



## Spatial domain detection using contrastive self-supervised learning for spatial multi-omics technologies

Jianing Yao, Jinglun Yu, Brian Caffo, et al.

*Genome Res.* published online May 20, 2025

Access the most recent version at doi:[10.1101/gr.279380.124](https://doi.org/10.1101/gr.279380.124)

---

**P<P** Published online May 20, 2025 in advance of the print journal.

**Accepted Manuscript** Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

**Open Access** Freely available online through the *Genome Research* Open Access option.

**Creative Commons License** This manuscript is Open Access. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International license), as described at <http://creativecommons.org/licenses/by/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Published by Cold Spring Harbor Laboratory Press

# Spatial domain detection using contrastive self-supervised learning for spatial multi-omics technologies

Jianing Yao<sup>1</sup>, Jinglun Yu<sup>2</sup>, Brian Caffo<sup>1</sup>, Stephanie C. Page<sup>3</sup>, Keri Martinowich<sup>3,4,5</sup>,  
Stephanie C. Hicks<sup>1,6,7,8\*</sup>

<sup>1</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, MD, USA

<sup>2</sup>Department of Electrical and Computer Engineering, Johns Hopkins University, MD, USA

<sup>3</sup>Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA

<sup>4</sup>The Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA

<sup>5</sup>Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

<sup>6</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

<sup>7</sup>Center for Computational Biology, Johns Hopkins University, Baltimore, MD, USA

<sup>8</sup>Malone Center for Engineering in Healthcare, Johns Hopkins University, MD, USA

\*Corresponding author: [shicks19@jhu.edu](mailto:shicks19@jhu.edu)

## Running title:

Spatial clustering with contrastive autoencoders

## Abstract

Recent advances in spatially-resolved single-omics and multi-omics technologies have led to the emergence of computational tools to detect or predict spatial domains. Additionally, histological images and immunofluorescence (IF) staining of proteins and cell types provide multiple perspectives and a more complete understanding of tissue architecture. Here, we introduce Proust, a scalable tool to predict discrete domains using spatial multi-omics data by combining the low-dimensional representation of biological profiles based on graph-based contrastive self-supervised learning. Our scalable method integrates multiple data modalities, such as RNA, protein, and H&E images, and predicts spatial domains within tissue samples. Through the integration of multiple modalities, Proust consistently demonstrates enhanced accuracy in detecting spatial domains, as evidenced across various benchmark datasets and technological platforms.

## 26 **Keywords**

27 Spatially-resolved transcriptomics; immunofluorescence images; contrastive learning; graph-based  
28 autoencoder; multi-modal learning; spatial clustering.

## 29 **Introduction**

30 Spatially-resolved multi-omics technologies enable the profiling of multiple omic measurements, such as the  
31 transcriptome and proteome, in individual tissue sections, leading to an improved understanding of regulatory  
32 mechanisms along spatial coordinates (Liu et al. 2020; Nature 2021; Zhang et al. 2023; Li 2023). These spatial  
33 technologies have already revolutionized our understanding of human tissue architecture and the impact on  
34 tissue architecture from disease (Asp et al. 2019; Maniatis et al. 2019; Chen et al. 2020; Ji et al. 2020).  
35 Examples of these types of multi-omics technologies include measuring RNA and protein (DBiT-seq (Liu et  
36 al. 2020) or the 10x Genomics Visium Spatial Proteogenomics (SPG)  
37 (<https://www.10xgenomics.com/products/spatial-gene-and-protein-expression>) platform). In addition, the  
38 standard 10x Genomics Visium platform (<https://www.10xgenomics.com/products/spatial-gene-expression>)  
39 could also be viewed as a multi-omics technology through the integration of transcriptome data with a paired  
40 brightfield image after the tissues are stained with hematoxylin and eosin (H&E).

41 A standard step in the analysis of spatial multi-omics data is to identify discrete spatial domains. These  
42 domains can be further investigated for potential markers of tissue architecture corresponding to morphology  
43 (Maynard et al. 2021) or unique niche-specific domains that might appear in complex diseases, such as cancer  
44 (Denisenko et al. 2024). However, similar to single-cell data, it remains challenging to leverage supervised  
45 learning approaches to predict discrete spatial domains because of cell segmentation. Therefore, most existing  
46 tools used in practice today identify discrete spatial domains with either unsupervised or self-supervised  
47 learning approaches (Li et al. 2022b; Hu et al. 2021b; Zeng et al. 2022; Long et al. 2023).

48 To identify discrete spatial domains from spatial multi-omics data, one approach is to ignore the  
49 spatial information entirely, consider only one of the omic data modalities (e.g. RNA), and apply unsupervised  
50 clustering methods used for single-cell, such as *k*-means (Lloyd 1982), Louvain, and Leiden (Traag et al.

2019) algorithms. However, these approaches assume independence between the spatial coordinates and often lead to discontinuous or incoherent spatial domains (Maynard et al. 2021). A second approach is to continue with only one omic data modality, but to incorporate spatial information to account for the correlation of molecular information between the spatial coordinates. Some examples of these methods include (i) unsupervised learning approaches (BayesSpace (Zhao et al. 2021), Giotto (Dries et al. 2021), STAGATE (Dong and Zhang 2022), CCST (Li et al. 2022a)) and (ii) self-supervised learning approaches (GraphST (Long et al. 2023), SpaceFlow (Ren et al. 2022), ConGI (Zeng et al. 2023), CAST (Tang et al. 2024)). In particular, the methods using contrastive self-supervised learning aim to maximize the similarity between adjacent spatial coordinates and dissimilarity between non-adjacent spatial coordinates, while also showing great promise in their ability to detect discrete spatial domains using only one data modality. A third set of tools aim to leverage more than one omic data modality, while also leveraging the spatial coordinates. SpaGCN (Hu et al. 2021a) combines gene expression, spatial information, and histology image for spatial clustering using a graph convolutional neural network. However, this tool is designed to work only with hematoxylin and eosin (H&E) images. In this work, we aimed to address these limitations.

Inspired by the graph-based autoencoder and contrastive self-supervised learning frameworks for single omic data, here, we introduce Proust, a computationally scalable algorithm using contrastive self-supervised learning to predict discrete domains specifically designed spatial multi-omics data. We introduce an overview of the algorithm and compare the performance of our method to existing domain detection algorithms. By combining the lower-dimensional representation of multi-omic features that aggregate local tissue context through graph-based autoencoders, Proust identifies more biologically accurate and coherent tissue structures compared to existing state-of-the-art methods. In addition, we demonstrate how Proust can be used to detect discrete spatial domains in spatial tissue sections. Finally, we show how our method is computationally efficient and scalable in terms of memory and time, and we provide open-source software implemented in Python.

## 75    **Results**

### 76    **Overview of Proust to detect spatial domains integrating multiple data modalities**

77    Proust is a graph-based contrastive self-supervised learning framework to predict discrete spatial domains  
 78    using spatial multi-omics data. For the purposes of clarity, we describe how Proust can detect spatial domains  
 79    with two omic data modalities (RNA and proteins) (**Fig. 1**). However, these ideas can be generalized to other  
 80    types of multi-omics, for which we give examples of using RNA and brightfield images. Considering RNA  
 81    and protein, Proust takes as input gene expression counts, multi-channel immunofluorescence (IF) images, and  
 82    spot-level spatial position. The first step involves constructing a neighborhood graph structure based on the  
 83    relative distance between spots. Next, graph-based convolutional autoencoders are trained separately for gene  
 84    expression and extracted image features, aggregating genomic and protein information from neighboring  
 85    locations. Furthermore, the framework uses contrastive self-supervised learning (**Supplemental Fig. S1**) to  
 86    refine the latent embedding that maximizes similarities between adjacent spots while minimizing those  
 87    between non-adjacent spots. The reconstructed gene expression and image features obtained from the graph-  
 88    based decoder are used to extract the top principal components (PCs), which are used in conjunction with a  
 89    model-based clustering algorithm, `mclust` (Scrucca et al. 2016), to identify spatial domains.

### 90    **Proust increases accuracy to detect spatial domains and marker genes**

91    Proust was first applied to a CK-p25 mouse coronal brain tissue data set (Welch et al. 2022), measured on the  
 92    10x Genomics Visium Spatial Proteogenomics (SPG) platform (**Fig. 2A**). This dataset includes transcriptome-  
 93    wide gene expression and an immunofluorescence (IF) image measuring  $\gamma$ H2AX, a protein involved in DNA  
 94    repair. The resulting predicted spatial domains from Proust were compared to other methods GraphST,  
 95    SpaGCN, and STAGATE that employ graph convolutional neural networks (**Supplemental Fig. S2**). In one of  
 96    the CK-p25 mouse brain tissue replicates, Proust identified spatial domains 12, 15, 19, and 18, corresponding  
 97    to well-known mouse hippocampus (HPC) subfields, including CA1, CA2&3, dentate gyrus, and other HPC  
 98    subfields, respectively, as per the Allen Brain Atlas (<https://mouse.brain-map.org/static/atlas>). Additional  
 99    analysis showed that Proust consistently detected similar hippocampal subfields despite minor adjustments in

the predefined number of clusters (**Supplemental Fig. S3**). Proust’s latent embeddings showed distinct HPC regions that were well-separated from other brain regions (**Fig. 2B**). On the other hand, the compared methods did not detect these finer granularity subfields in the HPC, instead only identifying the HPC broadly (**Fig. 2C**). As noted in the original paper,  $\gamma$ H2AX is associated with the enrichment of reactive microglia (RM) (Welch et al. 2022). To evaluate the expression of marker genes within the spatial domains identified by Proust, we conducted a sub-regional analysis. This analysis revealed expression levels of reactive microglia marker genes, including cystatin *Cst7*, major histocompatibility complex (MHC) class I gene *H2-d1*, galectin gene *Lgals3bp*, and lipase *Lpl* in identified domains (**Fig. 2D** and **Supplemental Fig. S4**). These genes showed increased expression within the HPC subregions detected by Proust, aligning with areas of  $\gamma$ H2AX<sup>+</sup> capture, which suggests that Proust’s ability to integrate gene expression data with protein information supports the identification of more refined spatial structures. By incorporating the additional protein channel, Proust demonstrates its capacity to reveal spatially-resolved gene expression patterns that correspond with known subfields in the hippocampus, such as the dentate gyrus and CA regions.

Next, to evaluate the performance of the spatial domains detected by Proust, we used four Visium SPG human dorsolateral prefrontal cortex (DLPFC) brain tissue slices obtained from neurotypical donors (Huuki-Myers et al. 2024), each with paired multiplexed IF images stained for nuclei and four cell types (**Fig. 3A** and **Supplemental Fig. S5**). The four tissue sections included manually annotated spatial domains for white matter (WM) along with six morphological domains (Layers 1-6 or L1-6). Using the adjusted Rand index (ARI) as a performance measure, we compared the similarity of these manual annotations to the predicted spatial domains. We evaluated the performance of Proust along with five existing clustering methods that are commonly used for spatial domain detection, namely GraphST (Long et al. 2023), SpaGCN (Hu et al. 2021a), STAGATE (Dong and Zhang 2022), BayesSpace (Zhao et al. 2021), and *k*-means (Lloyd 1982).

It was found that the two GNN-based autoencoder and contrastive self-supervised learning frameworks (Proust and GraphST) resulted in the highest ARI, followed by SpaGCN, BayesSpace, STAGATE, with *k*-means showing the lowest adjusted Rand index (ARI) across the four tissue sections (**Fig. 3B**). However, Proust, which integrated both RNA and protein data modalities, outperformed the other methods in recognizing more coherent and biologically meaningful gray and white matter layers (**Fig. 3C** and

**Supplemental Fig. S6**). We also compared the Silhouette scores (Rousseeuw 1987) of Proust and three other GCN-based methods to assess the degree of separation between clusters in the human DLPFC Visium SPG samples (**Supplemental Fig. S7** and see **Supplemental Note S1** for comments on the Silhouette index).

Using Proust domains, we found that the UMAP (Uniform Manifold Approximation and Projection) (McInnes et al. 2018) embeddings revealed separate cortical layers ordered in known morphological layers (**Fig. 3D**). Using previously known marker genes for the cortical layers (Maynard et al. 2021), Proust domains resulted in laminar-specificity with the known marker genes (*AQP4* for L1, *HPCAL1* for L2 and L3, *RORB* for L4, *PCP4* for L5, *KRT17* for L6, and *MOBP* for WM) (**Figures 3E** and **Supplemental Fig. S8**), which also had more similar expression distributions in the manually annotated domains than GrahST (**Fig. 3F** and **Supplemental Fig. S9-S10**). However, using a data-driven approach, differentially expressed genes using Proust domains led to a more biologically meaningful hierarchical clustering of identified layers with white matter and L6 are grouped together, as opposed to the manual annotation, where instead white matter and L1 are grouped together (**Fig. 3G-H**).

## **Proust flexibly weights data modalities to detect spatial domains**

One advantage of Proust is the flexibility to weight multi-omics profiles (or data modalities), such as RNA and protein (**Fig. 1**), to detect spatial domains particularly in the context of either healthy or disease tissue. Here, we use as an example a 10x Genomics Visium SPG dataset profiling human inferior temporal cortex tissue sections collected from individuals with late-stage Alzheimer's disease (AD) (Kwon et al. 2023). In this dataset, the IF images contained five protein channels, namely nuclei (DAPI), amyloid-*beta* ( $A\beta$ ), hyperphosphorylated tau (pTau), microtubule-associated protein 2 (MAP2), and astrocytes (GFAP) (Kwon et al. 2023) (**Fig. 4A** and **Supplemental Fig. S11**). Also, the role of protein information is different from previous datasets, as some protein channels, are not useful to identify morphological domains. For example,  $A\beta$  is sparsely distributed throughout the tissue associated with AD pathology. Furthermore,  $A\beta$  and pTau are only detected at the protein level. In this section, we demonstrate how Proust can flexibly weigh the multiple data modalities to accurately identify spatial domains, even in diseased tissue.

For example, Proust can use all five protein channels to create a hybrid profile (top 30 PCs from the reconstructed gene expression and the top 5 PCs from the reconstructed extracted image features) to identify spatial domains (**Fig. 4B** and **Supplemental Fig. S12**), which can be used to visualize marker genes associated with the white and gray matter (**Fig. 4C**). Alternatively, we can compare predicted spatial domains from Proust (i) only using gene expression (and ignoring proteins) (**Fig. 4D**, columns 1 and 2 using  $k=2$  or  $k=4$ , respectively) and (ii) using gene expression and only two pathology-related protein channels ( $A\beta$  and pTau) (**Fig. 4D**, columns 3 and 4 using  $k=4$  or  $k=7$ , respectively).

In the latter case, Proust can weigh the RNA and protein information separately by controlling the number of PCs extracted from each data modality. This can lead to detected spatial domains corresponding to known morphology in healthy tissue, pathologies associated with disease, or both. For instance, in the tissue slice from Br3880, Proust identified spotty areas (cluster 3) and a layer (cluster 2), which are visually correlated to  $A\beta$ - and pTau- captured areas when clustering by a hybrid profile for  $k = 4$  clusters. In contrast, spatial domains corresponding to only cortical layers were detected when clustering by genes alone with the same number of clusters (**Fig. 4D**). Upon increasing the number of clusters to 7, Proust distinguished additional sub-layers within the gray matter with higher precision while retaining regions associated with  $A\beta$  and pTau (**Supplemental Fig. S13**). These results demonstrated that Proust is able to detect relevant spatial domains of interest by leveraging different protein channels and flexibly adjusting the number of PCs from each data modality into the hybrid profile. Finally, it was also observed that Proust's performance improved when the broad and connected spatial pattern is evident in IF images to complement expression information.

### **Proust accurately detects spatial domains with expression and histology images**

Next, we demonstrate how Proust can be generalized to other types of multi-omics, specifically with gene expression and H&E brightfield images, rather than IF staining, measured on the 10x Genomics Visium Spatial Expression platform. To evaluate the performance of Proust, we compared the predicted spatial domains to results from five existing clustering algorithms on  $N=12$  Visium human dorsolateral prefrontal cortex (DLPFC) tissue slices that have manually annotated spatial domains to be used as a gold standard (Maynard et al. 2021). The three RGB channels were included separately at the pixel level in the image feature

extraction steps and autoencoder model training. Proust achieved the highest ARI with a median (across  $N=12$  tissue sections) value of 0.60 and exhibited comparable performance to that of GraphST (median ARI = 0.53) (**Fig. 5A**). This suggests that Proust was also effective when applied to histology images. Additionally, the cortical layers segmented by Proust were more biologically consistent with manual annotations and exhibited greater spatial continuity, whereas other methods tended to produce more fragmented results (**Fig. 5B** and **Supplemental Fig. S14**). In particular, Proust was able to identify thinner layers, such as Layer 2, and provide coherent sub-layers within the gray matter, as demonstrated in samples 151509 and 151674. Although Proust and GraphST yielded similar ARIs, the UMAP plots of latent embeddings from Proust displayed a clearer separation of adjacent clusters and reflected more nuanced distinctions in identified spatial domains (**Fig. 5C** and **Supplemental Fig. S15**). These results suggest that Proust can also effectively extract histology image features that distinguish neighboring layers and refine the detection of coherent spatial regions and functional domains.

## Discussion

Proust is a novel framework that utilizes spatially-resolved transcriptomics, IF staining for protein channels, and spatial location information for the identification of spatial domains in individual tissue sections. Inspired by GraphST, Proust leverages a similar GCN-based autoencoder and contrastive self-supervised learning framework to detect spatial domains. In settings where additional protein information is unavailable, it is expected that Proust and GraphST will perform similarly. However, in settings where protein information is available, we demonstrate here that Proust improves upon GraphST by combining transcriptomics and proteomics. Specifically, Proust captures complex dependencies and spatial patterns, allowing for spatial segmentation of tissue structures with high accuracy. In principle, Proust's framework could easily be extended and applied in other multi-omics settings including measuring RNA and chromatin accessibility (Spatial ATAC–RNA-seq) (Zhang et al. 2023) or measuring RNA and metabolomics (spatial metabolomics) (Alexandrov 2023).

In this work, we demonstrate how Proust outperformed five popular existing methods in layer segmentation on human and mouse brain datasets generated from the 10x Visium platform, identifying

biologically meaningful layers consistent with manual annotations in human DLPFC and improving the detection of mouse hippocampus structures enriched with proteins of interest. Additionally, Proust is an adaptable framework that allows users to adjust the weights assigned to gene and protein information as needed. By selecting different protein channels and adjusting top PCs of different data modalities used in the downstream clustering step, Proust can better capture the spatial domains enriched with specific proteins of interest. For instance, when assigning higher weights to IF protein images, Proust can detect regions that contain sufficient levels of protein associated with disease within identified broad biological layers, as shown in the Visium SPG human inferior temporal cortex dataset. Future work could focus on strategies to identify optimal weight adjustment in a data-driven way. Furthermore, Proust can incorporate histology information in addition to spatial transcriptomics when IF images are unavailable in the dataset, thus making it a versatile tool for spatial analysis using various image types. Overall, Proust's flexibility allows users to fine-tune their analysis and obtain more accurate results tailored to their specific research questions and data types.

The use of Proust has some limitations and caveats. We recognize the challenge of selecting appropriate weights between different data modalities, especially when there is no manual annotation or ground truth available. Users may find it difficult to decide how much weight to give to each modality, which can impact the clustering results. We suggest starting with the default values in Proust (30 PCs for gene expression and 5 PCs for images), which have shown to work well across a range of datasets. However, we recommend experimenting with different numbers of PCs depending on the specific dataset and research questions. We also provide the option for users to specify how much variance they want to capture from each modality. This allows the algorithm to automatically select the appropriate number of PCs based on the user's preferences.

Second, we assume that adjacent spots or tissue areas have similar biological profiles in the gene and protein space. This assumption of the universal law of geography led to the implementation of graph-based neural networks for feature learning for both data modalities. However, for some proteins, such as  $A\beta$ , expression may be disconnected on the spot level, especially in the late stages of progressive diseases, potentially making it challenging for Proust to accurately identify all regions where the protein occurs, as individual spots may absorb dissimilar neighboring information during model training. To address this issue, a

potential solution is to enhance the graph structure with weighted edges based on the similarity of spot-level protein information or to incorporate inter-modality contrastive learning to maximize the mutual information between gene expression and proteomics (Yuan et al. 2021; Zeng et al. 2023).

Another limitation is Proust is currently only designed to identify spatial domains within one tissue section. Further work is needed to extend Proust to multiple tissue sections. Other ways this framework could be extended is to explore the use of statistical inference here as a way to explore the uncertainty of the predicted spatial domains. Additionally, since Proust is designed for spot-level information extraction when spatial coordinates for spot centroids are provided in 10x Visium datasets, it fails to recover pixel-level protein information from IF images unless strong signals of associated spatially-resolved gene expression are present. Additional applications could be explored including technologies with subcellular resolution. Another caveat is that image processing as a form of noise reduction is a crucial step before utilizing Proust, as the method may be affected by pixel values that result from technical artifacts.

The average computational time and peak GPU memory usage for analyzing the four tested datasets are approximately 2 minutes and 1GB per tissue sample on a high-performance computing cluster. We anticipate that datasets with more spots and protein channels will require more memory and processing time (Supplemental Fig. S16 and Supplemental Note S2). Thus, we recommend enabling the GPU option during model training to achieve quicker performance.

## Methods

### Data availability

#### CK-p25 mouse coronal brain data set

The CK-p25 mouse coronal brain tissue data set (Welch et al. 2022) was measured on the 10x Genomics Visium SPG platform. The data are available from the 10x Genomics website <https://www.10xgenomics.com/resources/datasets/>. In this dataset, the IF images captured  $\gamma$ H2AX, a protein involved in DNA repair. To enhance the  $\gamma$ H2AX gray-scale image and reduce noise, we established a

threshold of 6 standard deviations above the mean pixel values and applied a max filter with a  $10 \times 10$  image box to facilitate the detection of protein-rich regions by Proust.

## DLFPC Visium SPG

The human DLFPC brain tissue was measured on the 10x Genomics Visium SPG platform (Huuki-Myers et al. 2024). The source data are also publicly available from the Globus endpoint ‘jhpce#spatialDLFPC’ that are also listed at <http://research.libd.org/globus/>.

## ITC

The human inferior temporal cortex tissue sections collected from individuals with late-stage Alzheimer’s disease (AD) were profiled with the 10x Genomics Visium SPG platform (Kwon et al. 2023). The source data are also publicly available from the Globus endpoint ‘jhpce#Visium\_SPG\_AD’, which is also listed at <http://research.libd.org/globus>. In this dataset, the IF images contained five protein channels, namely nuclei (DAPI), amyloid-*beta* ( $A\beta$ ), hyperphosphorylated tau (pTau), microtubule-associated protein 2 (MAP2), and astrocytes (GFAP). We used the preprocessed grayscale images from VistoSeg (Tippani et al. 2023) and enhanced them in a manner similar to that employed in analyzing the Visium SPG mouse CK-p25 brain tissue dataset. To amplify the influence of protein information on spatial segmentation, we decreased the number of top PCs for transcriptomics to 10 while increasing those for proteomics to 10 from the default setting to create a hybrid biological profile.

## DLFPC Visium H&E

The human DLFPC brain tissue was measured on the 10x Genomics Visium Spatial platform with H&E images (Maynard et al. 2021). The raw data are publicly available from the Globus endpoint ‘jhpce#HumanPilot10x’, which is also listed at <http://research.libd.org/globus>.

## Spatially-resolved gene expression preprocessing

We evaluated the performance of Proust by testing it on four Visium datasets generated from the 10x Genomics platform. In Visium, the capture area measures  $6.5 * 6.5$  mm. The spots have a diameter of  $55 \mu\text{m}$

(equivalent to an area of  $2,375 \mu\text{m}^2$ ) and are spaced  $100 \mu\text{m}$  center-to-center in a honeycomb pattern (i.e., each spot is surrounded by six adjacent neighbors). Each tissue slide contains a total of 4992 sequenced spots, capturing around 35,000 genes. The raw gene count matrix is generated by aligning Visium sequencing data with the fiducial frame on the histology or immunofluorescence (IF) staining image of the tissue slice. Two-dimensional (2D) spatial positions are provided as pixel coordinates of spot centroids. The first three test datasets were generated using Visium SPG, which include multiplexed immunofluorescence images of proteins of interest from the corresponding tissue area. Additionally, we evaluated the transferability of Proust on a Visium dataset containing hematoxylin and eosin (H&E) images of the tissue slice.

To prepare the datasets for analysis, spots outside the tissue area are removed before applying Proust's standard preprocessing steps for gene expression using the SCANPY package (Wolf et al. 2018). The spatially-resolved gene expression information is represented by an  $N \times G$  matrix for  $N$  spots and  $G$  types of genes, with spot centroid locations indicated by 2-D spatial coordinates  $(x, y)$ . Spike-in and mitochondrial genes are filtered out, and those expressed in fewer than three spots are excluded. Raw gene counts are then (1) normalized by library size, (2) log-transformed, and (3) scaled to unit variance, with any values exceeding a standard deviation of 10 clipped. The top 3,000 highly variable genes (HVGs) are selected as input for the subsequent model training.

## Image feature extraction

Protein density can vary across different spatial regions, offering valuable information that can be integrated with spatial transcriptomics to enhance tissue architecture inference. Spot-level protein information is obtained by first splitting the full-resolution image into  $d \times d$  circumscribed grids centered on each spot centroid, where  $d$  is the ceiling rounding for the spot diameter. To improve computational efficiency, we reduce the size of each image grid to  $48 \times 48$  pixels for 10x Genomics datasets, as the spot diameter is usually larger than or around 100 pixels. This downsizing is achieved by using the OpenCV Python package (Bradski 2000), where the updated pixel value of the new interpolated image is calculated as a weighted average of input pixel values from a sliding window grid in the original image. The grid size is determined by the ratios of the original image's width and height to those of the resized image. To ensure that different modalities of data can be

jointly analyzed and achieve optimal model performance in identifying spatial patterns using gene expression and protein data, pixel values are normalized to the same range as the preprocessed gene expression prior to model training.

Next, as is common practice in image processing, we implement a simple convolutional neural network (CNN) autoencoder to extract protein features from interpolated image grids (**Supplemental Note S5**). The model takes the resized image grid  $I$  for each spot as input. The encoder consists of two fully-connected inner layers with a kernel size of  $5 \times 5$ , each followed by an average pooling layer with a kernel size of 2 and stride of 2. The decoder consists of two transposed convolutional layers to reconstruct latent representations and images with the same dimensions as those of the intermediate output from the encoder. A ReLU (Rectified Linear Unit) activation function is applied after each convolution step, which allows the CNN model to learn complex and abstract features from images by introducing non-linearity to the output. We train this model by minimizing the self-reconstruction loss of image grids using the following equation:

$$L_{recon} = \sum_{i=1}^{N_{spot}} \| I - H \|_2$$

where  $H$  is the reconstructed image grid. We use the Adam optimizer to minimize the reconstruction loss with an initial learning rate of  $1 \times 10^{-3}$ . The default number of iterations is set to 800. The latent embeddings produced by this CNN-based autoencoder capture continuous spot-level spatial patterns for each protein channel instead of describing pixel-level intensity. This approach enables us to focus on the broad changes in protein distribution within the stained tissue slice, which complements our analysis of spatially-resolved gene expression (**Supplemental Fig. S17** and **Supplemental Note S3**).

## Graph convolutional autoencoder

### Graph structure based on spatial information.

Incorporating spatial information along with biological features is crucial for identifying coherent spatial domains, as it allows for considering neighboring information from the nearest spots. To leverage the spatial information, Proust first converts the spot-level spatial coordinates into an undirected neighbor graph  $G =$

( $V, E$ ) with a pre-defined (or user-defined) neighbor number  $k$ . Here,  $G$  represents the graph,  $V$  represents spots, and  $E$  represents connected edges between each pair of spots  $(i, j) \in V$ . The structure of the graph  $G$  based on spatial proximity is measured using the adjacency matrix  $A \in \mathbb{R}^{N \times N}$ , where  $N$  is the number of spots. For a given spot  $i$ , the adjacency matrix assigns a value of 1 to  $A_{ij}$  if spot  $j$  is among the  $k$ -nearest neighbors selected based on the Euclidean distance; otherwise, it assigns a value of 0. To normalize the influences across spots, a symmetric normalized Laplacian matrix  $L = I_n + D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$  is constructed, which also takes into account the information from each spot itself. The degree matrix  $D$  is a diagonal matrix with elements  $D_{ii} = \sum_j A_{ij}$  being the number of edges (i.e., neighbors) attached to each spot. When constructing the graph structure, we set  $k = 6$  to reflect the hexagonal structure of spots in the Visium assay.

### GCN-based autoencoder to reconstruct transcriptomic and proteomic features.

We construct an autoencoder using a graph convolutional network (GCN) to learn a latent representation  $Z_i$  for spot  $i$  by aggregating information from neighboring spots  $j$ 's that share similar biological profiles. In this approach, the GCN-based autoencoder takes preprocessed biological features  $X$  and spatial information stored in  $L$  as input and outputs the reconstructed spot-level feature matrix  $H$ .

The  $(t + 1)$ -th layer representation in the encoder for spot  $i$  is constructed using a trainable weight matrix  $W_e^t$  and a trainable bias vector  $b_e^t$ , with a non-linear activation function ReLU denoted by  $\sigma(\cdot)$ . The encoder representation is formulated as follows:

$$Z_i^{t+1} = \sigma(LZ_i^t W_e^t + b_e^t).$$

We denote  $Z^0$  as the original feature matrix  $X$  as input and  $Z$  as the final output of the encoder. The latent representation  $Z$  is then fed into a decoder to reconstruct the feature space  $H$  iteratively. The  $(t - 1)$ -th layer representation in the decoder is constructed using a trainable weight matrix  $W_d^t$  and a trainable bias vector  $b_d^t$ , with the same activation function ReLU. The decoder representation is formulated as follows:

$$H_i^{t-1} = \sigma(LH_i^t W_d^t + b_d^t),$$

where the first inner layer  $H^{t_{max}}$  in the decoder is set as the final latent space  $Z$  from the encoder, and  $H^0$  is the reconstructed feature matrix  $H$ .

The autoencoder for gene expression data takes the preprocessed gene counts as input. On the other hand, the autoencoder for protein information takes the extracted image features as input, which are flattened for each channel before being fed into the encoder. This image autoencoder is three-dimensional, where the first dimension is the number of observations (i.e., spots), the second dimension is the number of protein or image (i.e., RGB channels for H&E histology images) channels, and the third dimension is the protein features. By training separate autoencoders for transcriptomic and proteomic data, the model can capture distinct patterns and features specific to each modality, which can then be integrated before the downstream clustering procedure.

### Contrastive self-supervised learning

Proust addresses the challenge of distinguishing spots from similar spatial domains by implementing contrastive self-supervised learning (CSL) adapted from Deep Graph Infomax (Veličković et al. 2019) to learn more discriminative reconstructed biological features (**Supplemental Fig. S1**). Contrastive self-supervised learning is a method for training deep neural networks to extract representations of data by comparing similar and dissimilar pairs of samples without relying on labeled data. This refinement step enables the learning of attributes that are common between data groups and attributes that separate one data group from another. In the data augmentation part, we generate a corrupted neighboring graph structure  $(X', A)$ , as opposed to the original structure  $(X, A)$ , by randomly shuffling biological features while preserving the distance-based graph representing the spatial proximity between pairs of spots. We then feed the real and corrupted graph structures into a shared GCN-based autoencoder to obtain corresponding spot-level latent representations  $Z$  and  $Z'$ , respectively. A local neighborhood context  $S_i$  for a given spot  $i$  is summarized with a read-out function defined as follows:

$$S_i = \mathcal{R}(Z_i) = \sigma\left(\frac{1}{k} \sum_{j=1}^k Z_j + Z_i\right),$$

where  $k$  is the number of nearest neighbors for each individual spot. To maximize the mutual information between spot embeddings and local neighborhood embeddings, we calculate a discriminative score for context-spot representation pairs using a simple bilinear scoring function as the following:

$$\mathcal{D}(Z_i, S_i) = \sigma(Z_i^T W S_i)$$

where  $W$  is a trainable scoring matrix and  $\sigma(\cdot)$  is the logistic sigmoid function. A positive pair formed by  $Z_i$  with the corresponding real local representation  $S_i$  (or corrupted  $Z'_i$  with  $S'_i$ ) will be assigned a higher probability score, whereas a negative pair formed by  $Z_i$  with the corrupted local representation  $S'_i$  (or corrupted  $Z'_i$  with real local representation  $S_i$ ) will be assigned a lower probability score. CSL refines the final latent embeddings from the encoder before reconstructed layers in the decoder. We train the GCN-based autoencoder with CSL by minimizing an overall loss combined with the self-reconstruction loss of the autoencoder and contrastive loss:

$$Loss = \alpha L_{recon} + \beta (L_{CSL} + L_{CSL_{corrupt}})$$

where

$$L_{CSL} = -\frac{1}{2N} \left( \sum_{i=1}^N \mathbb{E}_{(X,A)} [\log \mathcal{D}(Z_i, S_i)] + \sum_{j=1}^N \mathbb{E}_{(X',A)} [\log(1 - \mathcal{D}(Z'_j, S_j))] \right)$$

$$L_{CSL_{corrupt}} = -\frac{1}{2N} \left( \sum_{i=1}^N \mathbb{E}_{(X',A)} [\log \mathcal{D}(Z'_i, S'_i)] + \sum_{j=1}^N \mathbb{E}_{(X,A)} [\log(1 - \mathcal{D}(Z_j, S'_j))] \right)$$

Hence, this supplementary process helps the GCN-based autoencoder to reconstruct the updated biological features, which bring similar spots together and differentiate dissimilar ones while preserving the fundamental information from the original input matrix (**Supplemental Fig. S17**). The Adam optimizer is utilized for training this deep learning architecture with an initial learning rate of  $1 \times 10^{-3}$ . The default number of iterations is set to 600.

## 388 Clustering and refinement

389 After training GCN-based autoencoders with preprocessed transcriptomic and proteomic features, Proust  
 390 extracts the top principal components (PCs) from the reconstructed gene expression and extracted protein  
 391 features separately. These components are then concatenated to create a hybrid biological profile for cluster  
 392 assignment using a non-spatial clustering algorithm called `mclust` (Scrucca et al. 2016). The default number  
 393 of PCs used for transcriptomics is 30, and the default number of PCs used for proteomics is 5 (**Supplemental**  
 394 **Fig. S18**). Users can adjust the number of PCs to give adjusted weights for the two data modalities based on  
 395 their needs (**Supplemental Note S4**). We pre-set the cluster count in `mclust` to correspond with the number  
 396 of clusters in datasets that have manual annotations. For datasets without prior knowledge of the correct cluster  
 397 count, we evaluated various cluster numbers and chose the one yielding the highest Silhouette score  
 398 (Rousseeuw 1987). After clustering, Proust also offers an optional refinement step in which a given spot  $i$  is  
 399 relabeled to the most common spatial domain of its  $r$  nearest surrounding spots. The default setting for  $r$  is 10.

## 400 Evaluation and comparison

### 401 Adjusted Rand Index

402 To evaluate the performance of the Proust framework, we use the Adjusted Rand Index (ARI) to measure the  
 403 agreement between identified spatial domains and manual annotation for individual tissue slices. Let  $\hat{Y}$   
 404 represent the assigned spatial clusters and  $Y$  represent the ground truth of clusters of  $N$  spots. Then,

$$ARI = \frac{\sum_{ls} \binom{N_{ls}}{2} - \sum_l \binom{N_l}{2} \sum_s \binom{N_s}{2} / \binom{N}{2}}{[\sum_l \binom{N_l}{2} + \sum_s \binom{N_s}{2}] / 2 - \sum_l \binom{N_l}{2} \sum_s \binom{N_s}{2} / \binom{N}{2}}$$

405 where  $l$  and  $s \in m$  clusters,  $N_l = \sum_i^N I(\hat{y}_i = l)$ ,  $N_s = \sum_i^N I(y_i = s)$ ,  $N_{ls} = \sum_i^N I(\hat{y}_i = l)I(y_i = s)$ .  $I(\cdot)$  is an  
 406 indicator function that follows  $I(a = b) = 1$  when  $a = b$ , otherwise 0. The ARI ranges from 0 to 1, wherein a  
 407 higher value indicates a better match between clustering results with the manual annotation.

### 408 Existing methods

409 We benchmark Proust against the following existing methods for spatial domain detection:

410 **GraphST:** GraphST is a graph contrastive self-supervised learning framework that incorporates spatial  
 411 information and gene expression for spatial clustering (Long et al. 2023). We followed the tutorial with default  
 412 parameter settings and set  $r = 50$  during refinement.

413 **SpaGCN:** SpaGCN combines gene expression, spatial information, and histology image for spatial clustering  
 414 using graph convolutional neural network (Hu et al. 2021a). We followed the tutorial to use SpaGCN with the  
 415 default parameter settings. The ‘histology’ option is disabled when no histology information is available in the  
 416 dataset.

417 **STAGATE:** STAGATE is a method that utilizes an autoencoder and graph attention mechanism to learn a  
 418 latent representation by incorporating spatial information and gene expression (Dong and Zhang 2022). We  
 419 followed the tutorial with default parameter settings and set the cell type-aware module ‘alpha’ to 0.

420 **BayesSpace:** BayesSpace is a spatial clustering method that utilizes a Markov random field with a Bayesian  
 421 framework. The method assigns greater importance to adjacent spots by incorporating a prior that considers  
 422 the spatial proximity of the spots (Zhao et al. 2021). We followed the tutorial with default parameter settings  
 423 and set the number of iterations ‘nrep’ to 10000.

424 ***k*-means:** *k*-means is a clustering algorithm that partitions a set of  $n$  data points into  $k$  clusters, where each  
 425 data point belongs to the cluster with the nearest centroid profile (Lloyd 1982). As the only method compared  
 426 that is not specifically designed for ST data, *k*-means is used as the baseline comparison. We followed the  
 427 tutorial from the bluster R/Bioconductor package with default parameter settings (Lun 2022).

428 **STACI:** This newly developed method (Zhang et al. 2022) jointly analyzes spatial transcriptomics and  
 429 chromatin imaging data with over-parameterized graph-based autoencoders. However, it is not evaluated here  
 430 as it is not available as a package at the time.

## 431 **Software Availability**

432 The Proust algorithm is implemented in Python and is available on GitHub ([github.com/JianingYao/proust](https://github.com/JianingYao/proust))  
 433 and as Supplemental Code. We used Proust version 1.0 for the analyses in this manuscript. The code to

434 reproduce all preprocessing, analyses, and figures in this manuscript is available on GitHub  
435 ([github.com/JianingYao/proust\\_paper](https://github.com/JianingYao/proust_paper)).

## 436 **Competing interest statement**

437 The authors declare that they have no competing interests.

## 438 **Acknowledgements**

439 We would like to acknowledge Erik Nelson, Sang Ho Kwon, and Kasper Hansen for their helpful comments,  
440 feedback, and suggestions on the methodology and manuscript. We would like to acknowledge Kristen  
441 Maynard for the manual annotations of the Visium H&E images. Erik Nelson, Sang Ho Kwon, and Kristen  
442 Maynard are from Johns Hopkins School of Medicine and Lieber Institute for Brain Development. Kasper  
443 Hansen is from the Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics. We also  
444 thank the maintainers of the Joint High-Performance Computing Exchange (JHPCE) compute cluster at Johns  
445 Hopkins Bloomberg School of Public Health for providing essential computing resources.

## 446 **Funding**

447 Research reported in this publication was supported by the National Institute of Mental Health (NIMH) of the  
448 National Institutes of Health (NIH) under the awards number U01MH122849 (KM) and R01MH126393  
449 (KM), and also supported by National Institute of General Medical Sciences (NIGMS) of the NIH under the  
450 award number R35GM150671 (SCH). This project was also supported by CZF2019-002443 and CZF2018-  
451 183446 (SCH) from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community  
452 Foundation. All funding bodies had no role in the design of the study and collection, analysis, and  
453 interpretation of data and in writing the manuscript.

## 454 **Author Contributions**

455 JYao developed the Proust framework, performed analyses, created figures, and drafted text. JYu implemented  
456 Proust in a software package and provided input on the methodological framework and text. BC provided

advice on the architecture of Proust and edited the text. SCP and KM provided input on the biological interpretation of the data used, figures, and text. SCH supervised the project; provided input on the methodological framework, software implementation, analyses, figures, and text; and drafted text. All authors approved the final manuscript.

## References

- Alexandrov T. 2023. Spatial metabolomics: from a niche field towards a driver of innovation. *Nat Metab* 5: 1443–1445. doi: [10.1038/s42255-023-00881-0](https://doi.org/10.1038/s42255-023-00881-0).
- Asp M, Giacomello S, Larsson L, Wu C, Fürth D, Qian X, Wärdell E, Custodio J, Reimegård J, Salmén F, et al. 2019. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* 179: 1647–1660.e19. doi: [10.1016/j.cell.2019.11.025](https://doi.org/10.1016/j.cell.2019.11.025).
- Bradski G. 2000. The OpenCV Library. *Dr Dobb's Journal of Software Tools*.
- Chen WT, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, Corthout N, Qian X, Laláková J, Kühnemund M, Voytyuk I, et al. 2020. Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* 182: 976–991.e19. doi: [10.1016/j.cell.2020.06.038](https://doi.org/10.1016/j.cell.2020.06.038).
- Denisenko E, De Kock L, Tan A, Beasley AB, Beilin M, Jones ME, Hou R, Muirí DÓ, Bilic S, Mohan GRKA, et al. 2024. Spatial transcriptomics reveals discrete tumour microenvironments and autocrine loops within ovarian cancer subclones. *Nat Commun* 15: 2860. doi: [10.1038/s41467-024-47271-y](https://doi.org/10.1038/s41467-024-47271-y).
- Dong K, Zhang S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun* 13: 1739. doi: [10.1038/s41467-022-29439-6](https://doi.org/10.1038/s41467-022-29439-6).
- Dries R, Zhu Q, Dong R, Eng CHL, Li H, Liu K, Fu Y, Zhao T, Sarkar A, Bao F, et al. 2021. Giotto: A toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* 22: 78. doi: [10.1186/s13059-021-02286-2](https://doi.org/10.1186/s13059-021-02286-2).

- 479 Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, Lee EB, Shinohara RT, Li M. 2021a. SpaGCN:  
 480 Integrating gene expression, spatial location and histology to identify spatial domains and spatially  
 481 variable genes by graph convolutional network. *Nat Methods* **18**: 1342–1351. doi: [10.1038/s41592-](https://doi.org/10.1038/s41592-021-01255-8)  
 482 [021-01255-8](https://doi.org/10.1038/s41592-021-01255-8).
- 483 Hu J, Schroeder A, Coleman K, Chen C, Auerbach BJ, Li M. 2021b. Statistical and machine learning methods  
 484 for spatially resolved transcriptomics with histology. *Comput Struct Biotechnol J* **19**: 3829–3841. doi:  
 485 [10.1016/j.csbj.2021.06.052](https://doi.org/10.1016/j.csbj.2021.06.052).
- 486 Huuki-Myers LA, Spangler A, Eagles NJ, Montgomery KD, Kwon SH, Guo B, Grant-Peters M, Divecha HR,  
 487 Tippi M, Sriworarat C, et al. 2024. A data-driven single-cell and spatial transcriptomic map of the  
 488 human prefrontal cortex. *Science* **384**: eadh1938. doi: [10.1126/science.adh1938](https://doi.org/10.1126/science.adh1938).
- 489 Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, Guo MG, George BM, Mollbrink A,  
 490 Bergenstr hle J, et al. 2020. Multimodal Analysis of Composition and Spatial Architecture in Human  
 491 Squamous Cell Carcinoma. *Cell* **182**: 497–514.e22. doi: [10.1016/j.cell.2020.05.039](https://doi.org/10.1016/j.cell.2020.05.039).
- 492 Kwon SH, Parthiban S, Tippi M, Divecha HR, Eagles NJ, Lobana JS, Williams SR, Mak M, Bharadwaj RA,  
 493 Kleinman JE, et al. 2023. Influence of Alzheimer’s Disease Related Neuropathology on Local  
 494 Microenvironment Gene Expression in the Human Inferior Temporal Cortex. *GEN Biotechnol* **2**: 399–  
 495 417. doi: [10.1089/genbio.2023.0019](https://doi.org/10.1089/genbio.2023.0019).
- 496 Li J, Chen S, Pan X, Yuan Y, Shen HB. 2022a. Cell clustering for spatial transcriptomics data with graph  
 497 neural networks. *Nat Comput Sci* **2**: 399–408. doi: [10.1038/s43588-022-00266-5](https://doi.org/10.1038/s43588-022-00266-5).
- 498 Li X. 2023. Harnessing the potential of spatial multiomics: A timely opportunity. *Signal Transduct Target*  
 499 *Ther* **8**: 234. doi: [10.1038/s41392-023-01507-3](https://doi.org/10.1038/s41392-023-01507-3).
- 500 Li Y, Stanojevic S, Garmire LX. 2022b. Emerging Artificial Intelligence Applications in Spatial  
 501 Transcriptomics Analysis. *Comput Struct Biotechnol J* **20**: 2895–2908. doi:  
 502 [10.1016/j.csbj.2022.05.056](https://doi.org/10.1016/j.csbj.2022.05.056).

- 503 Liu Y, Yang M, Deng Y, Su G, Enniful A, Guo CC, Tebaldi T, Zhang D, Kim D, Bai Z, et al. 2020. High-  
 504 Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**: 1665–  
 505 1681.e18. doi: [10.1016/j.cell.2020.10.026](https://doi.org/10.1016/j.cell.2020.10.026).
- 506 Lloyd S. 1982. Least squares quantization in PCM. *IEEE Trans Inf Theory* **28**: 129–137. doi:  
 507 [10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489).
- 508 Long Y, Ang KS, Li M, Chong KLK, Sethi R, Zhong C, Xu H, Ong Z, Sachaphibulkij K, Chen A, et al. 2023.  
 509 Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST.  
 510 *Nat Commun* **14**: 1155. doi: [10.1038/s41467-023-36796-3](https://doi.org/10.1038/s41467-023-36796-3).
- 511 Lun A. 2022. bluster: Clustering Algorithms for Bioconductor. doi: [10.18129/B9.bioc.bluster](https://doi.org/10.18129/B9.bioc.bluster).
- 512 Maniatis S, Äijö T, Vickovic S, Braine C, Kang K, Mollbrink A, Fagegaltier D, Andrusivová Ž, Saarenpää S,  
 513 Saiz-Castro G, et al. 2019. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral  
 514 sclerosis. *Science* **364**: 89–93. doi: [10.1126/science.aav9776](https://doi.org/10.1126/science.aav9776).
- 515 Maynard KR, Collado-Torres L, Weber LM, Uytingco C, Barry BK, Williams SR, Catallini JL2nd, Tran MN,  
 516 Besich Z, Tippani M, et al. 2021. Transcriptome-scale spatial gene expression in the human  
 517 dorsolateral prefrontal cortex. *Nat Neurosci* **24**: 425–436. doi: [10.1038/s41593-020-00787-0](https://doi.org/10.1038/s41593-020-00787-0).
- 518 McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: Uniform Manifold Approximation and Projection. *J*  
 519 *Open Source Softw* **3**: 861. doi: [10.21105/joss.00861](https://doi.org/10.21105/joss.00861).
- 520 Nature. 2021. Method of the Year 2020: Spatially Resolved Transcriptomics. *Nat Methods* **18**: 1–1. doi:  
 521 [10.1038/s41592-020-01042-x](https://doi.org/10.1038/s41592-020-01042-x).
- 522 Ren H, Walker BL, Cang Z, Nie Q. 2022. Identifying multicellular spatiotemporal organization of cells with  
 523 SpaceFlow. *Nat Commun* **13**: 4076. doi: [10.1038/s41467-022-31739-w](https://doi.org/10.1038/s41467-022-31739-w).
- 524 Rousseeuw PJ. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J*  
 525 *Comput Appl Math* **20**: 53–65. doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).

- 526 Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. mclust 5: Clustering, Classification and Density Estimation  
 527 Using Gaussian Finite Mixture Models. *R J* **8**: 289–317. doi: [10.32614/RJ-2016-021](https://doi.org/10.32614/RJ-2016-021).
- 528 Tang Z, Luo S, Zeng H, Huang J, Sui X, Wu M, Wang X. 2024. Search and match across spatial omics  
 529 samples at single-cell resolution. *Nat Methods* **21**: 1818–1829. doi: [10.1038/s41592-024-02410-7](https://doi.org/10.1038/s41592-024-02410-7).
- 530 Tippani M, Divecha HR, Catallini JL, Kwon SH, Weber LM, Spangler A, Jaffe AE, Hyde TM, Kleinman JE,  
 531 Hicks SC, et al. 2023. VistoSeg: Processing utilities for high-resolution images for spatially resolved  
 532 transcriptomics data. *Biol Imaging* **3**: e23. doi: [10.1017/S2633903X23000235](https://doi.org/10.1017/S2633903X23000235).
- 533 Traag VA, Waltman L, Van Eck NJ. 2019. From Louvain to Leiden: guaranteeing well-connected  
 534 communities. *Sci Rep* **9**: 5233. doi: [10.1038/s41598-019-41695-z](https://doi.org/10.1038/s41598-019-41695-z).
- 535 Veličković P, Fedus W, Hamilton WL, Liò P, Bengio Y, Hjelm RD. 2019. Deep Graph Infomax. In  
 536 *International Conference on Learning Representations*. doi: [10.48550/arxiv.1809.10341](https://doi.org/10.48550/arxiv.1809.10341).
- 537 Welch GM, Boix CA, Schmauch E, Davila-Velderrain J, Victor MB, Dileep V, Bozzelli PL, Su Q, Cheng JD,  
 538 Lee A, et al. 2022. Neurons burdened by DNA double-strand breaks incite microglia activation  
 539 through antiviral-like signaling in neurodegeneration. *Sci Adv* **8**: eabo4662. doi: [10.1126/sci-](https://doi.org/10.1126/sci-adv.abo4662)  
 540 [adv.abo4662](https://doi.org/10.1126/sci-adv.abo4662).
- 541 Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome*  
 542 *Biol* **19**: 15. doi: [10.1186/s13059-017-1382-0](https://doi.org/10.1186/s13059-017-1382-0).
- 543 Yuan X, Lin Z, Kuen J, Zhang J, Wang Y, Maire M, Kale A, Faieta B. 2021. Multimodal Contrastive Training  
 544 for Visual Representation Learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern*  
 545 *Recognition (CVPR)*, pp. 6991–7000. doi: [10.1109/CVPR46437.2021.00692](https://doi.org/10.1109/CVPR46437.2021.00692).
- 546 Zeng Y, Yin R, Luo M, Chen J, Pan Z, Lu Y, Yu W, Yang Y. 2023. Identifying spatial domain by adapting  
 547 transcriptomics with histology through contrastive learning. *Brief Bioinform* **24**: bbad048. doi:  
 548 [10.1093/bib/bbad048](https://doi.org/10.1093/bib/bbad048).

- 549 Zeng Z, Li Y, Li Y, Luo Y. 2022. Statistical and machine learning methods for spatially resolved  
550 transcriptomics data analysis. *Genome Biol* **23**: 83. doi: [10.1186/s13059-022-02653-7](https://doi.org/10.1186/s13059-022-02653-7).
- 551 Zhang D, Deng Y, Kukanja P, Agirre E, Bartosovic M, Dong M, Ma C, Ma S, Su G, Bao S, et al. 2023. Spatial  
552 epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**: 113–122. doi:  
553 [10.1038/s41586-023-05795-1](https://doi.org/10.1038/s41586-023-05795-1).
- 554 Zhang X, Wang X, Shivashankar GV, Uhler C. 2022. Graph-based autoencoder integrates spatial  
555 transcriptomics with chromatin images and identifies joint biomarkers for Alzheimer’s disease. *Nat*  
556 *Commun* **13**: 7480. doi: [10.1038/s41467-022-35233-1](https://doi.org/10.1038/s41467-022-35233-1).
- 557 Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, Williams SR, Uytingco CR, Taylor SEB,  
558 Nghiem P, et al. 2021. Spatial transcriptomics at subspot resolution with BayesSpace. *Nature*  
559 *Biotechnol* **39**: 1375–1384. doi: [10.1038/s41587-021-00935-2](https://doi.org/10.1038/s41587-021-00935-2).

## 560 **Figure Legends**

561 **Figure 1: Overview of Proust for detecting discrete domains using spatial multi-omics data.** For the  
562 purposes of clarity, we introduce Proust with two specific omic data modalities: RNA and protein, but these  
563 ideas can be generalized to other multi-omics, such as RNA and brightfield images. First, Proust constructs a  
564 graph structure based on the Euclidean distance between spatial coordinates. Next, graph-based convolutional  
565 autoencoders are trained separately for gene expression and protein information extracted from an  
566 immunofluorescence (IF) image. The latent embeddings are refined using contrastive self-supervised learning  
567 (CSL). The top principal components (PCs) from the reconstructed gene and image features are concatenated  
568 to create a hybrid profile for downstream clustering analysis.

570 **Figure 2: Proust improves the detection of hippocampal spatial domains in CK-p25 mouse coronal brain**  
571 **tissue by integrating gene expression with proteins of interest. (A)** From left to right: (i) annotation of  
572 mouse hippocampus subfields from the Allen Reference Brain Atlas, (ii) merged DAPI and  $\gamma$ H2AX

immunofluorescence images, and (iii) IF staining of  $\gamma$ H2AX. **(B)** UMAP representation of spots colored by spatial domains detected by `mclust` using Proust's latent embeddings. **(C)** Predicted spatial domains by Proust, GraphST (Long et al. 2023), SpaGCN (Hu et al. 2021a), and STAGATE (Dong and Zhang 2022) with  $k = 20$  domains. **(D)** Spatial expression level of four RM marker genes across the entire tissue slice and boxplots of corresponding marker genes stratified by  $k=20$  domains identified by Proust. Hippocampal subregions are depicted in orange; other regions are depicted in gray.

### **Figure 3: Proust improves the accuracy of predicting spatial domains compared to existing methods.**

Data from sample Br6432 from Visium SPG human DLPFC dataset (Huuki-Myers et al. 2024), unless noted otherwise. **(A)** Immunofluorescence images of five protein channels: nuclei (DAPI), neurons (NeuN), oligodendrocytes (OLIG2), astrocytes (GFAP), and microglia (TMEM119). **(B)** Boxplot of adjusted Rand index (ARI) across four samples. **(C)** Manual annotation of tissue slice from donor Br6432 and predicted spatial domains by the six methods. Labels do not indicate corresponding biological layers assigned by the algorithms. **(D)** UMAP visualization of spots from donor Br6432 colored by Proust predictions. **(E)** Stacked violin plot of marker gene distribution for white matter and sub-layers of gray matter based on literature in each spatial domain assigned by Proust. Red rectangles are highlighted marker genes in (F). **(F)** Violin plots of marker gene expression for Proust and manually annotated domains. **(G)** Heatmaps of the top 5 differentially expressed genes (centered and scaled) across layers from Proust and manual annotations. A dendrogram on the right shows hierarchical clustering. **(H)** Selected cluster-based marker genes expression and visualization of individual clusters identified by Proust. Layers were annotated according to the laminar organization indicated by the manual annotation.

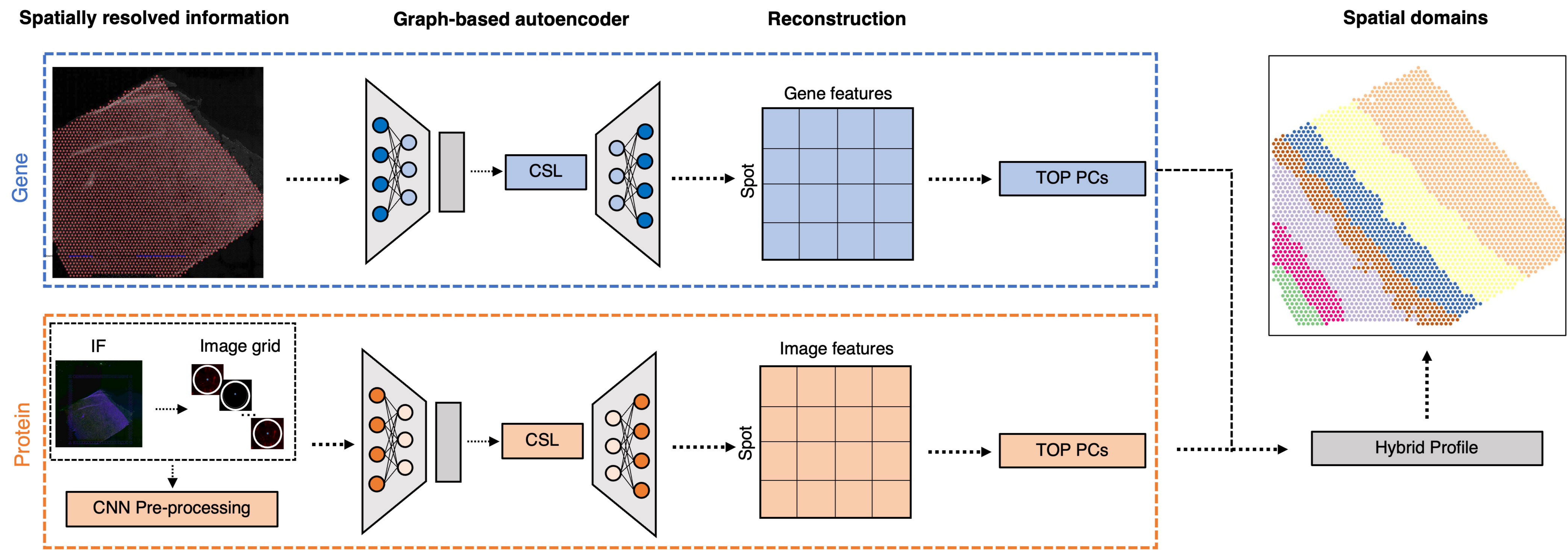
### **Figure 4: Proust achieves distinct spatial domain detection with different protein channels and weights assigned to transcriptomics and proteomics on Visium SPG human inferior temporal cortex tissue slices from donor Br3880 and Br3854.**

**(A)** Immunofluorescence staining images of  $A\beta$  and pTau. **(B)** Proust clustering result using five protein channels (DAPI,  $A\beta$ , pTau, MAP2, and GFAP), top 30 PCs from

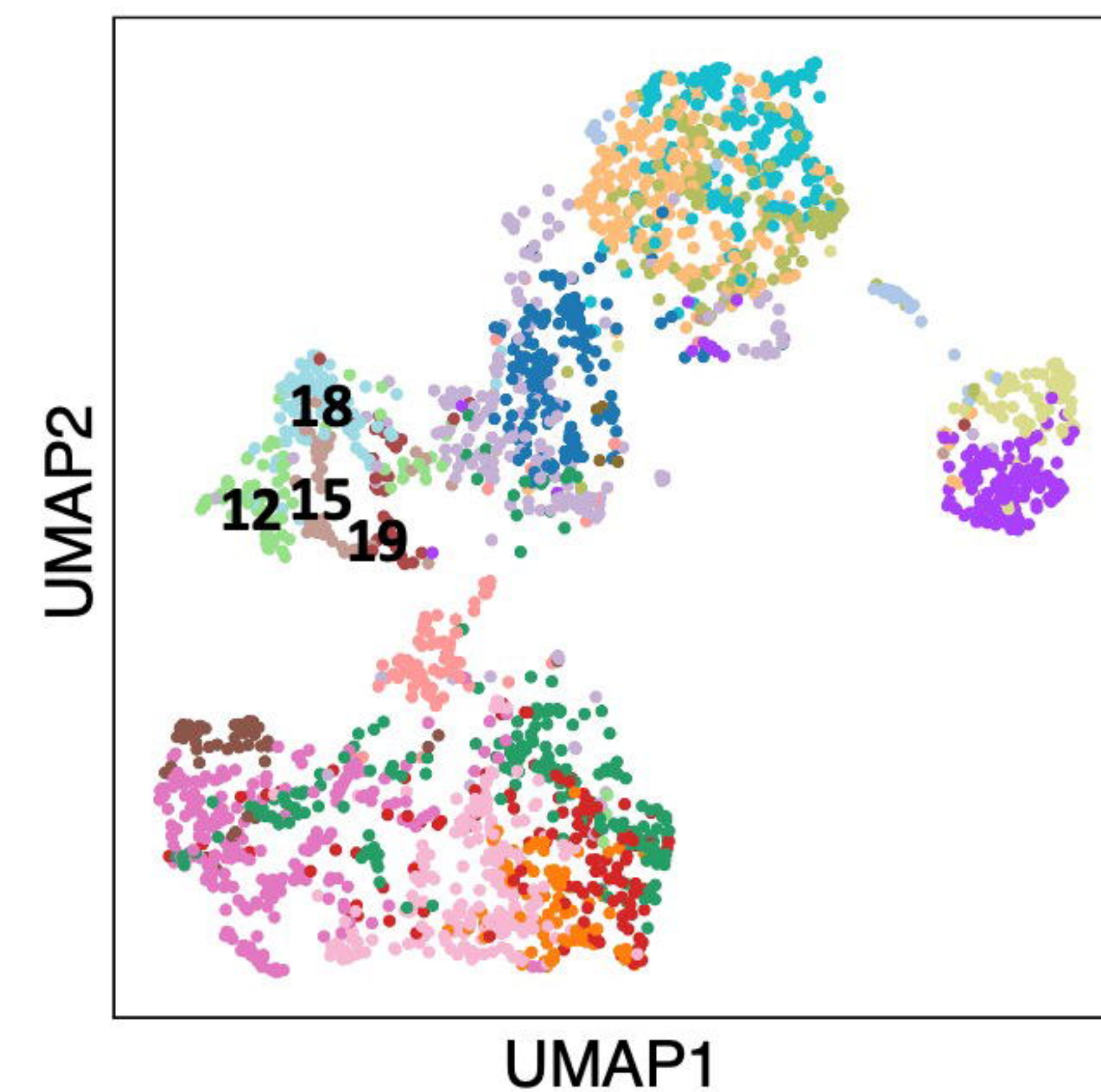
599 reconstructed gene expression, top 5 PCs from reconstructed extracted image features, and  $k = 7$  clusters. **(C)**  
 600 Stacked violin plot of the distribution of marker genes (*MOBP* for oligodendrocytes/WM, *SNAP25* for  
 601 neurons/gray matter) in each spatial domain assigned by Proust. **(D)** Proust clustering result using two protein  
 602 channels ( $A\beta$  and pTau). The first two columns show clustering results using transcriptomics only when  $k = 2$   
 603 and  $k = 4$  clusters, respectively. The last two columns show clustering results using a hybrid profile of  
 604 transcriptomics and proteomics, with the top 10 PCs from reconstructed gene expression and the top 10 PCs  
 605 from reconstructed extracted image features when  $k = 4$  and  $k = 7$  clusters, respectively.

606

607 **Figure 5: Evaluating and comparing the performance of Proust in layer segmentation with other**  
 608 **popular existing methods on the Visium human DLPFC dataset that contains H&E images. (A)** Boxplot  
 609 of clustering accuracy in 12 DLPFC samples across Proust and five other existing methods based on adjusted  
 610 rand index (ARI). **(B)** Manual annotation of tissue slices 151509 and 151674 and spatial domains assigned by  
 611 the six methods. **(C)** UMAP visualization of reduced dimensions from Proust and GraphST for 151509 and  
 612 151674.

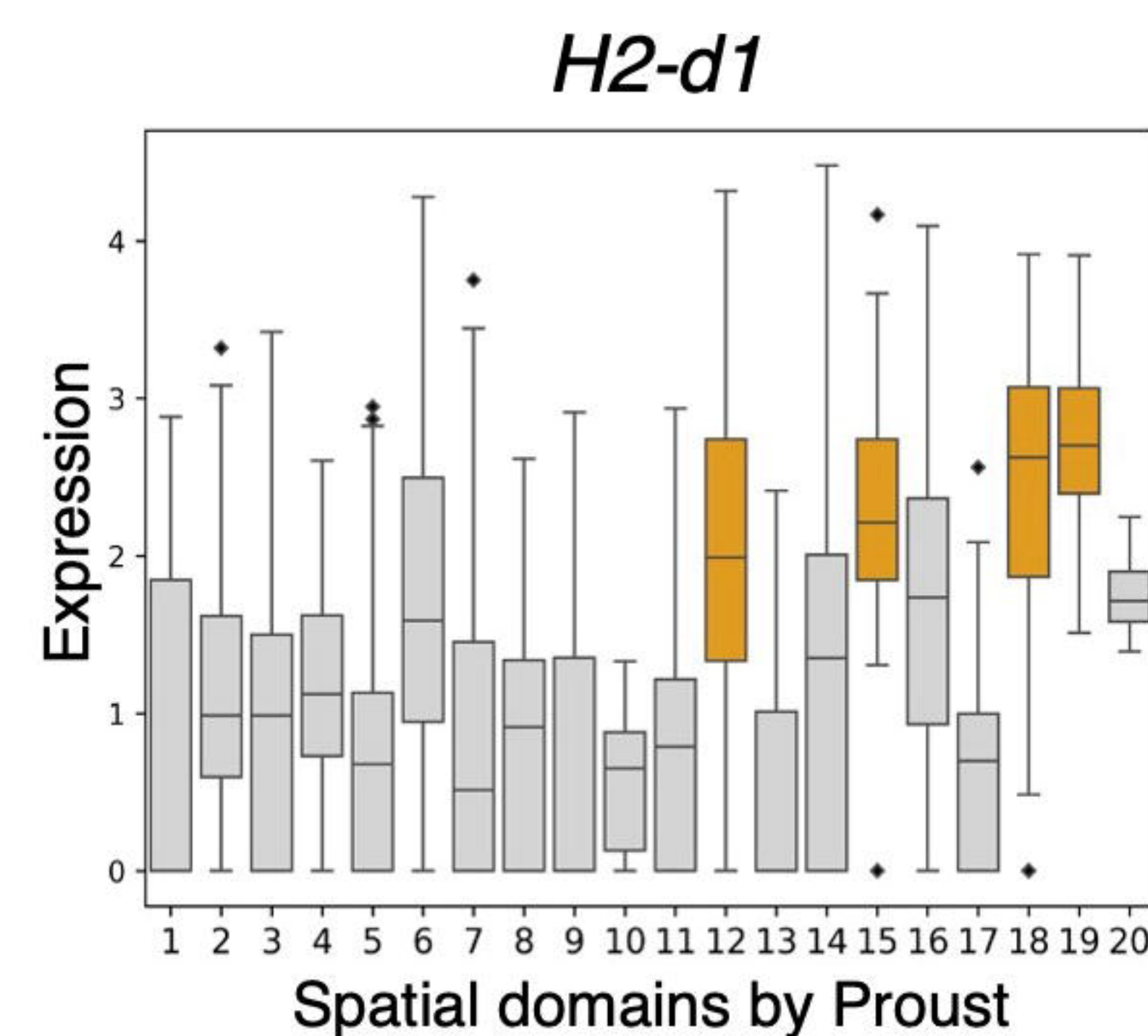


## Proust

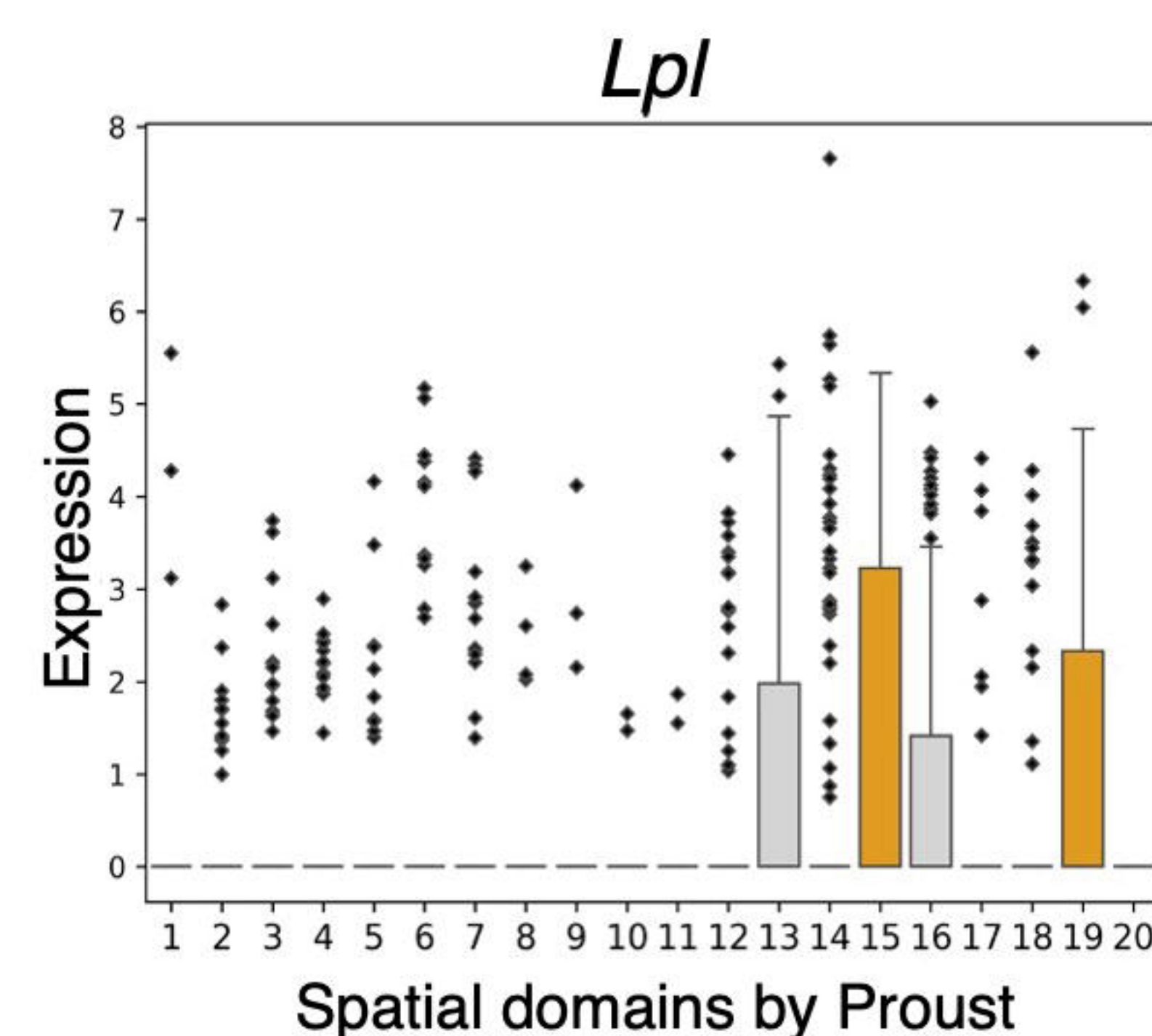


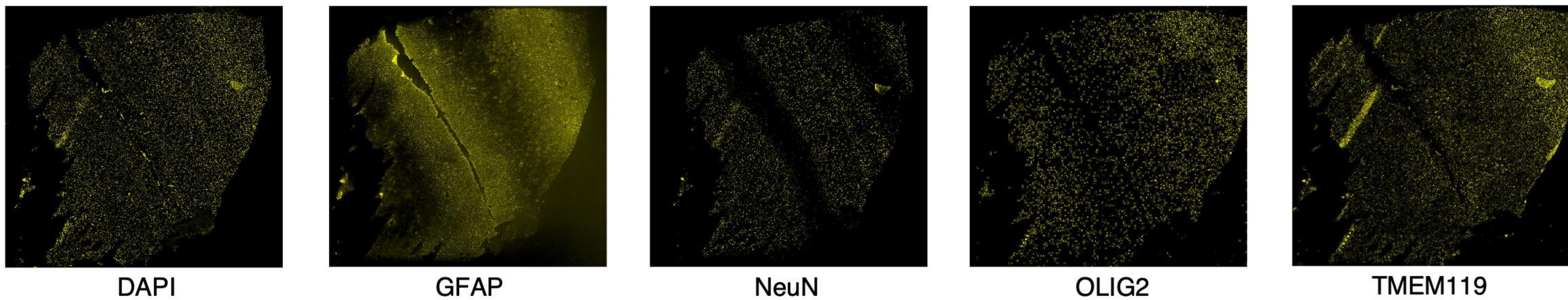
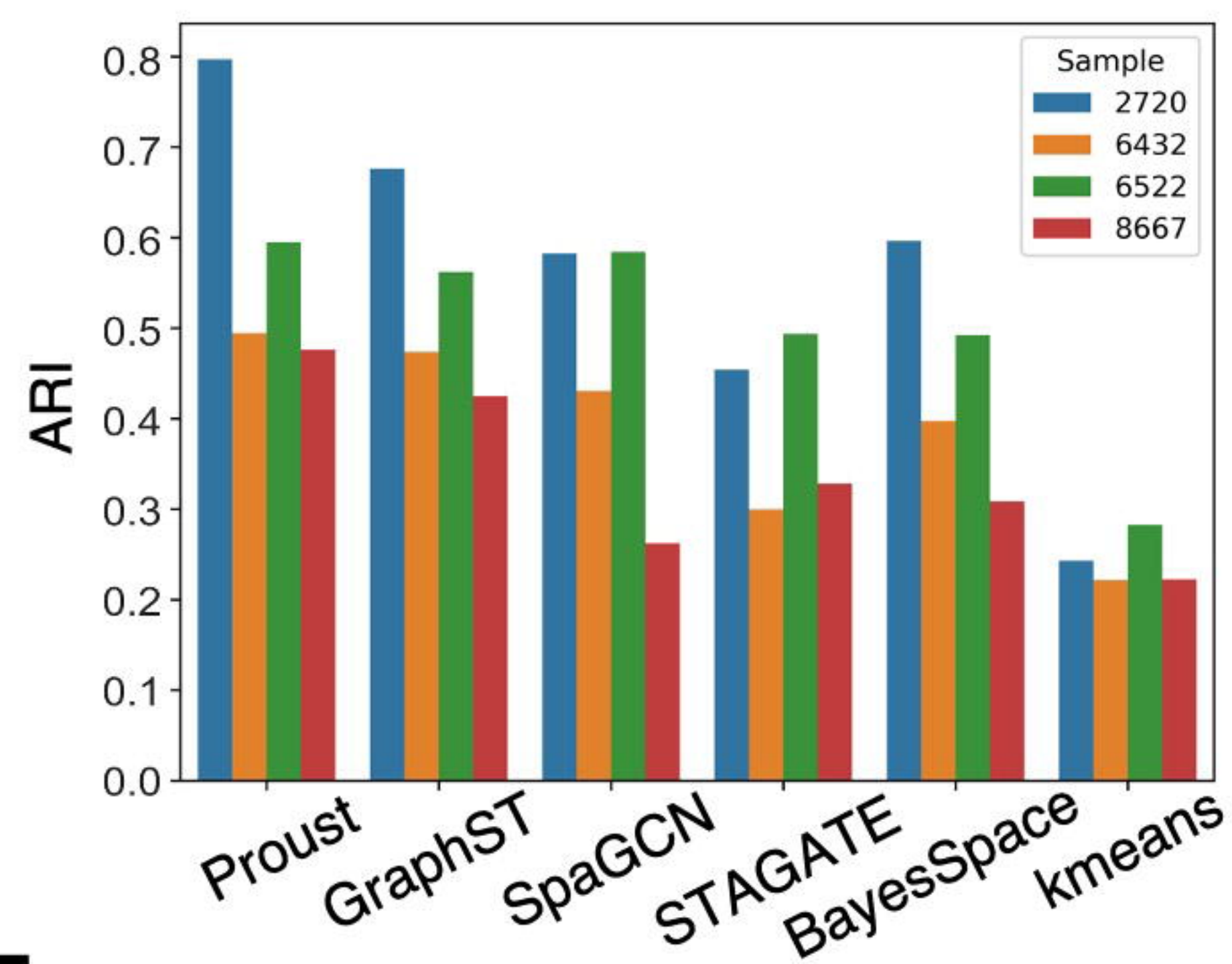
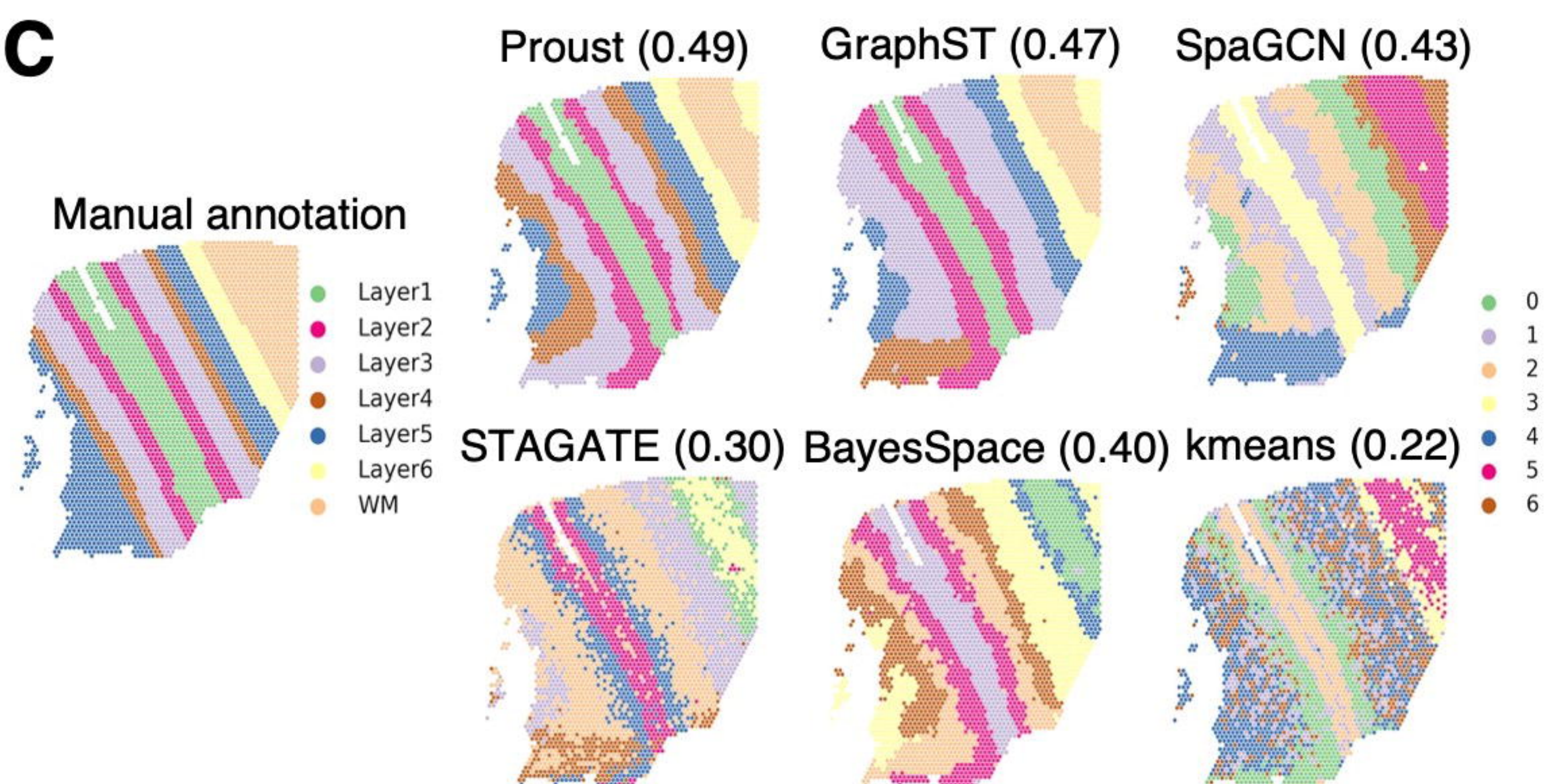
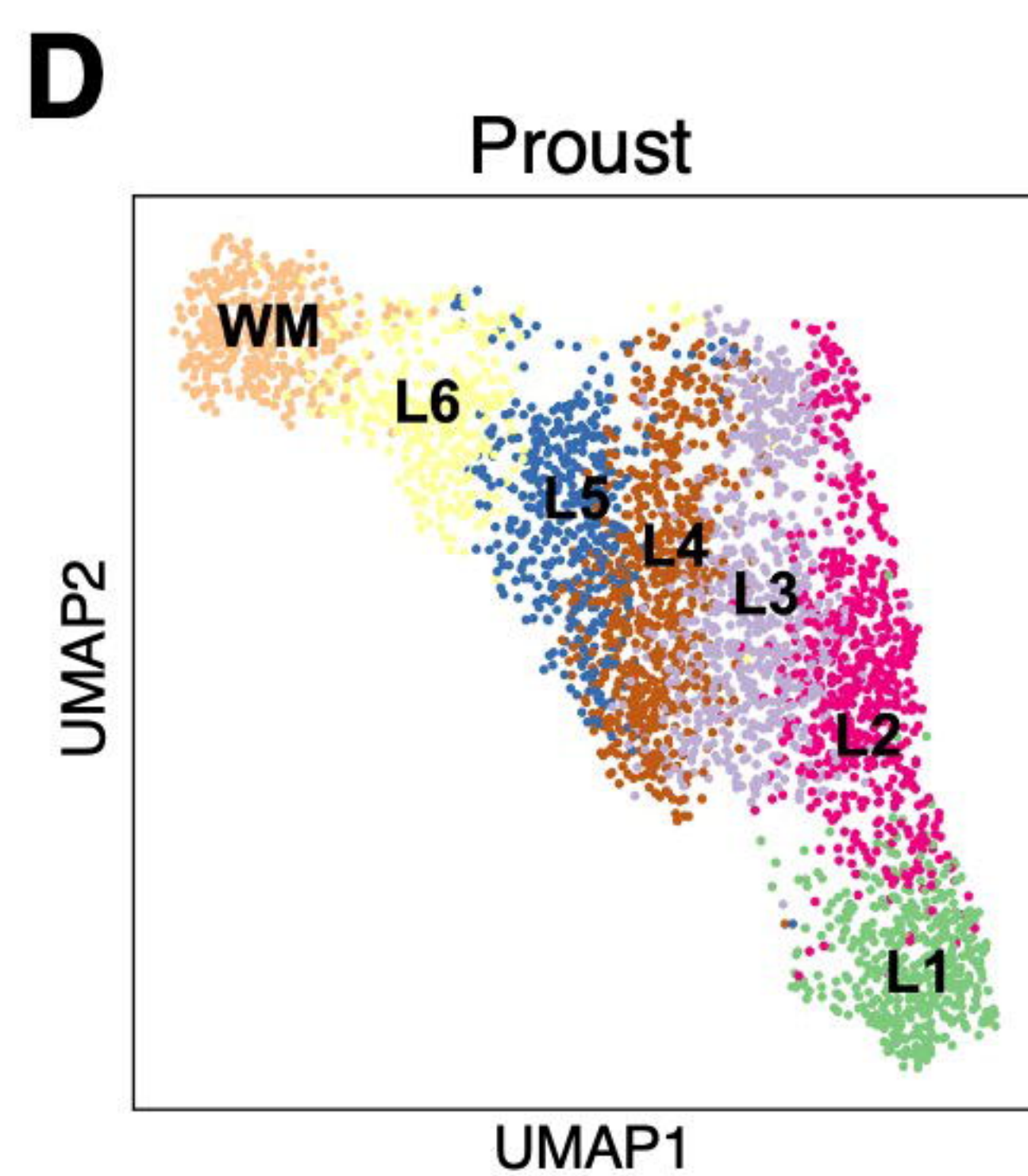
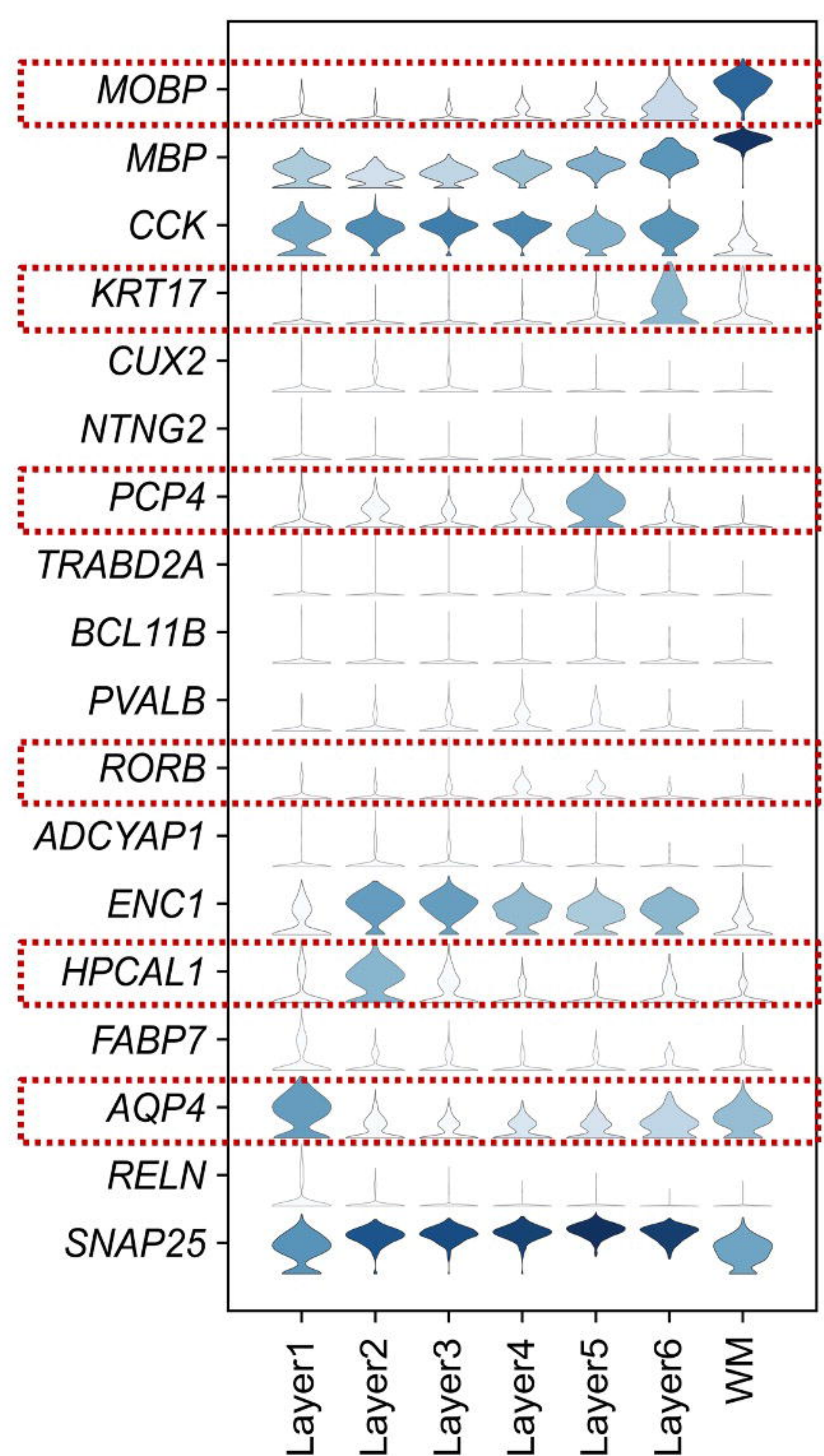
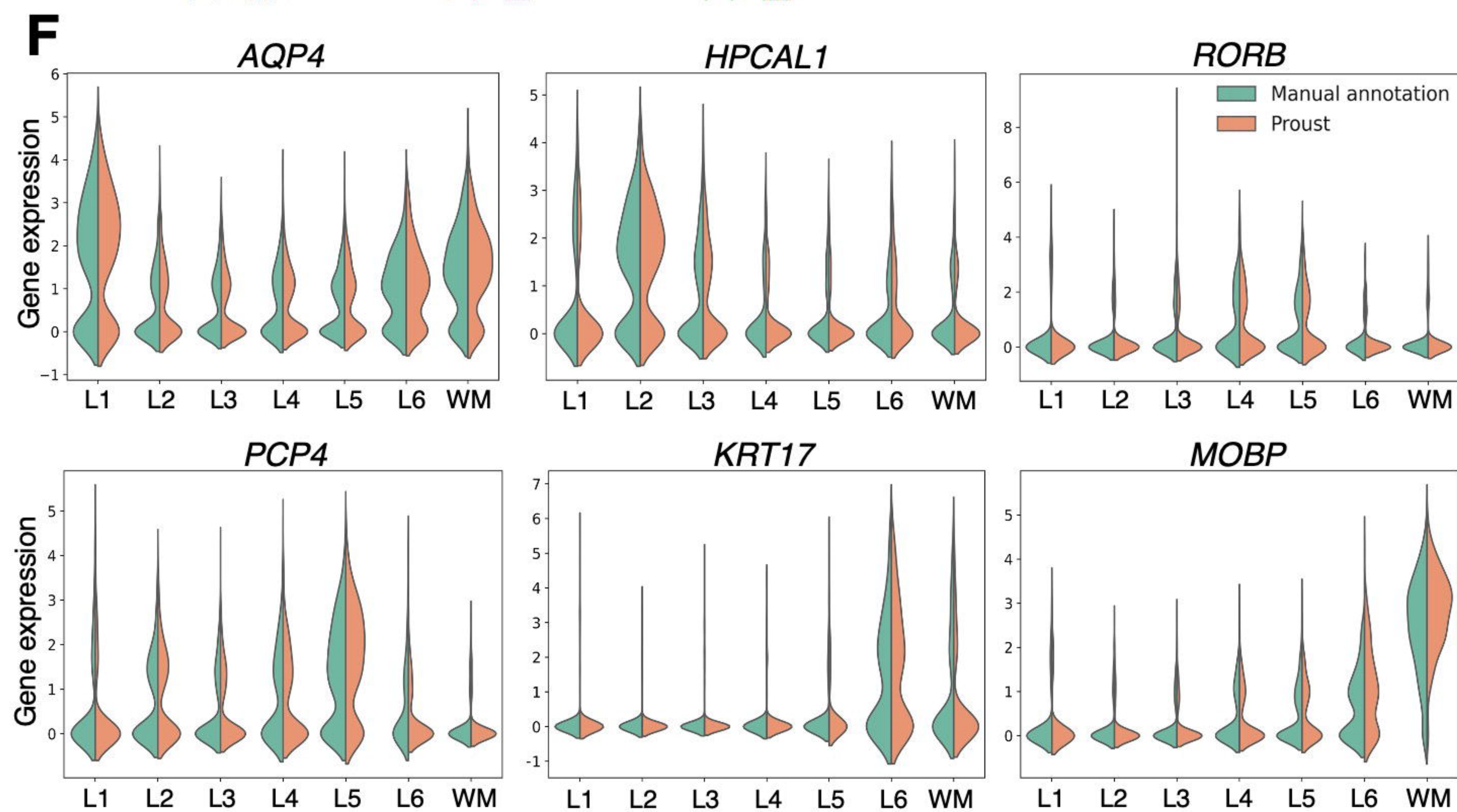
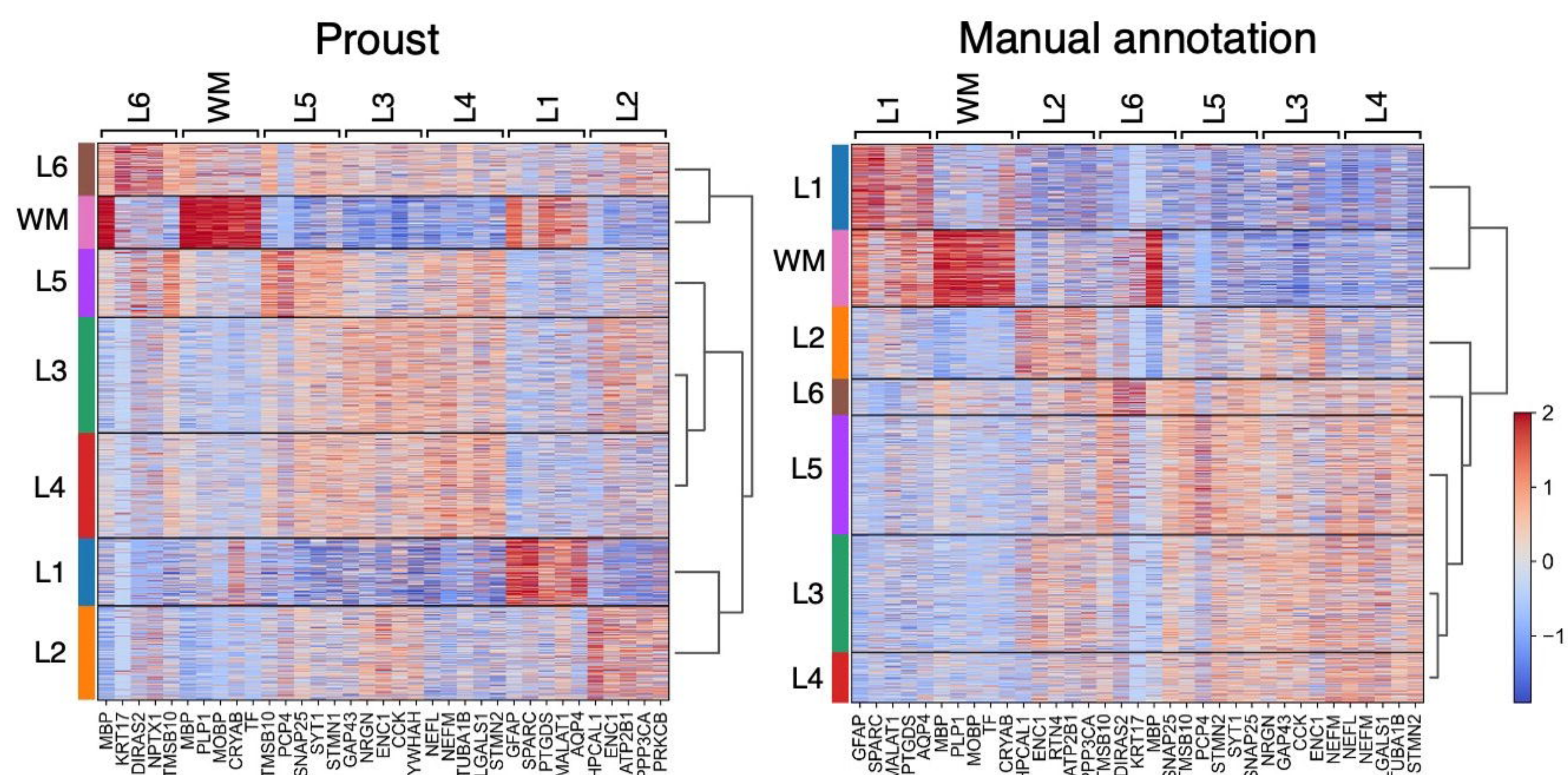
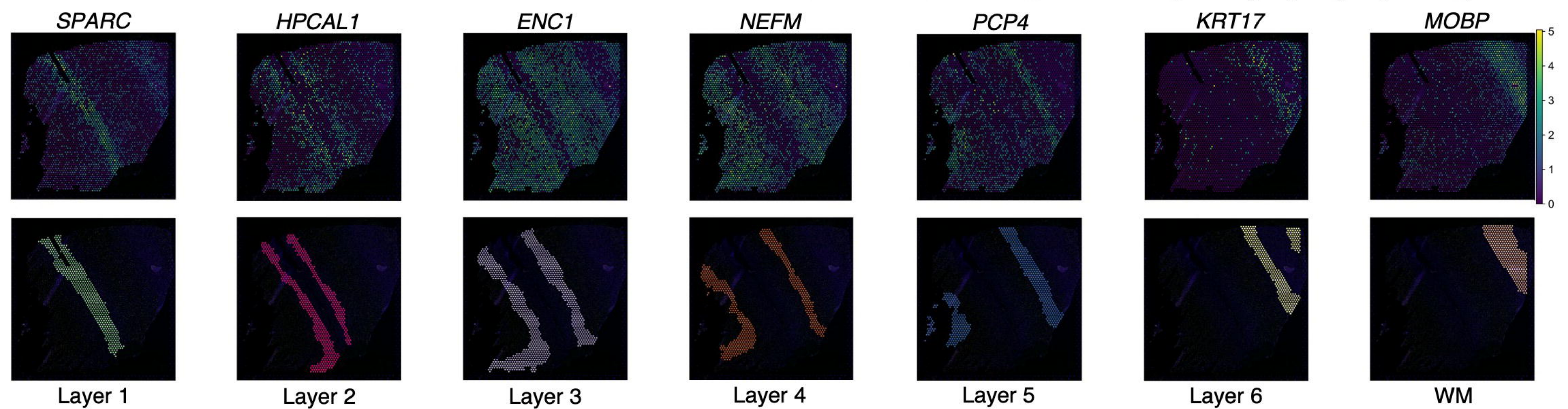
# STAGATE

*H2-d1*



*Lpl*



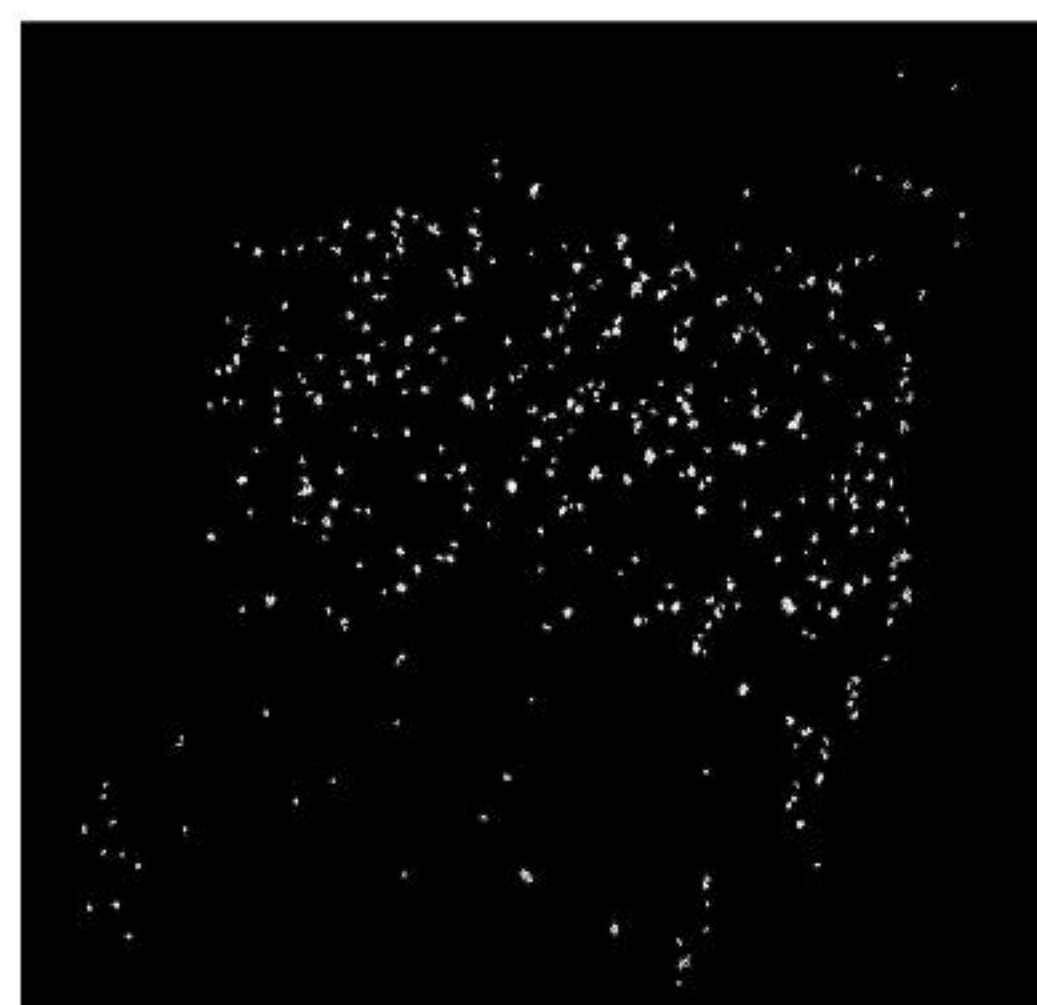
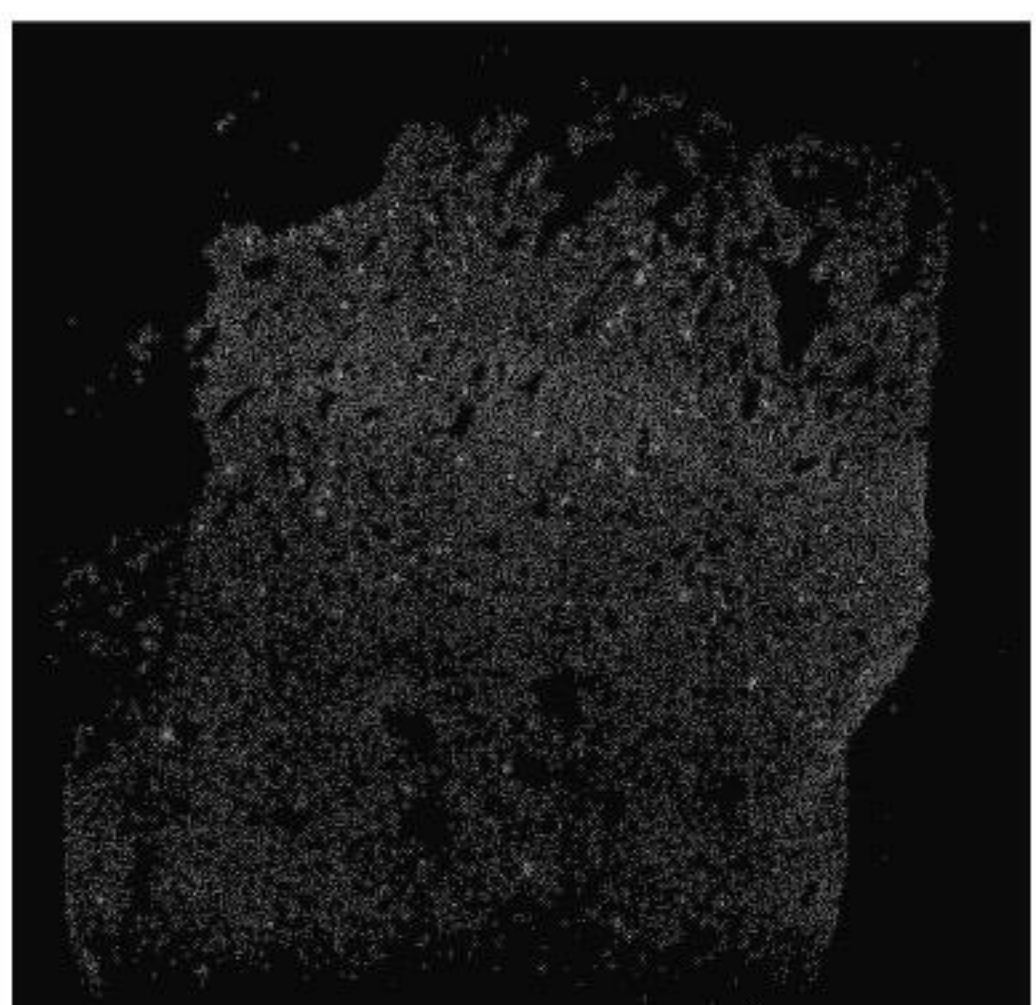
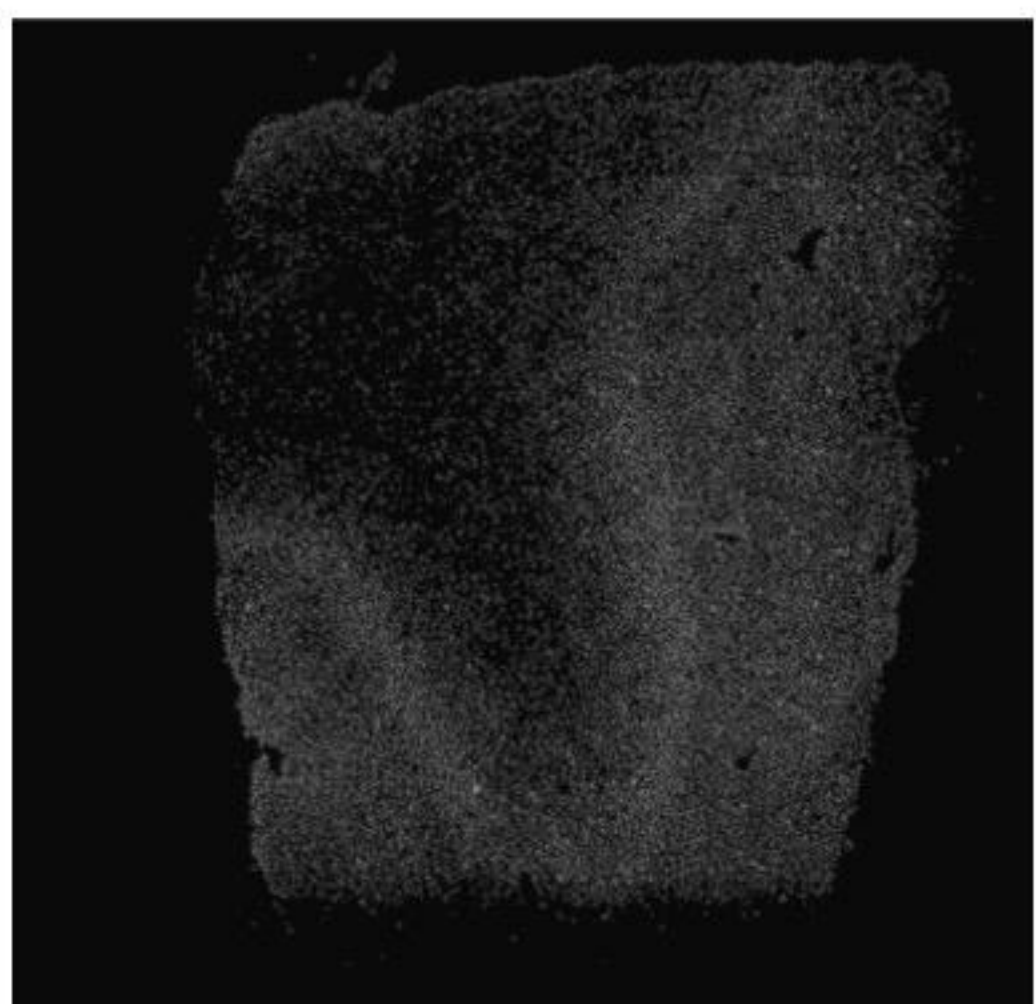
**A****B****C****D****E****F****G****H**

**A****A $\beta$** 

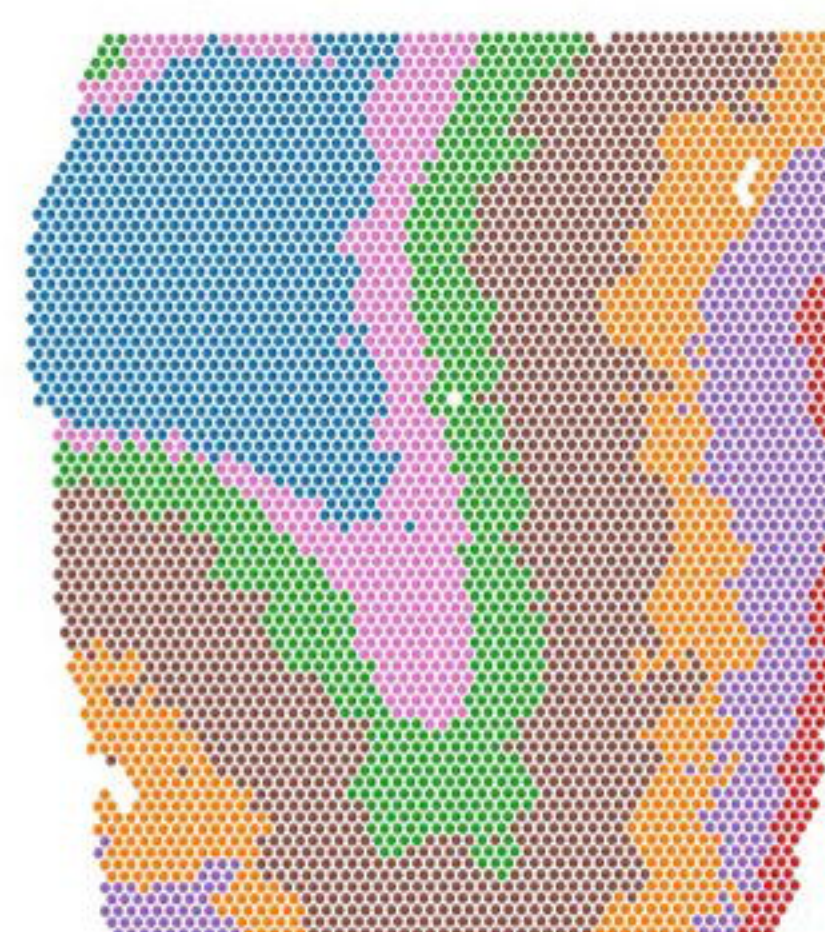
Br3880



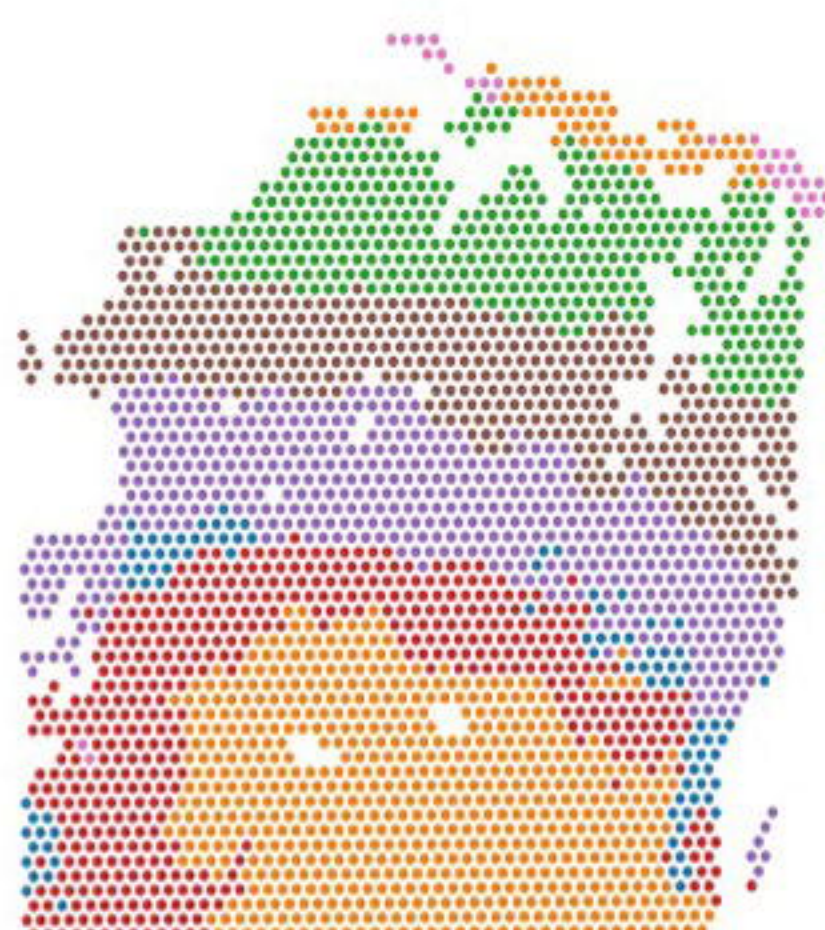
Br3854

**pTau****B**

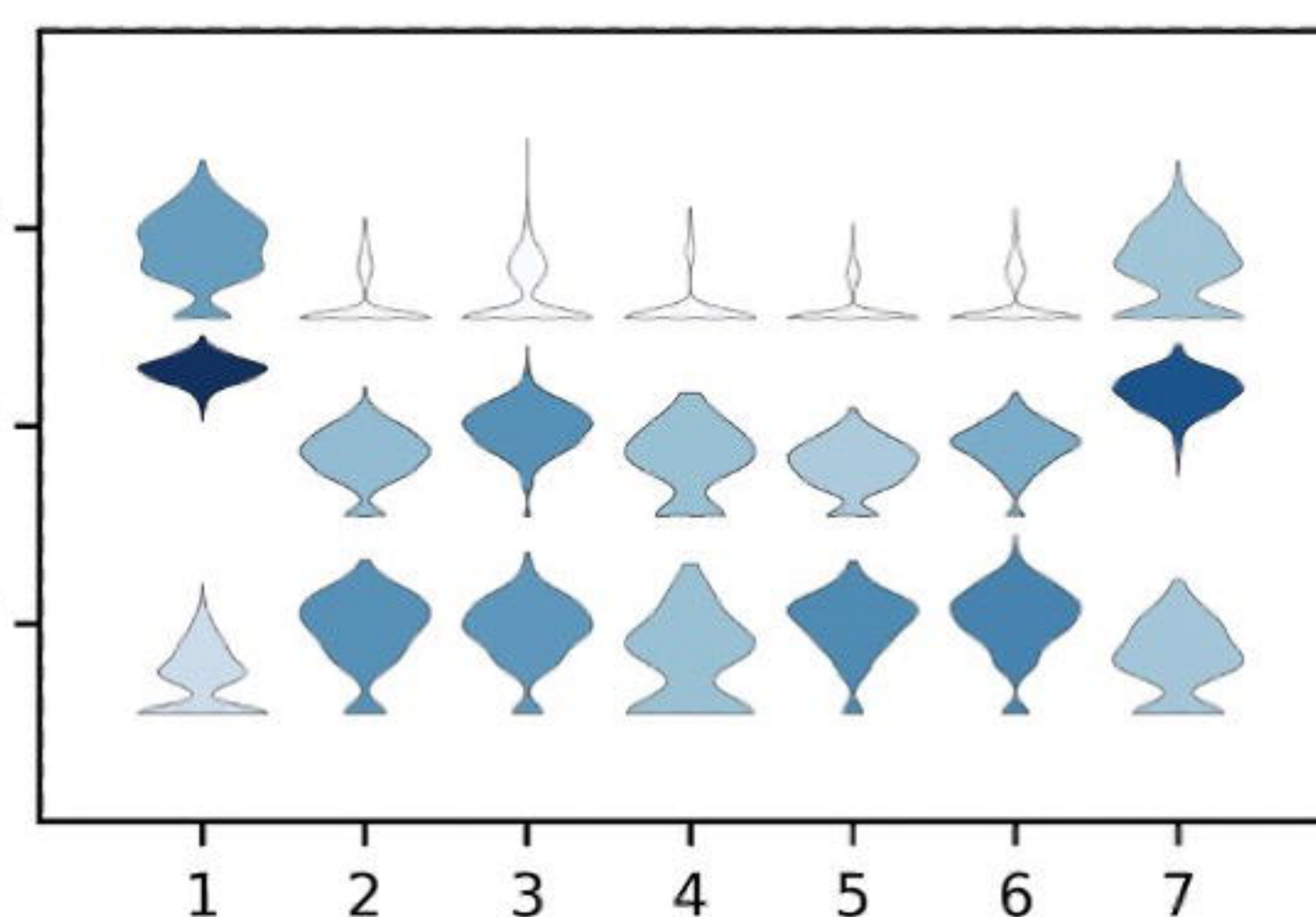
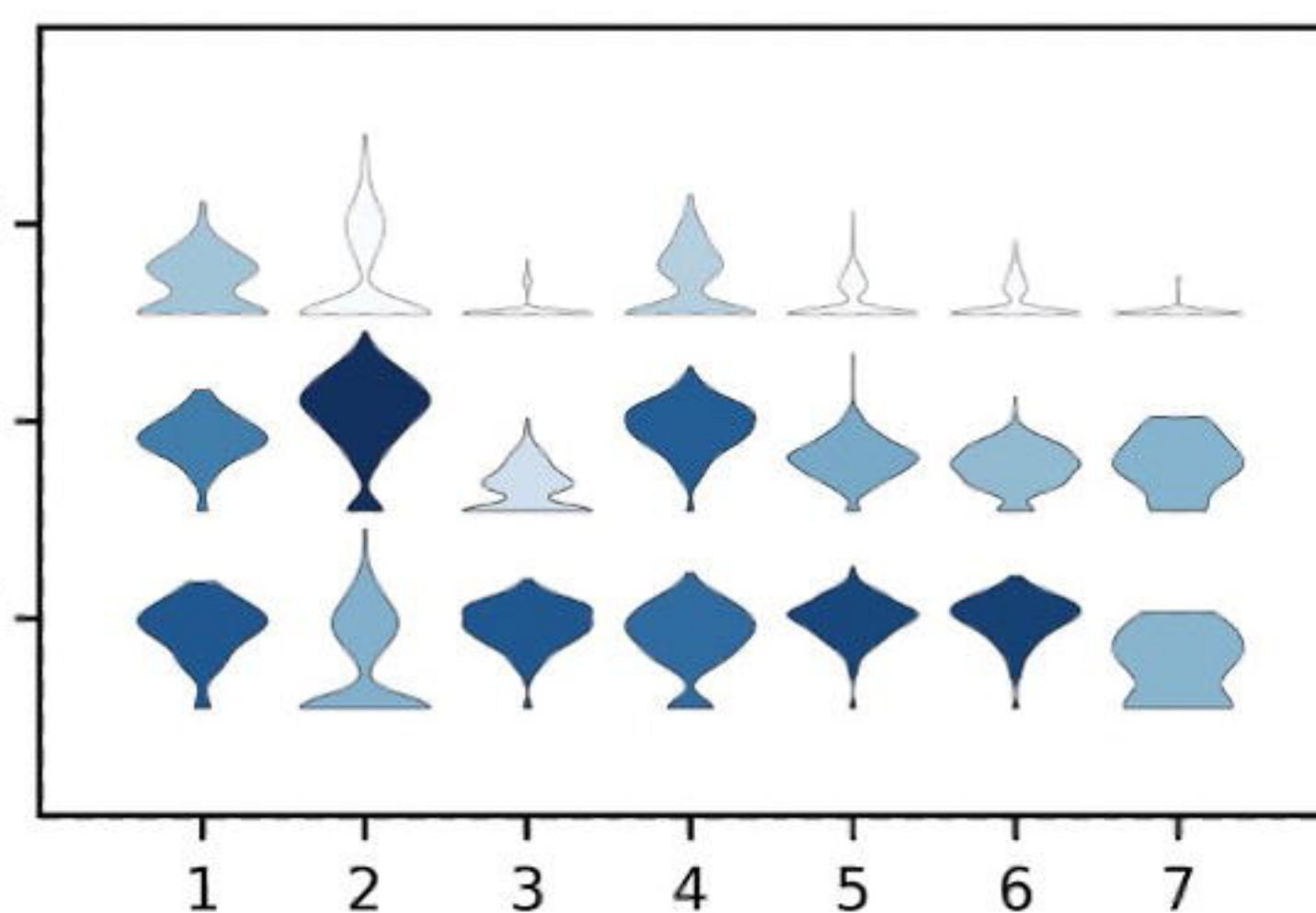
Cluster with 5 proteins



- 1
- 2
- 3
- 4
- 5
- 6
- 7

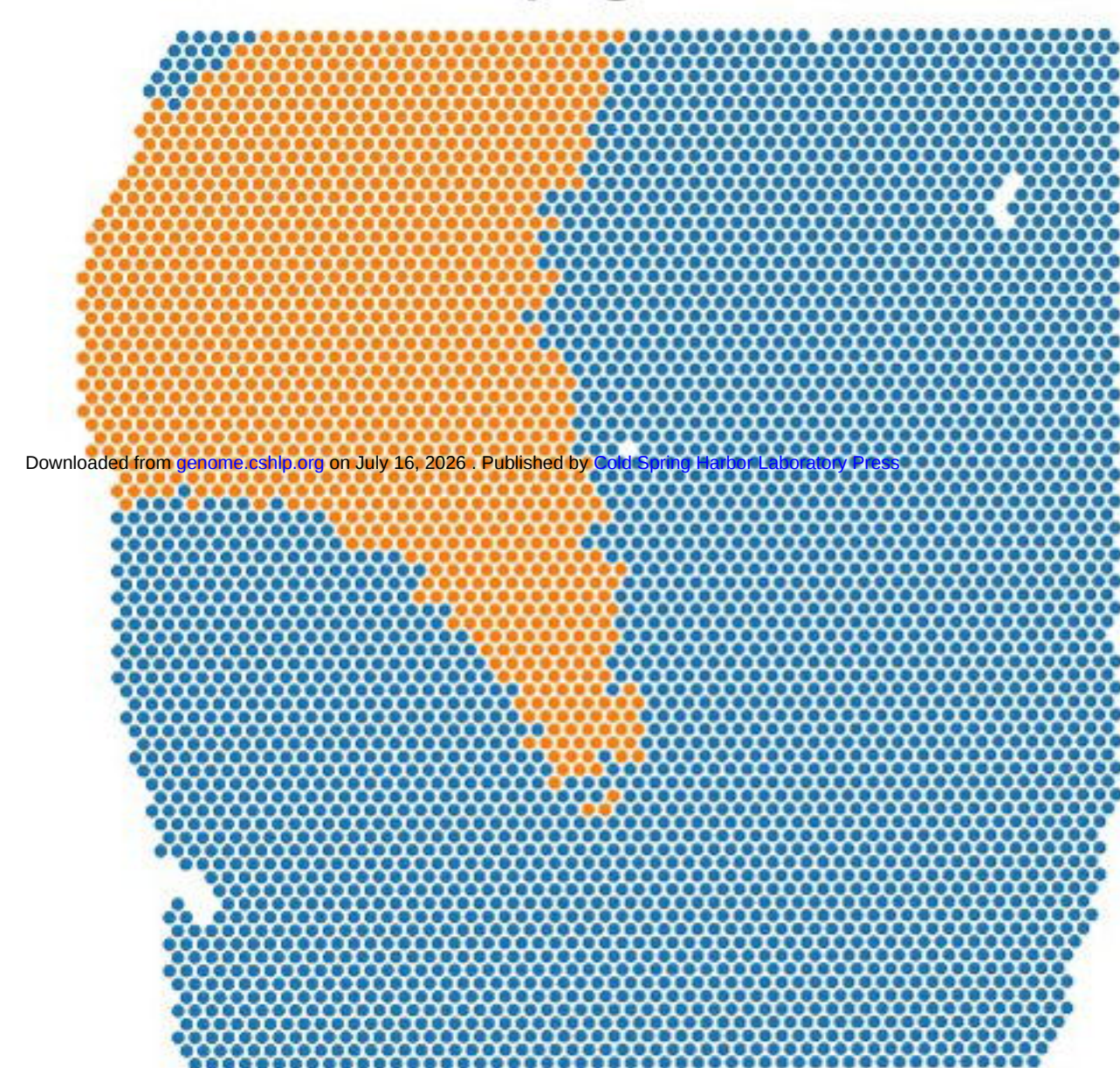


- 1
- 2
- 3
- 4
- 5
- 6
- 7

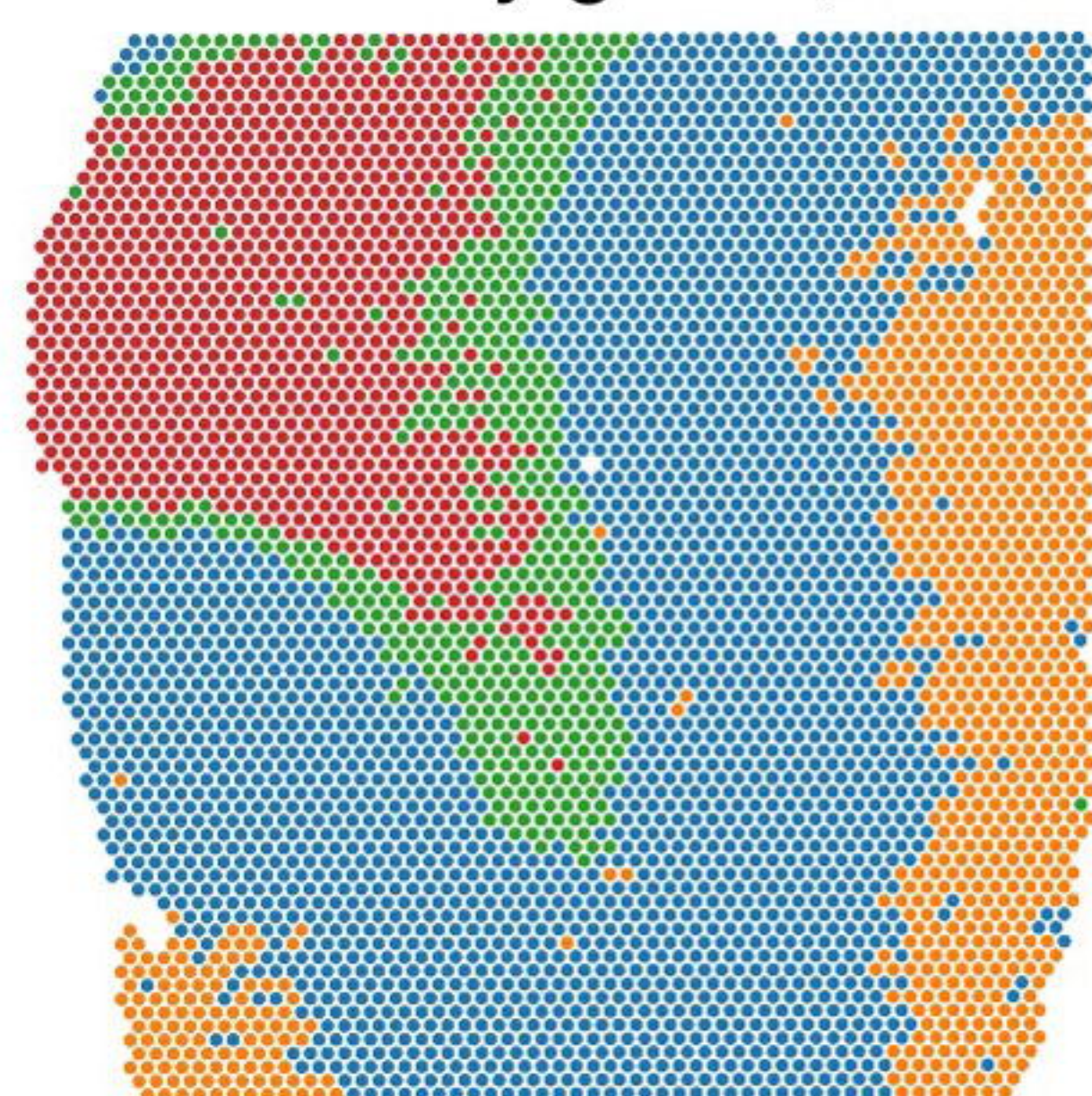
**C***MOBP*  
*MBP*  
*SNAP25*Median expression  
in group  
*MOBP*  
*MBP*  
*SNAP25*Median expression  
in group  
**D**

Cluster by genes, k = 2

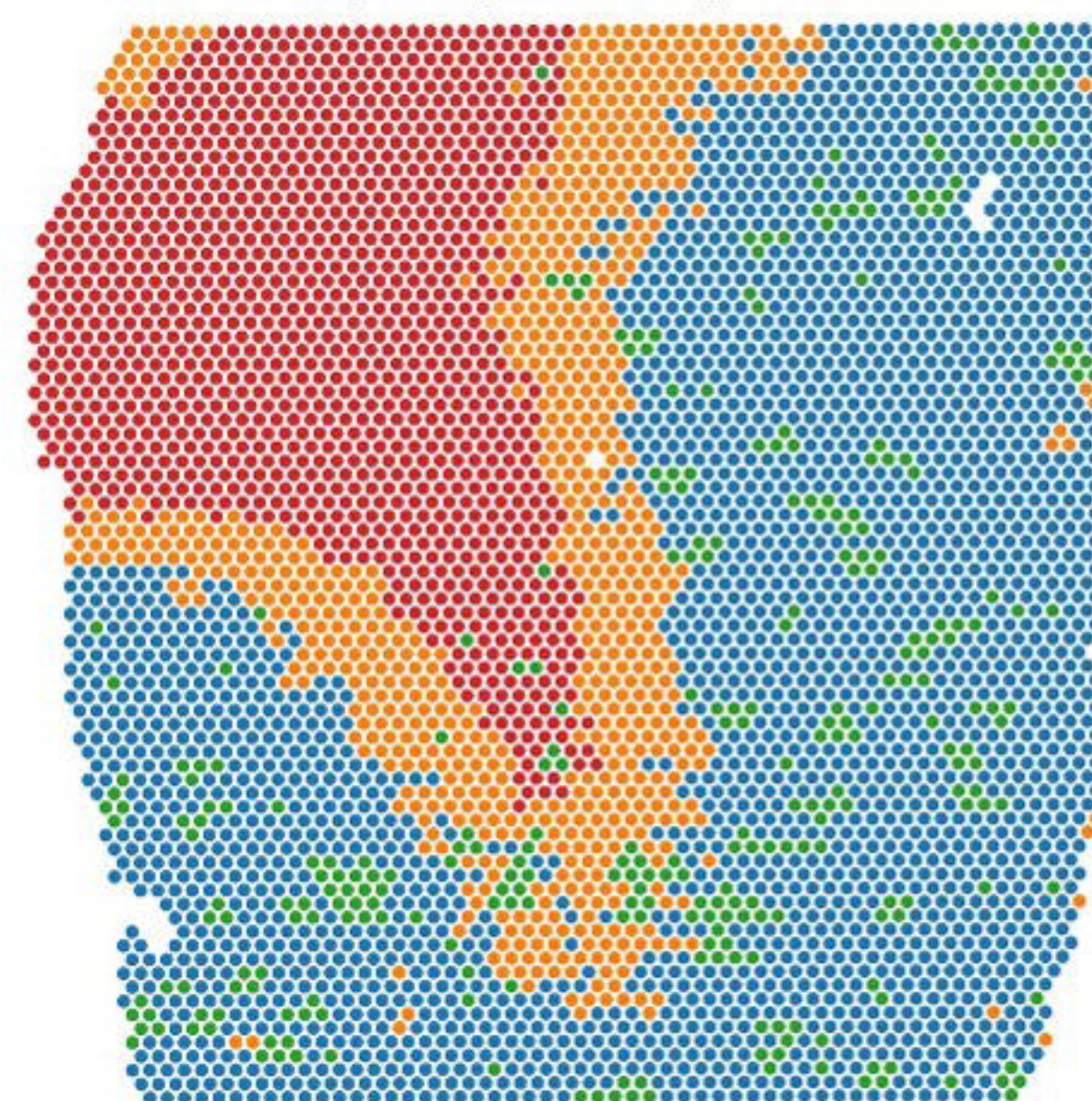
Br3880



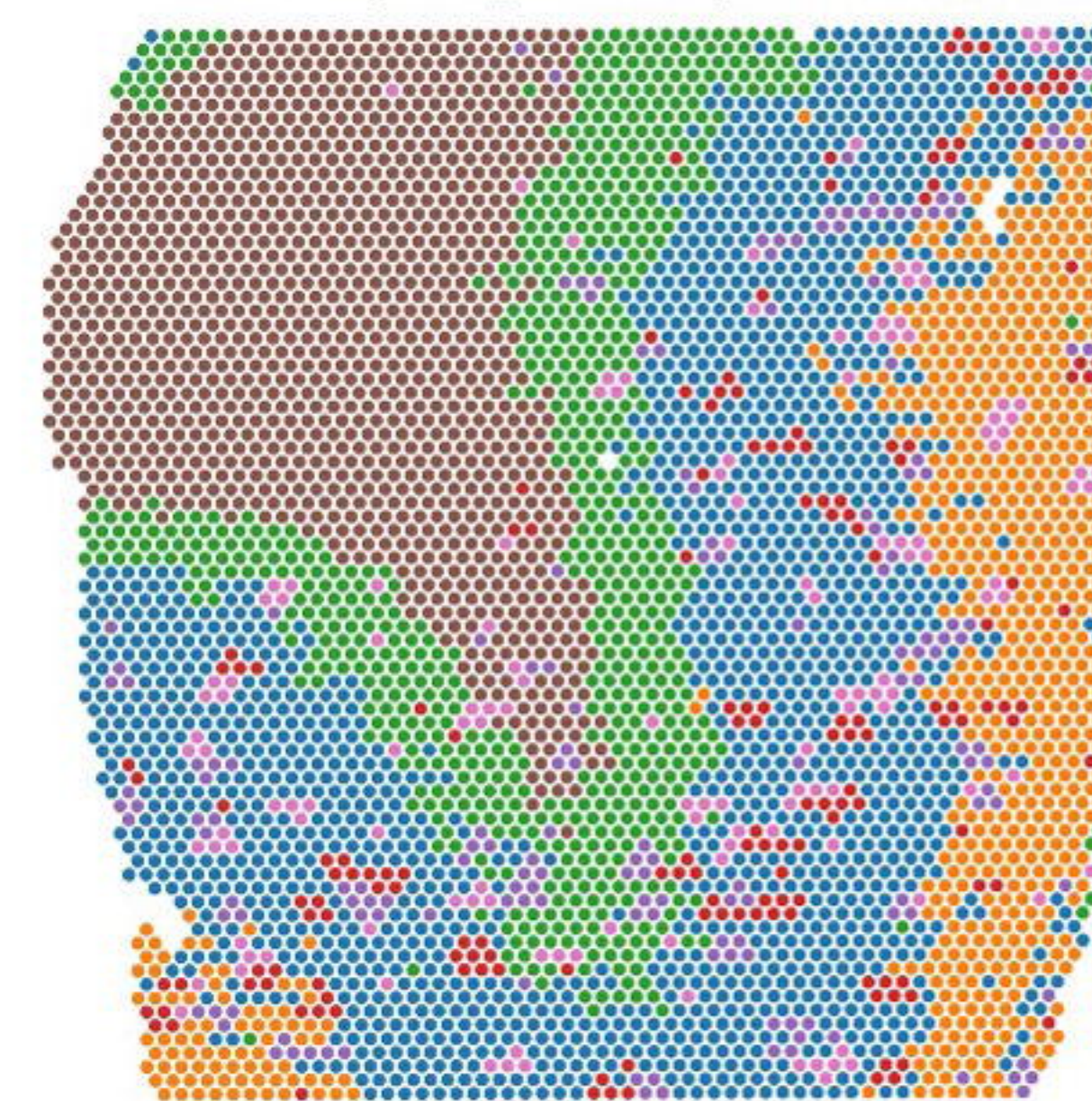
Cluster by genes, k = 4



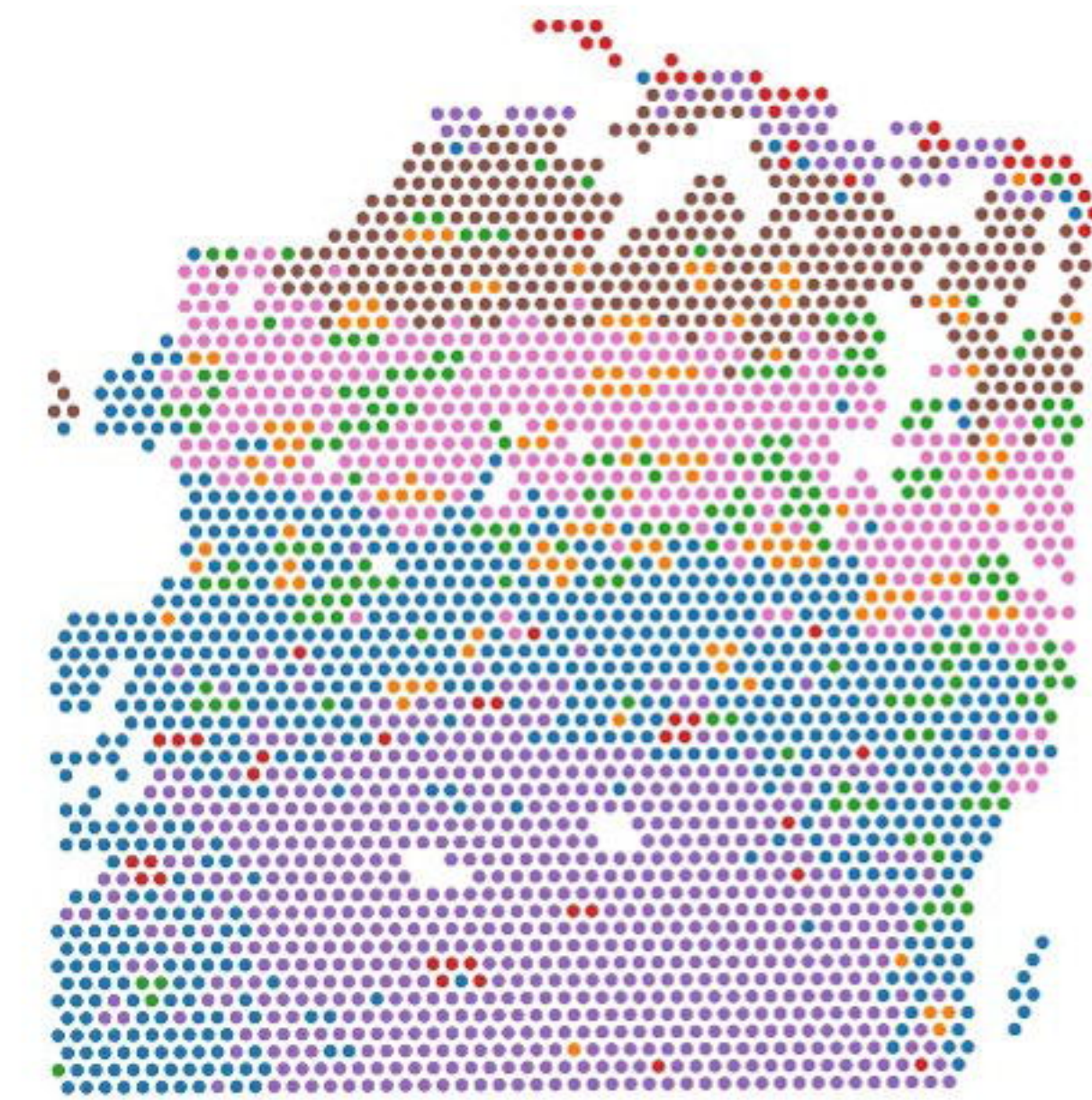
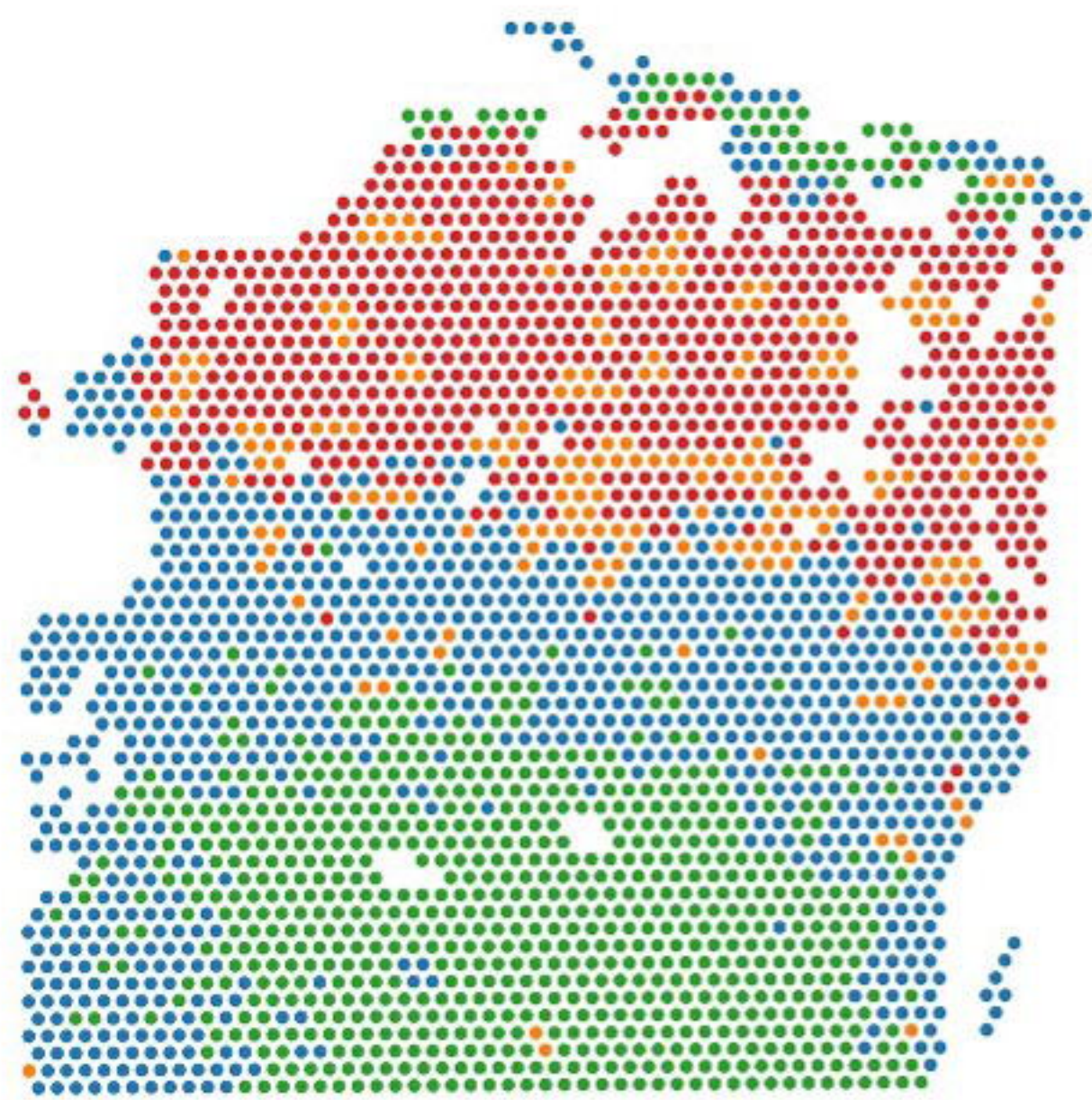
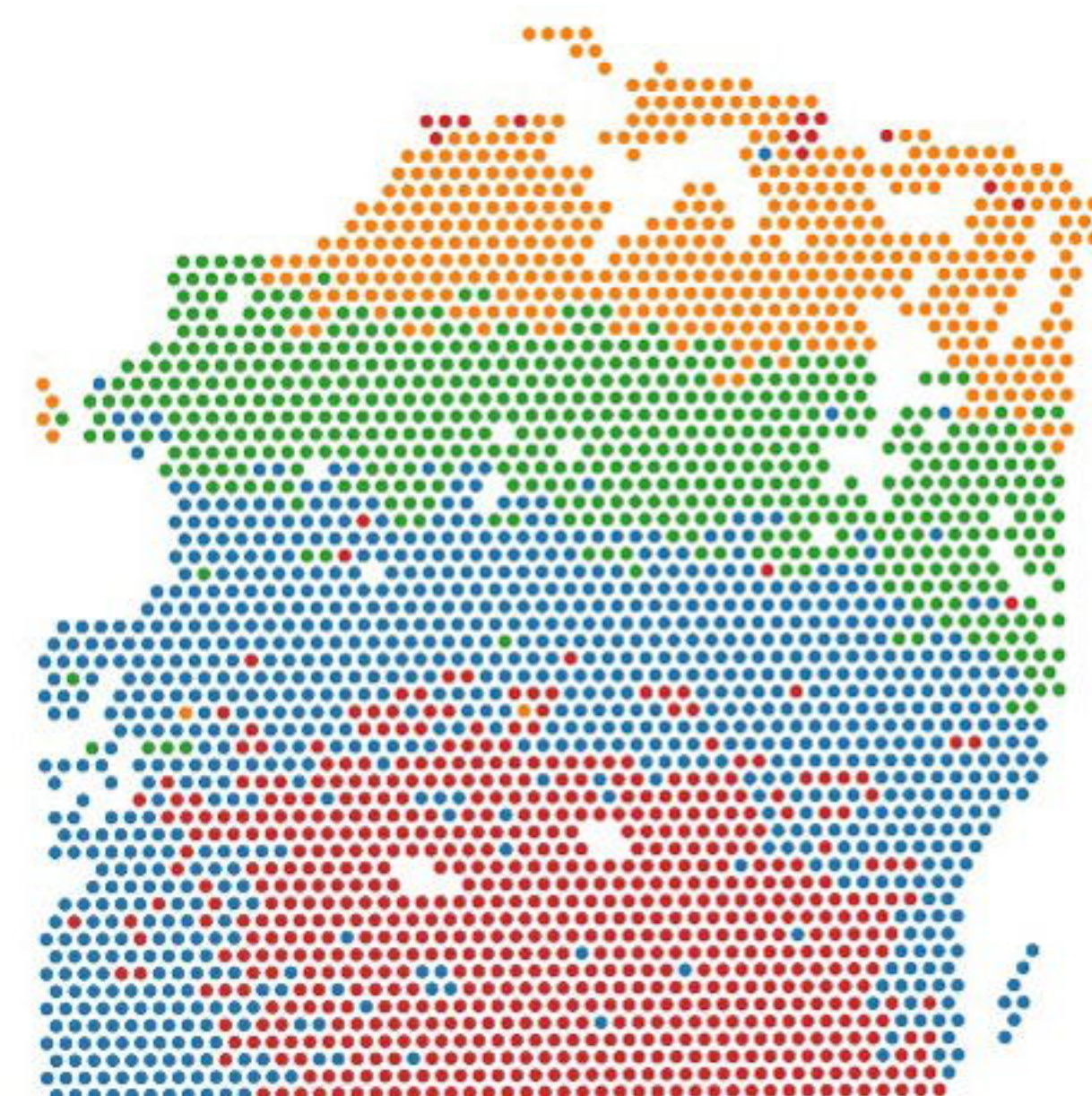
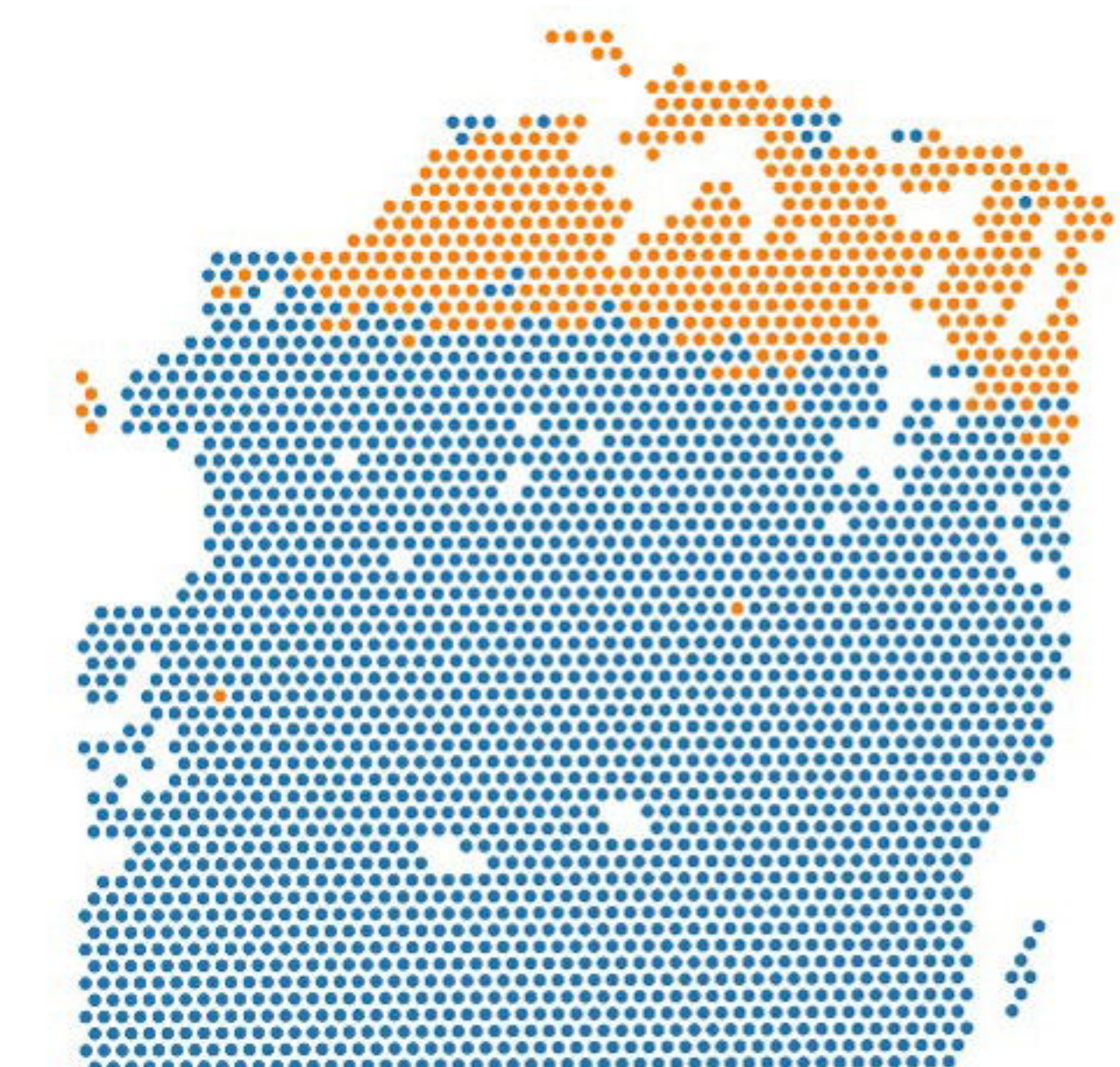
Cluster by hybrid profile, k = 4



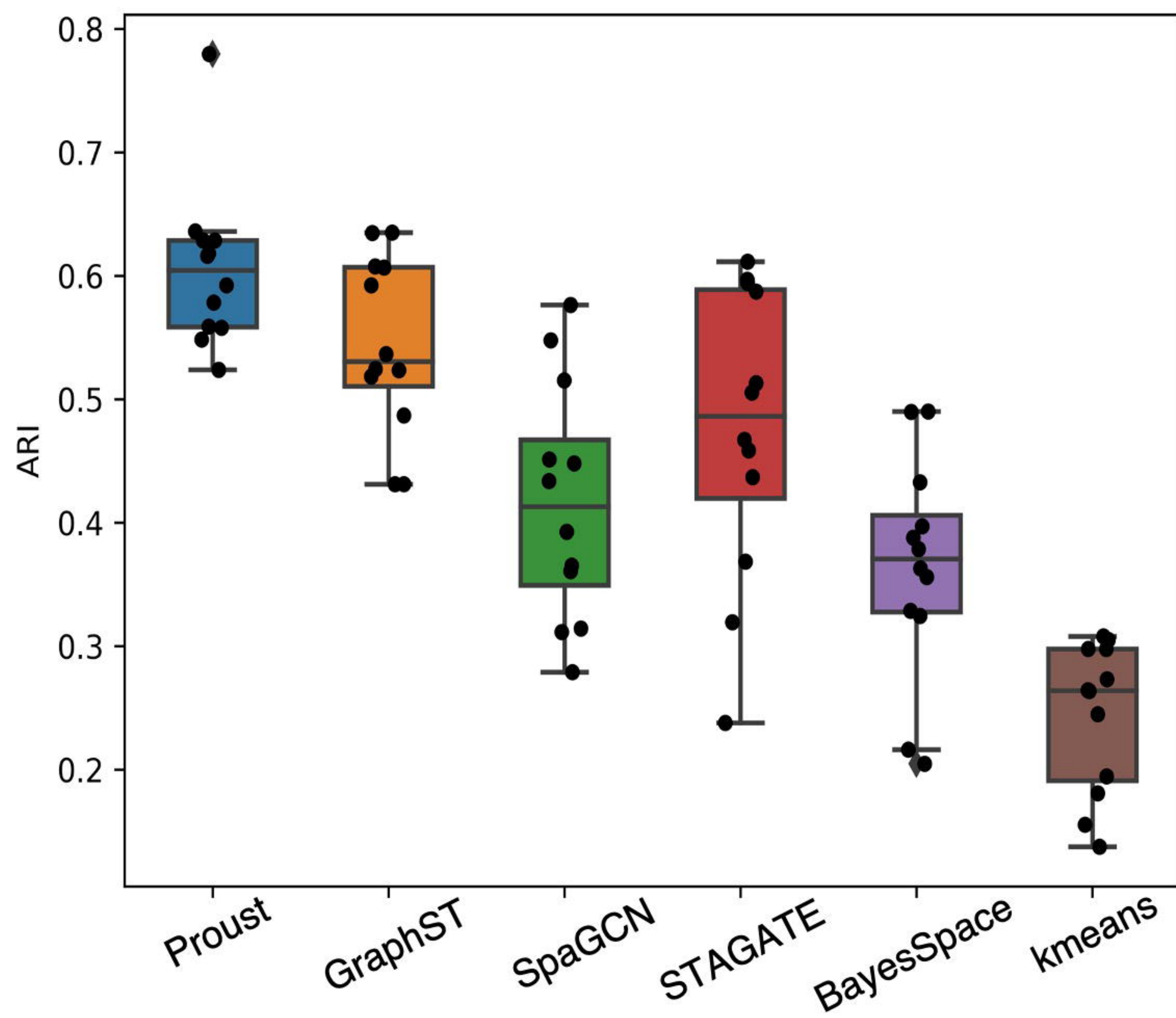
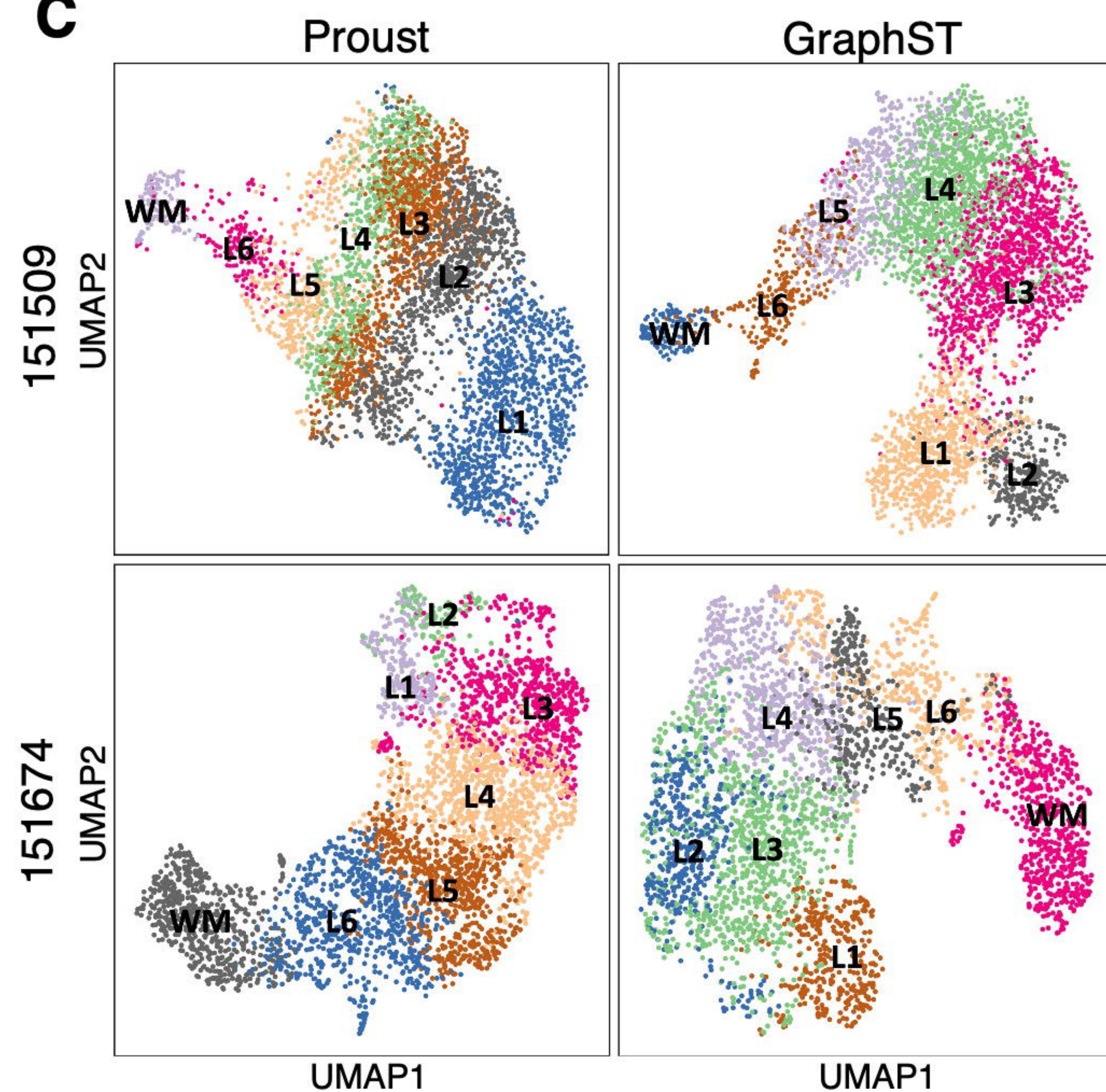
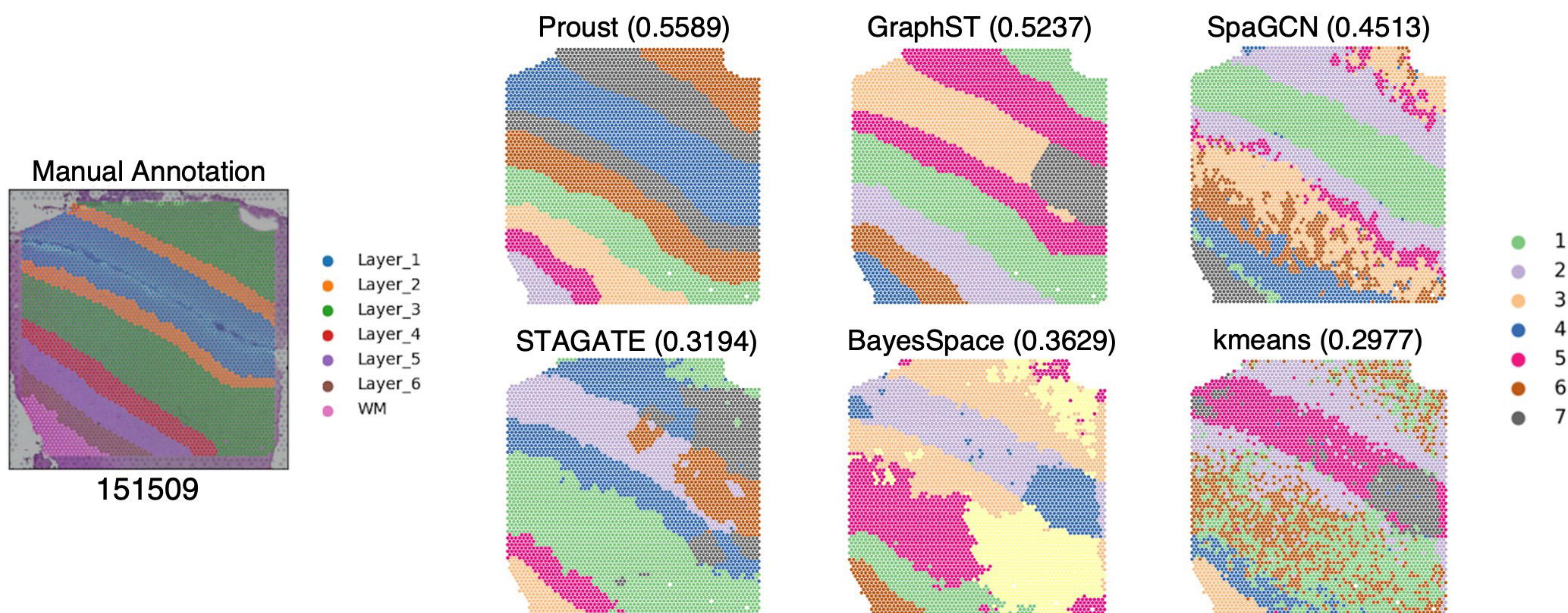
Cluster by hybrid profile, k = 7



Br3854



- 1
- 2
- 3
- 4
- 5
- 6
- 7

**A****C****B**

Downloaded from [genome.cshlp.org](https://genome.cshlp.org) on July 16, 2026. Published by Cold Spring Harbor Laboratory Press

