gaussian kernel propagates information across neighboring positions, any local maximum of gene expression in the gene expression vector field would contain signal from all nearby mRNAs, and therefore local maxima can be considered as the representative transcriptome of its containing cell (**Figure S1C**). These local maxima are filtered to remove artifacts, and can be further restricted to informative

parts of the tissue image using an optional “input mask” (Methods). Then, both the local maxima vectors and the vector field are then normalized (Methods).

After this first step, SSAM can be operated in either *guided* mode or *de novo* mode. The guided mode SSAM assigns cell types to the representative vectors using a given set of cell-type gene expression signatures. In guided mode, the following second step can be skipped and continued to the third step.

In the *de novo* mode, SSAM identifies cell-type gene expression signatures utilizing clustering algorithm and classifies the representative gene expression vectors to known cell types (**Figure 1B**). First, SSAM initially clusters the representative vectors using a reimplementation of the Seurat clustering method (Butler et al., 2018) (Methods). After clustering, SSAM excludes representative vectors that are lowly correlated to the cluster medoid, (i.e., the representative object of a cluster whose average dissimilarity to all the objects in the cluster is minimal). After filtering the lowly correlating vectors, SSAM models the representative gene expression profile for each cell type identified from clustering as the centroid of each cluster (i.e. the unweighted mean gene expression profile of a cluster) (**Figure S1A**). Next, clusters with highly similar expression profiles can be merged and dubious clusters removed. To assist users in selecting clusters for merging or removal, SSAM generates ‘diagnostic plots’ for each cluster (Methods).

In the third step, SSAM generates a cell-type map image (**Figure 1C**). Here, SSAM maps the computed cell-type signatures in *de novo* mode (or the given cell-type signatures in *guided* mode) to the normalized vector field. Each position is assigned a single cell type based on the highest Pearson’s correlation with the respective cell-type signature (**Figure S2A**). This assignment of cell-type signatures to the vector field is used to generate the cell-type map, which can be used to investigate the spatial distribution of cell types in the profiled tissue. In cases where the user is interested in a particular region of the tissue, an “output

mask" can be applied to restrict reporting to a specific region of the image.

In the fourth step, SSAM identify tissue domains by defining local neighborhoods of similar cell-type composition (**Figure 1D**). SSAM computes the cell-type compositions in a circular(2D)/spherical(3D) sliding window over the cell-type map, and clusters them using agglomerative hierarchical clustering (**Figure S2B**). Clusters with high correlation to each other are then merged into a single tissue domain signature, and the cell-type composition of each domain is calculated.

In the following sections, we provide results of SSAM applied to three multiplexed FISH datasets obtained using different methodologies. We reanalyze two previously published datasets, profiled by osmFISH (Codeluppi et al., 2018) and MERFISH (Moffitt et al., 2018), and provide biological insights of a newly generated multiplexed smFISH dataset.

SSAM improves astrocyte and ventricle detection in the mouse brain somatosensory cortex (SSp)

To demonstrate utility of SSAM, we first analyzed published osmFISH data, where the transcripts of 33 cell-type marker genes were localized in 2D space of the mouse brain somatosensory cortex (SSp) (Codeluppi et al., 2018).

The osmFISH dataset was analyzed using both the guided and *de novo* modes of SSAM (**Figure 2, 3**). For the guided mode, two sets of pre-determined cell-type signatures were used to generate cell-type maps: one obtained by segmentation of the osmFISH data and another from scRNA-seq (Marques et al., 2016; Zeisel et al., 2015). The resultant cell-type maps were very similar to the one previously published (**Figure S4E**).

Next, we tested a completely *de novo* cell-typing approach. The resultant 30 cell-type signatures (**Figure 2A, B**) were consistent with those identified from the segmentation-based

clustering (**Figure S4C**) and scRNA-seq based clustering (**Figure S4D**) (Codeluppi et al., 2018). As such, the SSAM *de novo* cell-type signature clusters were named after the closest matching segmentation-based cluster. Despite the difference of normalization and clustering methods, we find that the clustering of cell-type signatures are comparable. The domain analysis result based on SSAM *de novo* cell-type map correlated well with the known cortical layers of the tissue, consistent with results reported in the previous study (**Figure 3A**).

All of the cell-type maps generated by SSAM showed high density of *Mfge8* expressing astrocytes (Astrocyte Mfge8) in visual inspection. The tissue domains inferred from the *de novo* cell-type map also showed high contributions from the *Mfge8* expressing astrocytes (**Figure 3B**), confirming what we observed in visual inspection. Generally, we found that Mfge8 expressing astrocytes contributed 7-14 % of each of the tissue layers, in contrast to the significantly fewer numbers of Astrocyte Mfge8 cells called in the previous study (Codeluppi et al., 2018). Comparison of high-resolution images of DAPI and poly-A signals with KDE generated Mfge8 expression implicates that the poly-A signal was not strong enough to discriminate presence of Mfge8 expression astrocyte cells from the background, while the DAPI images clearly supported the existence of Mfge8 expressing astrocytes identified by SSAM (**Figure 2E**). The clear DAPI signal but low poly-A signal for these Mfge8 expressing astrocytes implicates that they would have a lower mRNA content compared to other cells. To confirm this we investigated the total counts of mRNA molecules of astrocytes compared to other cell types from mouse brain scRNA-seq data (Zeisel et al., 2018). We found that astrocytes exhibited significantly less mRNA molecules than other cell classes (**Figure S4B**). Our observation reveals the inadequacy of the watershed segmentation algorithm applied to poly-A signal when not considering cells with a low total mRNA content. In addition, elements of the protruding processes of astrocytes were recovered in our cell-type map (**zoom panel, Figure 2C**).

SSAM also reconstructed a more complete structure of the ventricle, composed of

ependymal (yellow) and choroid plexus cell types (teal), compared to the results of the previous study (**Figure 2D**). The Poly-A and DAPI signals confirm the existence of both cell types in the ventricle area, but since ependymal and choroid plexus cells are small and tightly packed, and exhibit relatively lower DAPI and poly-A signal, the performance of watershed algorithm was insufficient to identify cells in the area.

SSAM confirms diversity of inhibitory and excitatory neuron cell types and localization in the hypothalamic preoptic region (POA)

To demonstrate the performance of SSAM for three-dimensional *in situ* transcriptomics data, we applied SSAM to previously published MERFISH data, where 135 transcripts were localized in 3D space of the hypothalamic preoptic region (POA) of a mouse brain (Moffitt et al., 2018). In this section, we demonstrated SSAM on a single layer of the MERFISH data at the posterior region of the mouse POA.

We first tested both SSAM guided mode and *de novo* mode. For guided mode, the previously known cell-type signatures obtained by segmentation and scRNA-seq were used. Both in guided mode and *de novo* mode, since the input mRNAs are located in 3D space, SSAM analysis were performed in 3D space accordingly. The resulting cell-type maps on the x-y plane at the center of slice on the z-axis (at 5 μm) were visually similar to the previous study (**Figure S5G**). Also, the SSAM cell-type signatures showed high correlation to the cell-type signatures from both the segmentation-based clusters and scRNA-seq clusters (**Figure S5E, F**). While we missed some cell types identified in the previous study by restricting our analysis to one out of twelve slices of the dataset, we found a similarly large number of cell-type signatures of inhibitory and excitatory neurons compared to Moffitt *et al* (25 vs 39 inhibitory and 13 vs 31 excitatory neurons). We also observed similar tissue localization patterns for inhibitory and excitatory cell types (**Figure 4D, E**), validating the computational approach adopted by SSAM to identify de novo cell-type signatures. The generated tissue domain map clearly shows the structure of tissue simplified to several domains consisting of

region with mainly inhibitory neurons, excitatory neurons, oligodendrocytes and the ventricle structure (**Figure 5A, 5B**).

Finally, the reconstructed three-dimensional cell-type map is visualized in movies of the turntable-rotating cell-type map (**Movie 1**). The whole cell-type map and the cell-type specific maps for inhibitory / excitatory neurons and astrocytes by sweeping in the z-direction with a scale of 1 μm (**Movie 2, 3, 4**) were generated, demonstrating the size and shape difference of cells giving rise to the cell-type signal identified by SSAM.

Compared to the osmFISH dataset, which was from a 2D image, the MERFISH dataset was from a 3D image. Despite the difference of dimensionality, SSAM is still able to successfully process the data and produce meaningful results. More importantly, the analyses in this section was performed with almost the same procedure and parameters used for the osmFISH data analysis: same lattice spacing, same bandwidth, same cluster refining threshold, circular window size among others. Therefore, we set the parameters as the default values. This implies that one can easily analyze their own multidimensional *in situ* transcriptomics dataset with little effort and generate accurate and meaningful results rapidly using the default parameters of SSAM.

SSAM defines rare cell types and cortical sub-layering in the adult mouse visual cortex (VISp)

To further demonstrate that SSAM can be used for rapid and robust analysis of *in situ* transcriptomics data, we applied SSAM to unpublished multiplexed smFISH data of the mouse primary visual cortex (VISp) generated as part of the SpaceTx consortium (Perkel, 2019). In total, the expression of 22 genes was quantified *in situ* (Methods).

The VISp region on the tissue was manually defined to restrict analysis to relevant cell types (**Figure S6D**), and local maxima vectors were only selected within the defined VISp region

input mask (**Figure S6A**). SSAM was performed in both guided mode and de novo mode (**Figure S7D**). The guided mode of SSAM was performed with scRNA-seq data (Tasic et al., 2018). For the *de novo* run, the name of cell-type signature clusters were assigned with the name of closest correlated cluster in the scRNA-seq data (**Figure 6A,B**). Then the tissue domains were identified based on *de novo* cell-type map (**Figure 7**), with the result showing the laminar structure of the VISp region. We found that there were two different layer 4 (L4) neuronal clusters determined by SSAM. Interestingly, both of them showed the highest correlation to the single L4 IT type identified via scRNA-seq, but their spatial locations show a clear difference (**Figure 6C, S7C**). We named the cluster mapping to superficial region of L4 layer as 'L4 IT Superficial'. This finding adds context to the previously observed heterogeneity of the L4 IT cell type (Tasic et al., 2018), where the heterogeneity could be related to superficial and deep localization in layer 4.

The cell-type map generated by SSAM guided mode were visually similar to that of *de novo* mode, except for the cell types found in the layer 2 (L2) (**Figure S7D**). We found that the majority of cell types found in L2 were assigned to the VLMC type in SSAM guided mode. We observed that this type was actually a neuronal type in L2. This cell type showed high expression of *Alcam*, a marker gene of the VLMC cell type, but low expression of other genes. Due to the limited number of genes profiled in the multiplexed smFISH experiment, lack of other neuronal marker genes led to incorrect high correlation of this type VLMC. However, SSAM properly assigned the centroid to be L2 neurons in *de novo* mode.

There is one type mapped in the cell-type map generated by SSAM guided mode (yellow spot, **Figure S7D**), that was not found in the initial try of SSAM *de novo* mode. The corresponding type found in the scRNA-seq data is Sst Chodl, which is known to be a rare neuronal type related to long-range projection and sleep-active neurons (Gerashchenko et al., 2008; Tasic et al., 2016; Tomioka et al., 2005). Therefore, we manually verified whether SSAM detected the vectors corresponding to this cell type. We found that there were two

high Chodl expressing vectors that were in close proximity to each other (**Figure S7A**), but due to the limitation of the clustering method employed, these vectors were not clustered correctly. By comparing the centroid of the two vectors to the scRNA-seq data, we found that this cell-type signature showed highest correlation to the Sst Chodl cluster in scRNA-seq data (**Figure S6C, S6E**). In addition, the centroids of these types were found in layer 5, consistent with the localization of Sst Chodl types to L5 and L6 as previously reported (Tasic et al., 2016). Thus, these two vectors were manually rescued (**Figure S7A, B**), and its centroid added to the list of signatures identified by SSAM. Also, in the cell-type map the centroid is clearly mapped to the yellow spot region but not anywhere else (**Figure 6C**), confirming that its gene expression signature is unique to other areas in the vector field. This finding is consistent with the expectations that the Sst Chodl cell type has a very distinct expression signature.

The application of SSAM to this previously undescribed dataset, using default parameters, clearly demonstrates that it is feasible to rapidly and robustly identify cell types and tissue structures without segmentation. This example also demonstrates the possible use case of employing both guided mode and de novo mode of SSAM - the former is helpful to quickly identify known rare cell types in tissue, and the latter can be used to identify new cell-type clusters not observed in the scRNA-seq data.

Discussion

We describe a segmentation-free computational framework for processing *in situ* transcriptomics data and demonstrate its performance on three different adult mouse brain datasets: the somatosensory cortex (SSp) profiled by osmFISH, the hypothalamic preoptic region (POA) by MERFISH, and the visual sensory cortex (VISp) by multiplexed smFISH. We find that the cell-type signatures and maps generated by SSAM for both osmFISH and MERFISH datasets were similar to the previously reported ones, validating the underlying methodology of SSAM. Based on this, we successfully determined cell types and

constructed cell-type and tissue domain maps in the multiplexed smFISH mouse VISp dataset.

In the osmFISH dataset our method outperforms the original segmentation-based cell-type map reconstruction in cases that were limited by the segmentation process. In the MERFISH dataset we show that SSAM is able to identify diverse populations of cell types and that SSAM is scalable to 3D image data by reconstructing plausible tissue structures in 3D. For the VISp multiplexed smFISH data, SSAM identified a rare cell type and elucidated a suspected spatial heterogeneity of cell types in the cortex without segmenting a single cell.

SSAM is a reasonable alternative to segmentation based analysis, especially in difficult to segment tissues or when DAPI or poly-A images are not available. However, for some questions it is important to distinguish between cells to e.g. delineate growth arising from increasing cell size vs cell proliferation or to investigate multinucleation in cardiomyocytes or cytotrophoblast cells. In cases such as these, we also postulate the use of SSAM as a complementary method to segmentation-based analysis in two ways. First, the output of SSAM can be compared to validate that the segmentation process did not introduce artifacts. Secondly, to use the SSAM output as an input for the segmentation process to refine the segmentation procedure for different domains or cell-type signals.

SSAM identified a distinct cell-type signature of Aldoc-expressing astrocytes that had low expression of Gfap and Mfge8. When looking closely at the localization of their signal they corresponded to specific subcellular compartmentalization in astrocytes that express high levels of Mfge8, which could also be due to localization of these mRNAs to different parts of the cell in astrocytes, due to the internal subcellular localization of the gene in Astrocytes. Such an intracellular spatial organization of the transcriptome is often an important form of post-transcriptional regulation (Flynn et al., 2019) and imaging-based methods can reveal this organization (Battich et al., 2013). Thus, SSAM can be used to identify and investigate

organization of mRNAs.

One important parameter of SSAM is the kernel bandwidth of the KDE. We initially rationalized the bandwidths FWTM as 10.7 μm being close to the average size of cells in the mouse brain SSp tissue image (11.6 μm), and therefore applicable for the other brain tissue images, which is demonstrated by SSAMs performance in identifying neurons, microglia and astrocytes of different sizes. However, these tissues contain cells that exhibit a range of sizes which makes it hard to postulate that the FWTM of the kernel is a one-size-fits-all just because of its size. In fact, we believe the bandwidth needs only to be sufficient to smooth the gene expression signal over the majority of the cell body in order to identify an local maxima vector which represents that cell's gene expression profile. Future applications of SSAM will need to show whether this may be a parameter that would need to be optimized for other tissue types that consist of cells of varying sizes and densities.

In our clustering analysis of all three datasets, we observed that some clusters showed moderate mixed expression of signature genes from different cell types. Possible causes include: (1) different cells can overlap at different z location if the thickness of the section is comparable to the cell size; (2) clustering of the vectors might not be perfect and can include vectors in nearby clusters; (3) the gene expression estimated by the KDE algorithm is smoothed and the gene expression of one cell type can contaminate cells with different cell types located nearby. We found that (1) and (2) are of major importance for this phenomenon, but (3) also becomes noticeable in the closely packed small cells. For example, we found that relatively higher expression of the *Foxj1* gene (a signature gene of ependymal cells) is detected in the choroid plexus signature, compared to that of segmentation-based osmFISH centroid. Such signal 'contamination', caused by the KDE spreading signal into adjacent cells, can be controlled by the bandwidth value of KDE - the smaller the bandwidth, the lower the contamination; however, use of very low bandwidths break one of the primary assumptions of SSAM in that the smoothed KDE signal should

represent cells, and not subcellular features. In this paper, we showed that this phenomenon is not so critical as to hinder detection of cell types in our examples when using a bandwidth of 2.5 µm (thus making this the default values for the bandwidth for KDE). However, it is recommended trying different bandwidths when it is expected that the average cell size deviates significantly from 10 µm.

Currently, the field of *in situ* transcriptomics is advancing rapidly and more than 10,000 genes can be simultaneously profiled using FISH-based methods (Eng et al., 2019; Xia et al., 2019). The high number of genes detected in large volumes opens up the potential for *in situ* transcriptomics methods to at least partially replace single cell RNA sequencing at large scale, placing SSAM as the first generic and segmentation-free pipeline to rapidly and precisely reconstruct tissue structure independent of the underlying imaging technique. Moreover, since the only required input data for SSAM is mRNA locations, it is highly adaptable to spatially resolved transcriptomics technologies beyond FISH methods, e.g. *in situ* or intact tissue sequencing (Ke et al., 2013; Lee et al., 2015; Wang et al., 2018), composite *in situ* imaging (Cleary et al., 2019), Slide-seq (Rodriques et al., 2019), and Spatial Transcriptomics (Ståhl et al., 2016; Vickovic et al., 2019b).

Also, the modular nature of the SSAM framework allows for easy incorporation of new features such as spatial differential gene expression analysis (Svensson et al., 2018b), pseudo-time analysis to infer differentiation trajectories (Angerer et al., 2016; Haghverdi et al., 2016; Qiu et al., 2017; Trapnell et al., 2014) and RNA velocity analysis to analyze the speed of transcriptional reprogramming or flux (La Manno et al., 2018) that is particularly applicable to the recently published intronSEQFISH technique (Shah et al., 2018).

In summary, we present a novel algorithm, SSAM, to analyze cell types based on mRNA locations. Although not required, SSAM can make use of both prior defined cell-type signatures and segmentation. SSAM not only reproduces cell-type maps comparable to

segmentation-based approaches, but can improve them when image based cell segmentation is the limiting factor. This not only places SSAM as an independent method to segmentation-based approaches, but also a complementary one. SSAM is written as a Python library, with some core analysis functions wrapped up with external C functions to speed up the computation. The package is available as an easily installable Python package, and can easily be extended with existing *in situ* transcriptomics pipelines, e.g. starfish (https://github.com/spacetx/starfish) or Giotto (Dries et al., 2019). SSAM is accompanied with a notebook outlining all the steps presented in this paper. Taken together, we present a novel, flexible and robust method for fully automated cell-type and tissue domain analysis that is readily applicable to virtually any *in situ* transcriptomics methods including all imaging and *in situ* sequencing methods.

Acknowledgements

We thank Sten Linnarsson and Jeffrey Moffitt for providing support and access to the osmFISH and MERFISH datasets, respectively. We also thank Yue Zhuo, Ed Lein, Jeremy Miller, Ambrose Carr, Nagarajan Paramasivam, Stephen Krämer, Zuguang Gu, Daniel Hübschmann, Luca Tosti, and Christian Conrad for helpful discussions and comments on data analysis. The authors also thank Bianca Hennig for designing Figure 1, and assistance in improving figures. The preliminary analysis of multiplexed smFISH data occurred during the SpaceTx SpaceJam Hackathon at the Allen Institute for Brain Science, which was organized by Ed Lein, and generously supported by the Chan Zuckerberg Initiative.

Author contributions

JP, WC, RE, NI conceived the study.

BT, EG, TN.N, BL acquired and interpreted the multiplexed smFISH data.

JP, WC, ST, TN.N, NI performed data analysis.

LE.B, MS, BL, BT provided critical comments and discussions.

RE, NI supervised the study.

All authors commented on and critically revised the manuscript.

Declaration of interests

The authors declare no competing interests.

Figure titles and legends

Figure 1. Schematic diagram of the SSAM computational workflow for cell type and tissue domain definition based on gene expression data.

(A) In step 1, SSAM converts mRNA locations into a vector field of gene expression values. For this, SSAM applies a Gaussian KDE to mRNA locations for each gene and projects the resulting mRNA density values to a square lattice which represents coordinates in the tissue. The mRNA density estimated per each gene are stacked to produce a "gene expression vector field" over the lattice. The gene expression vector field is analogous to a 2D/3D image where each pixel/voxel encodes the averaged gene expression of the unit area. Further details of the application of KDE can be found in **Figure S1A**.

(B) In step 2, cell-type signatures are identified *de novo*. First, the gene expression profile at probable cell locations are identified as the local regions in the gene expression vector field where the signal is highest. These local maxima of gene expression signals are identified and used for *de novo* cell type identification by cluster analysis. Alternatively, previously defined cell-type signatures can be used. Further details on local maxima selection can be found in **Figure S1B**.

(C) In step 3, a cell-type map is generated. For this, the cell-type signatures are mapped onto the gene expression vector field and cell types are assigned based on Pearson's correlation between each cell-type expression signature to the vector field to define cell-type distribution *in situ*. Further details about creating the cell-type map can be found in **Figure S2A**.

(D) In step 4, the tissue domains are identified. The tissue domain signatures are identified using a sliding window to sample the cell-type neighborhood around local maxima. The tissue domain map is created by mapping these signatures onto the cell-type map. Further details on creating the tissue domain map can be found in **Figure S2B**.

See also **Figure S1, S2.**

Figure 2. SSAM improves astrocyte and ventricle detection in the mouse SSp region.

(A) Gene expression heatmap showing cell-type specific expression of marker genes. Rows show z-score normalized gene expression and columns show the gene expression patterns of filtered local maxima vectors (representative of gene expression within a cell). The top annotation shows the cell types and coloring based on the best correlating segmentation-based cell-type signature from Codeluppi *et al*. The colors of the top annotation correspond to the cell type legend in **Figure 2B**.

(B) A t-SNE map of cell-type signatures with distinct expression. Cell-type clusters are visualized as a 2D t-SNE embedding of filtered local maxima vectors. Cell-type annotation and coloring are based on the best correlating segmentation-based cell-type signature from Codeluppi *et al*. The cell-type legend is grouped by cell-type classes labels shown in the tSNE plot, and are based on groupings by Codeluppi *et al.,* (**Figure S3B,C**).

(C) The SSAM de novo cell-type map showing spatial organization of the cell types signatures in the gene expression vector field. Inset shows a zoom in of the highlighted tissue region. The colors of the cell types correspond to the cell-type legend in **Figure 2B**.

(D) SSAM improves the reconstruction of the ventricle. The upper left 2 panels show the DAPI and Poly-A signal around the ventricle area, showing tightly packed cells (occlusion) and lower signal in the ventricle structure compared to surrounding cells. The lower left 2 panels show the KDE gene expression signature for Foxj1 (the marker for ependymal cells) and Ttr (the marker for choroid plexus cells). The upper right 2 panels show the cell-type maps reconstructed by SSAM, showing a more complete reconstruction, and by Codeluppi *et al.*, which misses parts of the ventricle structure. The bottom right 2 panels show the reconstructions of only the ependymal (yellow) and choroid plexus (teal) cell types by SSAM and Codeluppi *et al.*

(E) SSAM has increased sensitivity of astrocyte detection. The far left upper and lower panels show DAPI and Poly-A signal for a region in the tissue. The middle left upper and lower panels show the overlap of Mfge8 signal (a marker for one astrocyte) with DAPI and Poly-A signals, showing that Mfge8 signal corresponds with low Poly-A signal, but with

higher DAPI signal. The top right 2 panels show the cell-type signals for Mfge8 expressing astrocytes by SSAM and Codeluppi *et al.*, showing that SSAM detect much more astrocyte cell types. The bottom right 2 panels shows the overlay of Mfge8 signal with the cell-type calls by SSAM and Codeluppi *et al.*, showing the astrocyte signals detected by SSAM correspond well with Mfge8 signal.

See also **Figure S3, S4**.

Figure 3. SSAM identifies cortical layer tissue domains in the mouse SSp cortex.

(A) Tissue domain map generated by SSAM. Tissue domain signatures were identified from clustering local cell-type composition over sliding 100 μm circular windows, and projected back onto the cell-type map. The reconstruction shows the various cortical layers.

(B) Cell-type composition within each tissue domain. The plots show that each domain consists of 7-14% Astrocyte Mfge8 cell types, apart from the ventricle, which instead shows a majority of choroid plexus and ependymal cell types. The colors in the pie charts correspond to the cell-type legend in **Figure 2B.**

Figure 4. SSAM confirms rich diversity of inhibitory and excitatory neuron cell types and localization in the posterior hypothalamic POA.

(A) Gene expression heatmap showing cell-type specific expression of marker genes. Rows show z-score normalized gene expression and columns show the gene expression patterns of filtered local maxima vectors (representative of gene expression within a cell). The bottom row of the top annotation shows the cell types. Due to a rich diversity of various inhibitory and excitatory neurons captured, the cell types were grouped into classes. The top row of the top annotation shows the cell classes which are named and colored based on the best cell-type signatures and cell classes from Moffitt *et al*. The colors of the cell classes top annotation correspond to the cell-type legend in **Figure 4B**.

(B) A tSNE map of cell-type signatures with distinct expression. Cell-type clusters are

visualized as a 2D t-SNE embedding of filtered local maxima vectors. Cell-type annotation and coloring are based on the best correlating segmentation-based cell-type signature from Moffitt *et al*. The tSNE map clearly shows the distinct cluster of different inhibitory and excitatory cell-type signatures. Cell types are grouped into classes based on groupings by Moffitt *et al.*

(C) The SSAM de novo cell-type map showing spatial organization of the cell types signatures in the gene expression vector field. Below left and right a zoom in of the highlighted tissue regions of the ventricle structure and clusters of oligodendrocyte cell types. The colors of the cell types correspond to the cell-type legend in **Figure 4B**.

(D) Spatial localization of various inhibitory cell-type signatures. We found a number of inhibitory cell types which both matched expression signature and tissue localization described by Moffitt *et al*. The cell-type clusters and names (and corresponding cell type from Moffitt *et al*) are: C39 Inhibitory Coch (I-12), C16 Inhib Arhgap36 (I-13), C45 Inhib Isr4 (I-15), C34 Inhib Calcr (I-14), and C14 Inhib Gda (I-23).

(E) As panel D, but for excitatory cell types. Shown are: C19 Excitatory Cbln1,Cbln2 (E-19), C42 Excitatory Omp (E-16), C25 Excitatory Necab1,Gda (E-9), C8 Excitatory Necab1 (E-14), and C36 Excitatory Col25a1 (E-24).

See also **Figure S5**.

Figure 5. SSAM identifies enriched inhibitory and excitatory tissue domains in the posterior hypothalamic POA.

(A) Tissue domain map generated by SSAM. Tissue domain signatures were identified from clustering local cell-type composition over sliding 100 µm circular windows, and projected back onto the cell-type map. The ventricle was manually removed from the tissue domain reconstruction. The reconstruction shows while the distribution of inhibitory and excitatory regions is intermingled, there are domains with enrichment of either of these cell types.

(B) Cell-type composition within each tissue domain. The plots shows the composition ratio

of approximately 5:1 of inhibitory to excitatory cell types in the inhibitory tissue domain and vice versa. The colors in the pie charts correspond to the cell-type legend in **Figure 4B.**

Figure 6. SSAM identifies a new cell type in L4 and confirms rare Sst Chodl cell type in the mouse VISp region.

(A) Gene expression heatmap showing cell-type specific expression of marker genes. Rows show z-score normalized gene expression and columns show the gene expression patterns of filtered local maxima vectors (representative of gene expression within a cell). The top annotation shows the cell types and coloring based on the highest correlating single cell RNA-seq based cell-type signature from previous result (Tasic et al., 2018). The colors of the top annotation correspond to the cell-type legend in **Figure 6B**.

(B) A tSNE map of cell-type signatures with distinct expression. Cell-type clusters are visualized as a 2D t-SNE embedding of filtered local maxima vectors. Cell-type annotation and coloring are based on the best correlating segmentation-based cell-type signature from previous result (Tasic et al., 2018).

(C) The SSAM de novo cell-type map showing spatial organization of the cell types signatures in the gene expression vector field. Lower images zoom in on the highlighted tissue regions of the new cell type found in the L4 superficial region (boxed in white), and rare Sst Chodl cell type. The colors of the cell types correspond to the cell-type legend in **Figure 6B**.

See also **Figure S6**.

Figure 7. Rare Sst Chodl cell type localizes to the L5b cortical layer of the mouse VISp region.

(A) Tissue domain map generated by SSAM. Tissue domain signatures were identified from clustering local cell-type composition over sliding 100 µm circular windows, and projected back onto the cell-type map. The reconstruction shows the various cortical layers within the adult mouse VISp, with very clear separation of the pia layer, and separation of layer 5 into 2

layers, 5a and 5b. Inset zooms into the location of the rare Sst Chodl cell type found in layer 5b.

(B) Cell-type composition within each tissue domain. The colors in the pie charts correspond to the cell-type legend in **Figure 6B**.

See also **Figure S7**.

Materials and Methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Roland Eils (roland.eils@charite.de).

METHOD DETAILS

Using Kernel Density Estimation to generate the gene expression vector field

We used the n-dimensional KDE algorithm to estimate the density of mRNAs in 2D and 3D. To compute Gaussian KDE, we used our own implementation of the KDE algorithm for rapid computation. Spatial distribution of the probability of mRNA presence is estimated using the kernel density estimation;

$$\hat{p}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \kappa_h(\mathbf{x} - \mathbf{x_i})$$

where κ_h is a kernel function with a fixed window size h. Here we use the Gaussian kernel:

$$\kappa_h(\mathbf{x}) = \frac{1}{(2\pi h^2)^{d/2}}\, e^{-\|\mathbf{x}\|/2h^2}$$

Note that each data point $\mathbf{x}$ lies within the respective image, hence the dimension d is either two or three. Ideally, the probability density at each lattice point must be evaluated by integration over the unit area. The lattice size is considered sufficiently fine-grained to capture all relevant information of the continuous Gaussian curve. To create a proper probability density, the lattice points are scaled to a sum of 1. Finally, the gene expression is estimated by multiplying each density by its total number of mRNA molecules.

Filtering of local maxima vectors

The full set of local maxima may contain spurious signals originating from the extracellular space due to small debris, image artifacts, and background noise. To eliminate such deficiencies, SSAM uses minimum expression threshold defined by the position of observable drop in the histogram of gene expressions (**Figure S3A, S5A, S6A**) as an initial selection criterion. After that, the local maxima is filtered once more with a minimum total gene expression threshold further reducing spurious local maxima (**Figure S3B, S5B, S6B**). Furthermore, we implemented an optional “input mask” feature to limit selection of local maxima to regions of the image containing informative data, e.g. a mask outlining the informative tissue area.

Normalization of local maxima vectors and the vector field

Since the gene expression profiles of local maxima vectors are representative of the transcriptomes of cells, we considered them to be analogous to the gene expression count matrix obtained from single cell RNA sequencing (scRNA-seq) using unique molecular identifiers (UMI). Therefore, we normalized the local maxima vectors of the vector field (which would be representative of single cells) using sctransform (Hafemeister and Satija, 2019), a normalization and regularization algorithm for UMI count data. After that, each vector of the vector field is normalized using sctransform, with the same parameters previously used to normalize the local maxima.

Clustering of representative gene expression vectors

The clustering algorithm implemented in SSAM is based on the source code of the R package Seurat (Butler et al., 2018). Here, we used the same algorithm reimplemented in Python. In short, an SNN network with correlation metric is built using a python package NetworkX (Hagberg et al., 2008). The weight of the network is calculated by a Jaccard similarity coefficient. A weight smaller than 1/15 was set to zero. Clustering was done by detecting communities in the network using a Louvain community detection algorithm

implemented in Python (python-louvain, https://python-louvain.readthedocs.io/). The resolution of the Louvain algorithm is set to 0.15.

SSAM diagnostic plots

To provide support to the user on whether to merge or remove clusters, SSAM generates a cluster-wise 'diagnostic plot', which consists of the following four panels: 1) location of the vectors originating from the cluster, 2) a map of the centroid embedded into the vector field, 3) the centroid of the cluster, 4) the location of the cluster in t-SNE or UMAP embedding. In the three applications in this paper, the clusters to be merged or removed often showed a mismatch between the location of vectors (panel 1) and the map of the centroid (panel 2). For sub cell types, the map typically does not clearly show the full shape of the cells but only fragments, but simultaneously having clear marker gene expression (panel 3). This usually indicates that there is another centroid that has higher correlation to the expression profile of the entire cell body. In such cases, such centroids are merged to the centroid with higher correlation. For dubious clusters, it is observed that vectors are usually located outside the tissue region or represent image artifacts (panel 1), the map clearly shows that the centroid is mapped to the artifacts (panel 2), or that the gene expression does not show any clear expression of marker genes (panel 3). Such clusters are removed thereafter. The remaining clusters are then identified by comparing cluster marker genes to known cell-type markers. Note that in many cases, the identity of clusters can be easily assigned by comparing the centroids of the clusters to the known cell-type signatures, e.g., from single cell RNA sequencing. Therefore, if such signatures are given, SSAM additionally shows the closest cell-type signature among the given signature in the diagnostic plot to help users easily assign classes to clusters. The diagnostic plots for osmFISH, MERFISH, and multiplexed smFISH data is available online in the Jupyter notebook uploaded to zenodo (http://doi.org/10.5281/zenodo.3478502).

SSAM analysis of osmFISH data

KDE was performed with a bandwidth of 2.5. The individual gene expression threshold and total gene expression threshold for selection of local maxima was 0.027 and 0.04, respectively (**Figure S3A, S3B**). Local maxima vectors were filtered once more with local k-nearest neighbor density with a threshold of 0.002 (**Figure S3C**). The selected local maxima vectors were passed to *sctransform* to determine normalization parameters, after which the whole vector field was normalized.

In SSAM guided mode, the mRNA count matrix of both the previously segmented cells and the scRNA-seq data were normalized by *sctransform*. The centroid of each of the annotated clusters was used to classify cell types in the vector field, generating a cell-type map guided by prior knowledge.

In SSAM *de novo* mode, initially the selected local maxima vectors were clustered using the Louvain algorithm with a resolution value of 0.15. 66 clusters were initially identified (**Figure S4A**). The sub-cell-type clusters were merged manually and spurious clusters were removed, resulting in a total of 30 clusters (**Figure 2A, 2B**). For each cluster, the vectors with insufficient correlation to its cluster medoid were excluded from the centroid calculation (**Figure S1B**). The cluster centroids were compared to that of the segmentation-based clustering result (**Figure S4B**) and scRNA-seq result (**Figure S4C**) using Pearson's correlation coefficient. The name of *de novo* clusters were determined based on the name of the highest correlated segmentation-based cluster for easy comparison of the gene expression signature. Note that clusters closest mapped to Inhibitory IC and Inhibitory CP cell types do not only appear in the internal capsule and caudoputamen, but also in the cortex. Therefore, we renamed these clusters to Inhibitory Kcnip2 (since Kcnip2 was the third most expressed gene for this cluster) and Inhibitory Rest, respectively.

For tissue domain analysis based on the *de novo* cell-type map, the radius of the circular

window was 100 μm. The cell-type proportions at each window were clustered using agglomerative clustering with 15 clusters as an initial estimate. Spatially connected clusters with a correlation coefficient higher than 0.6 are merged. The resulting domain map is resized to match the size of the cell-type map, after which the cells in different domains are colored.

Quantification of mRNA abundance in astrocytes and other brain cell types for osmFISH data interpretation

The "L5_All.loom" loom object containing scRNA-seq expression data of half a million cells from the mouse nervous system (Zeisel et al., 2018) was downloaded (http://mousebrain.org/downloads.html). Using Python, the total mRNA molecules per cell were extracted and aggregated by their level 2 class labels (astrocytes, immune, vascular, ependymal, neuronal, peripheral glia and oligodendrocyte cells). The total mRNA counts per class were log normalized and subsequently followed a normal distribution (tested using the Shapiro-Wilk test for normality, all *p-values* < 1 x 10e-4 for each class), therefore a Student's t-test was applicable. For each of the two classes of interest ('Astrocytes','Immune'), we performed independent log-space t-tests for unequal sample sizes and unequal variance against each of the other classes. Both astrocyte and immune cell classes have significantly lower mRNA molecule counts compared to other cell types (all *p-values* < 1 x 10e-12). While the distribution of mRNA counts in log space followed a normal distribution, the use of a Student's t-test for large numbers may be not appropriate. Hence, we also describe the difference in their distributions. For both astrocyte and immune cell classes, more than half of the cells of each classes exhibited a lower UMI count than the lowest quartile of any other cell class.

SSAM analysis of MERFISH data

KDE was performed with bandwidth 2.5. For local maxima selection the individual gene expression threshold was 0.0055, and total gene expression threshold was 0.0035 (**Figure**

S5A, S5B). The selected local maxima vectors were passed to *sctransform* to determine normalization parameters, after which the whole vector field was normalized.

When running SSAM in guided mode, both the mRNA count matrix for each previously segmented cell was obtained, as well as scRNA-seq data. Both signature sets were normalized using *sctransform*, and mapped onto the normalized vector field producing the guided cell-type maps.

For SSAM *de novo* mode, the selected vectors were clustered with resolution 0.15 of Louvain algorithm, resulting in 66 clusters (**Figure S5C**). By manual inspection, the sub cell-type clusters were merged, and spurious clusters were removed, resulting in a total of 50 clusters (**Figure 2A, 2B**). For each cluster, the vectors that did not have high correlation to its cluster medoid were excluded from the centroid calculation (**Figure S1B**). The centroids of the clusters are compared with that of the segmentation-based clustering result (**Figure S4B**) and scRNA-seq result (**Figure S4C**) using Pearson's correlation coefficient. The SSAM *de novo* clusters correlating best to inhibitory and excitatory neurons were named based on the most highly expressed genes of each cluster, and the other clusters were named based on the previous study (Moffitt et al., 2018).

Tissue domain analysis based on the cell-type map was performed with sliding spherical window with radius 100 μm. The cell-type proportions from each window are clustered using agglomerative hierarchical clustering with 20 clusters as an initial estimate, subsequently merging the clusters with correlation coefficient higher than 0.8. The resulting domain map was resized to match the size of the cell-type map, after which the cells in different domains were colored.

Comparison of localization of inhibitory and excitatory neurons

For a number of inhibitory and excitatory neuronal subtypes identified in the posterior POA

tissue image using SSAM *de novo* mode, we identified the best matching cell type based on Pearson correlation of their gene expression signatures (**Figure S5F**). We matched the following cell types: SSAM cluster 39 (C39) called Inhibitory Coch to Moffitt cluster I-12, C16 Inhibitory Arhgap36 to I-13, C45 Inhibitory Isr4 to I-15, C34 Inhibitory Calcr to I-14 , C14 Inhibitory Gda to I-23, C19 Excitatory Cbln1-Cbln2 to E-19, C42 Excitatory Omp to E-16, C25 Excitatory Necab1-Gda to E-9, C8 Excitatory Necab1 to E-14, and C36 Excitatory Col25a1 to E-24. For these cell types we checked the tissue localizations reported in the previous studies figures 5a, 5c, 5e, 6b, 6d, and S17 (Moffitt et al., 2018). Visually comparing localization of these neurons computed by SSAM reconstruction and those taken from the original publication revealed very similar patterns of localization (**Figure 4D,E**).

3D modelling of MERFISH cell-type maps

Firstly, the connected components in 3D were determined using a python package called 'connected-components-3d' (https://github.com/seung-lab/connected-components-3d). Components comprising fewer than 100 voxels were removed. After this, the voxels filling connected components were removed, and only the contours were used for the vertex of the 3D models. For each vertex the vertex normal was calculated by simple physics simulation, assuming that the direction of vertex normal vector is the same as the force vector when there are pulling forces between all of the contour voxels. The surface of the objects are reconstructed using screened Poisson reconstruction algorithm (Kazhdan and Hoppe, 2013; Kazhdan et al., 2006) using default parameters. The number of vertices was reduced to 5% of the total number of vertices using 'vtkQuadricDecimation' function (Garland and Heckbert, 1997; Hoppe, 1999) of VTK library (Schroeder et al., 2006). Finally the objects are merged into one file. Each scene of the rotating movie was created using Meshlab (Cignoni et al., 2008).

VISP multiplexed smFISH data generation

Multiplexed smFISH data of mouse primary visual cortex (VISp) was generated as part of

the SpaceTx consortium. Tissue processing was carried out as previously described (Hodge et al., 2019), with some modifications.

Silanization of coverslips (#1.5, Thorlabs CG15KH) was performed by plasma cleaning for 30 min in a Plasma-Prep III (SPI 11050-AB), followed by vapor deposition of 3-aminopropyltriethoxysilane (APES, Sigma A3648) in a vacuum for 10 minutes. Coverslips were then washed in 100% methanol for 2 x 5 minutes, allowed to dry, and stored in a dust-free environment until use.

Fresh-frozen mouse brain tissue was sectioned at 10 µm onto silanized coverslips, let dry for 20 min at -20°C, then fixed for 15 min at 4 °C in 4% PFA in PBS. Sections were washed 3 × 10 min in PBS, then permeabilized and dehydrated with chilled 100% methanol at -20°C for 10 min and allowed to dry. Sections were stored at −80 °C until use. Frozen sections were rehydrated in 2X SSC (Sigma 20XSSC, 15557036) for 5 min, then treated 10 min with 8% SDS (Sigma 724255) in PBS at room temperature. Sections were washed 5 times in 2X SSC. Sections were then incubated in hybridization buffer (10% Formamide (v/v, Sigma 4650), 10% dextran sulfate (w/v, Sigma D8906), 200 µg/mL BSA (ThermoFisher AM2616), 2 mM ribonucleoside vanadyl complex (New England Biolabs S1402S), 1 mg/ml tRNA (Sigma 10109541001) in 2X SSC) for 5 min at 37°C. Probes were diluted in hybridization buffer at a concentration of 250 nM and hybridized at 37°C for 2 h. Following hybridization, sections were washed 2 × 10 min at 37°C in wash buffer (2X SSC, 20% Formamide), and 1 × 10 min in wash buffer with 5 µg/ml DAPI (Sigma 32670), then washed 3 times with 2X SSC. Sections were then imaged in Imaging buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 0.8% glucose (Sigma G8270), 30 U/ml pyranose oxidase (Sigma P4234), 50 µg/ml catalase (Abcam ab219092). Following imaging, sections were incubated 3 × 10 min in stripping buffer (65% formamide, 2X SSC) at 30°C to remove hybridization probes from the first round. Sections were then washed in 2X SSC for 3 × 5 min at room temperature before repeating the hybridization procedure.

The multiplexed smFISH image data was collected and processed using methods previously described (Hodge et al., 2019), except that images from different rounds of hybridization were registered in (x,y) based on the DAPI signal. The spot locations and raw data are available on request.

SSAM analysis of VISp multiplexed smFISH data

KDE was performed with bandwidth 2.5 μm. Local maxima were filtered with gene expression threshold of 0.027, and then filtered with total gene expression threshold of 0.2 (**Figure S6A, S6B**). Initially 30 clusters were obtained using Louvain algorithm with a resolution value of 0.15. The rare cell type (Sst Chodl) was rescued, hence a total of 31 clusters are considered for further analysis. By manual inspection, two pairs of the sub-cell-type clusters are merged, and three spurious clusters were removed, resulting in 26 clusters. The centroids of the clusters are compared with that of scRNA-seq result using Pearson's correlation coefficient (**Figure S6E**). The name of clusters were determined based on the name of highest correlated clusters found in the scRNA-seq data, except the newly found 'L4 IT Superficial' cluster.

Tissue domains were defined with a sliding circular window with radius 100 μm, on a square periodic lattice with spacing 10 μm over the cell-type map. Agglomerative clustering of the compositions of cell types within the windows was initially performed with 20 clusters. Clusters with Pearson's correlation higher than 0.8 were merged to result in nine clusters. Further, two clusters are merged since these are different parts of the Pia layer, and one cluster is removed since the cluster is mapped outside of the tissue region, resulting in a final set of seven clusters representing tissue domains (**Figure 4E**).

Plotting

The python packages Matplotlib 3.1.0 (Caswell et al., 2019) and Seaborn 0.9.0 (Waskom et

al., 2018) were used to draw 2D images, plots, and heatmaps. In SSAM, helper functions are included to easily generate plots.

Movies

Movies were generated by using Virtualdub (1.10.4-AMD64, http://www.virtualdub.org/). The H.264 codec was used to compress videos.

Software

Python version 3.7.0 was used throughout. The following python packages were used: *numpy, scipy, pandas, matplotlib, seaborn, scikit-learn, umap-learn, python-louvain, sparse, scikit-image*. R package *sctransform* was used for normalization and variance stabilization of the data.

DATA AND CODE AVAILABILITY

The source code of SSAM is available online at: https://github.com/eilslabs/ssam. A Jupyter notebook (https://github.com/eilslabs/ssam_example) outlines the commands used to download and pre-process the data, and to reproduce the results and figures of this study. The Jupyter notebooks also contain the extensive diagnostic plots used for parameter selection, and choice of removal or merging of clusters. All large files are available online from zenodo: http://doi.org/10.5281/zenodo.3478502.

The osmFISH data (Codeluppi et al., 2018) used within the study is available from http://linnarssonlab.org/osmFISH/availability/. The single cell RNA sequencing data of mouse somatosensory cortex (Marques et al., 2016; Zeisel et al., 2015) are available from http://loom.linnarssonlab.org/. The single cell RNA sequencing data (Zeisel et al., 2018) to compare total mRNA molecules between cell types is available from http://mousebrain.org/. The high resolution poly-A and DAPI images of osmFISH data (Codeluppi et al., 2018) were kindly provided by Sten Linnarsson. The MERFISH data (Moffitt et al., 2018) is available

from https://datadryad.org/handle/10255/dryad.192644. Mouse VISp multiplexed smFISH data is available from Zenodo: http://doi.org/10.5281/zenodo.3478502.

Supplemental Videos

Supplemental Video 1. MERFISH 3D cell-type map, turntable rotating

Supplemental Video 2. MERFISH 3D cell-type map, sweeping along z axis by 1 μm

Supplemental Video 3. MERFISH neuronal cells, sweeping along z axis by 1 μm

Supplemental Video 4. MERFISH astrocytes, sweeping along z axis by 1 μm

References

Angerer, P., Haghverdi, L., Büttner, M., Theis, F.J., Marr, C., and Buettner, F. (2016). destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics *32*, 1241–1243.

Battich, N., Stoeger, T., and Pelkmans, L. (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution. Nat. Methods *10*, 1127–1133.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Caswell, T.A., Droettboom, M., Hunter, J., Firing, E., Lee, A., Klymak, J., Stansby, D., de Andrade, E.S., Nielsen, J.H., Varoquaux, N., et al. (2019). matplotlib/matplotlib v3.1.0. Zenodo, http://dx.doi.org/10.5281/zenodo.2893252.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., and Ranzuglia, G. (2008). Meshlab: an open-source mesh processing tool. Eurographics Italian Chapter Conference *2008*, 129–136.

Cleary, B., Murray, E., Alam, S., Sinha, A., Habibi, E., Simonton, B., Bezney, J., Marshall, J., Lander, E.S., Chen, F., et al. (2019). Compressed sensing for imaging transcriptomics. bioRxiv, http://dx.doi.org/10.1101/743039.

Codeluppi, S., Borm, L.E., Zeisel, A., and La Manno, G. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. Nature Methods *15*, 932–935.

Dries, R., Zhu, Q., Eng, C.-H.L., Sarkar, A., Bao, F., George, R.E., Pierson, N., Cai, L., and Yuan, G.-C. (2019). Giotto, a pipeline for integrative analysis and visualization of single-cell spatial transcriptomic data. bioRxiv, http://dx.doi.org/10.1101/701680.

Eng, C.-H.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature *568*, 235–239.

Flynn, R.A., Smith, B.A.H., Johnson, A.G., Pedram, K., George, B.M., Malaker, S.A., Majzoub, K., Carette, J.E., and Bertozzi, C.R. (2019). Mammalian Y RNAs are modified at discrete guanosine residues with N-glycans. bioRxiv, http://dx.doi.org/10.1101/787614.

Garland, M., and Heckbert, P.S. (1997). Surface simplification using quadric error metrics. Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques 209–216.

Gerashchenko, D., Wisor, J.P., Burns, D., Reh, R.K., Shiromani, P.J., Sakurai, T., de la Iglesia, H.O., and Kilduff, T.S. (2008). Identification of a population of sleep-active cerebral cortex neurons. Proc. Natl. Acad. Sci. U. S. A. *105*, 10227–10232.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008) 11–15.

Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion

pseudotime robustly reconstructs lineage branching. Nat. Methods *13*, 845–848.

Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature *542*, 352–356.

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61–68.

Hodneland, E., Kögel, T., Frei, D.M., Gerdes, H.-H., and Lundervold, A. (2013). CellSegm - a MATLAB toolbox for high-throughput 3D cell segmentation. Source Code Biol. Med. *8*, 16.

Hoppe, H. (1999). New quadric metric for simplifying meshes with appearance attributes. Proceedings Visualization '99 (Cat. No.99CB37067) 59–510.

Jiang, J., Kao, P.-Y., Belteton, S.A., Szymanski, D.B., and Manjunath, B.S. (2019). Accurate 3D Cell Segmentation using Deep Feature and CRF Refinement. arXiv [cs.CV],.

Kazhdan, M., and Hoppe, H. (2013). Screened poisson surface reconstruction. ACM Trans. Graph. *32*, 29.

Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson surface reconstruction. Proceedings of the Fourth Eurographics Symposium on Geometry Processing 61–70.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods *10*, 857–860.

Kishi, J.Y., Beliveau, B.J., Lapan, S.W., West, E.R., Zhu, A., Sasaki, H.M., Saka, S.K., Wang, Y., Cepko, C.L., and Yin, P. (2019). SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. Nat. Methods *16*, 533–544.

Kong, J., Wang, F., Teodoro, G., Liang, Y., Zhu, Y., Tucker-Burden, C., and Brat, D.J. (2015). Automated cell segmentation with 3D fluorescence microscopy images. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) 1212–1215.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Ferrante, T.C., Terry, R., Turczyk, B.M., Yang, J.L., Lee, H.S., Aach, J., et al. (2015). Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat. Protoc. *10*, 442–458.

Lignell, A., Kerosuo, L., Streichan, S.J., Cai, L., and Bronner, M.E. (2017). Identification of a neural crest stem cell niche by Spatial Genomic Analysis. Nat. Commun. *8*, 1830.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods *11*, 360–361.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. *15*, e8746.

Maniatis, S., Äijö, T., Vickovic, S., Braine, C., Kang, K., Mollbrink, A., Fagegaltier, D., Andrusivová, Ž., Saarenpää, S., Saiz-Castro, G., et al. (2019). Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. Science *364*, 89–93.

Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha Falcão, A., Xiao, L., Li, H., Häring, M., Hochgerner, H., Romanov, R.A., et al. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science *352*, 1326–1329.

Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science *362*.

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. Ann. Math. Stat. *33*, 1065–1076.

Perkel, J.M. (2019). Starfish enterprise: finding RNA patterns in single cells. Nature *572*, 549–551.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). Science Forum: The Human Cell Atlas. Elife *6*.

Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463–1467.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. Ann. Math. Stat. *27*, 832–837.

Salmén, F., Vickovic, S., Larsson, L., Stenbeck, L., Vallon-Christersson, J., Ehinger, A., Häkkinen, J., Borg, Å., Frisén, J., Ståhl, P.L., et al. (2018). Multidimensional transcriptomics provides detailed information about immune cell distribution and identity in HER2+ breast tumors. bioRxiv, http://dx.doi.org/10.1101/358937.

Salvi, M., Morbiducci, U., Amadeo, F., Santoro, R., Angelini, F., Chimenti, I., Massai, D., Messina, E., Giacomello, A., Pesce, M., et al. (2019). Automated Segmentation of Fluorescence Microscopy Images for 3D Cell Detection in human-derived Cardiospheres. Sci. Rep. *9*, 6644.

Schroeder, W., Martin, K., and Lorensen, B. (2006). The Visualization Toolkit: An Object-oriented Approach to 3D Graphics. Kitware.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron *92*, 342–357.

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H.L., Koulena, N., Cronin, C., Karp, C., Liaw, E.J., et al. (2018). Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell *174*, 363–376.e16.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018a). Exponential scaling of single-cell RNA-seq in the past decade. Nat. Protoc. *13*, 599–604.

Svensson, V., Teichmann, S.A., and Stegle, O. (2018b). SpatialDE: identification of spatially variable genes. Nat. Methods *15*, 343–346.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346.

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature *563*, 72–78.

Thomas, R.M., and John, J. (2017). A review on cell detection and segmentation in microscopic images. 1–5.

Tomioka, R., Okamoto, K., Furuta, T., Fujiyama, F., Iwasato, T., Yanagawa, Y., Obata, K., Kaneko, T., and Tamamaki, N. (2005). Demonstration of long-range GABAergic connections distributed throughout the mouse neocortex. Eur. J. Neurosci. *21*, 1587–1600.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J.F., et al. (2019a). High-definition spatial transcriptomics for in situ tissue profiling. Nat. Methods *16*, 987–990.

Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J.F., Gould, J., et al. (2019b). High-density spatial transcriptomics arrays for in situ tissue profiling. bioRxiv, http://dx.doi.org/10.1101/563338.

Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science *361*.

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., Gemperline, D.C., Augspurger, T., Halchenko, Y., Cole, J.B., et al. (2018). mwaskom/seaborn: v0.9.0 (July 2018). Zenodo, http://dx.doi.org/10.5281/zenodo.1313201.

Xia, C., Fan, J., Emanuel, G., Hao, J., and Zhuang, X. (2019). Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc. Natl. Acad. Sci. U. S. A.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999–1014.e22.

Figure 1.

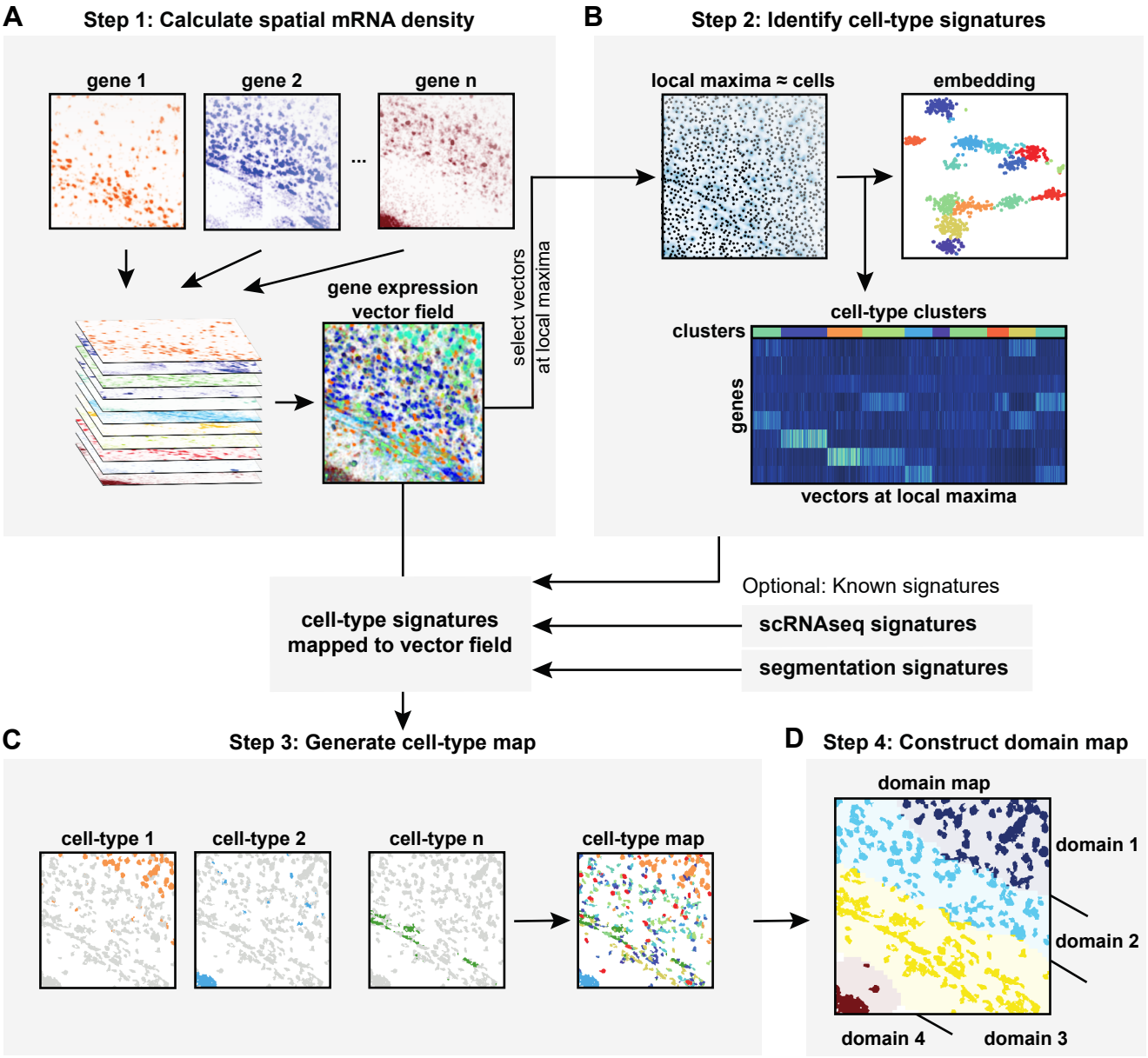

A
Step 1: Calculate spatial mRNA density
gene 1
gene 2
gene n
...
gene expression vector field
select vectors at local maxima
B
Step 2: Identify cell-type signatures
local maxima ≈ cells
embedding
cell-type clusters
clusters
genes
vectors at local maxima
Optional: Known signatures
cell-type signatures mapped to vector field
scRNAseq signatures
segmentation signatures
C
Step 3: Generate cell-type map
cell-type 1
cell-type 2
cell-type n
cell-type map
D
Step 4: Construct domain map
domain map
domain 1
domain 2
domain 4
domain 3

Figure 2.

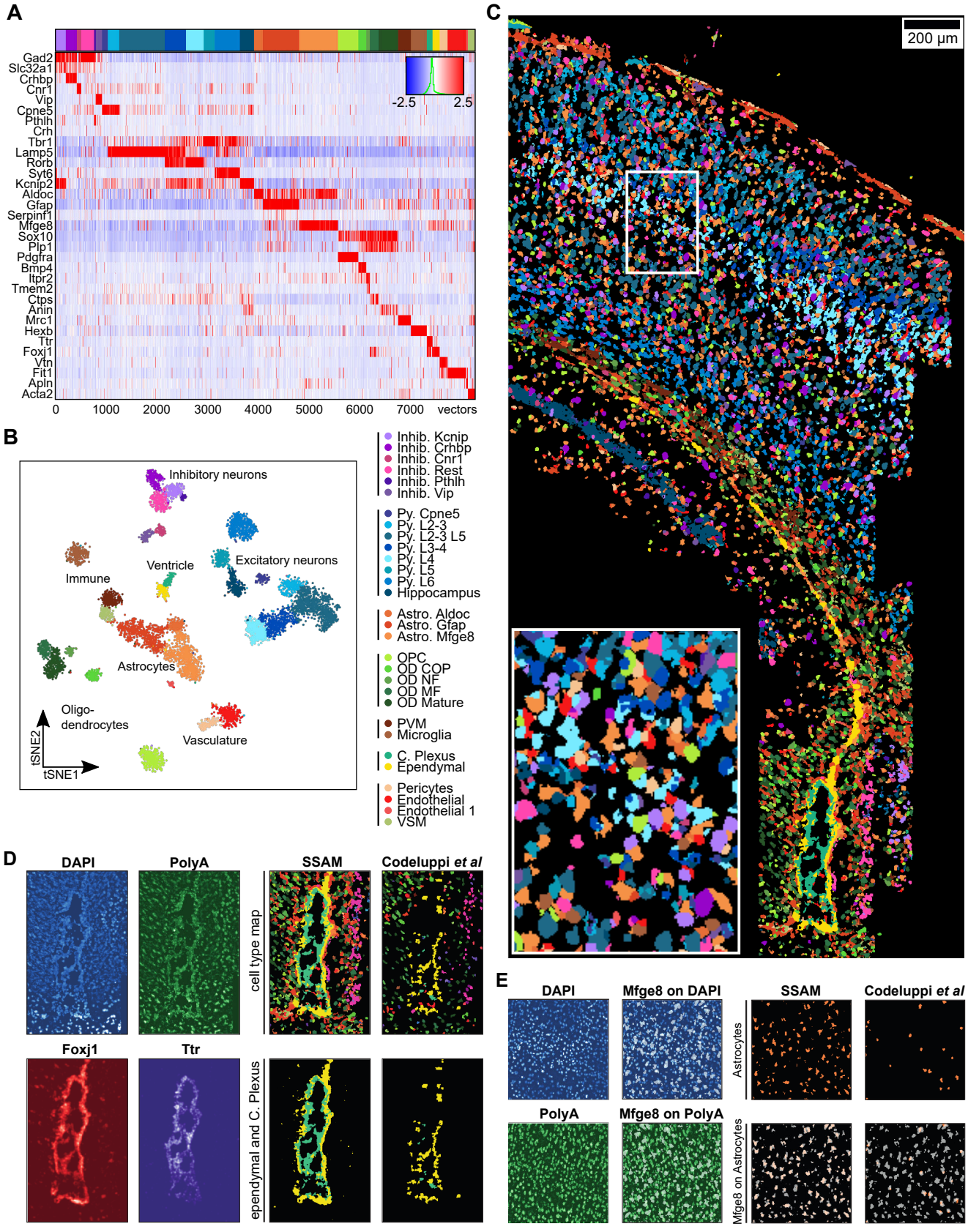

A
Gad2
Slc32a1
Crhbp
Cnr1
Vip
Cpne5
Pthlh
Crh
Tbr1
Lamp5
Rorb
Syt6
Kcnip2
Aldoc
Gfap
Serpinf1
Mfge8
Sox10
Plp1
Pdgfra
Bmp4
Itpr2
Tmem2
Ctps
Anln
Mrc1
Hexb
Ttr
Foxj1
Vtn
Fit1
Apln
Acta2
-2.5
2.5
0
1000
2000
3000
4000
5000
6000
7000
vectors
B
Inhibitory neurons
Excitatory neurons
Ventricle
Immune
Astrocytes
Oligo-
dendrocytes
Vasculature
tSNE2
tSNE1
Inhib. Kcnip
Inhib. Crhbp
Inhib. Cnr1
Inhib. Rest
Inhib. Pthlh
Inhib. Vip
Py. Cpne5
Py. L2-3
Py. L2-3 L5
Py. L3-4
Py. L4
Py. L5
Py. L6
Hippocampus
Astro. Aldoc
Astro. Gfap
Astro. Mfge8
OPC
OD COP
OD NF
OD MF
OD Mature
PVM
Microglia
C. Plexus
Ependymal
Pericytes
Endothelial
Endothelial 1
VSM
C
200 µm
D
DAPI
PolyA
SSAM
Codeluppi et al
cell type map
Foxj1
Ttr
ependymal and C. Plexus
E
DAPI
Mfge8 on DAPI
SSAM
Codeluppi et al
Astrocytes
PolyA
Mfge8 on PolyA
Mfge8 on Astrocytes

Figure 3.

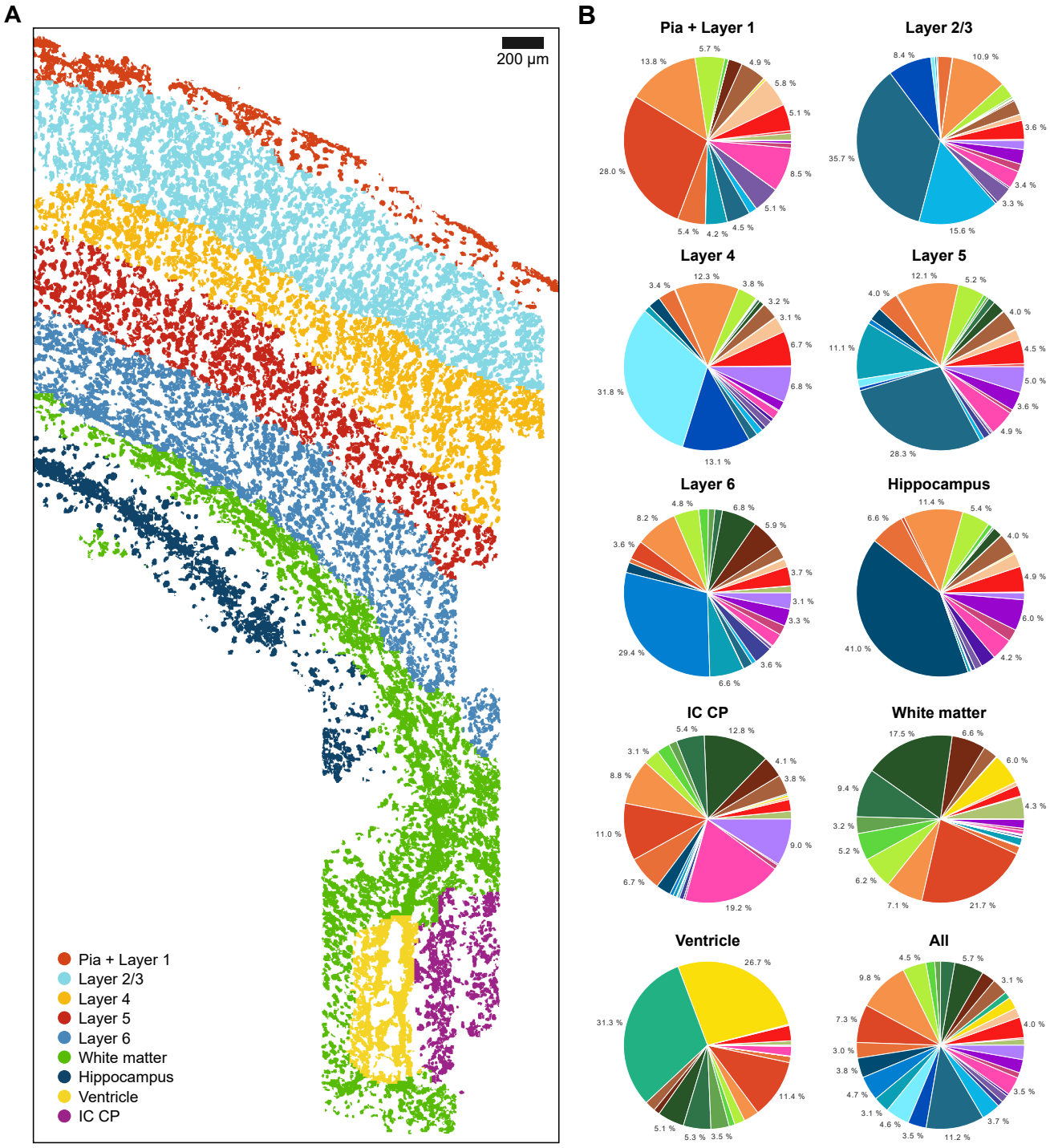

A
200 µm
Pia + Layer 1
Layer 2/3
Layer 4
Layer 5
Layer 6
White matter
Hippocampus
Ventricle
IC CP
B
Pia + Layer 1
Layer 2/3
Layer 4
Layer 5
Layer 6
Hippocampus
IC CP
White matter
Ventricle
All

Figure 4.

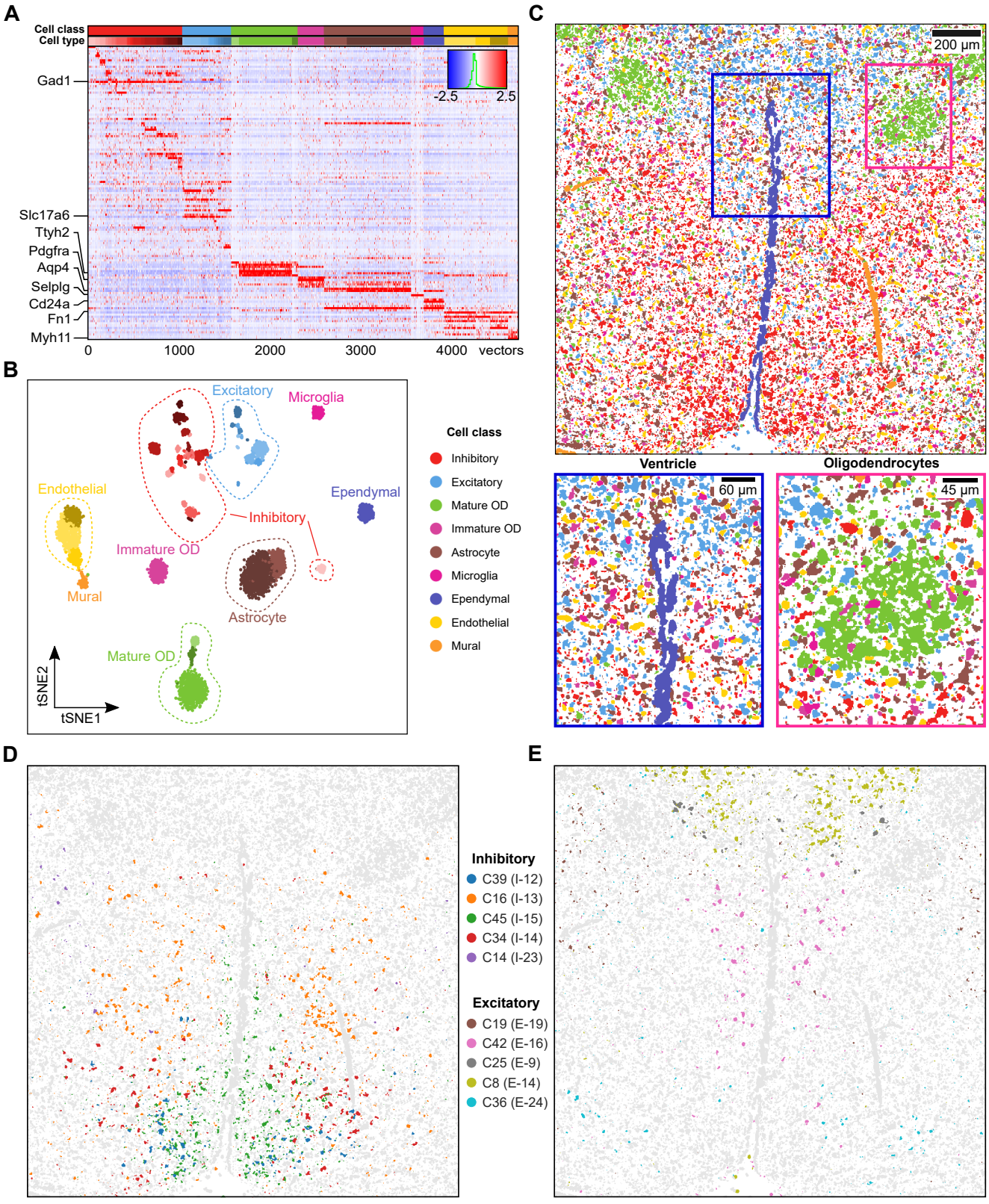

A
Cell class
Cell type
Gad1
Slc17a6
Ttyh2
Pdgfra
Aqp4
Selplg
Cd24a
Fn1
Myh11
-2.5
2.5
0
1000
2000
3000
4000
vectors
B
Excitatory
Microglia
Endothelial
Ependymal
Inhibitory
Immature OD
Mural
Astrocyte
Mature OD
tSNE2
tSNE1
Cell class
Inhibitory
Excitatory
Mature OD
Immature OD
Astrocyte
Microglia
Ependymal
Endothelial
Mural
C
200 μm
Ventricle
Oligodendrocytes
60 μm
45 μm
D
Inhibitory
C39 (I-12)
C16 (I-13)
C45 (I-15)
C34 (I-14)
C14 (I-23)
Excitatory
C19 (E-19)
C42 (E-16)
C25 (E-9)
C8 (E-14)
C36 (E-24)
E

Figure 5.

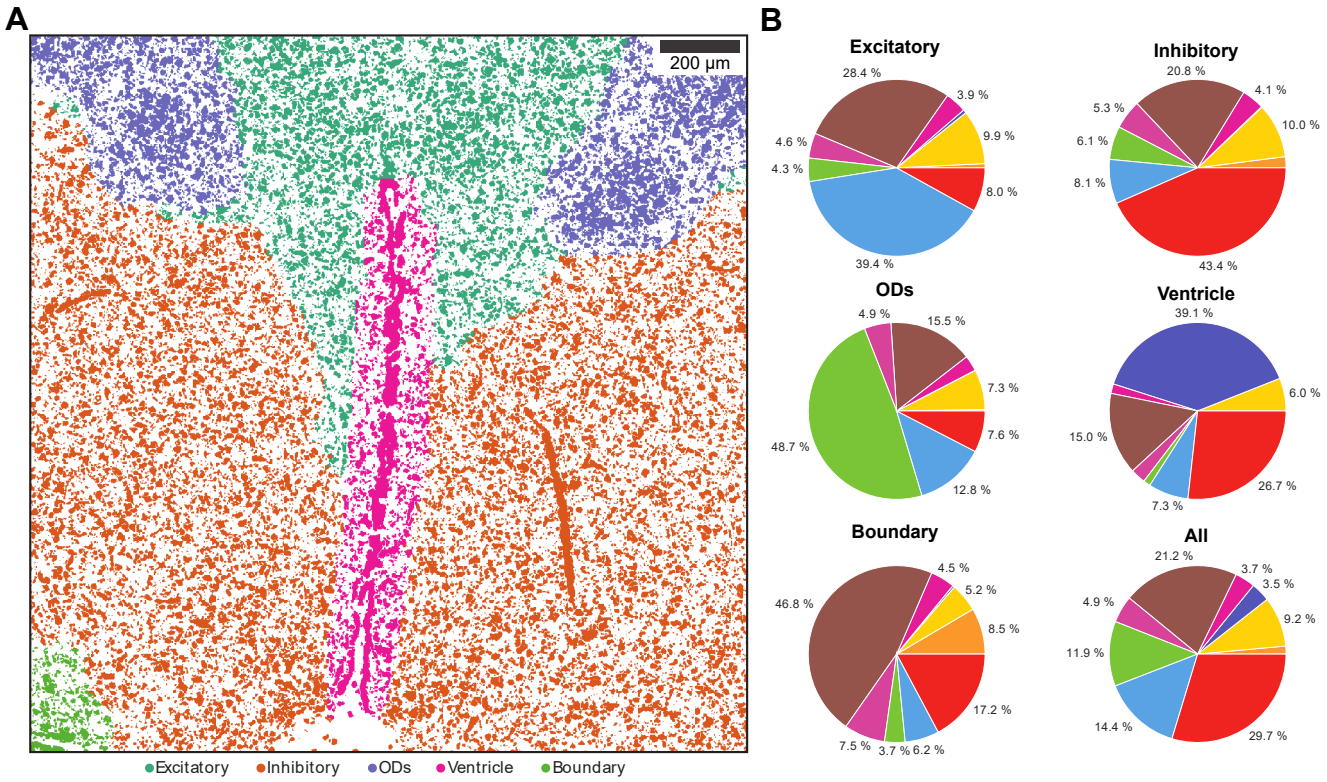

A
200 μm
Excitatory
Inhibitory
ODs
Ventricle
Boundary
B
Excitatory
28.4 %
3.9 %
9.9 %
8.0 %
39.4 %
4.3 %
4.6 %
Inhibitory
20.8 %
4.1 %
10.0 %
43.4 %
8.1 %
6.1 %
5.3 %
ODs
4.9 %
15.5 %
7.3 %
7.6 %
12.8 %
48.7 %
Ventricle
39.1 %
6.0 %
26.7 %
7.3 %
15.0 %
Boundary
4.5 %
5.2 %
8.5 %
17.2 %
6.2 %
3.7 %
7.5 %
46.8 %
All
21.2 %
3.7 %
3.5 %
9.2 %
29.7 %
14.4 %
11.9 %
4.9 %

Figure 6.

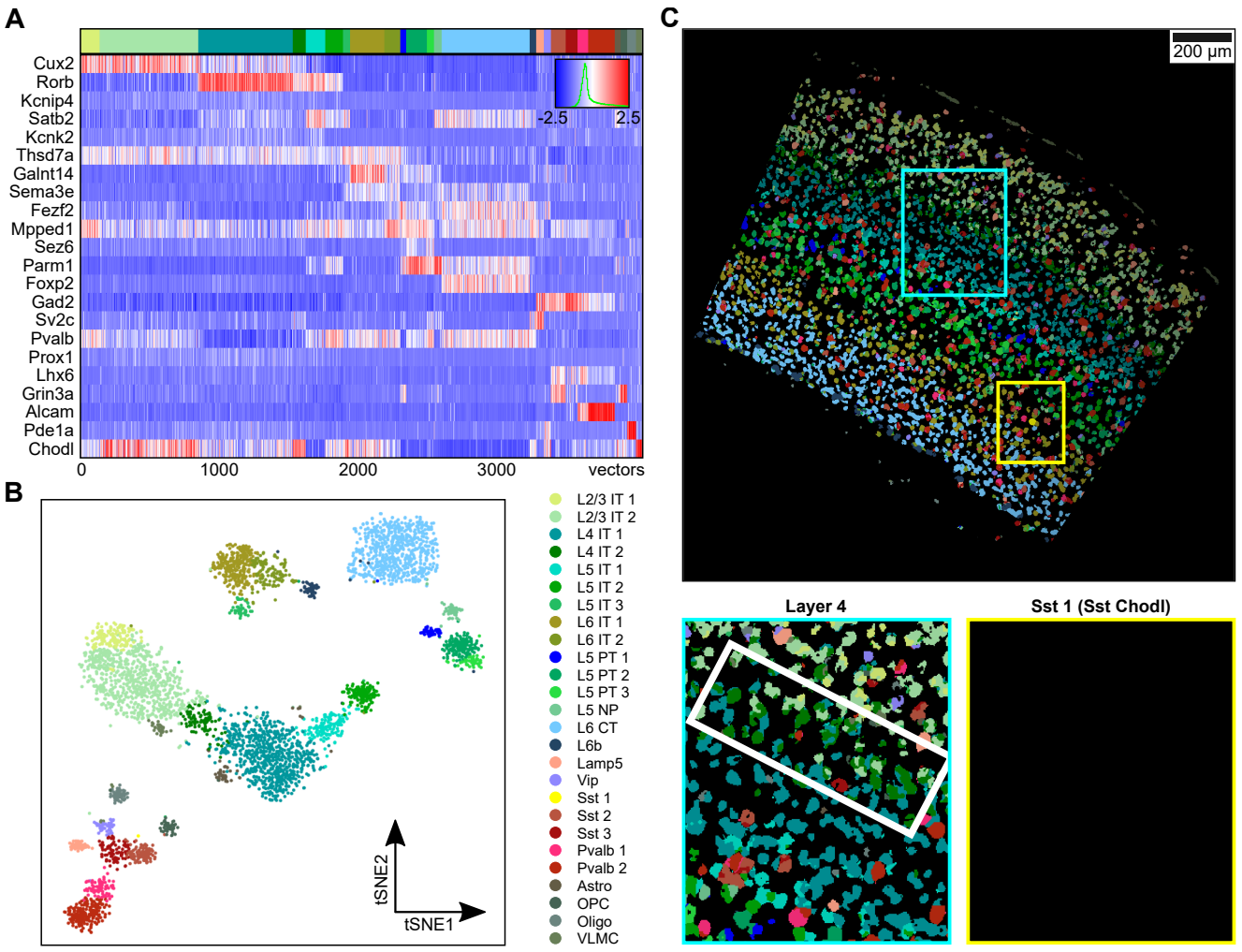
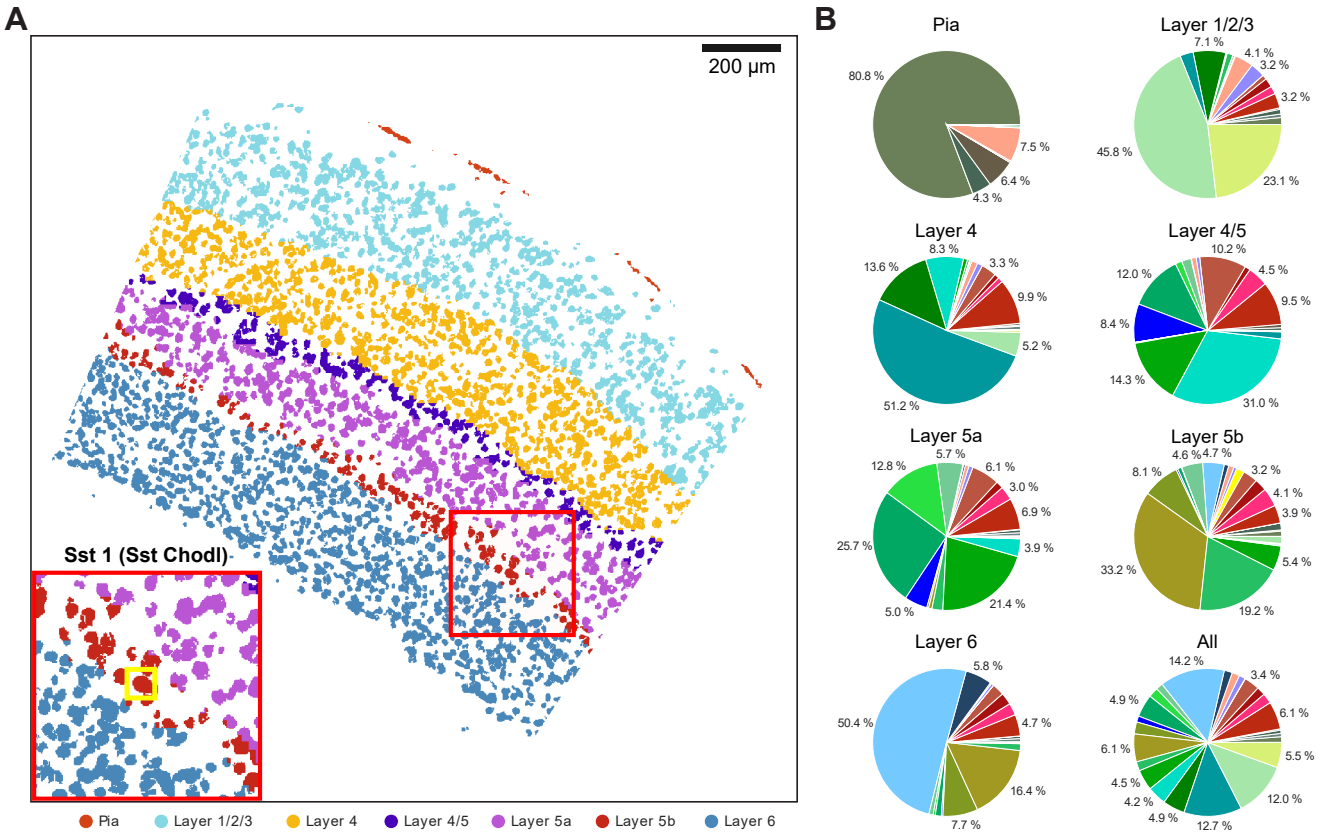

A
Cux2
Rorb
Kcnip4
Satb2
Kcnk2
Thsd7a
Galnt14
Sema3e
Fezf2
Mpped1
Sez6
Parm1
Foxp2
Gad2
Sv2c
Pvalb
Prox1
Lhx6
Grin3a
Alcam
Pde1a
Chodl
-2.5
2.5
0
1000
2000
3000
vectors
B
tSNE2
tSNE1
L2/3 IT 1
L2/3 IT 2
L4 IT 1
L4 IT 2
L5 IT 1
L5 IT 2
L5 IT 3
L6 IT 1
L6 IT 2
L5 PT 1
L5 PT 2
L5 PT 3
L5 NP
L6 CT
L6b
Lamp5
Vip
Sst 1
Sst 2
Sst 3
Pvalb 1
Pvalb 2
Astro
OPC
Oligo
VLMC
C
200 µm
Layer 4
Sst 1 (Sst Chodl)

Figure 7.


A
200 µm
Sst 1 (Sst Chodl)
Pia
Layer 1/2/3
Layer 4
Layer 4/5
Layer 5a
Layer 5b
Layer 6
B
Pia
80.8 %
7.5 %
6.4 %
4.3 %
Layer 1/2/3
7.1 %
4.1 %
3.2 %
3.2 %
45.8 %
23.1 %
Layer 4
8.3 %
13.6 %
3.3 %
9.9 %
5.2 %
51.2 %
Layer 4/5
10.2 %
12.0 %
4.5 %
9.5 %
8.4 %
14.3 %
31.0 %
Layer 5a
5.7 %
12.8 %
6.1 %
3.0 %
6.9 %
25.7 %
3.9 %
5.0 %
21.4 %
Layer 5b
4.6 %
4.7 %
8.1 %
3.2 %
4.1 %
3.9 %
33.2 %
5.4 %
19.2 %
Layer 6
5.8 %
50.4 %
4.7 %
16.4 %
7.7 %
All
14.2 %
3.4 %
4.9 %
6.1 %
6.1 %
5.5 %
4.5 %
4.2 %
4.9 %
12.0 %
12.7 %