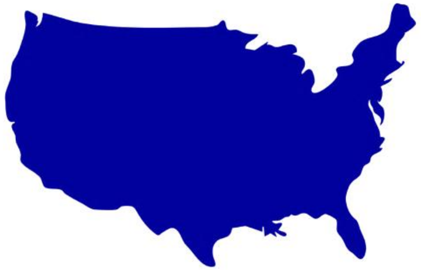
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

STAMarker: Determining spatial domain-specific variable genes with saliency maps in deep learning

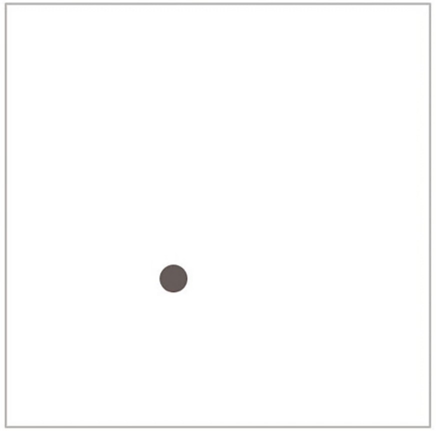
张驰浩
中国科学院数学与系统科学研究院

第三届数学生命科学大会
无锡 · 2023年10月21

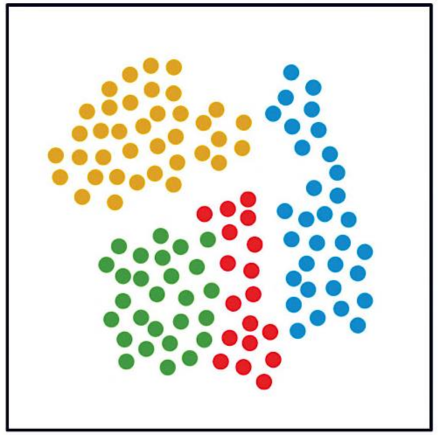
Bulk, Single Cells and Spatial Map



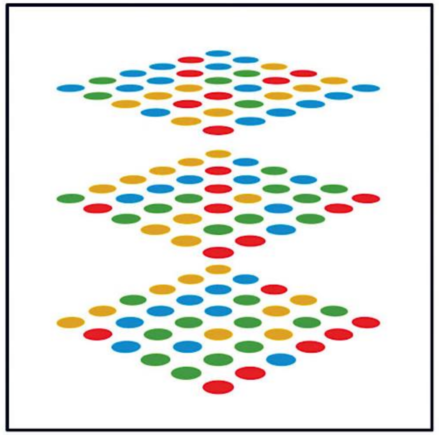
1st	New Jersey	14th	Oregon	27th	Texas	39th	New Mexico
2nd	Connecticut	15th	Louisiana	28th	Arkansas	40th	Maryland
3rd	Washington	16th	Missouri	29th	West Virginia	41st	Idaho
4th	California	17th	Alabama	30th	Arizona	42nd	Nevada
5th	Nebraska	18th	Florida	31st	Montana	43rd	North Dakota
6th	Mississippi	19th	Kentucky	32nd	Alaska	44th	Michigan
7th	Iowa	20th	Pennsylvania	33rd	Colorado	45th	South Carolina
8th	Tennessee	21st	North Carolina	34th	Ohio	46th	Maine
9th	Rhode Island	22nd	Delaware	35th	New Hampshire	47th	Virginia
10th	Kansas	23rd	Indiana	36th	Utah	48th	Wyoming
11th	Illinois	24th	Wisconsin	37th	New York	49th	South Dakota
12th	Massachusetts	25th	Minnesota	38th	Oklahoma	50th	Georgia
13th	Hawaii	26th	Vermont				



Bulk



Single Cells

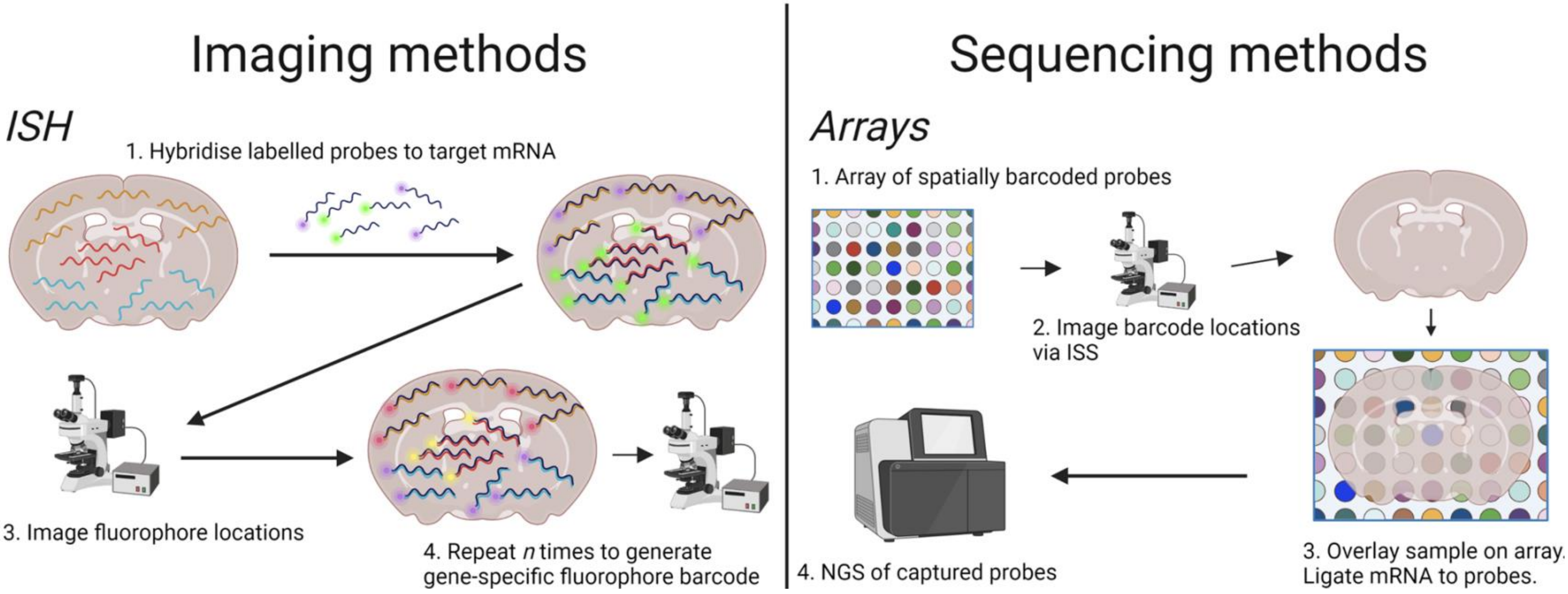


Spatial Map

The spatial organization of cells in multicellular systems is crucial for their proper function.

www.10xgenomics.com

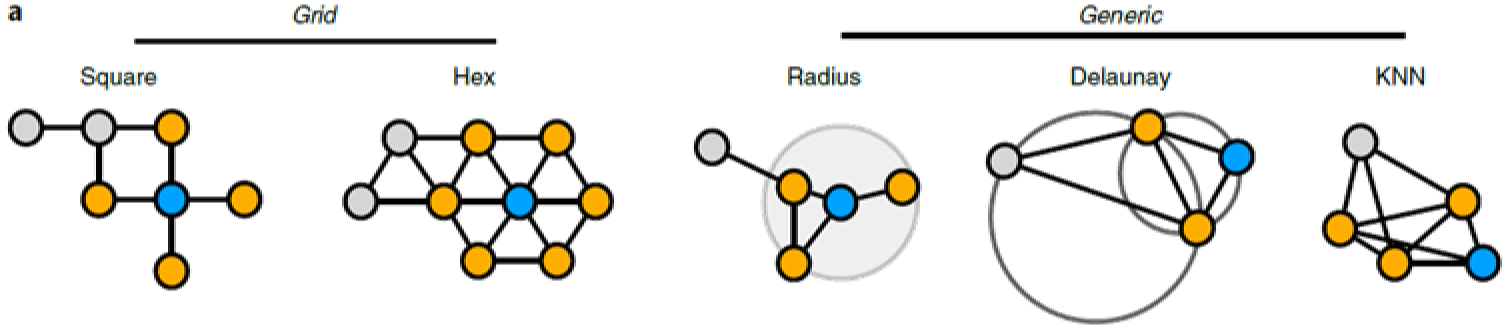
Overview of ST sequencing technology



Cameron G. Williams et. al., *Genome Medicine'2023*

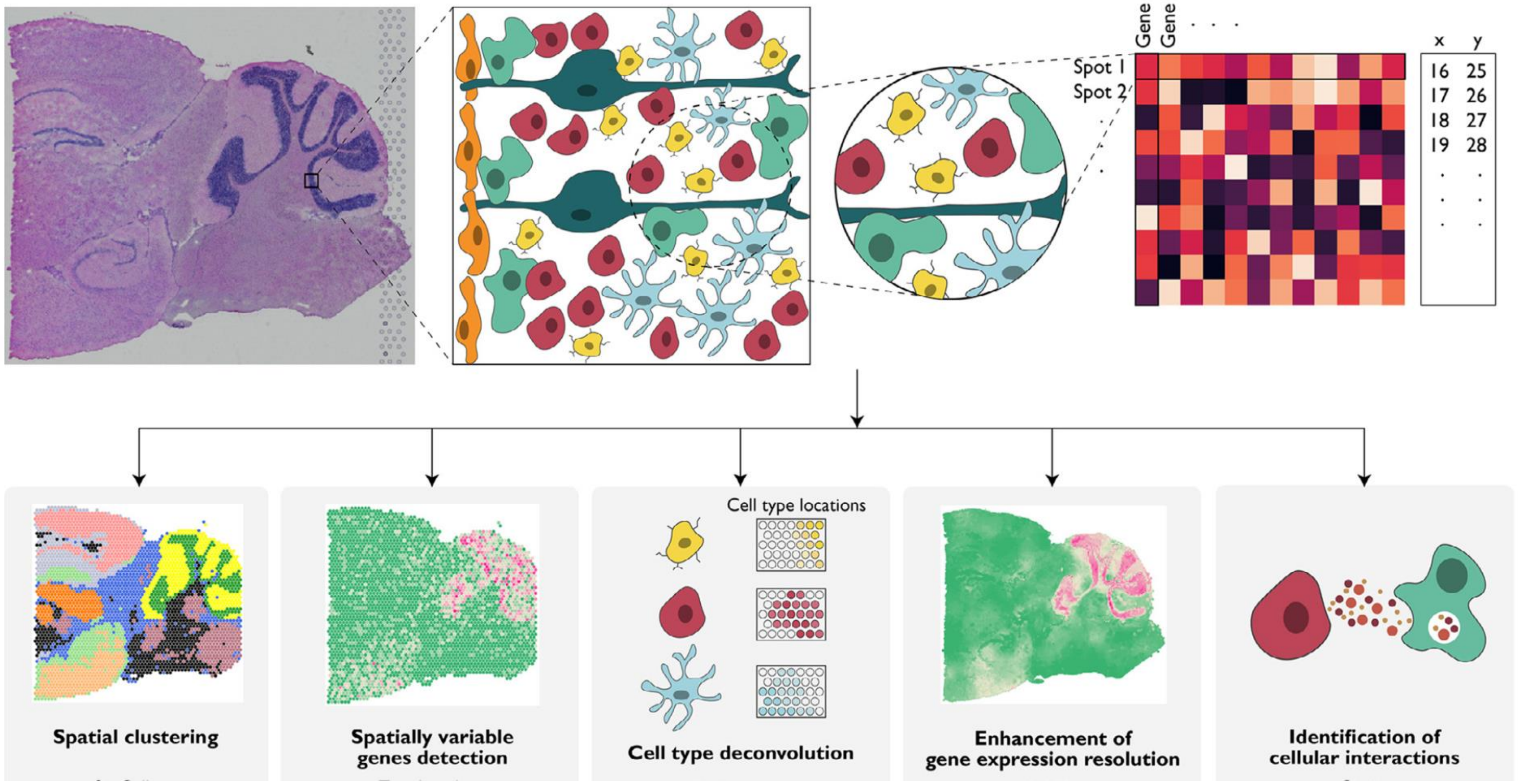
A glimpse of the ST sequencing technologies

	Spatial resolution	Point address	Spot configuration	Multiomics
ST	100µm	Spot	Square	No
Slide-seq	10µm	Bead	Random	No
HDST	2µm	Bead	Regular hexagon	No
10x Visium	55µm	Spot	Regular hexagon	No
DBiT-seq	10µm	Pixel	Square	Yes, with proteomics
Slide-seq V2	10µm	Bead	Random	No
Stereo-seq	220nm	NDB	Square	No
Seq-Scope	0.5–0.8µm	Pixel	Irregular	No
PIXEL-seq	≤1 µm	Pixel	Irregular	No



Palla et al., *Nature Methods* 2022

Computational problems of ST

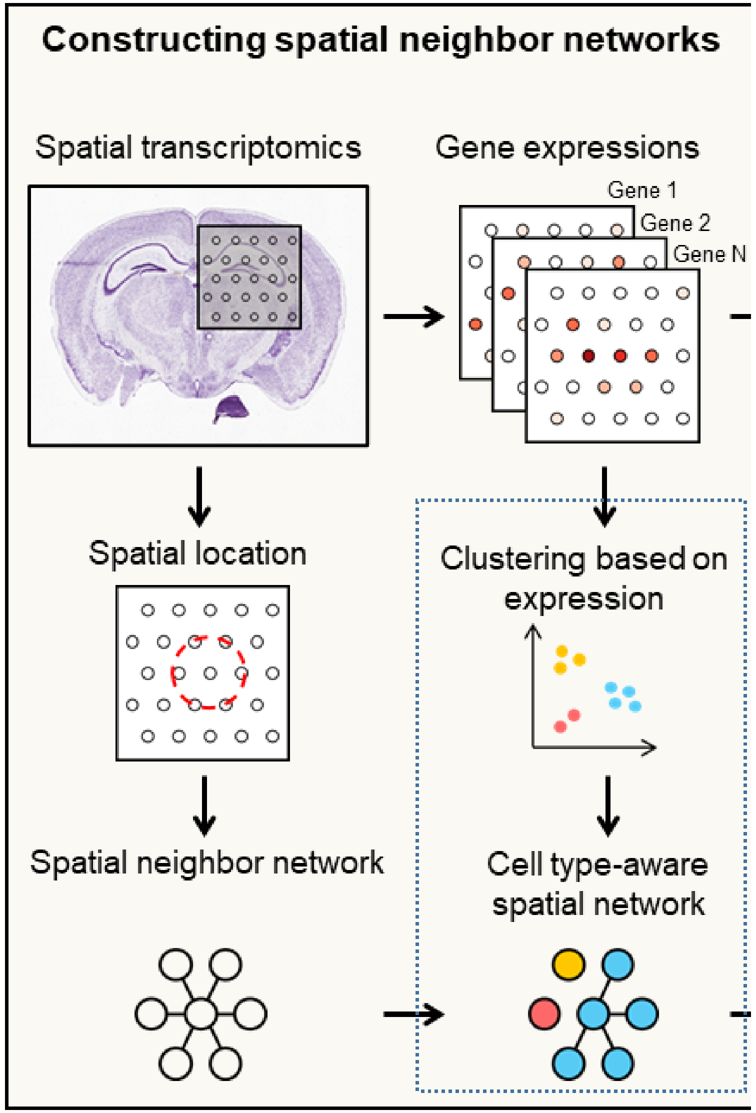


……

Larsson et al., *Nature Methods* 2021

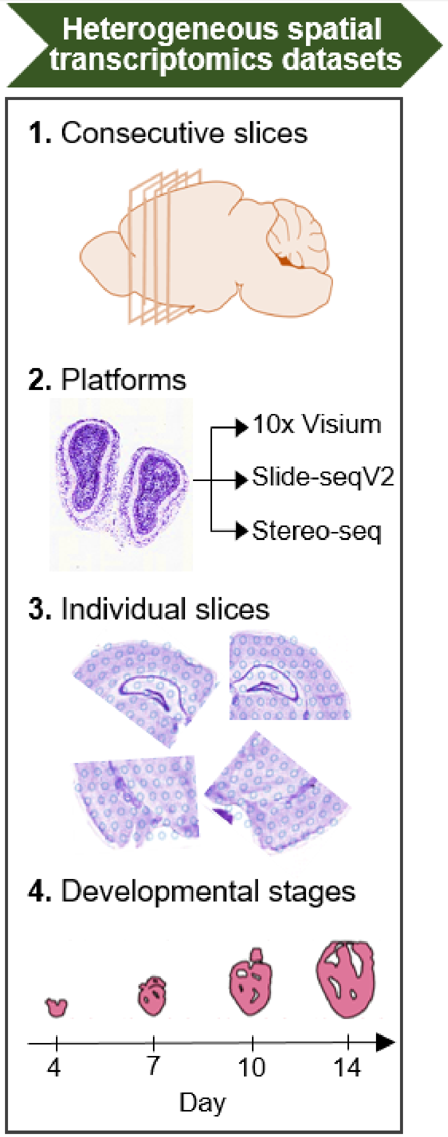
Hu et al., *Computational and Structural Biotechnology Journal* 2021

STAGATE: a graph attention auto-encoder for spatial domain identification



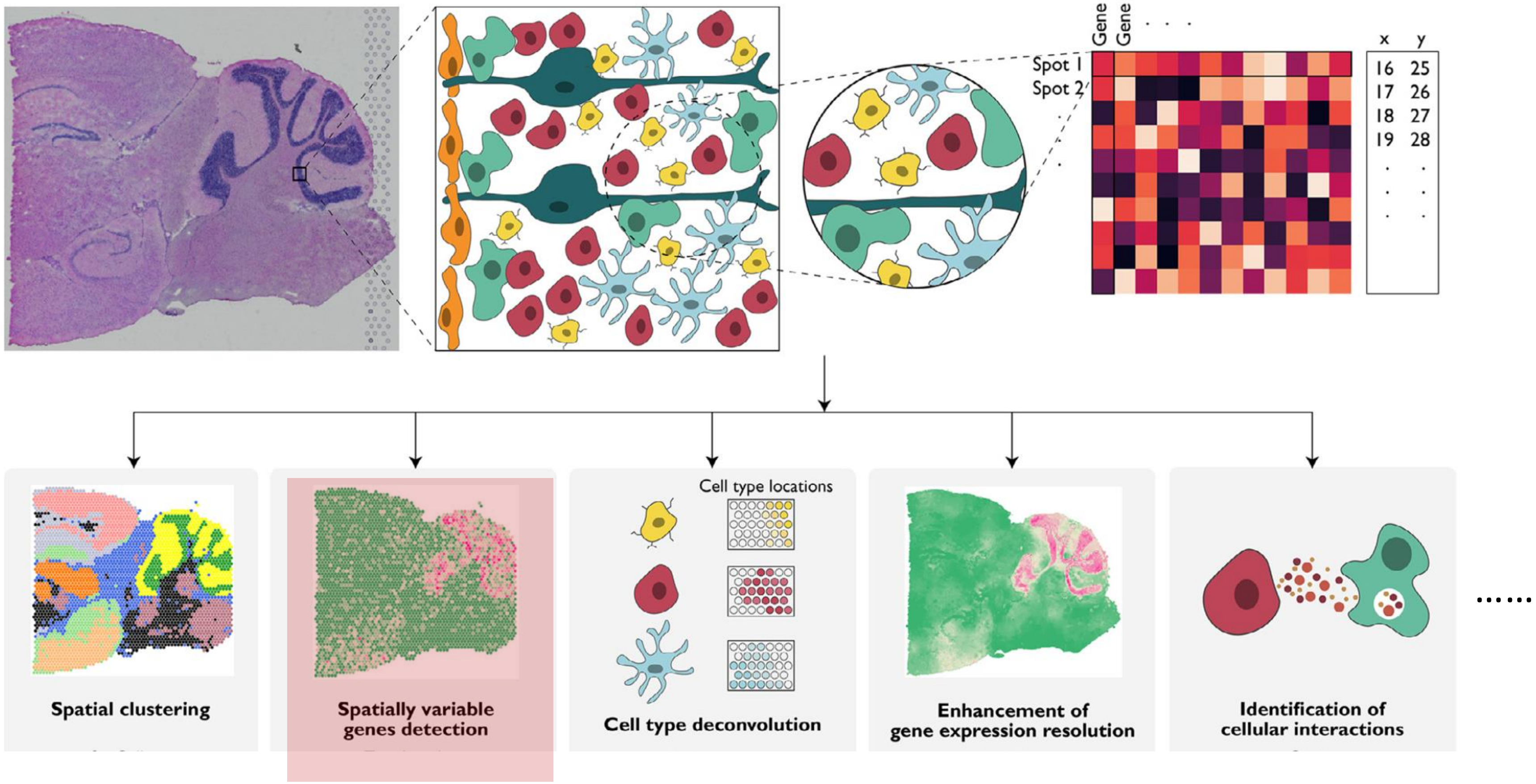
Dong K. and Zhang S., *Nature Communication'2022*

STAligner: aligning multiple slices of ST



Zhou X., Dong K., Zhang S. *Nature Computational Science'2023*

Computational problems of ST



Larsson et al., *Nature Methods* '2021

Hu et al., *Computational and Structural Biotechnology Journal* '2021

Statistical approaches for Spatial Variable Genes identification

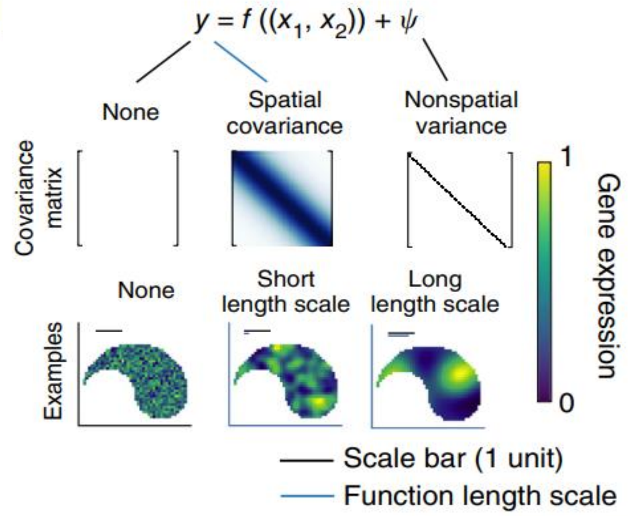
<u>SpatialDE</u>

Parametric model

Fit a Gaussian process regression for each gene

$$P(y\,|\,\mu,\sigma_s^2,\delta,\Sigma) = N(y\,|\,\mu\cdot 1,\sigma_s^2\cdot(\Sigma+\delta\cdot I))$$

Svensson V. et al., *Nature Methods'2018*

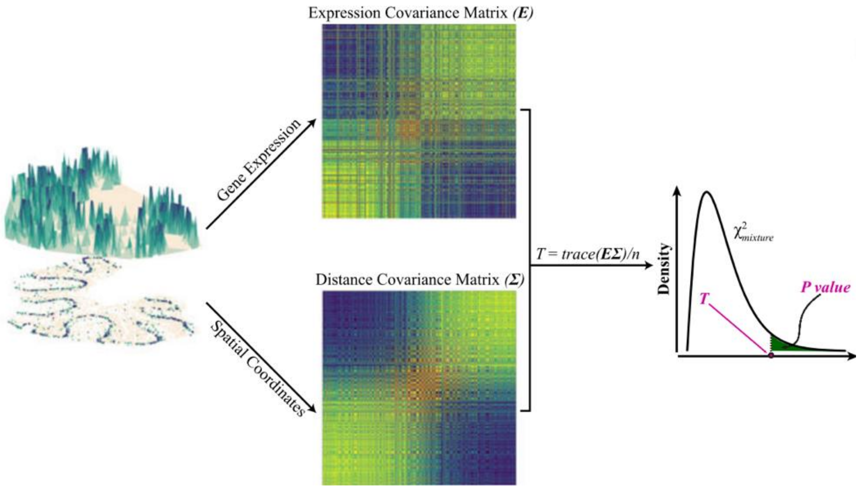


<u>SPARK-X</u>

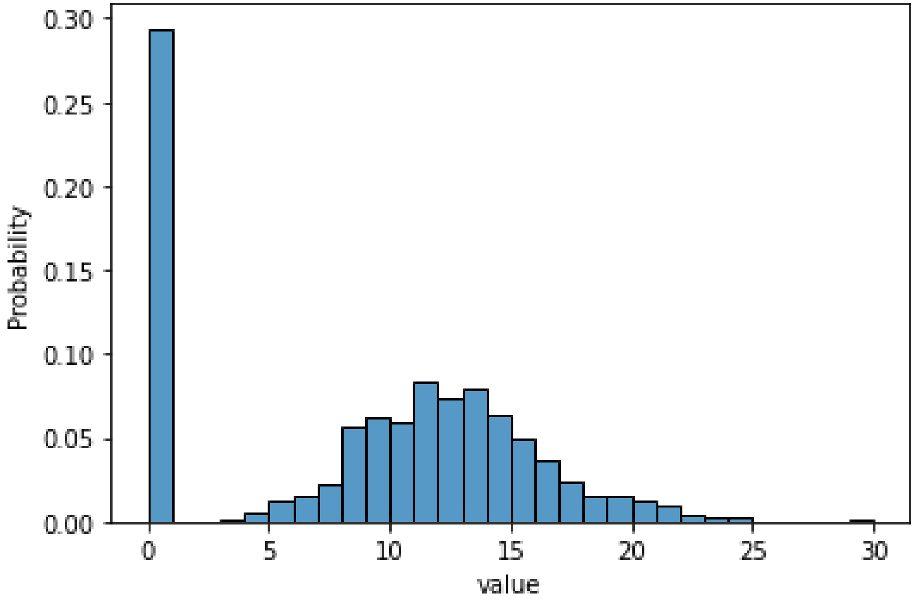
Nonparametric model

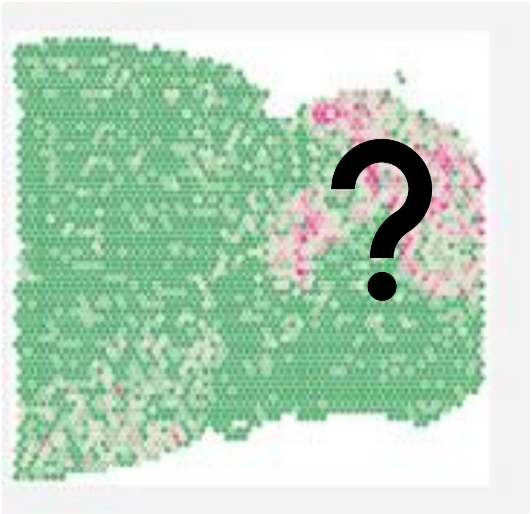
Test the dependence between expression covariance and spatial covariance for each gene

Zhu J et al., *Genome Biology'2021*



Challenges in SVGs identification





ST data is very sparse

- Sparsity causes instability in spatial domains and SVGs identification
- Perform hypothesis test individually weakens the power

Identified SVGs are not domain-specific

- Researchers may be interested in genes that display spatial patterns in one or several specific regions

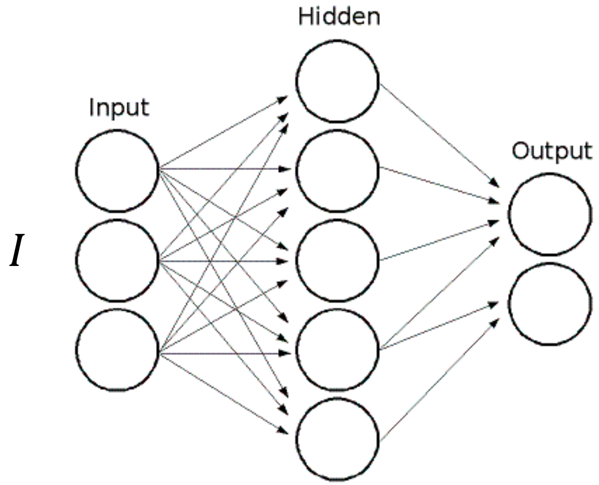
Key problem: robustly identify genes whose expressions display spatially varying patterns in specific regions (domain-specific Spatially Variable Genes)

Key Idea: genes contributing most to the tissue structures classification are potentially important to the corresponding spatial domains

STAMarker: Determining spatial domain-specific variable genes with saliency maps in deep learning

Zhang C., Dong K., Aihara K., Chen L. and Zhang S.
Nucleic Acid Research'2023

Saliency map measures the importance of the input for classification



$S_1(\mathrm{I})$ =0.3

$S_2(\mathrm{I})$=0.7

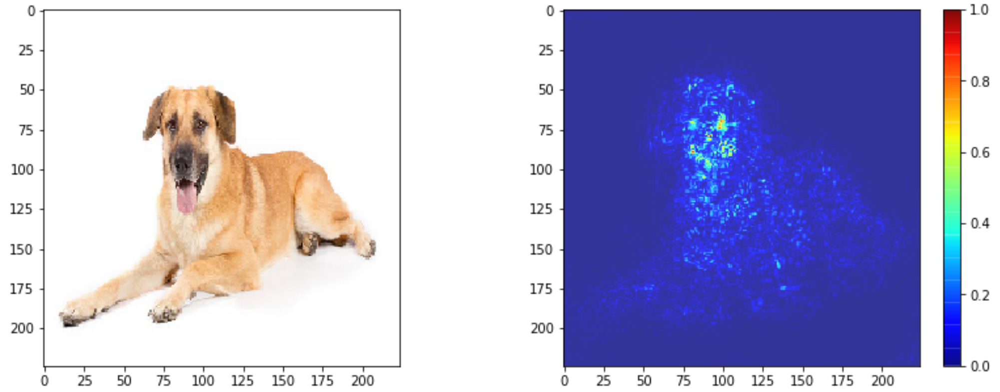
Measure the importance of the input I by gradient

$$w = \frac{\partial S_2}{\partial I}$$

- The saliency maps measure the contribution of each pixel

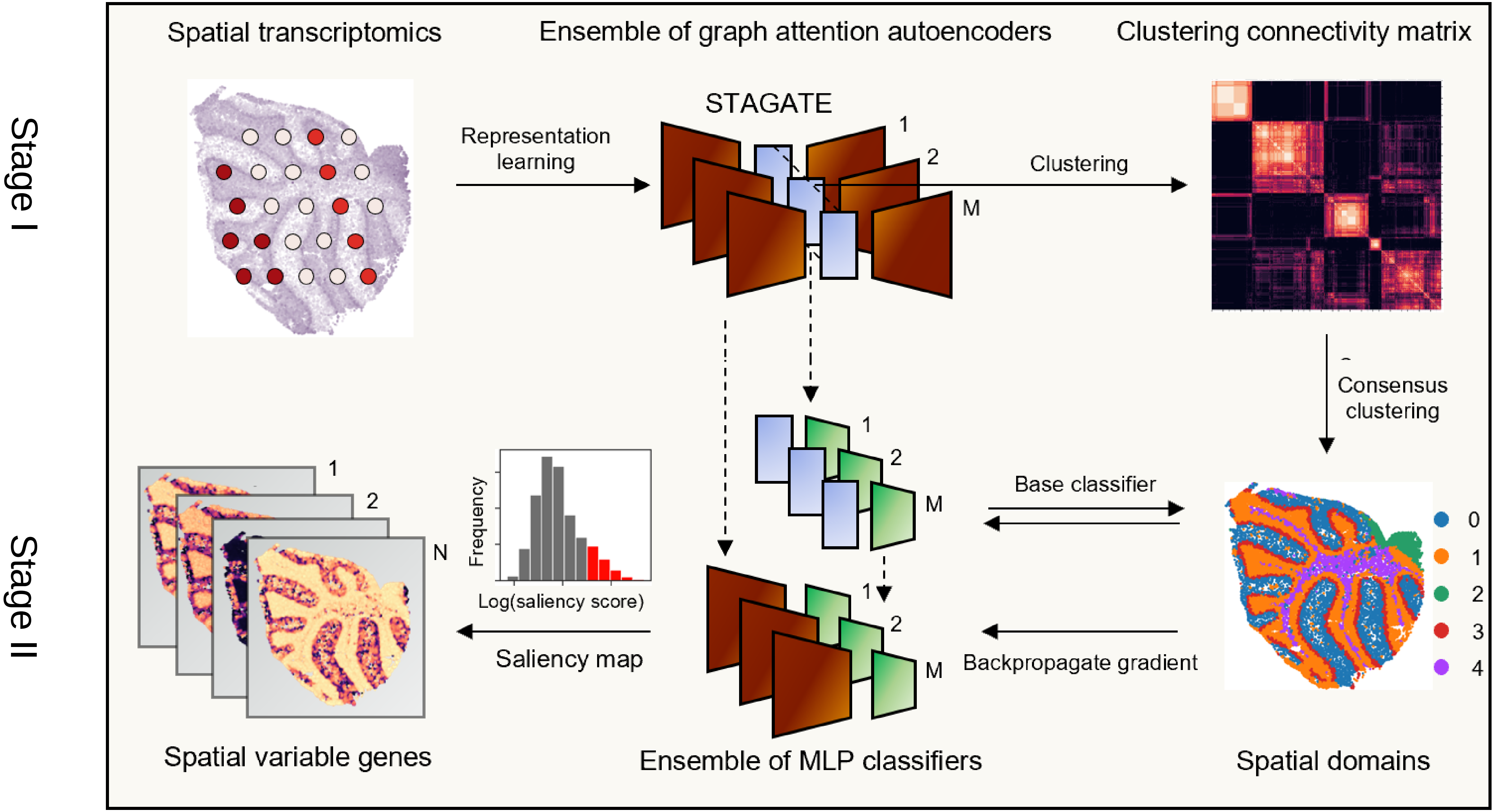
$$w = \frac{\partial S_c}{\partial I}$$

- Pixels with large values in w are more important for class c
- Consider all inputs at once

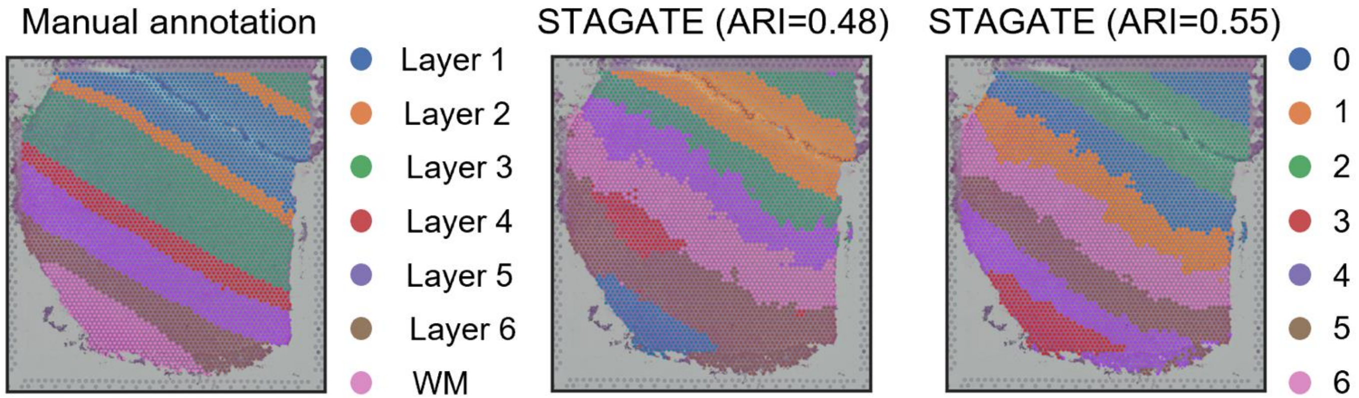


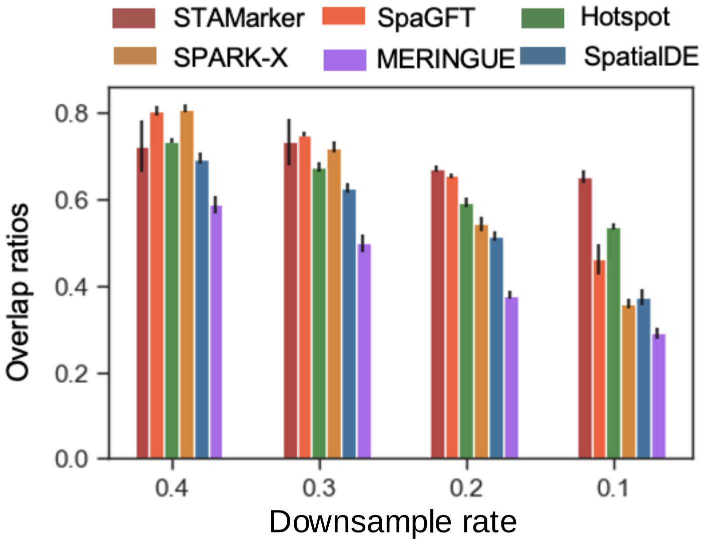
Simonyan K. et. al., *ICLR'2014*

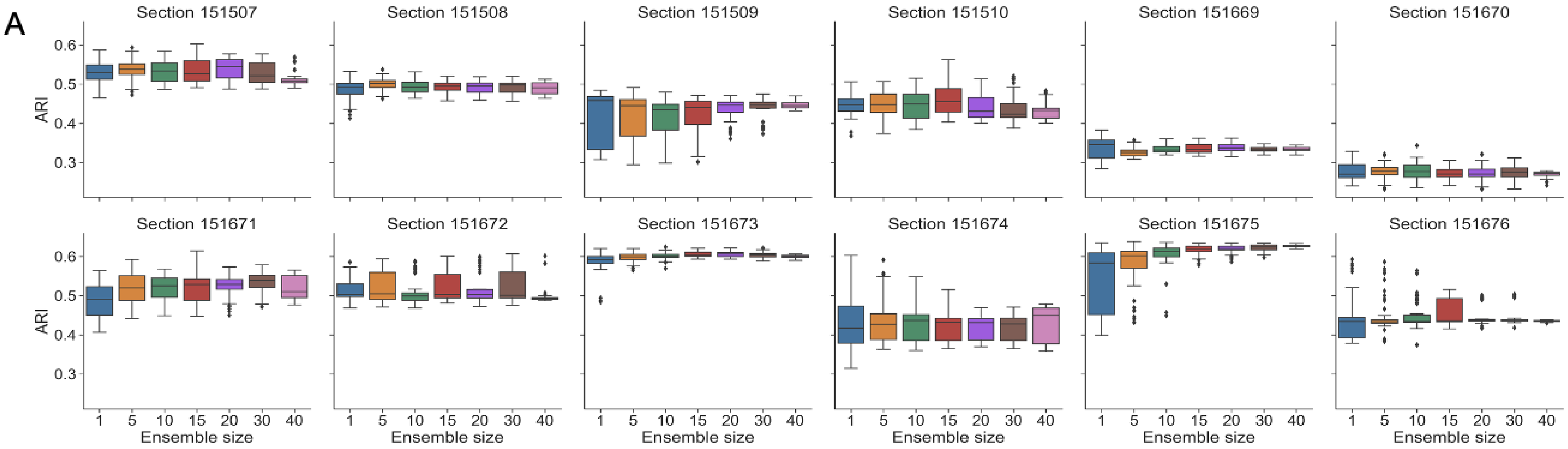
STAMarker: Identify spatial domain-specific variable genes



Ensemble enhances the spatial domain identification

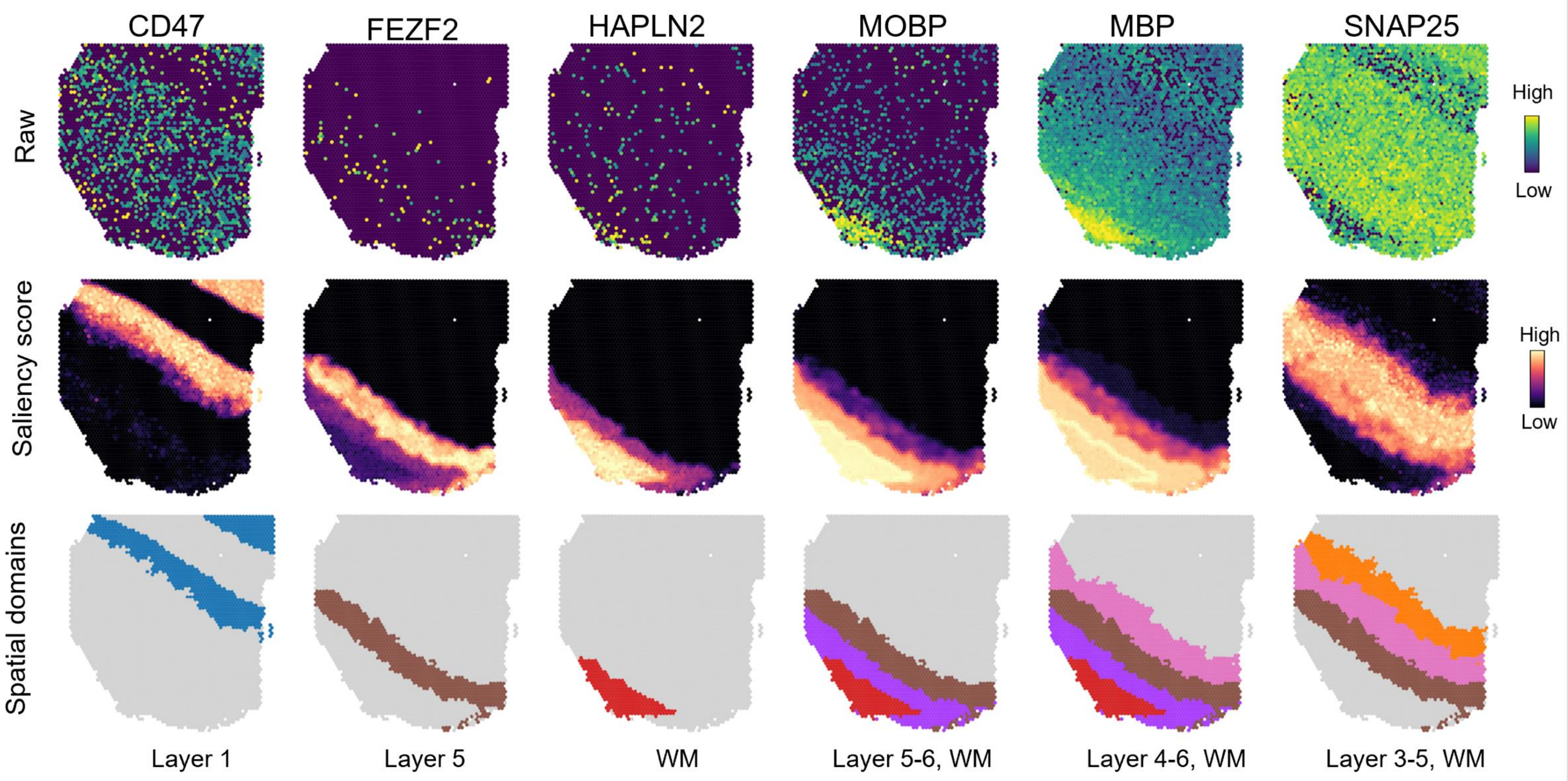






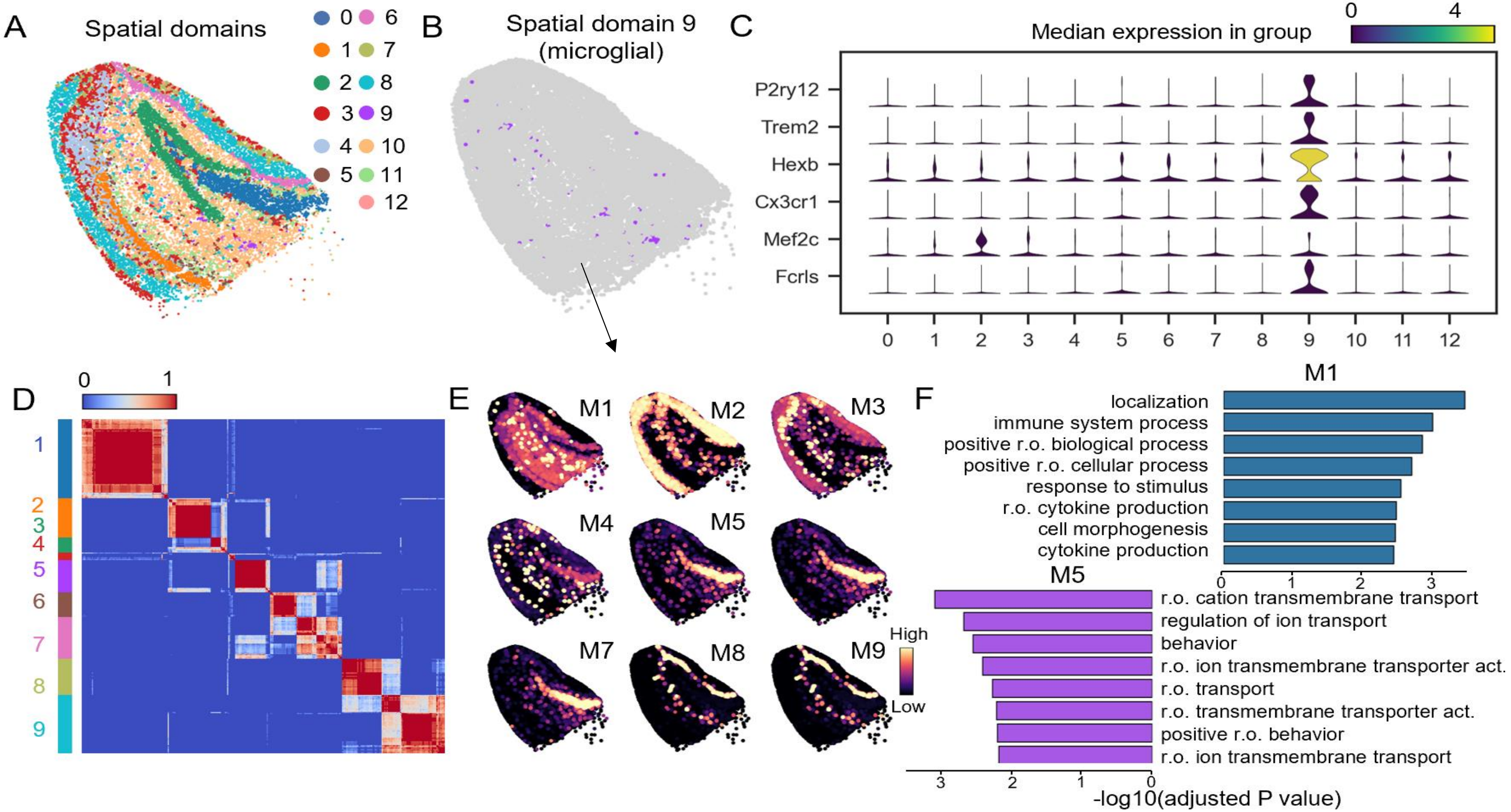
The human dorsolateral prefrontal cortex (DLPFC) dataset

Identify domain-specific spatial variable genes

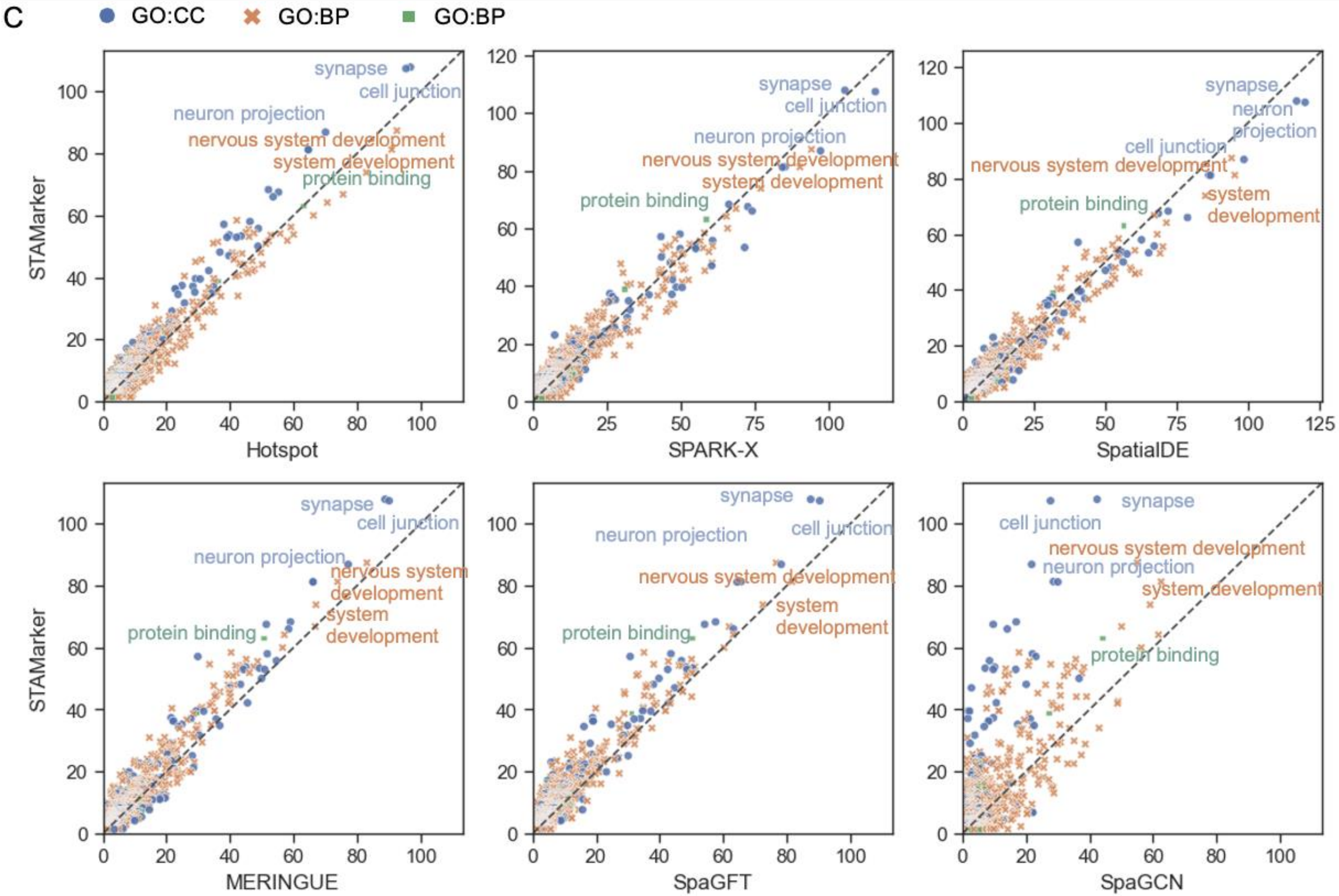


Enable fine-grained analysis on the mouse hippocampus dataset of Alzheimer’s disease

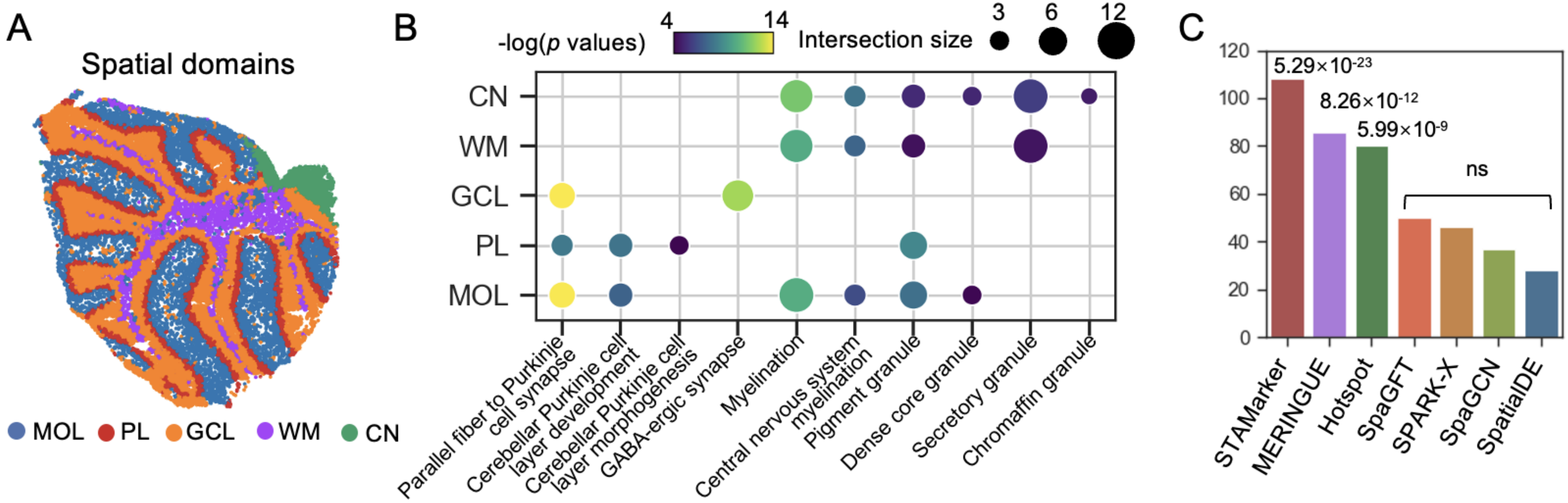
M1 corresponds to **microglial cells** and was specifically enriched in many GO terms related to the immune response terms



The SVGs identified by STAMarker tend to be more enriched in GO terms relevant to the tissues

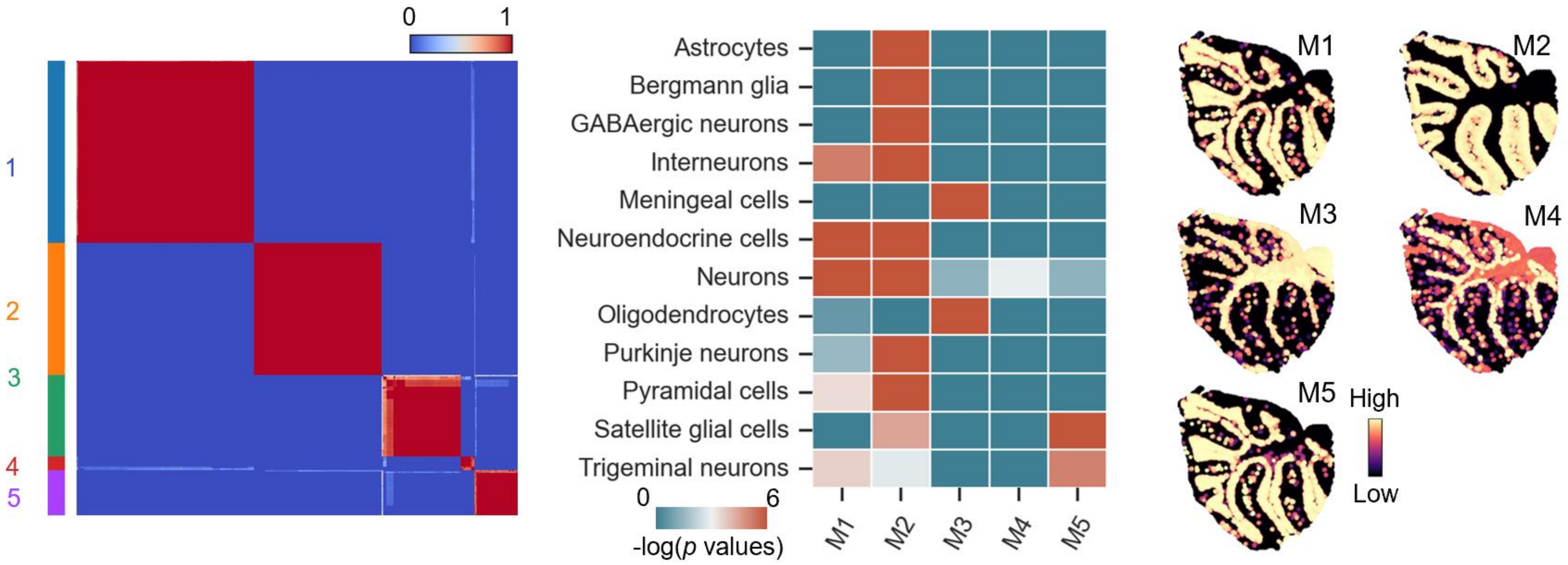


Uncovers spatial domain-specific SVGs on the mouse cerebellum data



Validate the identified SVGs with reference gene (Harmonizome database)

Identified SVGs modules characterize different cell type distribution



Take-home message

- STAMarker can identify domain-specific SVGs by considering all genes at once

- Compared with spatial domain-agnostic methods, STAMarker can investigate spatial domains of interest at a finer scale

Acknowledgement

AMSS

- Prof. Shihua Zhang
- Dr. Kangning Dong

SIBCB

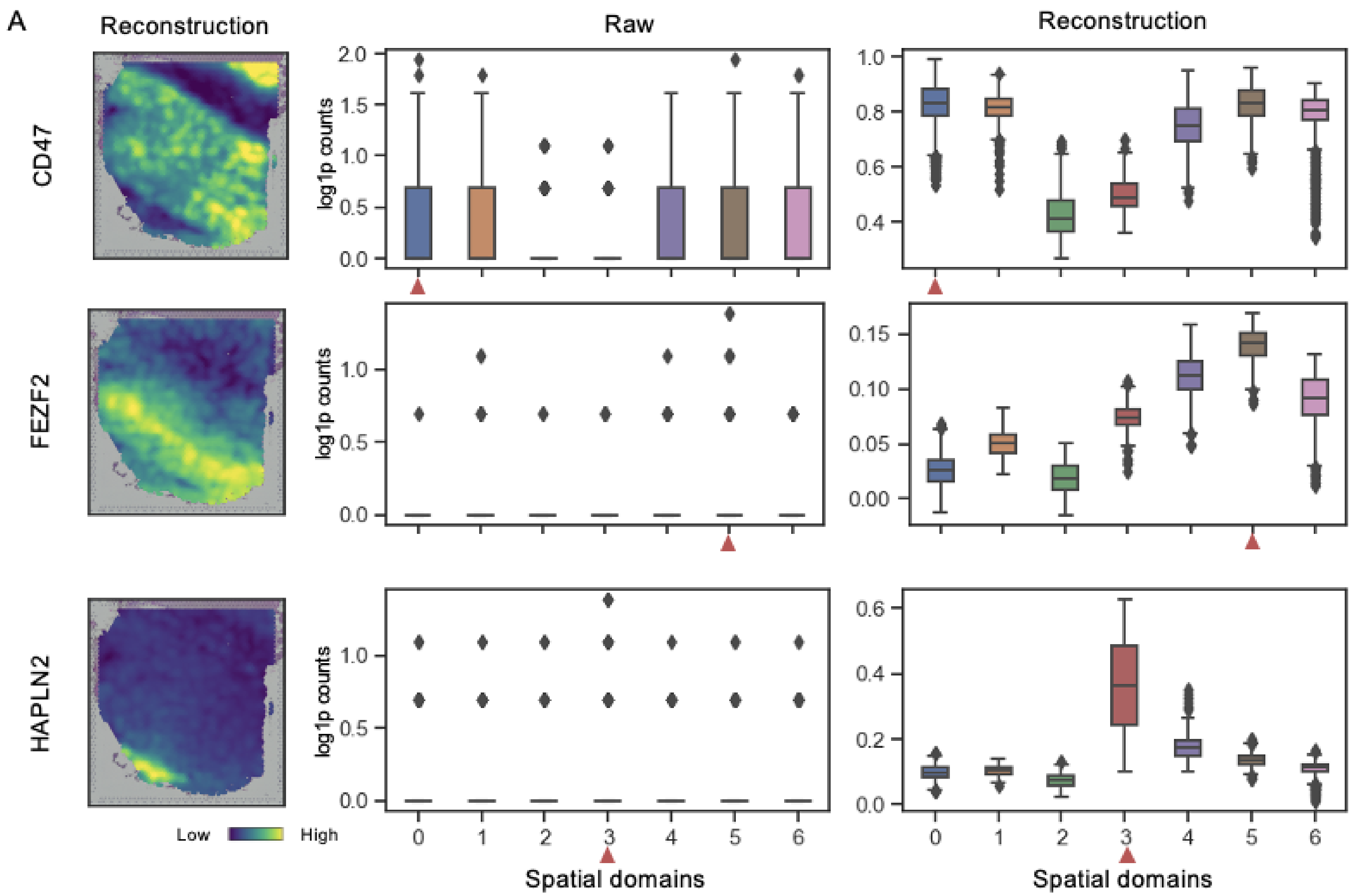
- Prof. Luonan Chen

UTokyo

- Prof. Aihara Kazuyuki.

Thanks for your patience!

Identify SVGs ignored by other methods



Agreement of identified SVGs across different datasets and platforms