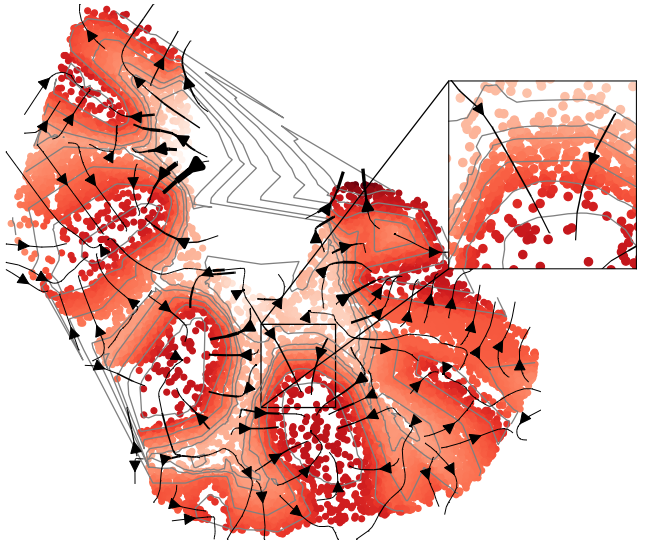
Mapping the topography of spatial gene expression with interpretable deep learning

Uthsav Chitra, Brian Arnold, Hirak Sarkar, Cong Ma, Sereno Lopez-Darwin, Kohei Sanno, Ben Raphael

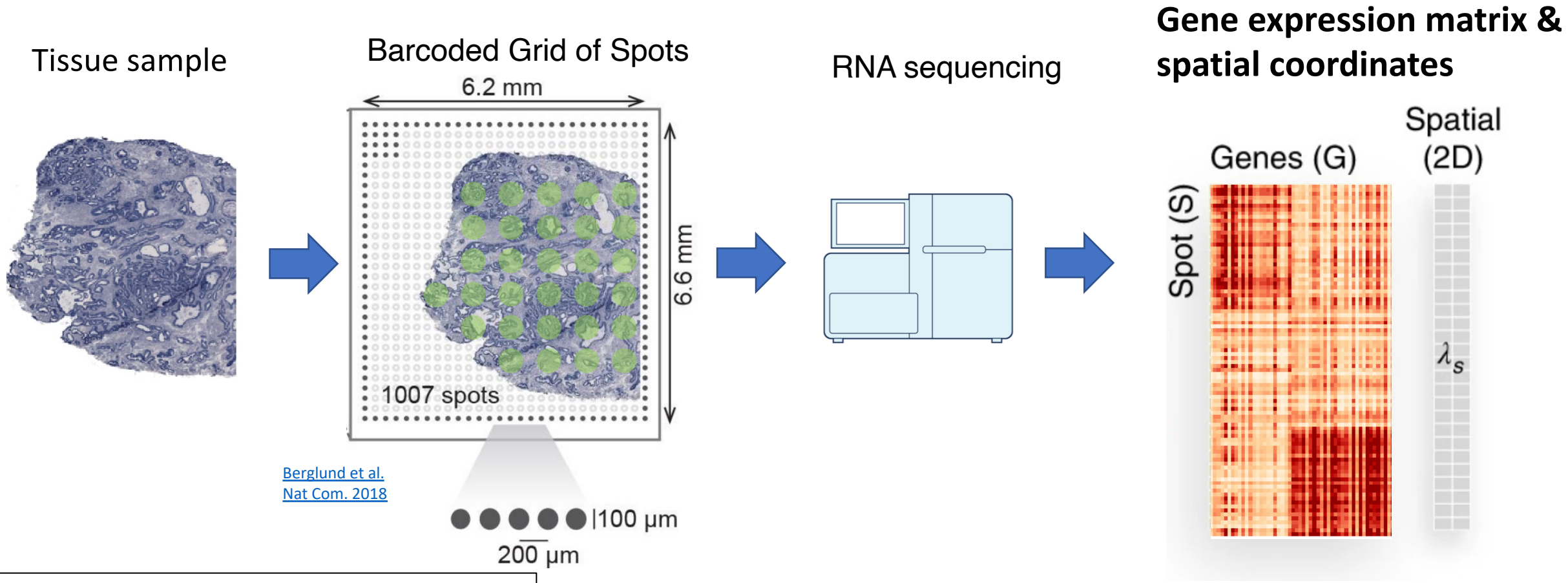
RECOMB 2024

May 2, 2024



PRINCETON UNIVERSITY

Spatially Resolved Transcriptomics (SRT/ST)



Berglund et al. Nat Com. 2018

Technologies: Slide-Seq, 10x Visium, MERFISH, STARMap, …

High-throughput: measure 1,000-20,000 genes at 1,000-10,000 spatial locations (each spot contains 1-20 cells)

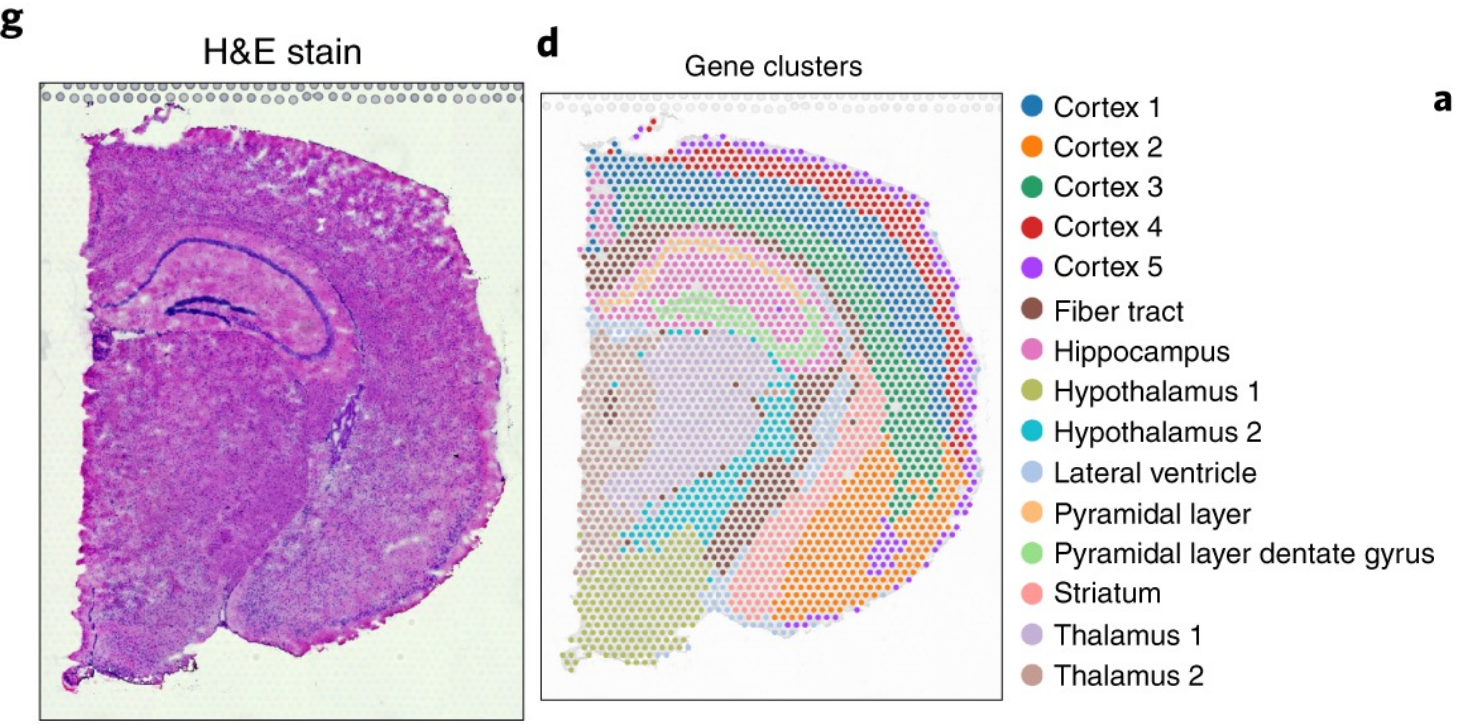
Editorial | Published: 06 January 2021

Method of the Year 2020: spatially resolved transcriptomics

Nature Methods **18**, 1 (2021) | Cite this article

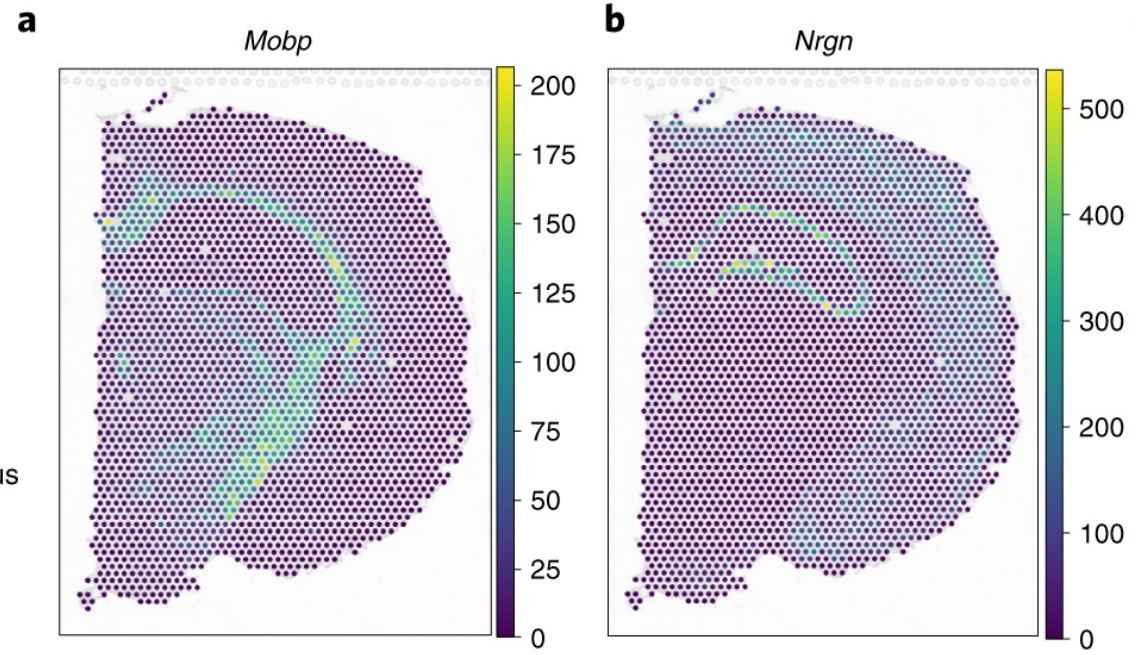
SRT enables study of tissue organization

Spatial domains



Marker genes

(genes differentially expressed in domains)



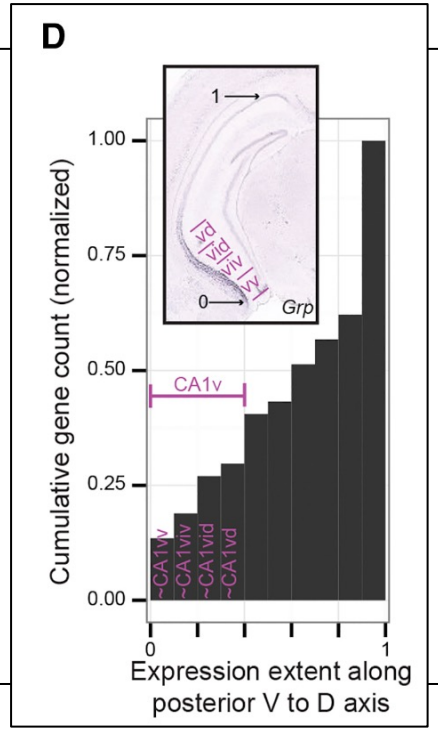
10X Genomics Visium data of mouse brain (Palla et al, Nature Methods 2022)

Spatial gradients are important for many biological processes

Continuous cell types/states

Article

Spatial Gene-Expression Gradients Underlie Prominent Heterogeneity of CA1 Pyramidal Neurons

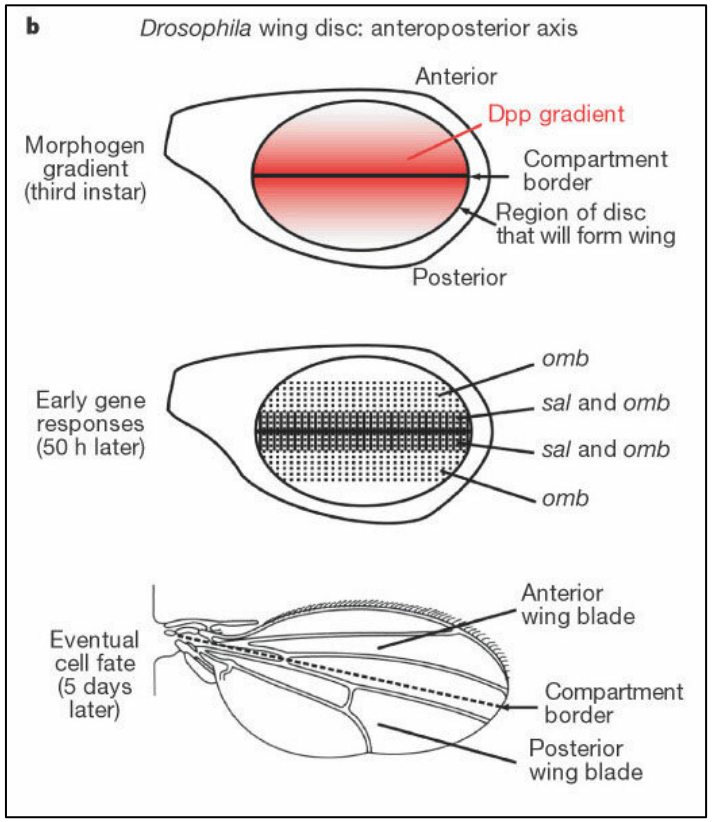


Review

What is a cell type and how to define it?

Hongkui Zeng[1,*]
[1]Allen Institute for Brain Science, Seattle, WA 98109, USA
*Correspondence: hongkuiz@alleninstitute.org
https://doi.org/10.1016/j.cell.2022.06.031

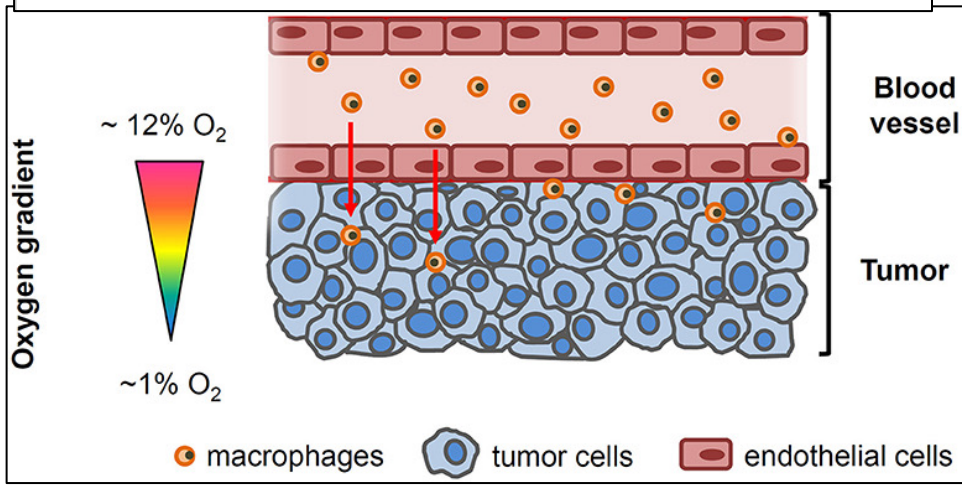
Morphogen gradients



Tumor microenvironment

Physical and Chemical Gradients in the Tumor Microenvironment Regulate Tumor Cell Invasion, Migration, and Metastasis

Madeleine J. Oudin[1] and Valerie M. Weaver[2,3,4,5]

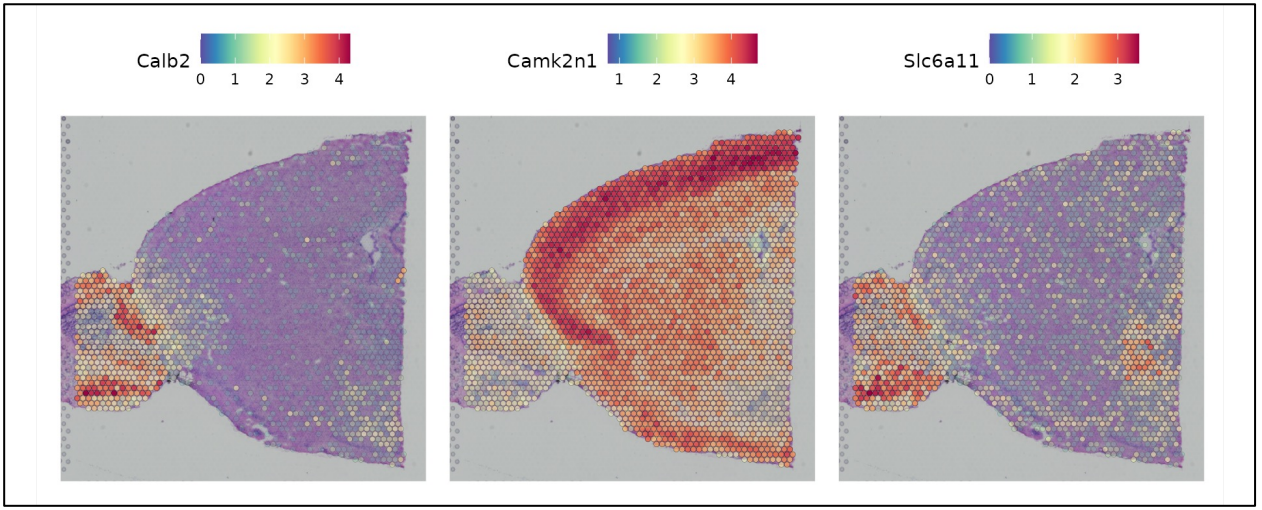


Campillo et al, Frontiers in Oncology 2019

Existing algorithms do not identify spatial gradients unsupervised

Many methods identify **a list of spatially varying genes**

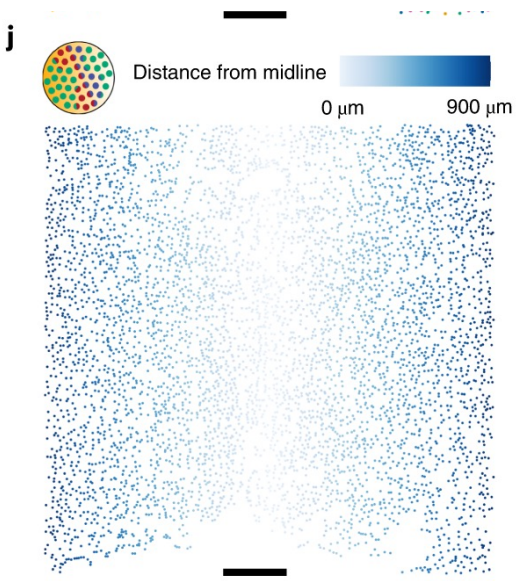
- ❌ Cannot distinguish between gradients vs discontinuous expression
- ❌ Cannot *quantify* gradients (e.g. diffusion coefficient)



Output of Seurat

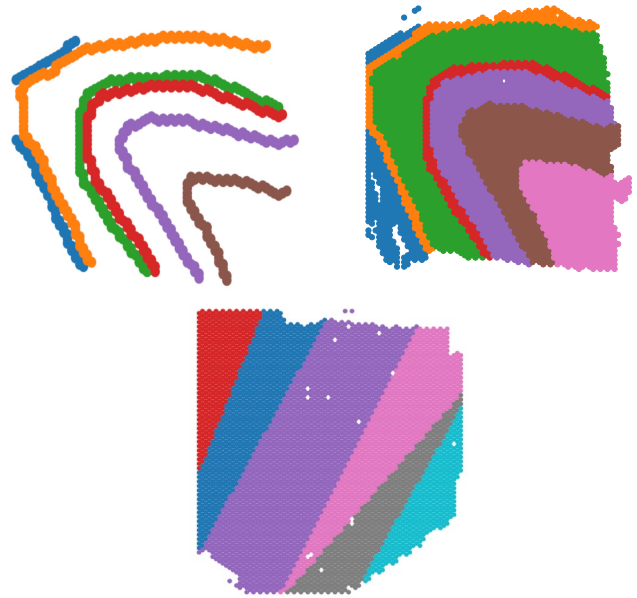
- uses Trendsceek (Edsgard et al, Nat Methods 2018)

A few methods quantify gradients but require **prior knowledge**



C-SIDE (Cable et al, Nat Methods 2022):

- **manually specified coordinate system**



Belayer (Ma*, Chitra* et al, RECOMB + Cell Systems 2022): either

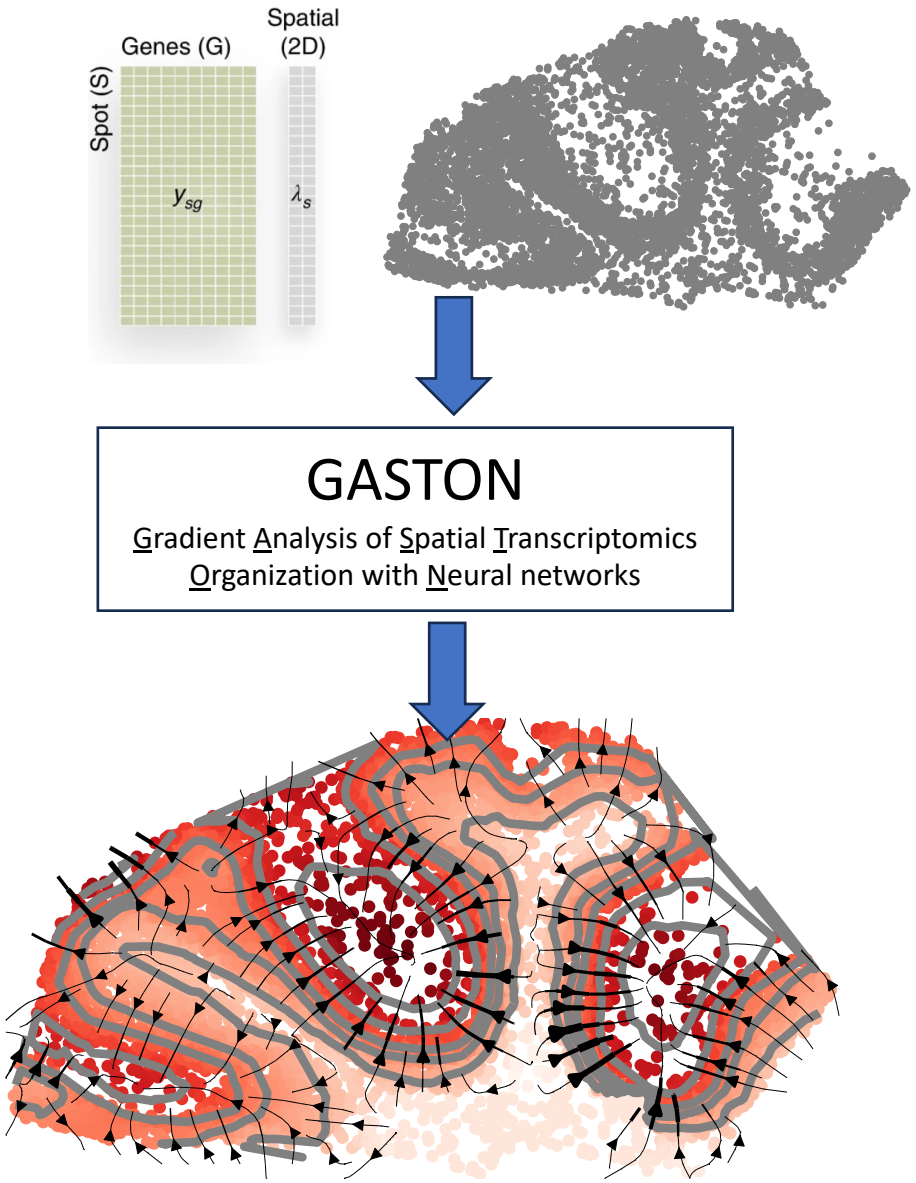
(1) **manually annotated layer boundaries** or

(2) **simple tissue geometry** (layers bounded by lines)

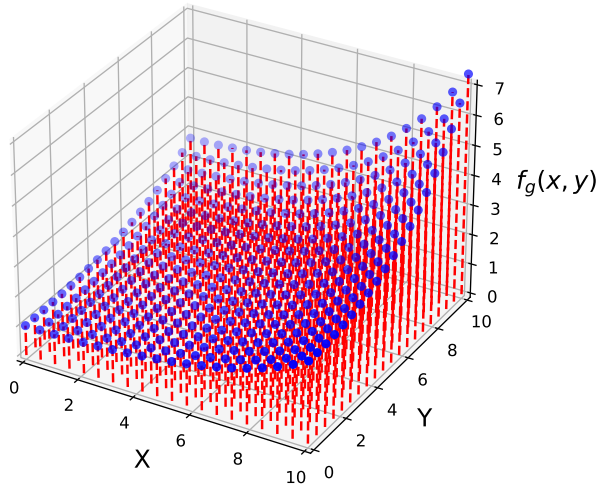
GASTON: an interpretable and unsupervised neural network for learning spatial gradients

1. Mathematical framework for **modeling** spatial gradients
   - Gene expression topography
2. Neural network for **learning** spatial gradients
3. Improved accuracy for detecting spatial domains + marker genes
   - GASTON identifies novel spatial expression gradients in brain + tumor

Spatially resolved transcriptomics data



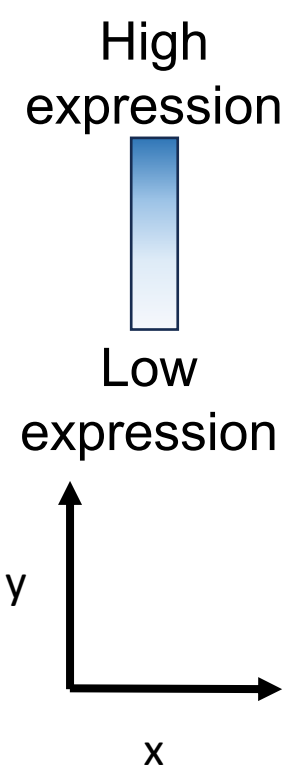
Spatial gradients and multivariable calculus

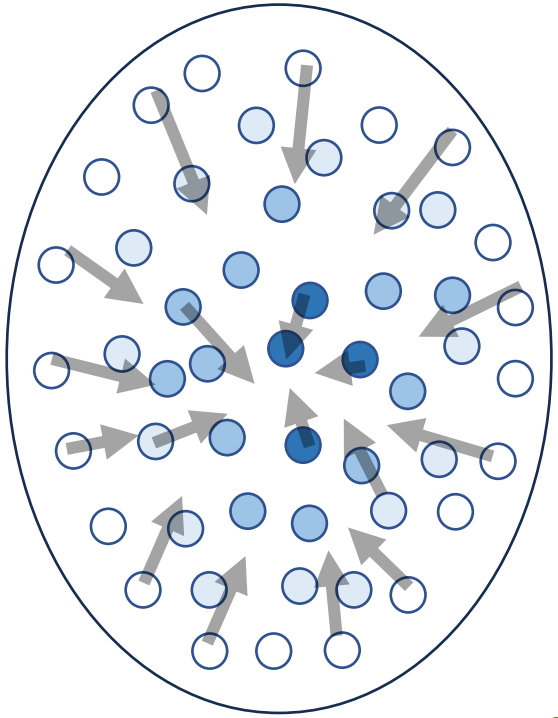


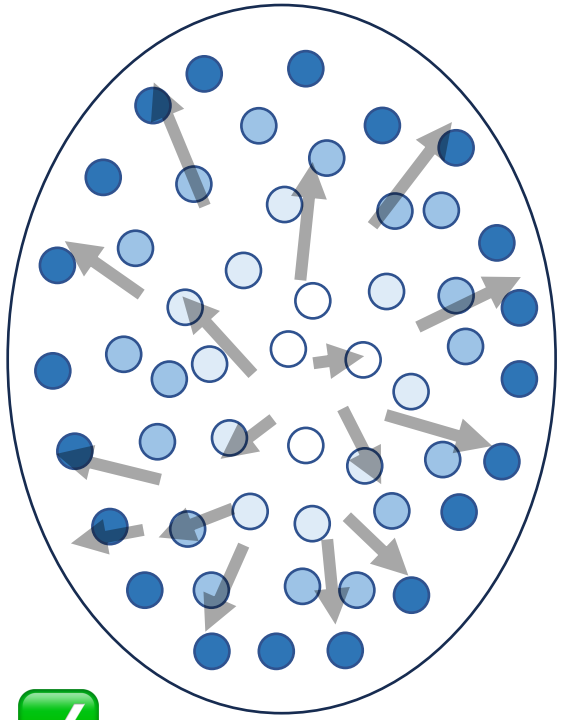
Gene expression function $f_g(x, y)$: $\mathbb{R}^2 \rightarrow \mathbb{R}$

Spatial gradients $\nabla f_g(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$

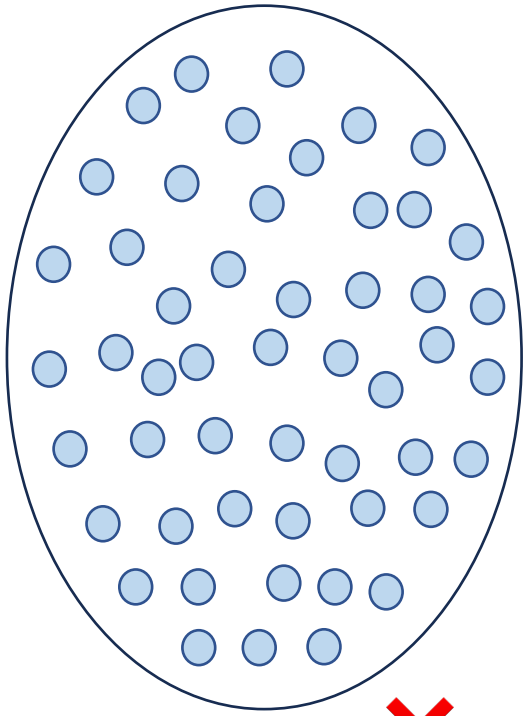
- Gradient $\nabla f_g(x, y)$ = direction (vector) of maximum increase
- Spatial gradients ∇f_g form <u>vector field</u>







Gradient ✅



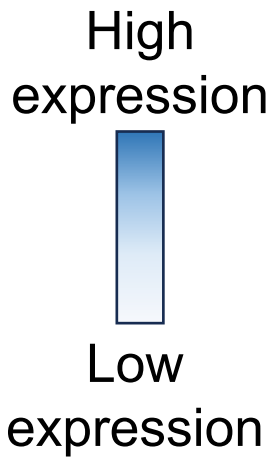
No gradient ❌

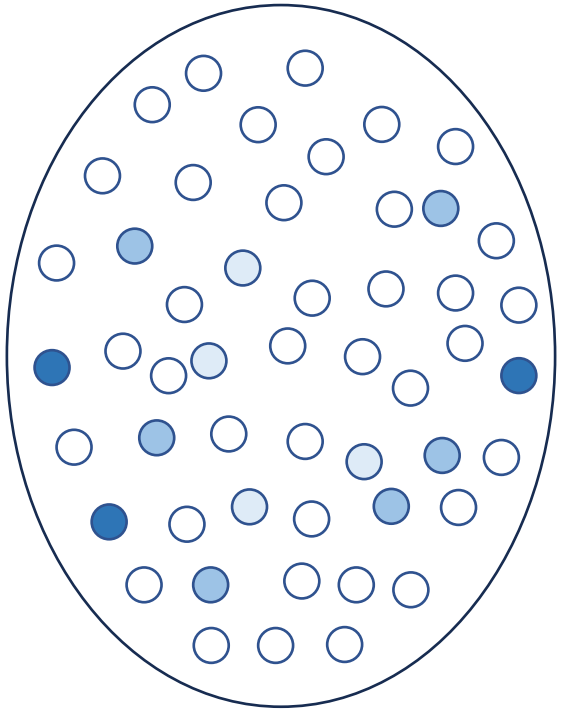
i.e. $\nabla f_g(x, y) = \vec{0}$

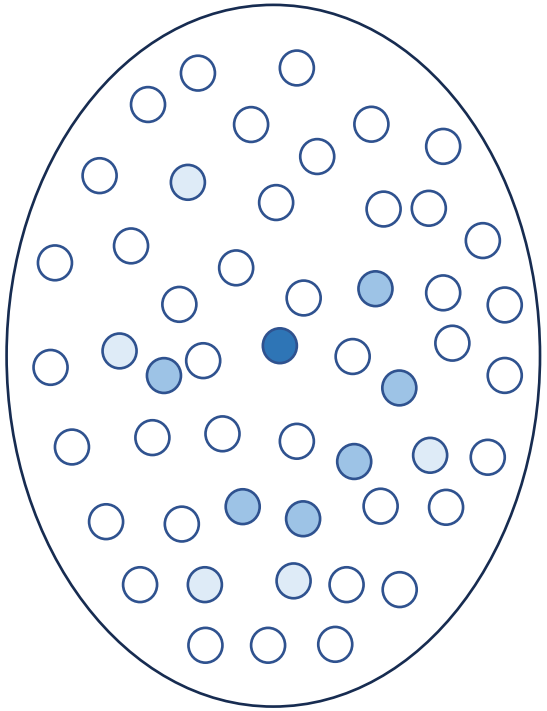
Challenge: modeling spatial gradients in sparse data

Spatial transcriptomics data contains **>70% zeros**

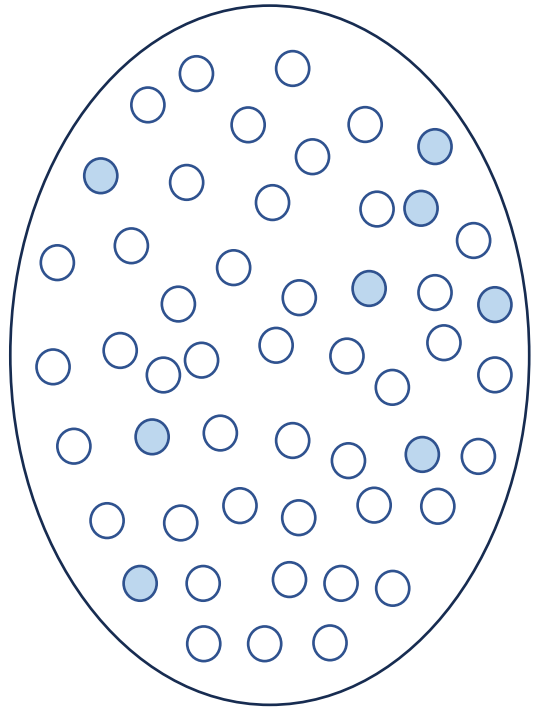
Hard to estimate **gene expression function** $f_g(x, y)$ and **gradient** $\nabla f_g(x, y)$



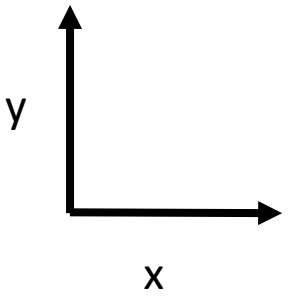




Gradient??



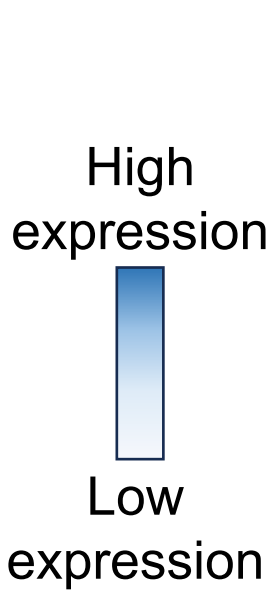
No gradient??

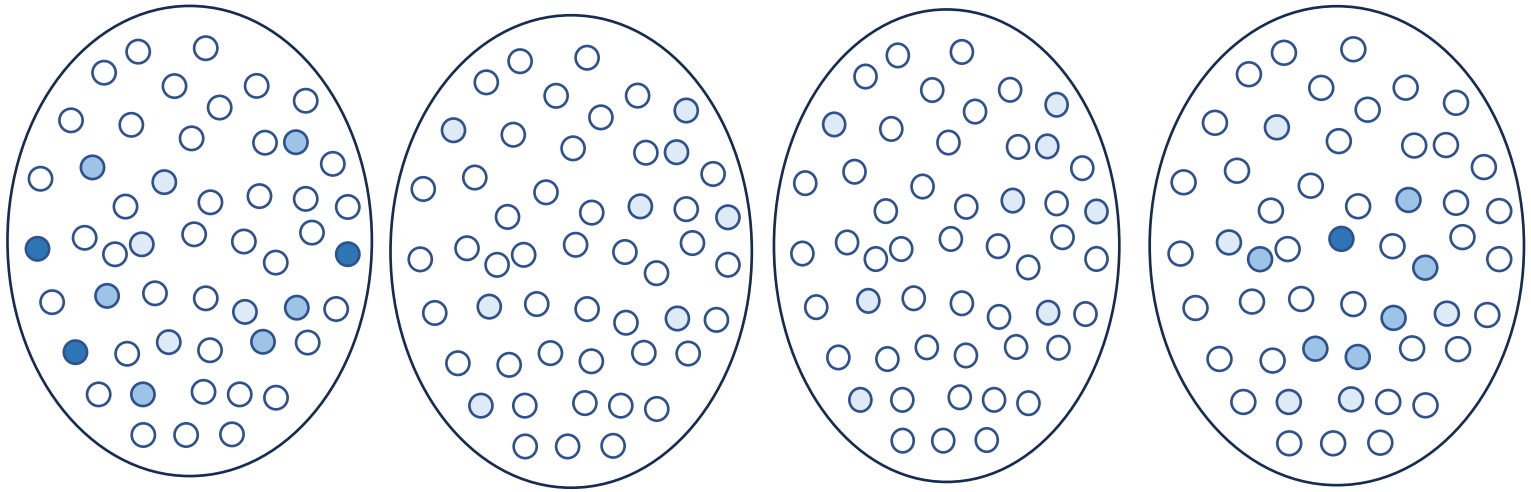


Challenge: modeling spatial gradients in sparse data

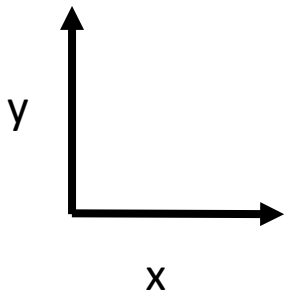
Spatial transcriptomics data contains **>70% zeros**

Hard to estimate **gene expression function** $f_g(x, y)$ and **gradient** $\nabla f_g(x, y)$





… **x 1,000-20,000 genes!**



Our approach: model a single gradient vector field **v** shared by all genes

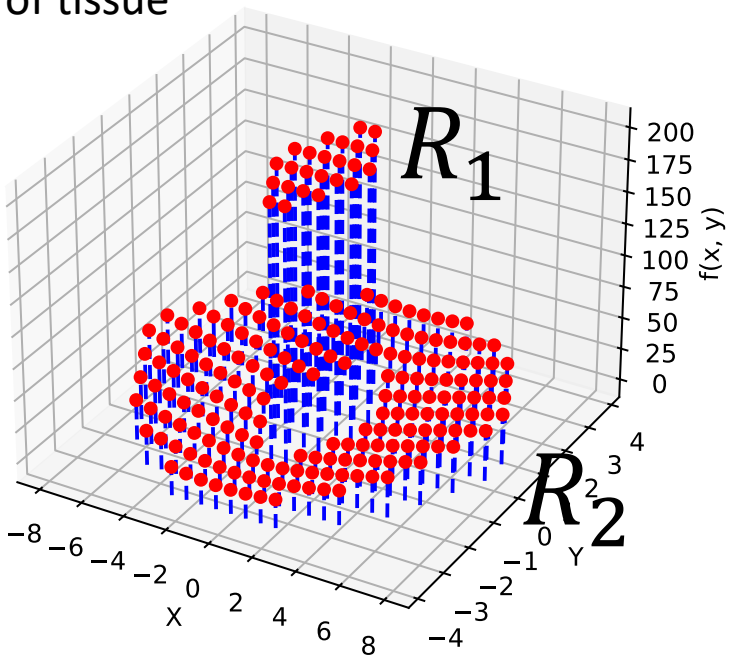
Mathematical assumptions

Gene expression function $f_g(x,y)\colon \mathbb{R}^2 \to \mathbb{R}$
Spatial gradients $\nabla f_g(x,y) : \mathbb{R}^2 \to \mathbb{R}^2$
- Gradient $\nabla f_g(x,y)$ = direction (vector) of maximum increase

Assumption 1: Piecewise continuous expression function

$$\mathbf{f}(x,y) = \sum_{p=1}^{P} \mathbf{f}_p(x,y) \cdot 1_{\{(x,y)\in R_p\}}$$

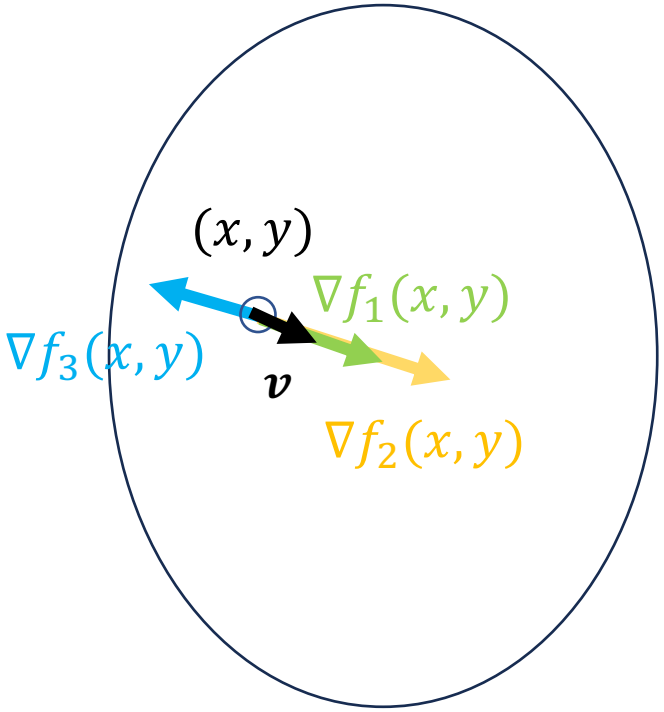
Pieces = different spatial domains of tissue



Assumption 2: Shared spatial gradients

$$\nabla f_g(x,y) = c_g \cdot \boldsymbol{v}(x,y)$$

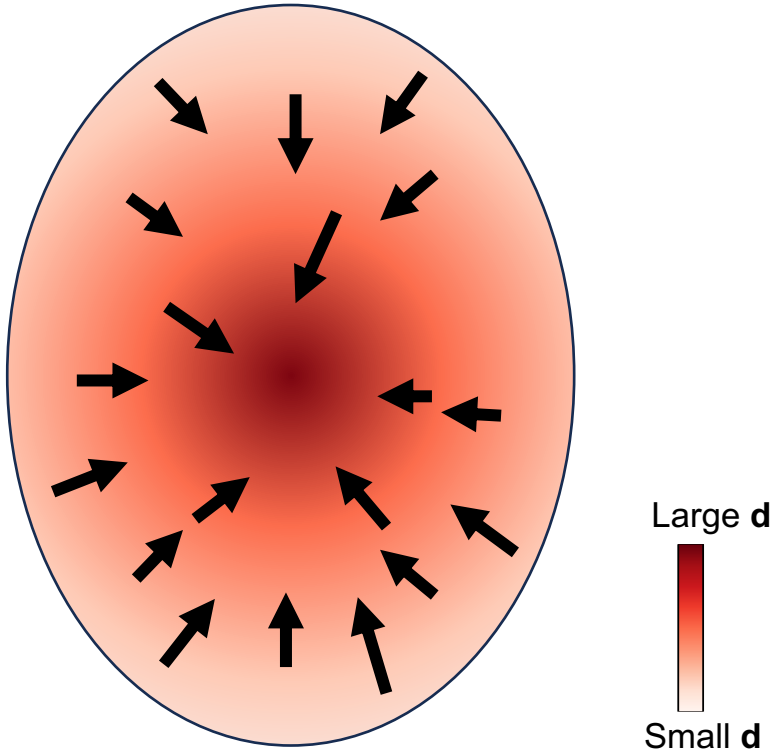
for all variable genes/components g



Assumption 3: Conservative gradient vector field

$\boldsymbol{v} = \nabla d$ (no "curl")

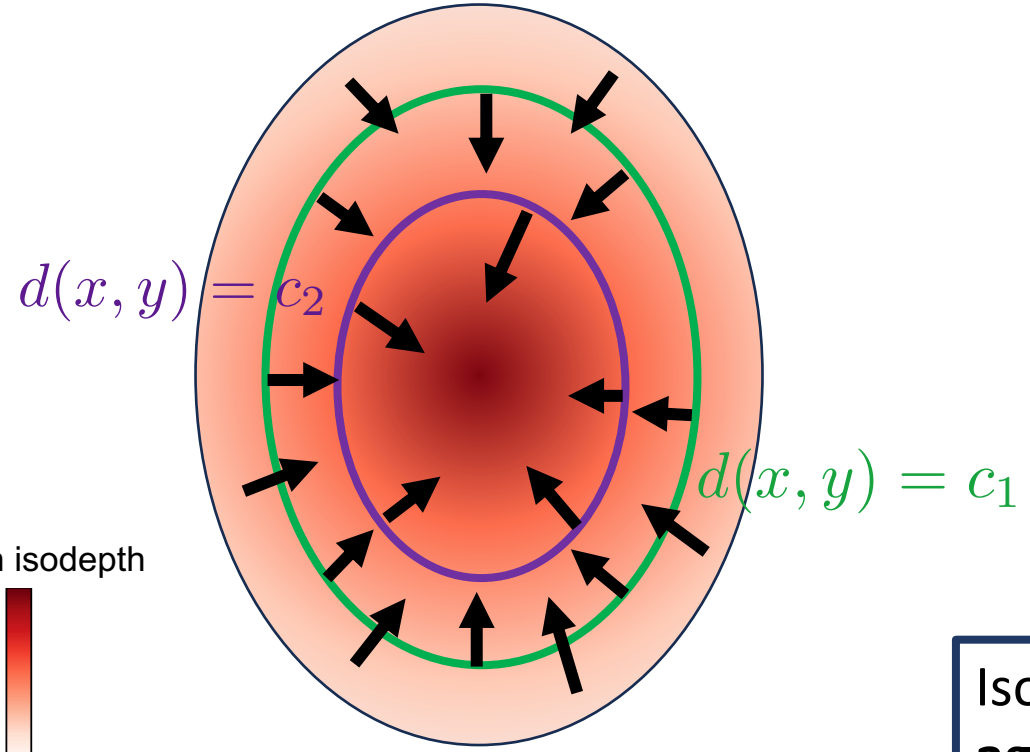
$d\colon \mathbb{R}^2 \to \mathbb{R}$ is (scalar) **potential function**

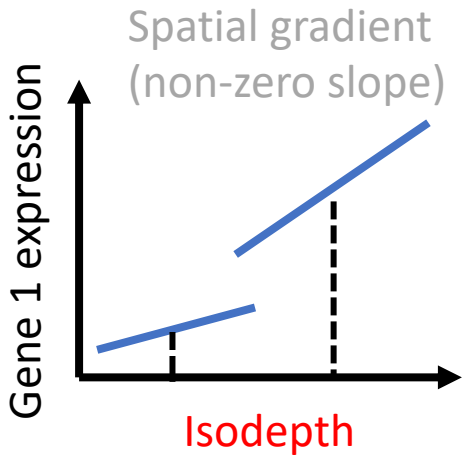


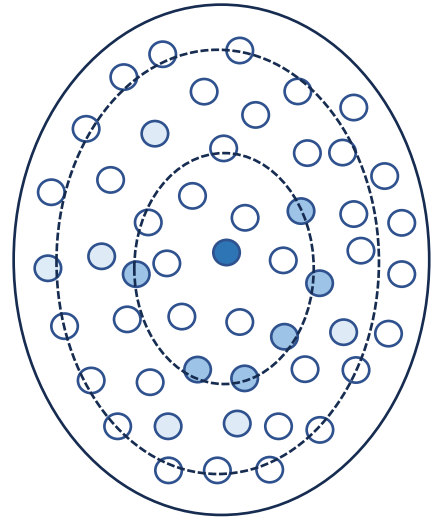
Isodepth: a 1-D **coordinate** describing shared spatial gradients

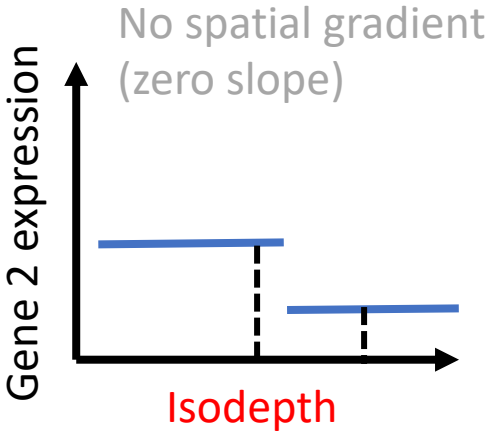
We prove: under our assumptions, gene expression $f_g(x, y)$ is **piecewise continuous** function of isodepth $\mathrm{d}(x, y)$:

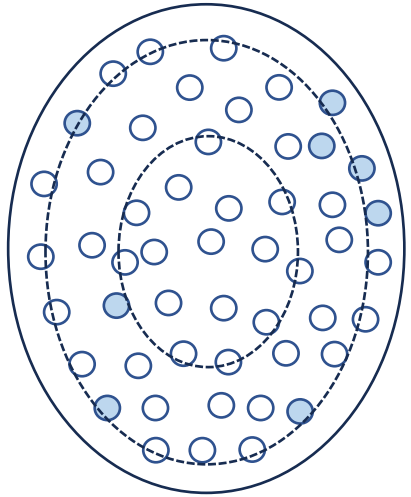
$$f_g(x, y) = h_g(d(x, y))$$





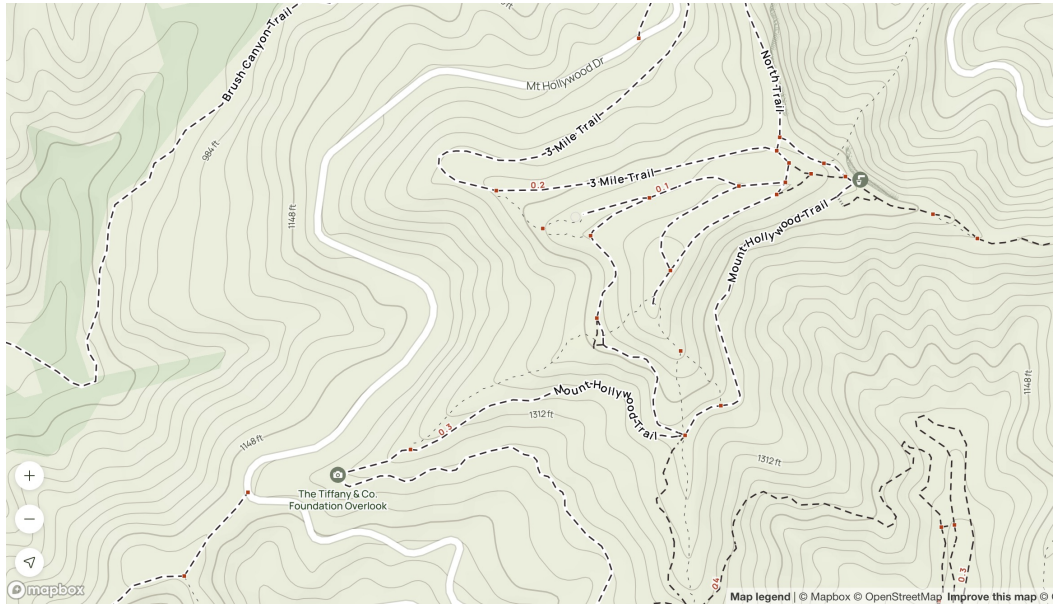


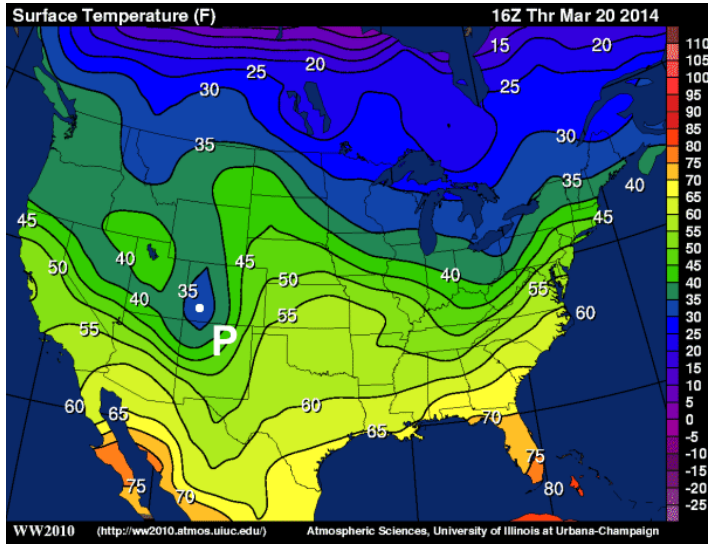




Isodepth accounts for sparsity by **aggregating expression across spots**

Isodepth describes “topography” of a tissue slice





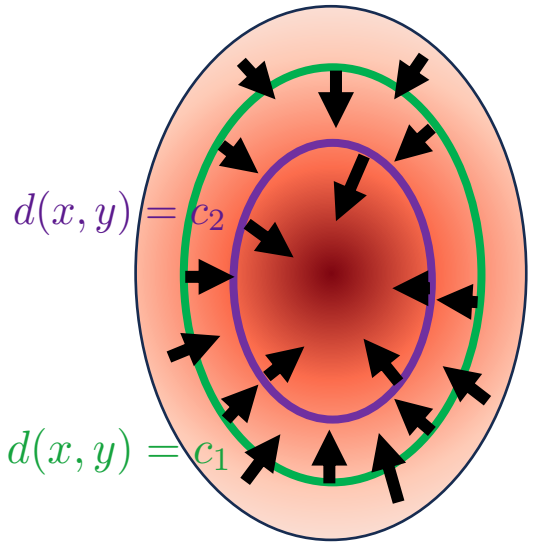
Isodepth: contours of equal depth $\mathrm{d}(\mathrm{x}, \mathrm{y}) = c$

- Generalizes *depth* in a layered tissues (e.g. skin)
- Contours surround spatial domains (like mountains in topographic map)

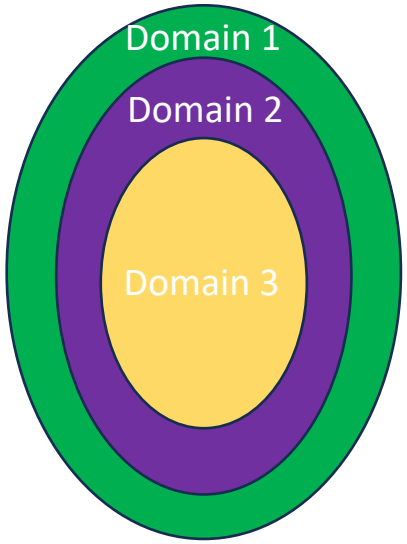
Spatial gradients $\nabla \mathrm{d}$ (gradient of isodepth)

- <u>Directions of maximum change in gene expression</u>

Isodepth and spatial gradients

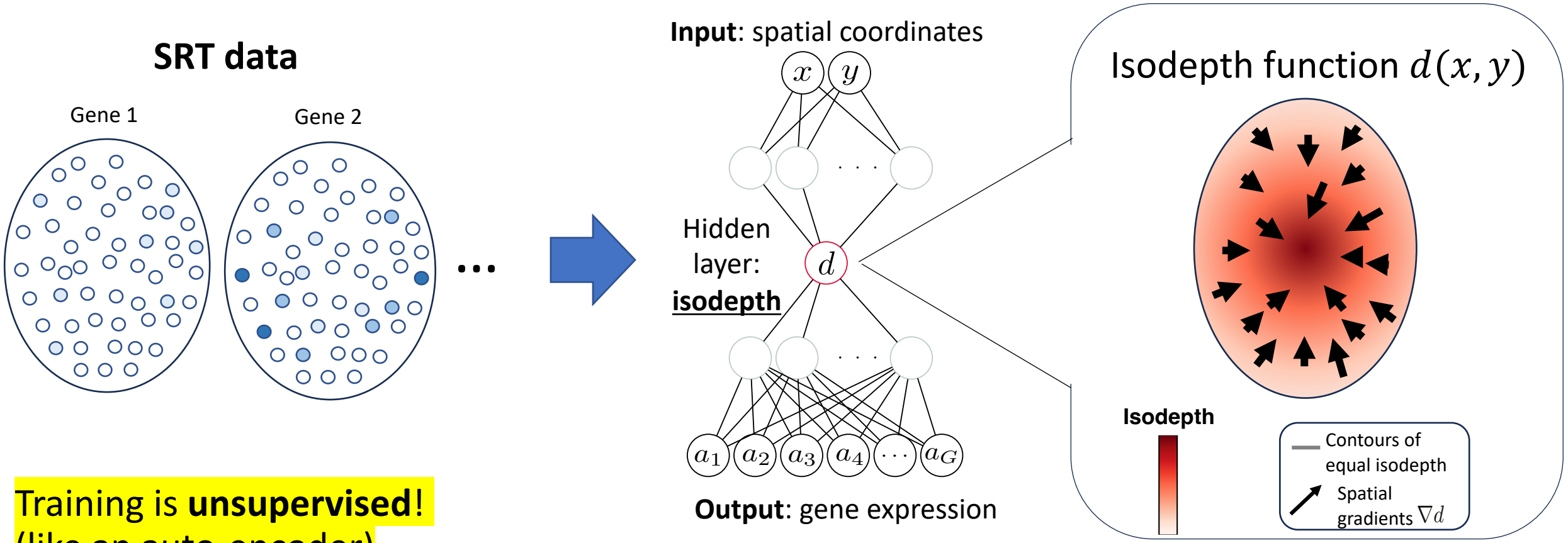


Spatial domains



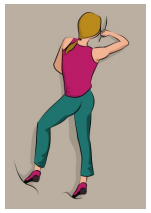
Spatial neural network learns isodepth and spatial gradients

Gene expression $f_g(x, y)$ is **piecewise continuous** function of isodepth $\mathrm{d}(x, y)$: $f_g(x, y) = h_g(d(x, y))$
Our approach: parametrize functions $\mathrm{h_g}, d$ with neural networks

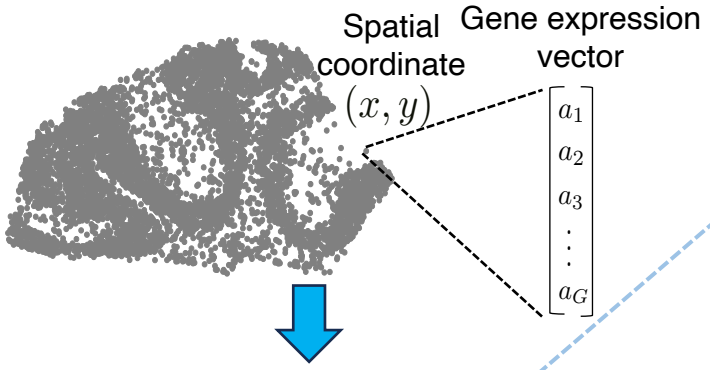


Training is **unsupervised**!
(like an auto-encoder)

GASTON



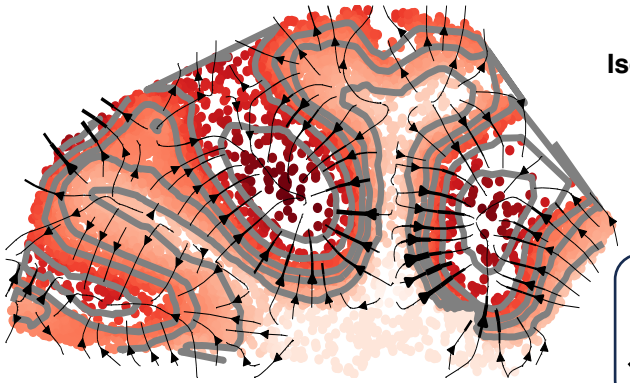
Spatially resolved transcriptomics data

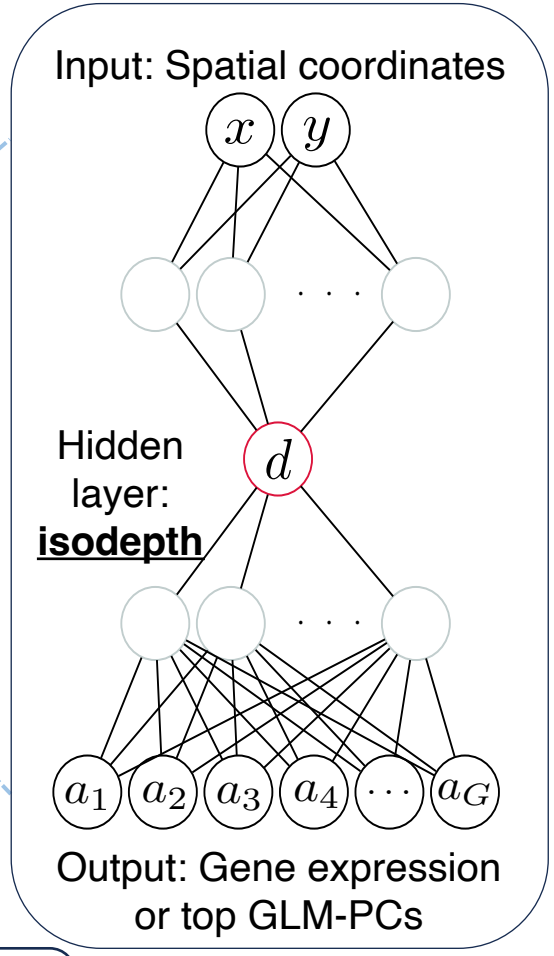


GASTON

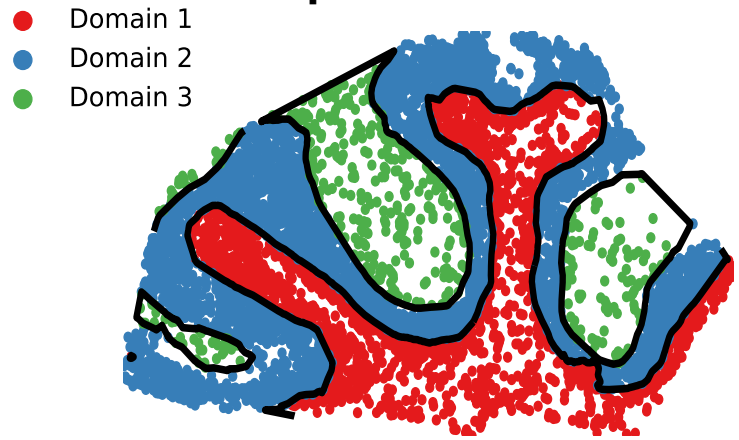
Gradient Analysis of Spatial Transcriptomics Organization with Neural networks

Topographic map of tissue slice

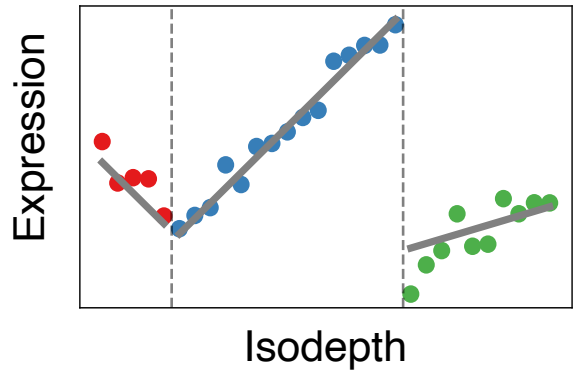




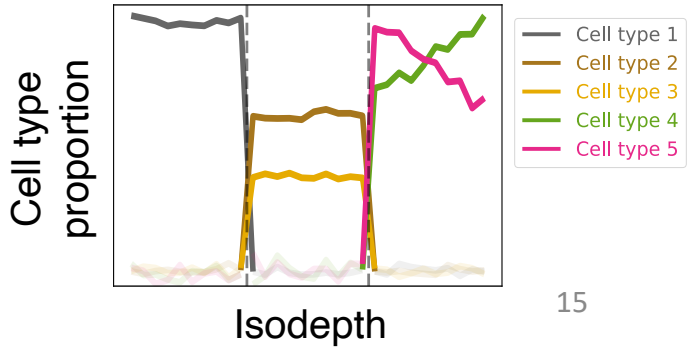
Spatial domains



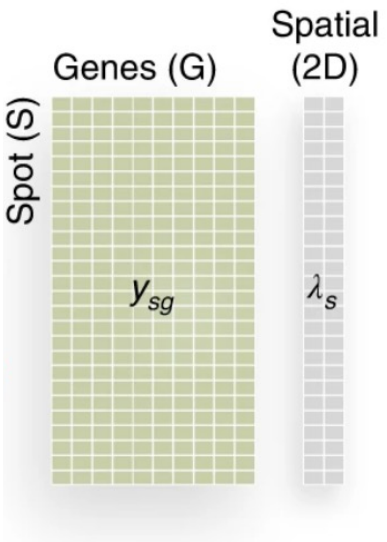
Continuous gradients and discontinuous variation in gene expression



Spatial variation in cell type organization



GASTON: Mouse Cerebellum (Slide-seqV2)

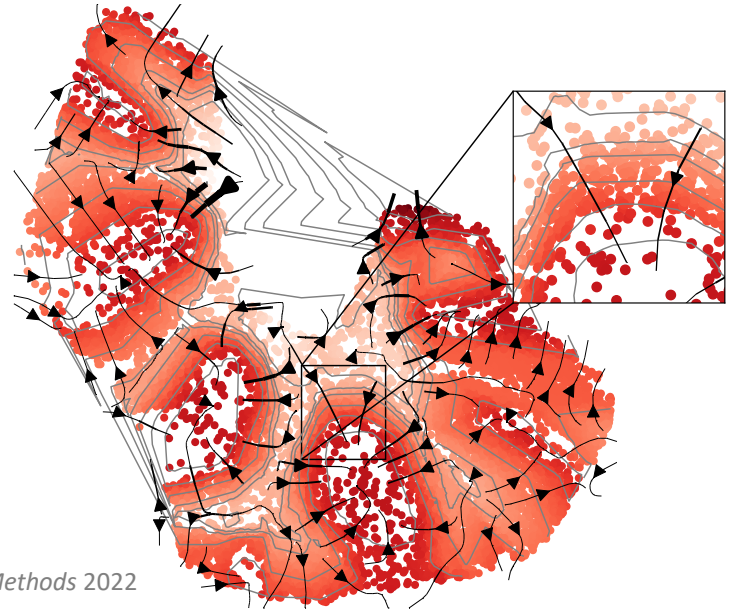


GASTON

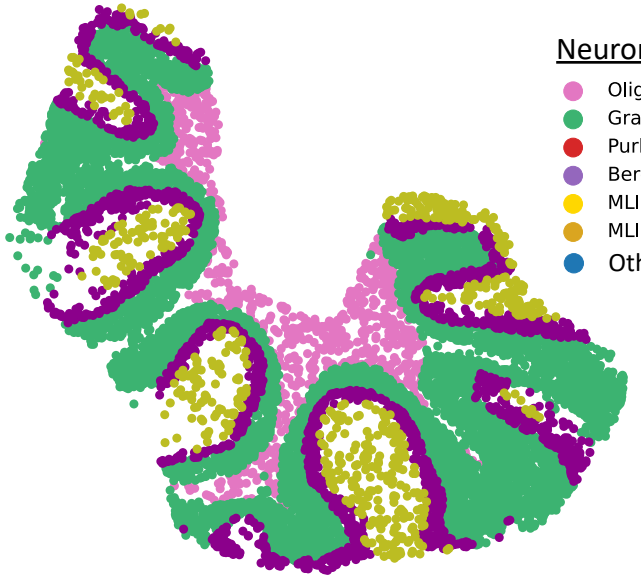
23,096 genes × 9,985 spots

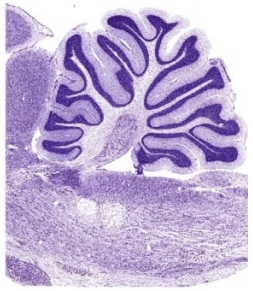
Cable et al., *Nature Methods* 2022

Topographic map

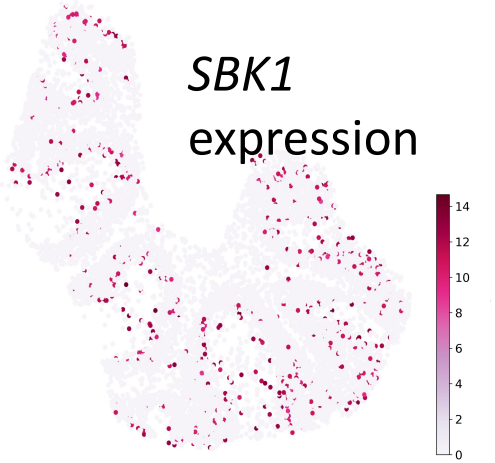


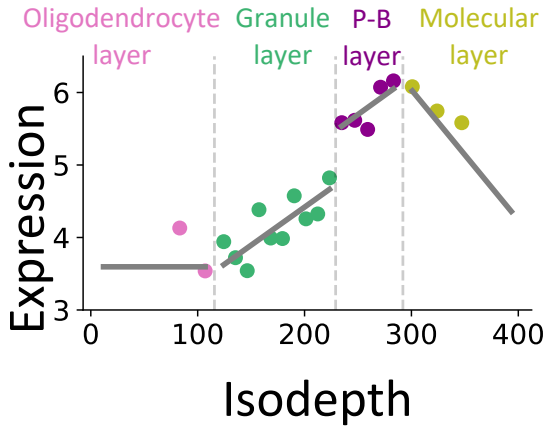
Spatial domains

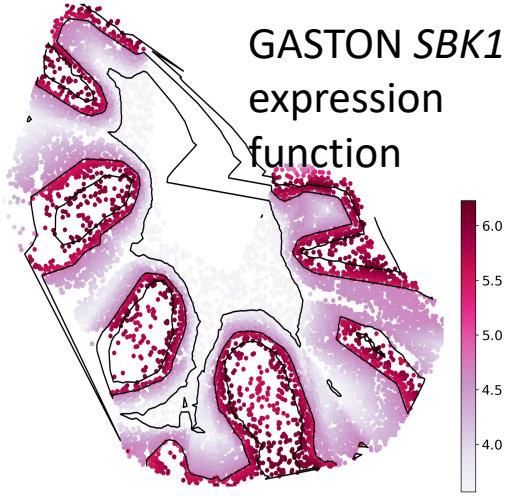




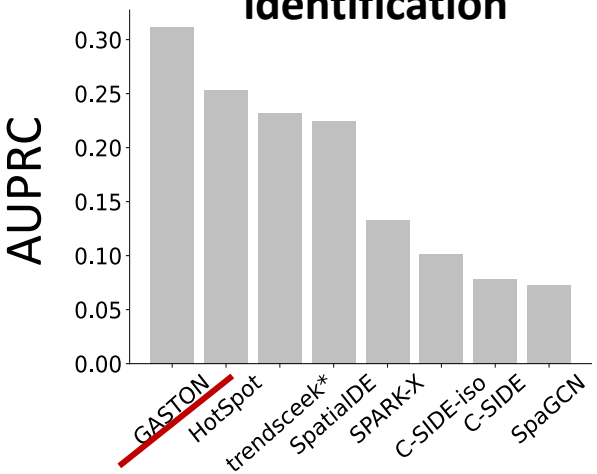
Baizer, Front. Hum. Neurosci. 2014







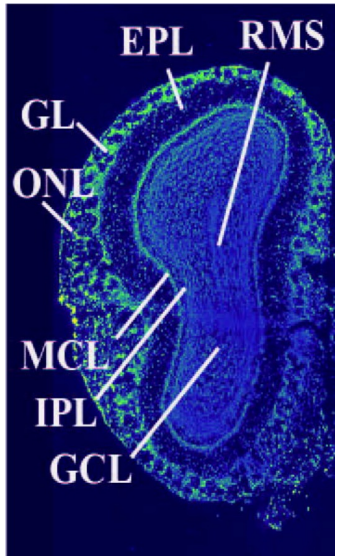
Marker gene identification



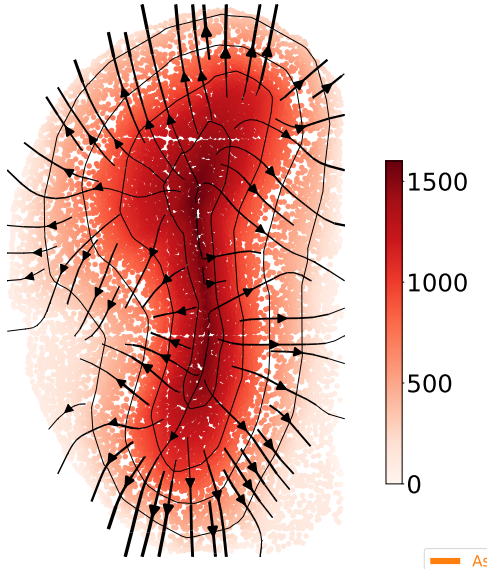
Olfactory bulb (Stereo-seq) 9,825 spots × 27,106 genes

- Olfactory nerve layer (ONL)
- Glomerular layer (GL)
- External plexiform layer (EPL)
- Mitral cell layer (MCL)
- Internal plexiform layer (IPL)
- Granule cell layer (GCL)
- Rostral migratory stream (RMS)

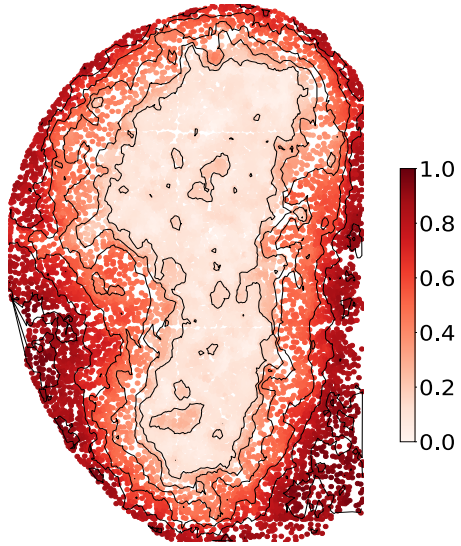
DAPI Stain



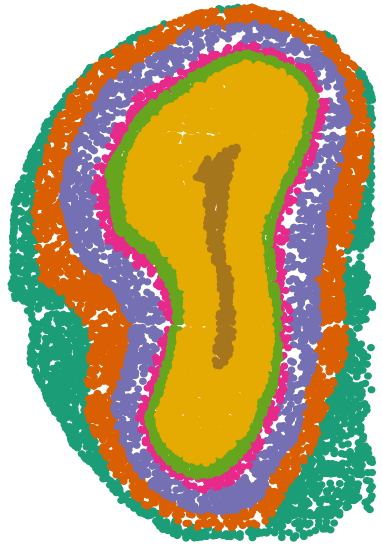
Isodepth and (negative) spatial gradients



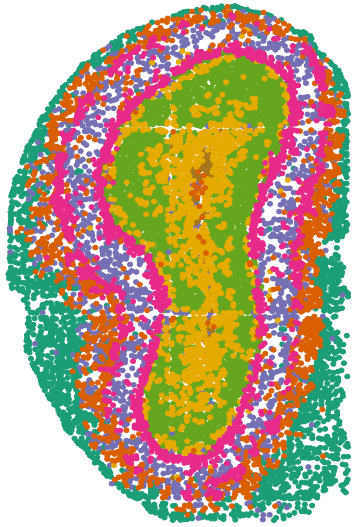
SpaceFlow (diffusion pseudotime)

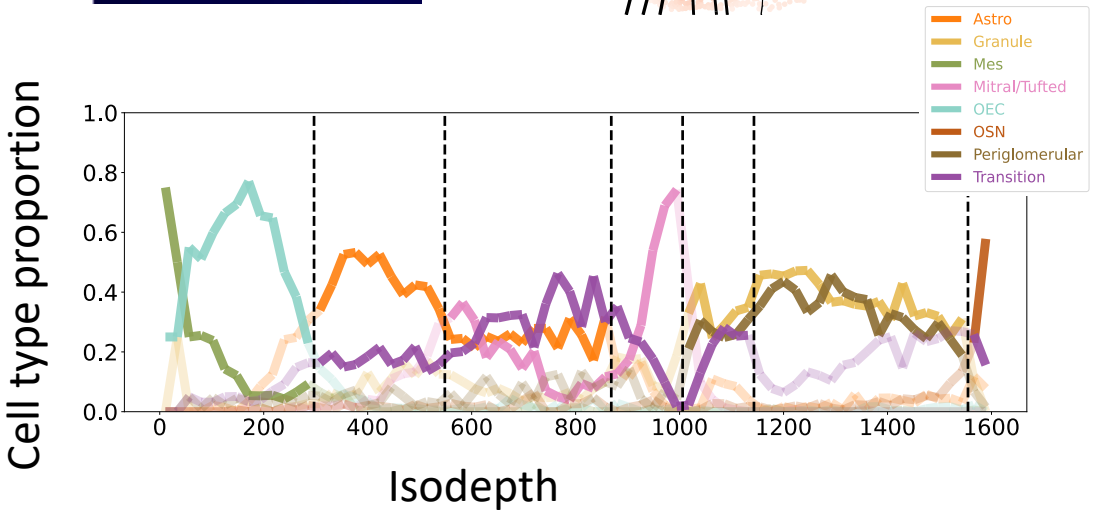


GASTON

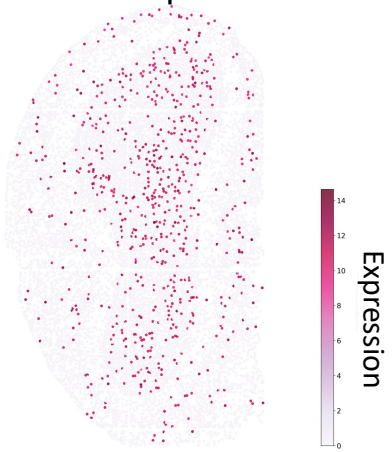


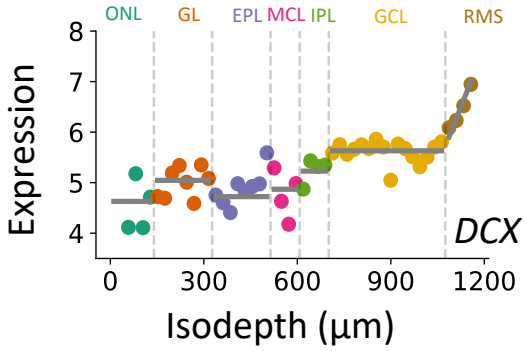
SpaGCN



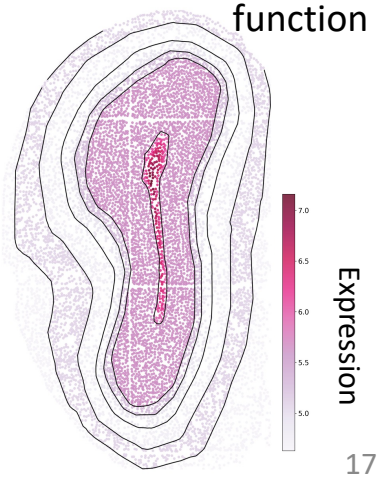


DCX expression



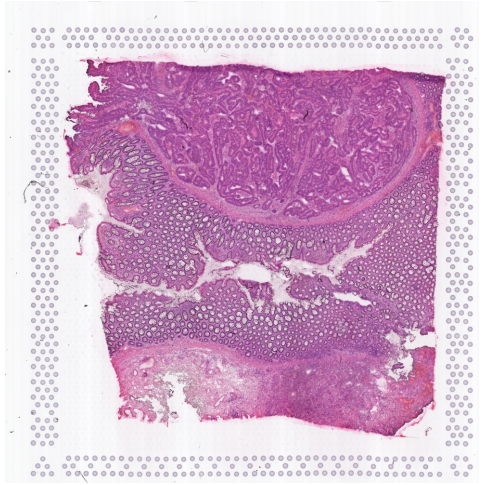


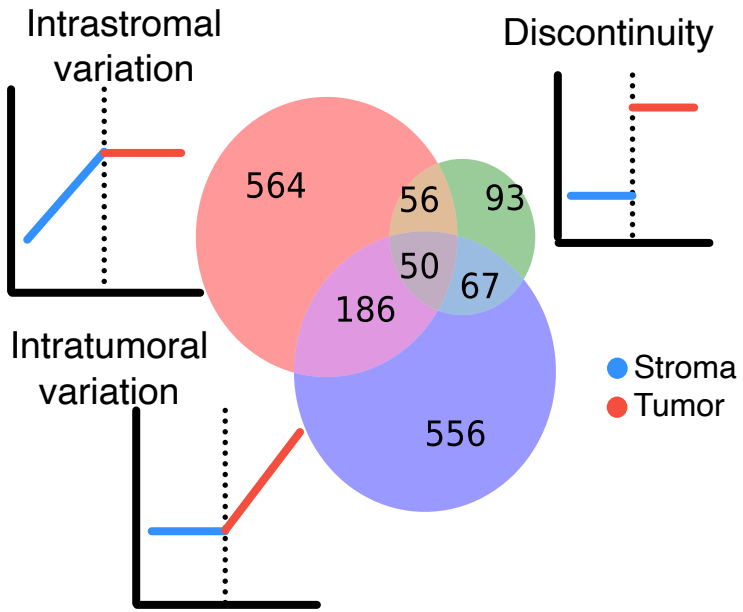
GASTON *DCX* expression function



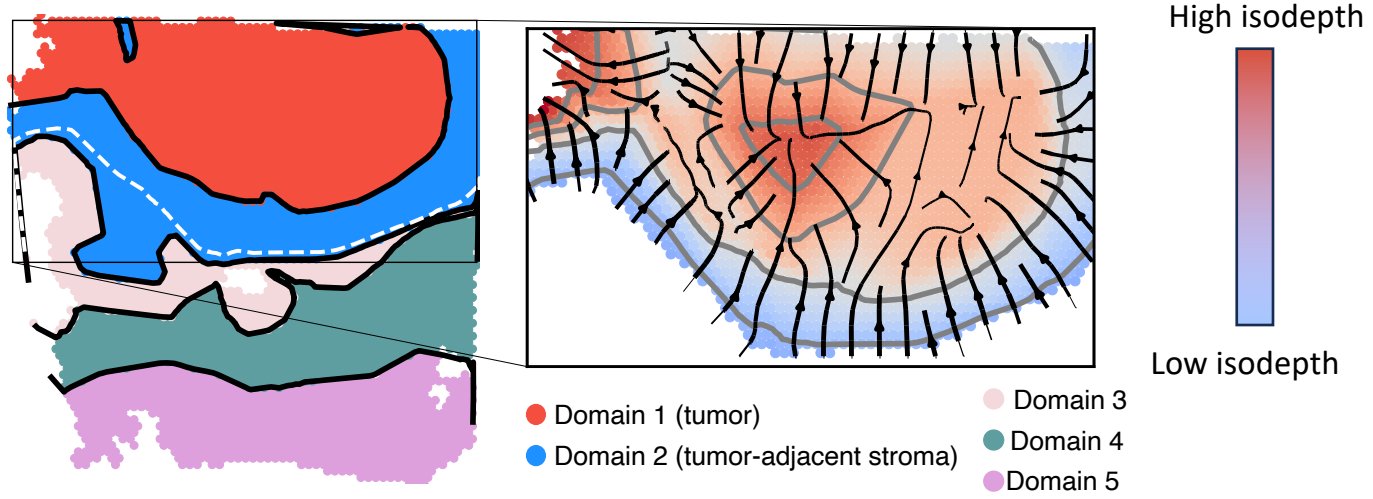
GASTON identifies gradients in tumor microenvironment

Colorectal tumor slice (stage IV)
(Wu et al, Cancer Discovery 2022)

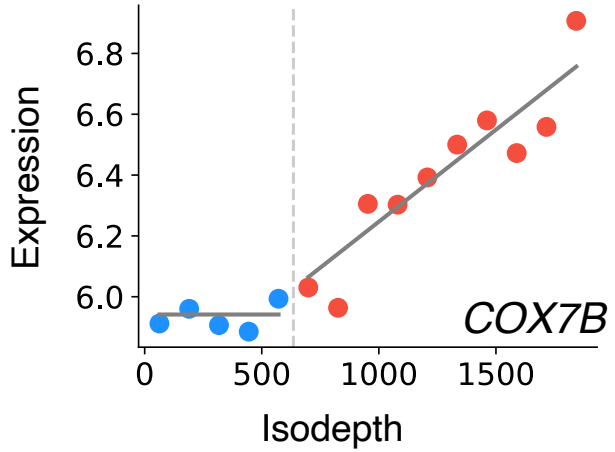




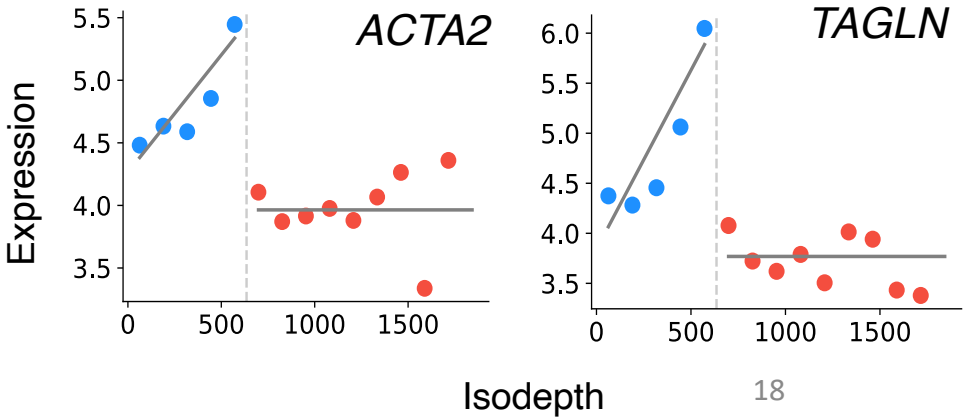
GASTON: spatial domains + isodepth



Intratumoral variation
aerobic metabolism



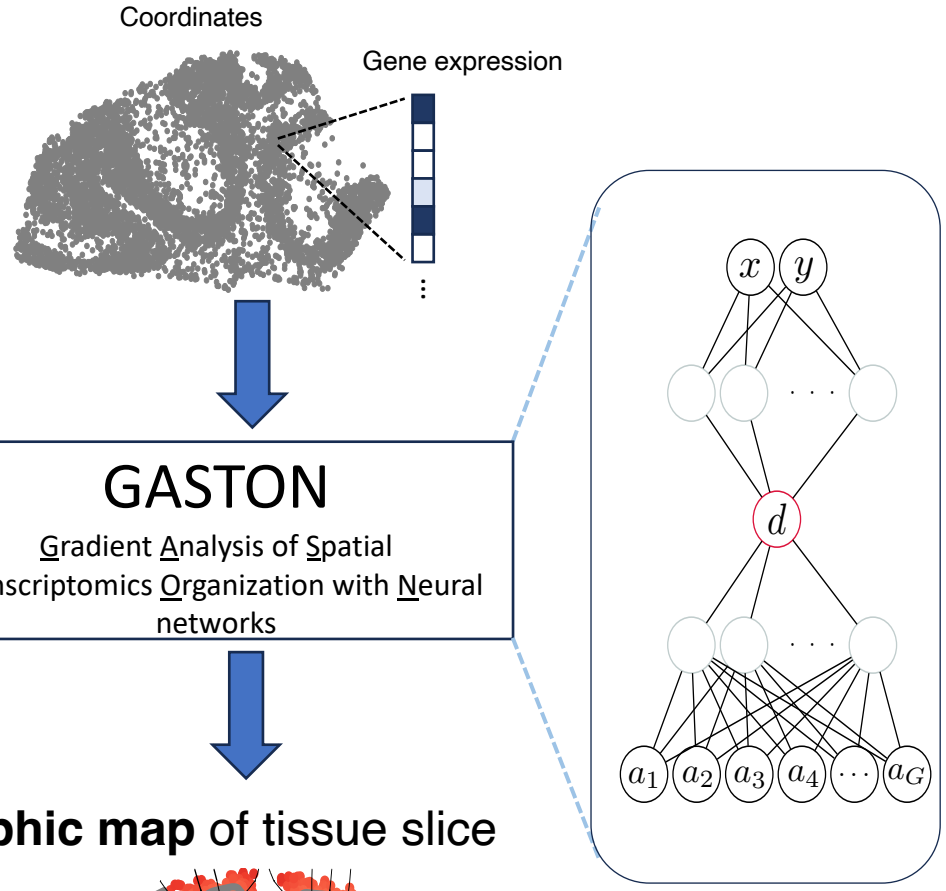
Discontinuity/Intrastromal variation-
epithelial-mesenchymal transition (EMT)



Summary: GASTON

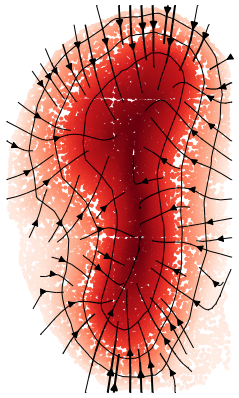
- Isodepth describes **topographic map** and **spatial gradients** of gene expression within tissue slice

- GASTON: **unsupervised** deep learning algorithm to learn isodepth
  - Uncovers spatial domains and gradients of gene expression/cell type

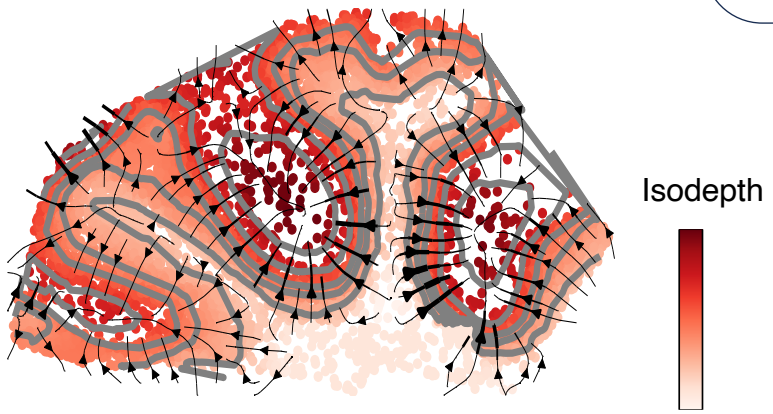


Paper

Code

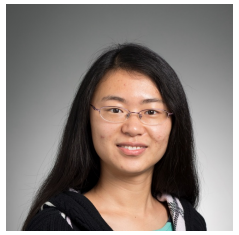


Topographic map of tissue slice



Acknowledgments

Ben Raphael



Cong Ma

Brian Arnold

Hirak Sarkar

Sereno Lopez-Darwin

Kohei Sanno

Other Raphael lab members:

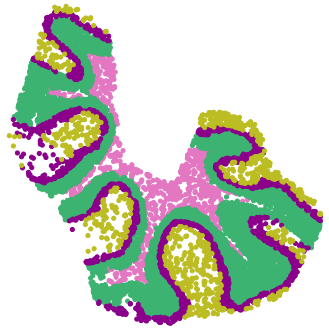
Hongyu Zheng
Palash Sashittal
Uyen Mai
Metin Balaban
Julian Gold
Richard Zhang
Ahmed Shuaibi
Alexander Strzalkowski
Henri Schmidt
Xinhao Liu
Akhil Jakatdar
Gary Hu
Peter Halmos
Gillian Chu
Clover Zheng

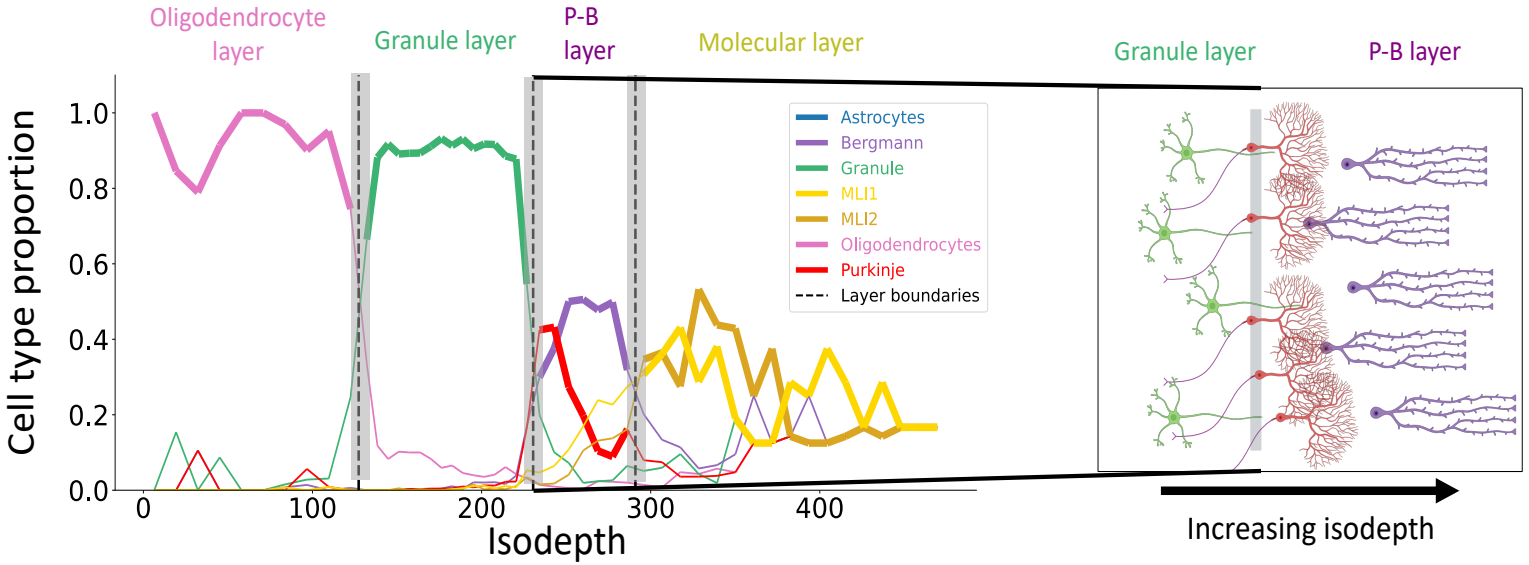
NSF GRADUATE RESEARCH FELLOWSHIP PROGRAM

Siebel Scholars

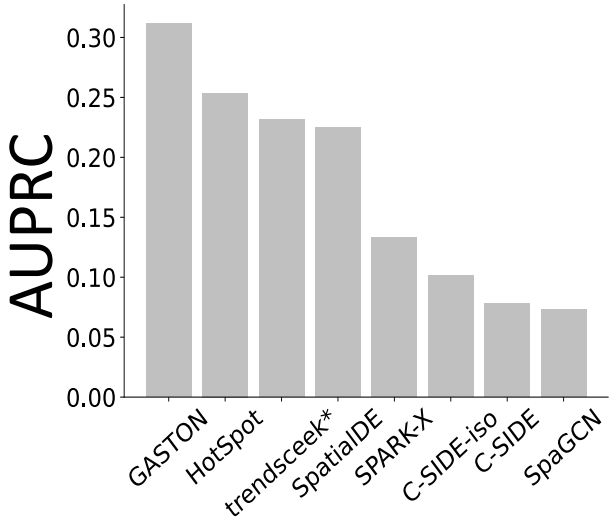
Cell type and gene expression gradients

GASTON

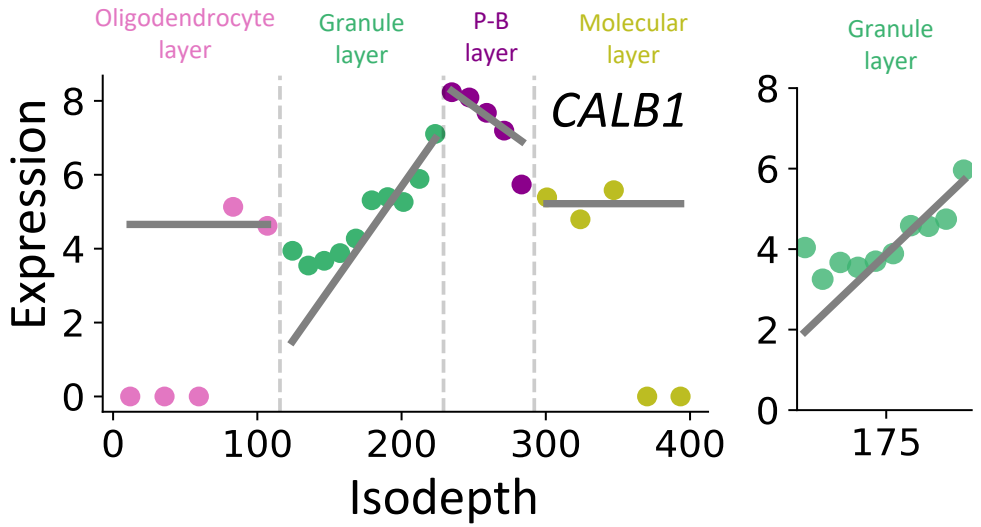




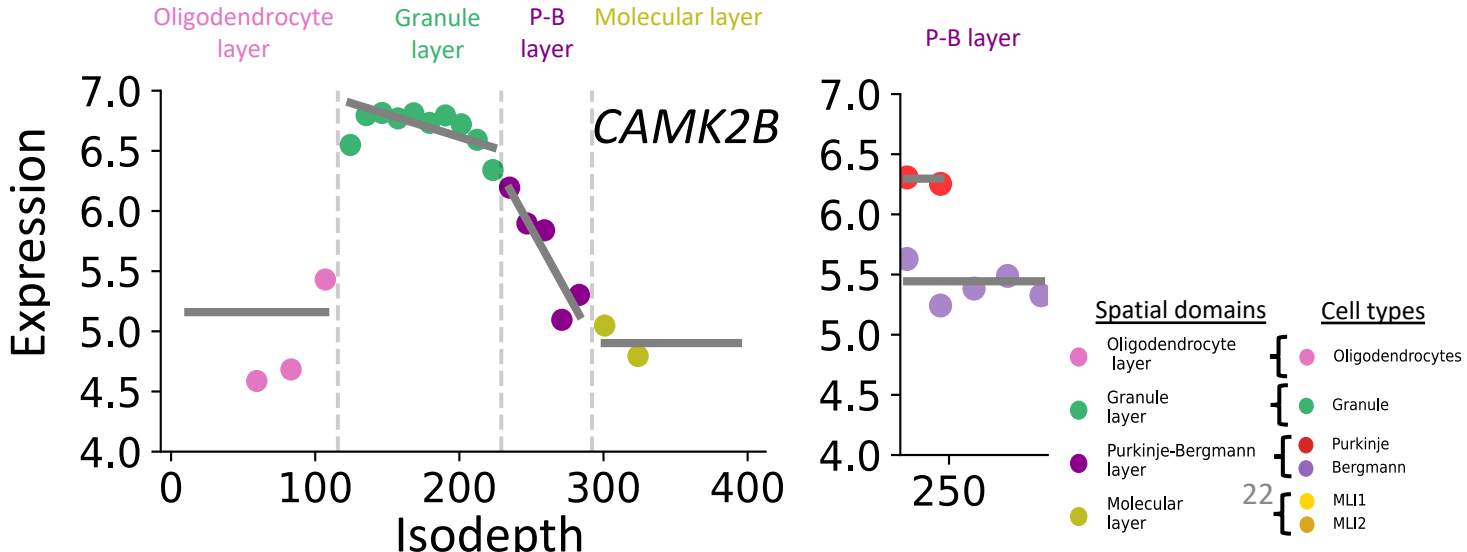
Marker gene identification



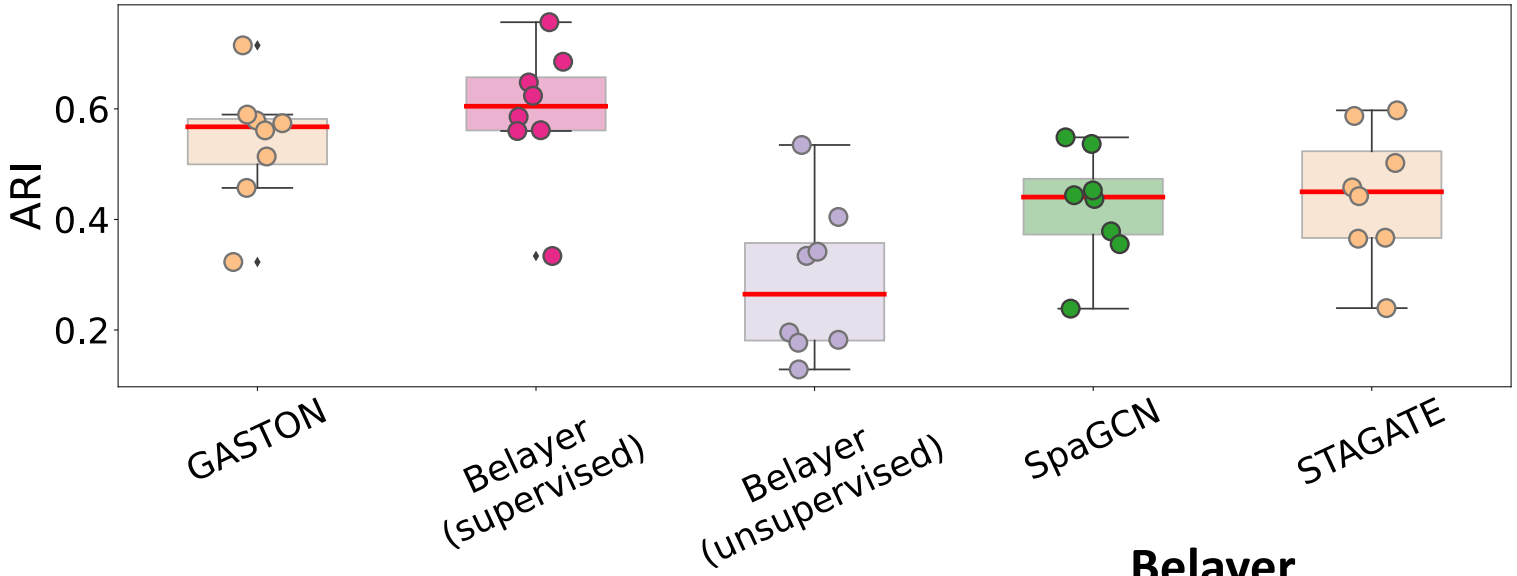
Cell type-attributable gradient



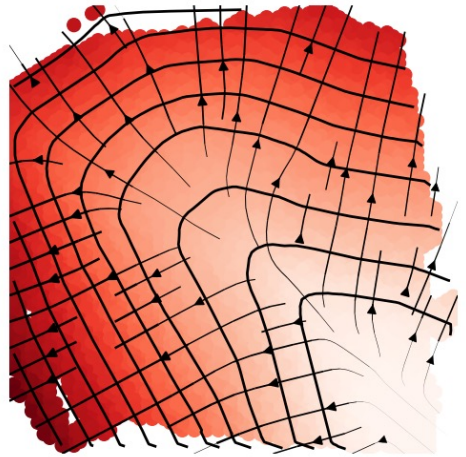
Other-attributable gradient



Human DLPFC: GASTON outperforms other neural networks and unsupervised Belayer

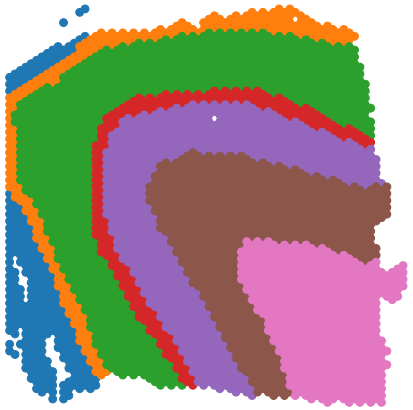


GASTON isodepth



0.99 correlation with (supervised) Belayer depth!

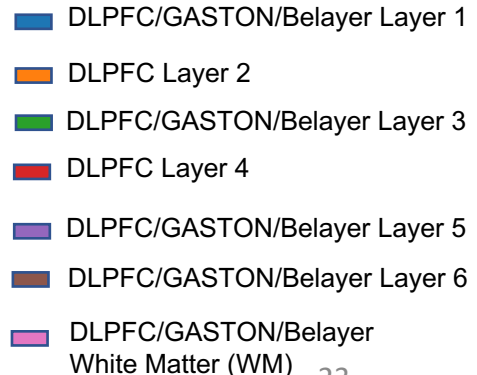
Manual annotation



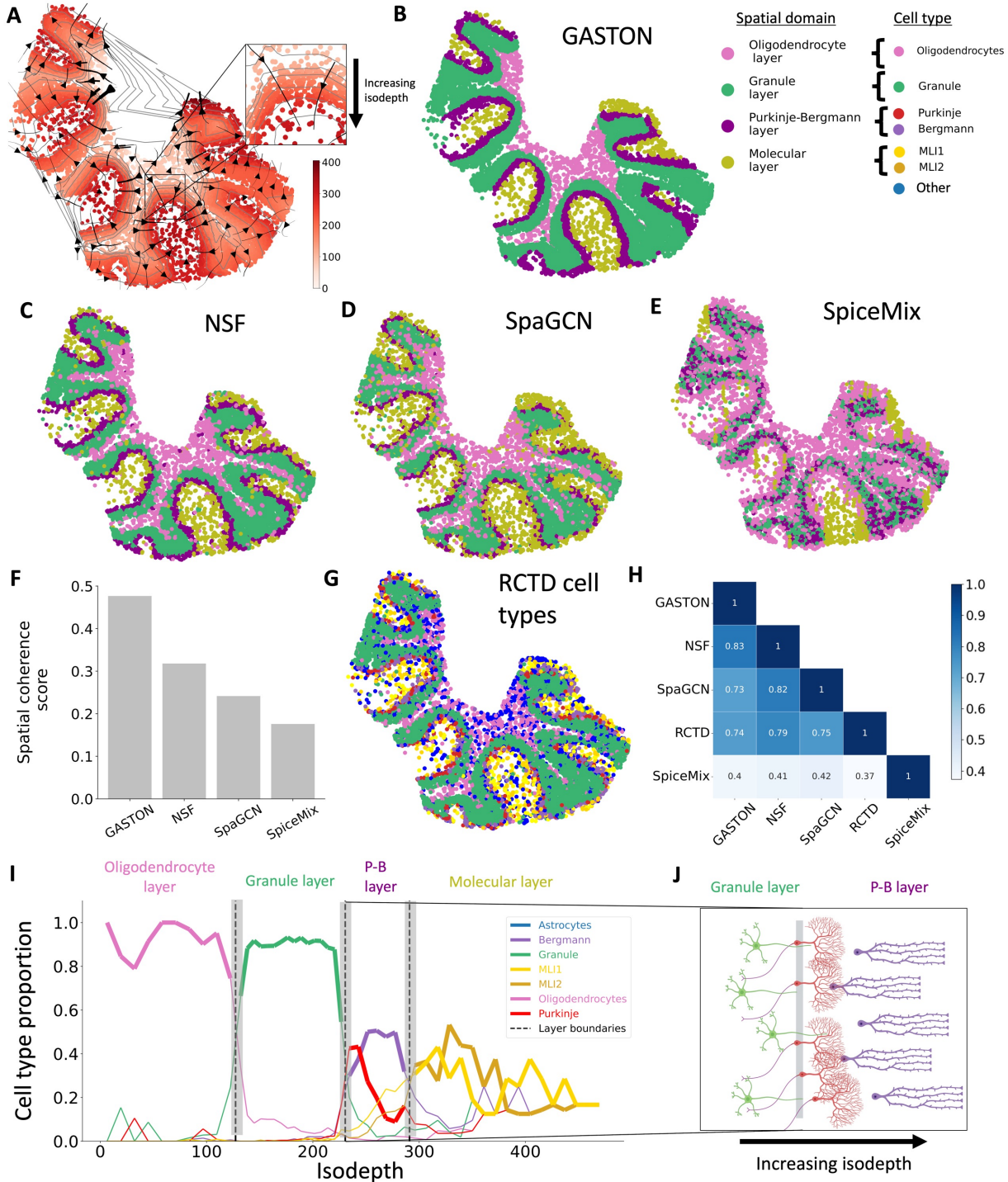
GASTON

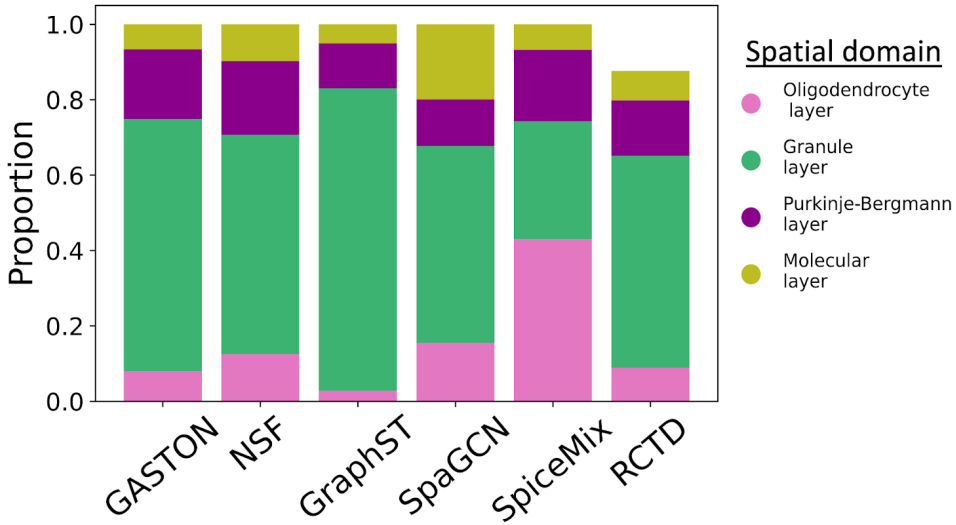
Belayer (supervised)

Belayer (unsupervised)

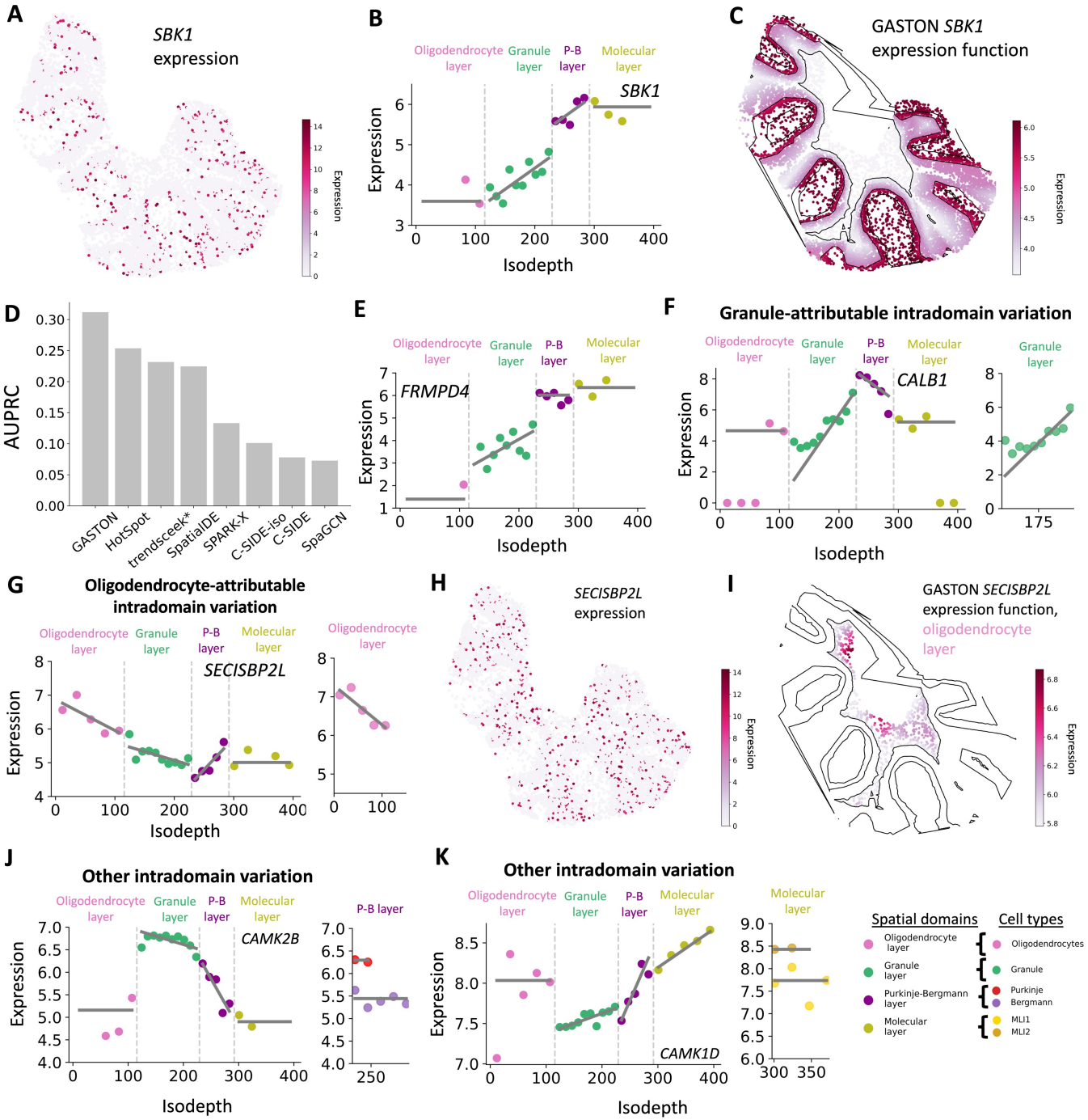


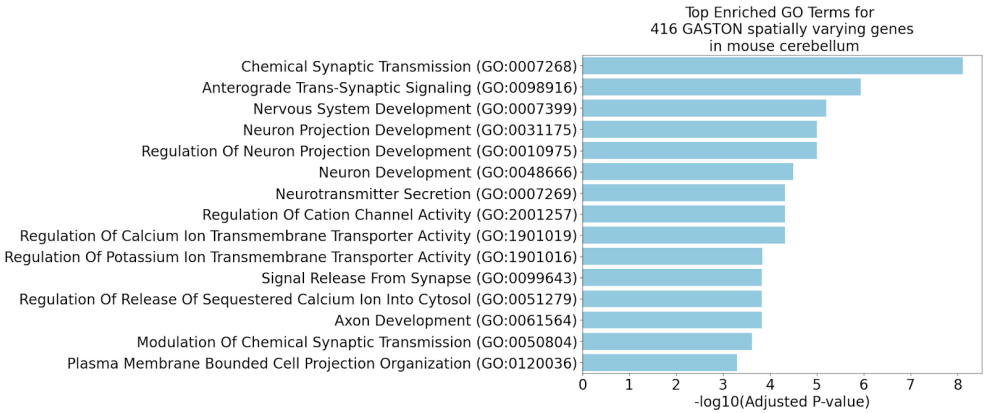
GASTON – cerebellum (spatial domains)



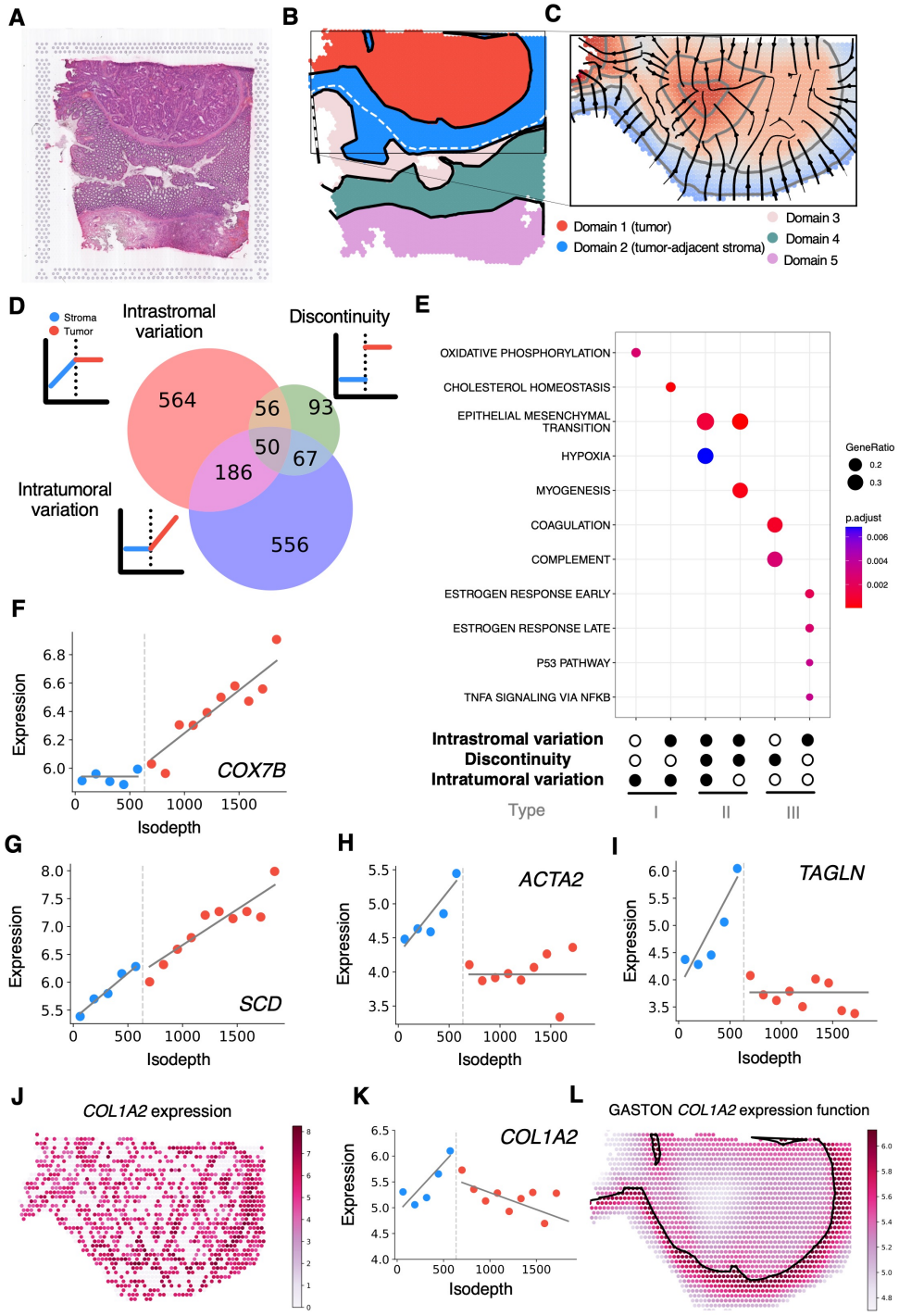


GASTON – cerebellum (spatial expression patterns)

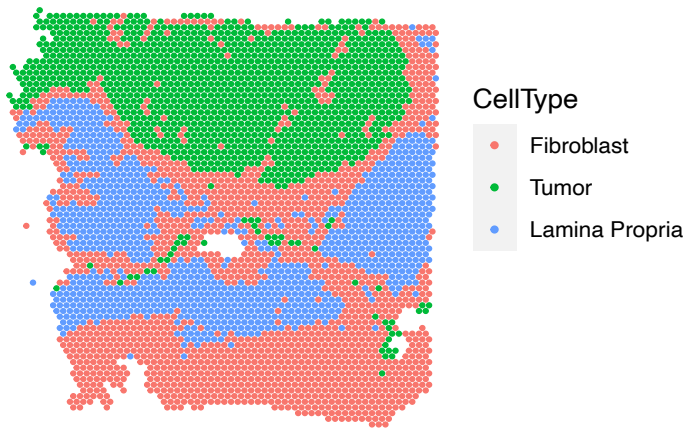




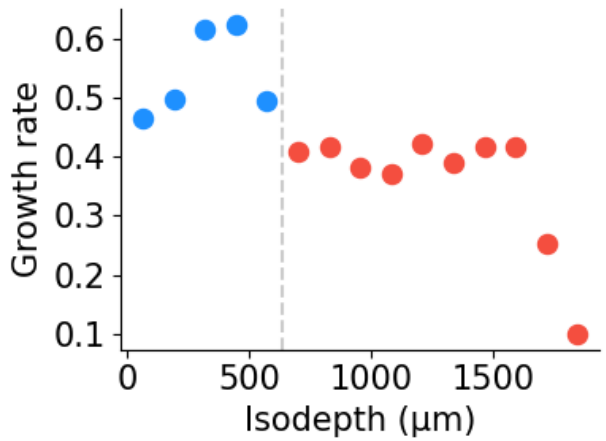
GASTON – colorectal tumor



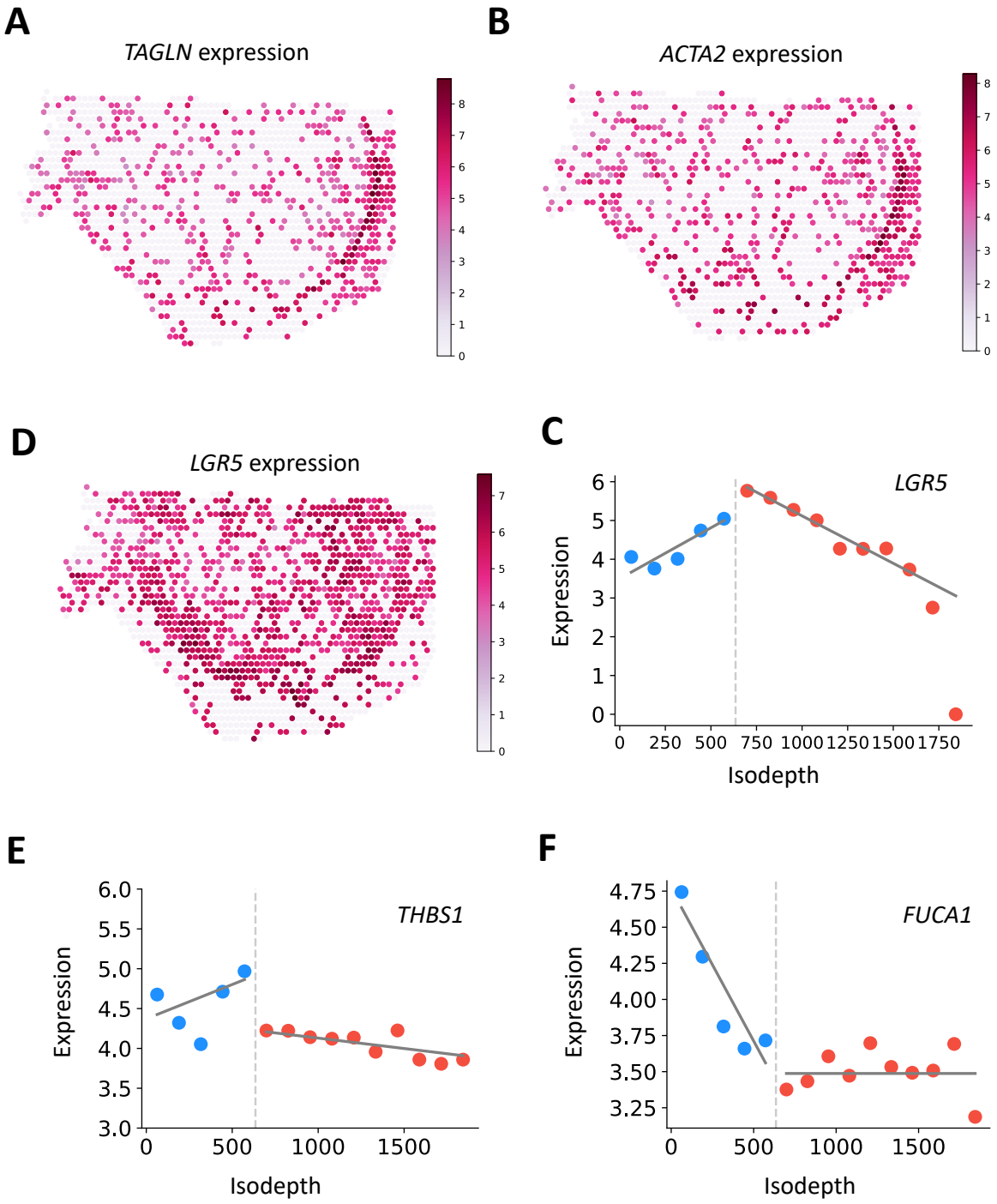
Seurat cell types



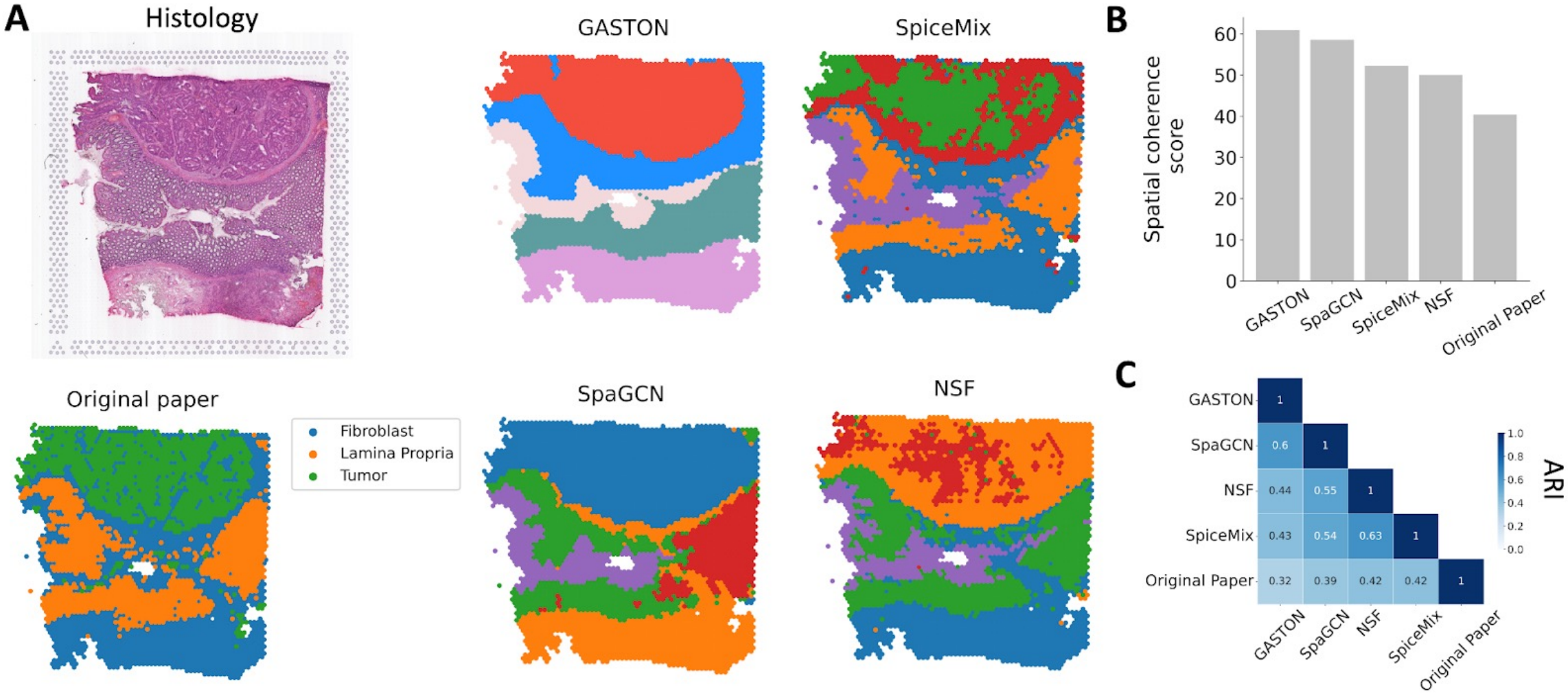
Tumor growth rate



GASTON – colorectal tumor (more patterns)



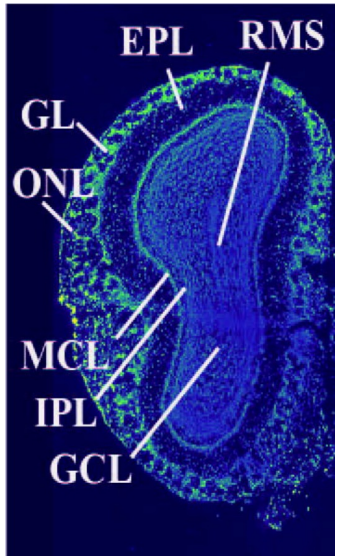
Comparison of domains on colorectal tumor



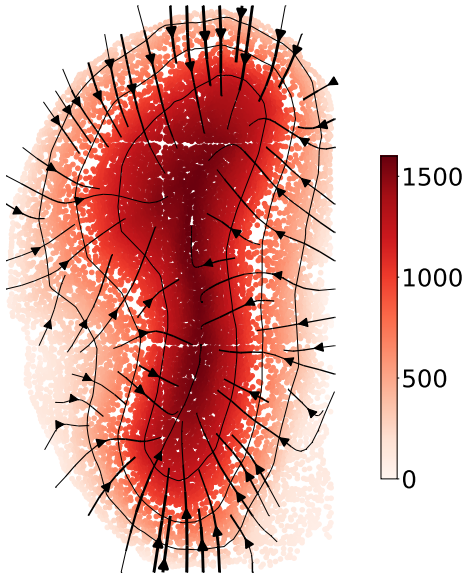
Olfactory bulb (Stereo-seq) 9,825 spots × 27,106 genes

- Olfactory nerve layer (ONL)
- Glomerular layer (GL)
- External plexiform layer (EPL)
- Mitral cell layer (MCL)
- Internal plexiform layer (IPL)
- Granule cell layer (GCL)
- Rostral migratory stream (RMS)

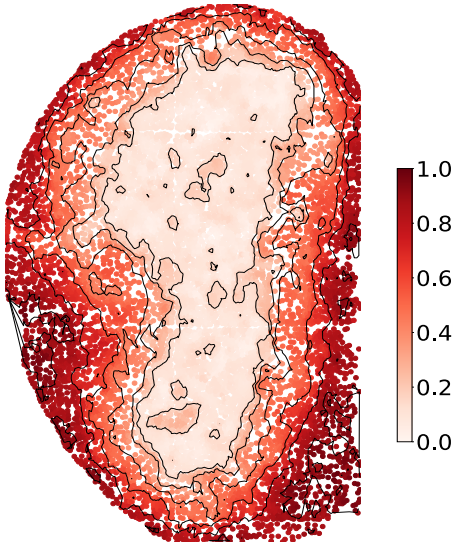
DAPI Stain



Isodepth and spatial gradients

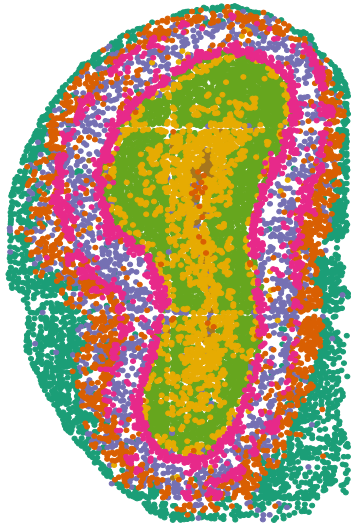


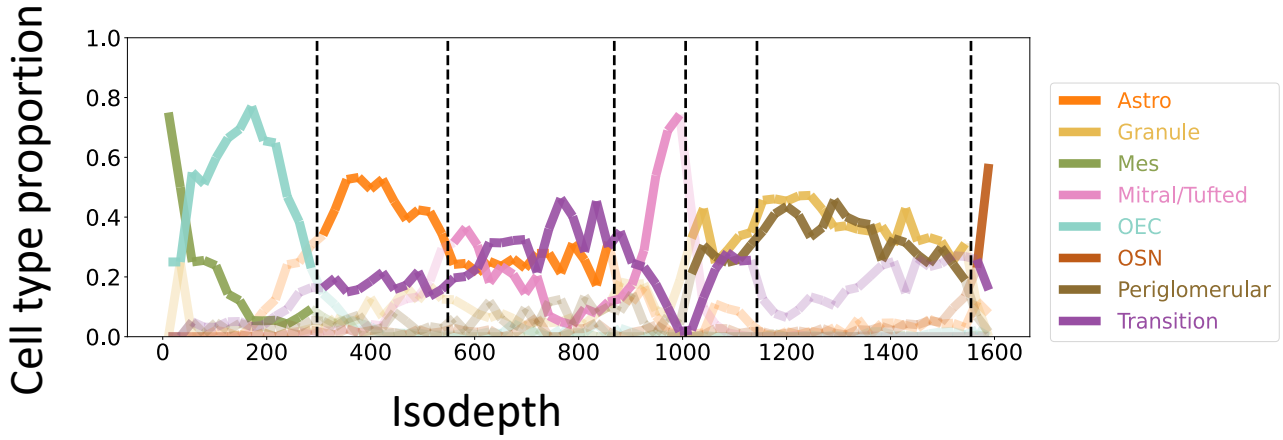
SpaceFlow (diffusion pseudotime)



GASTON

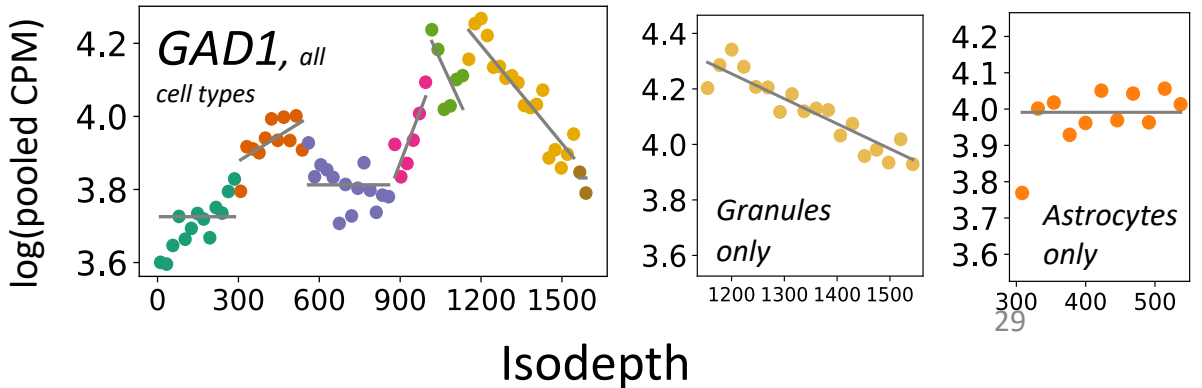
SpaGCN



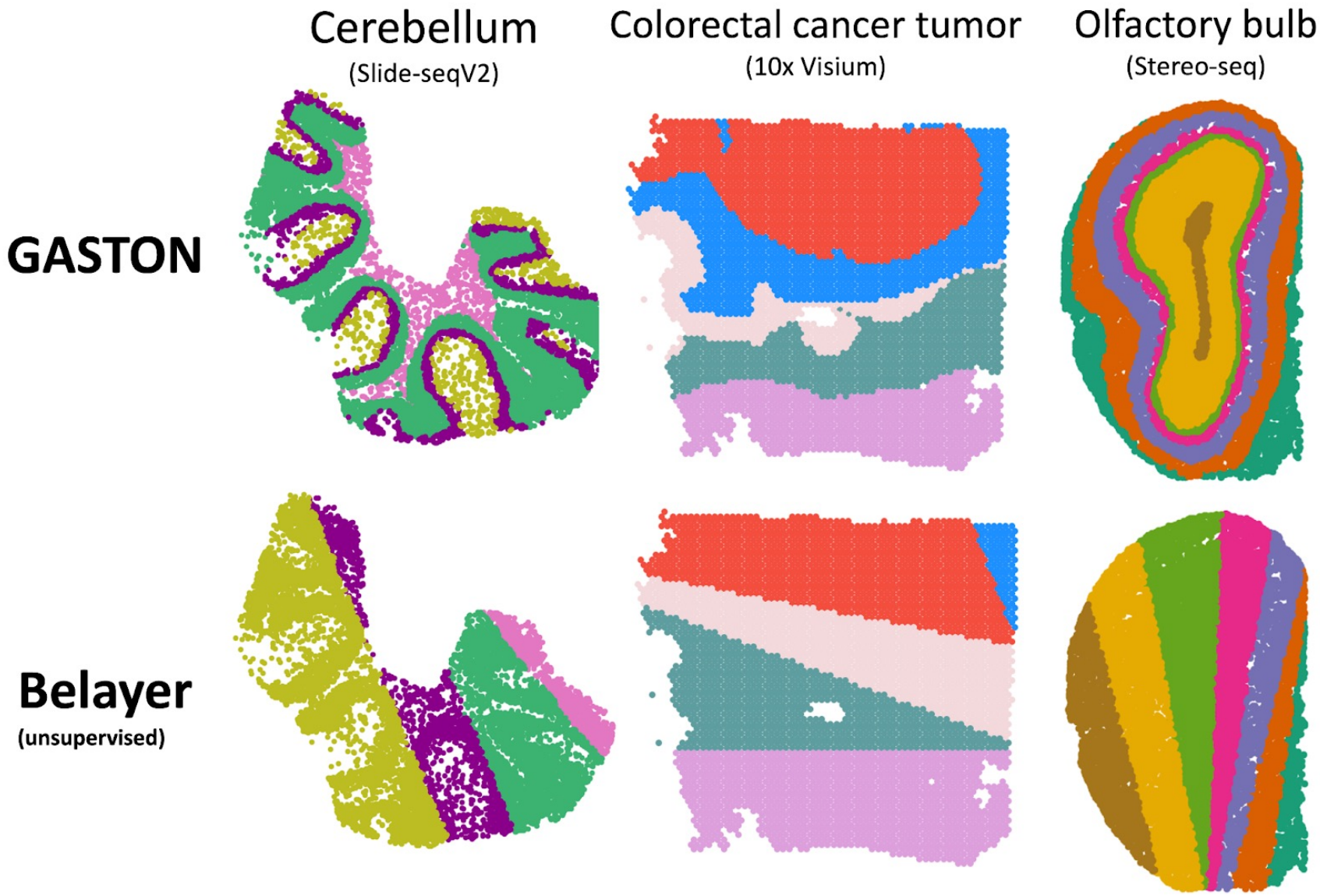


Cell type-*attributable* gradient

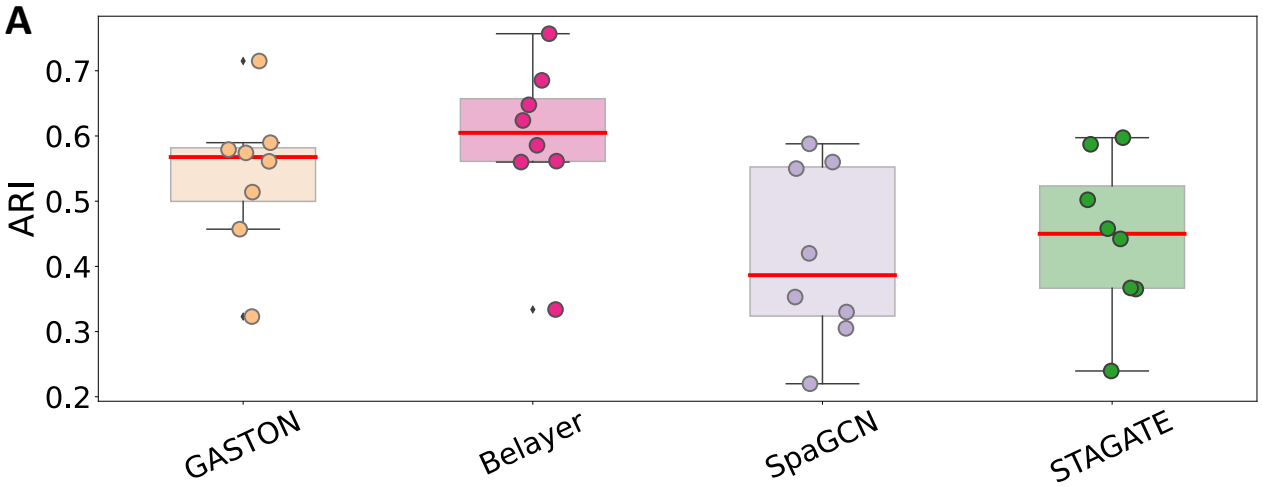
Other attributable gradient

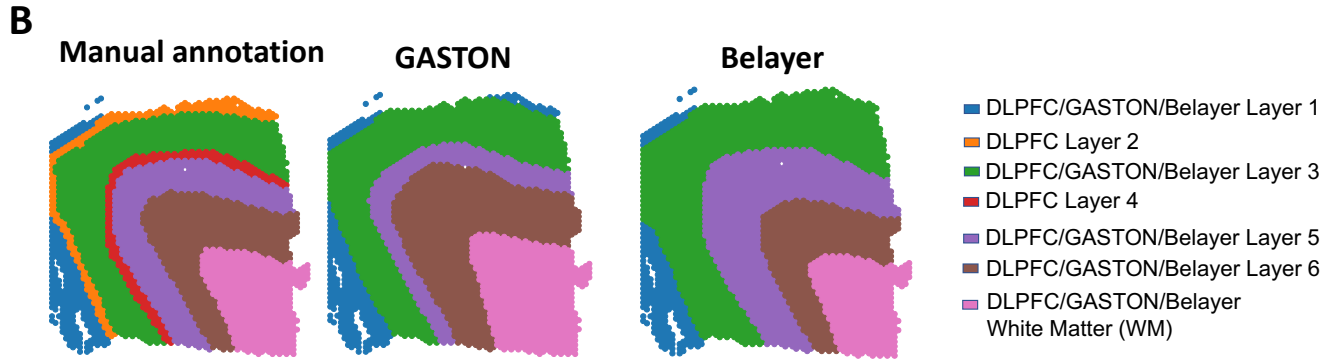


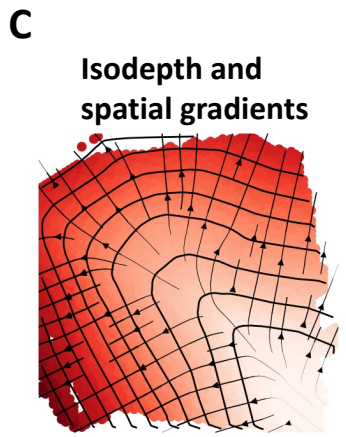
Comparison b/w GASTON and Belayer

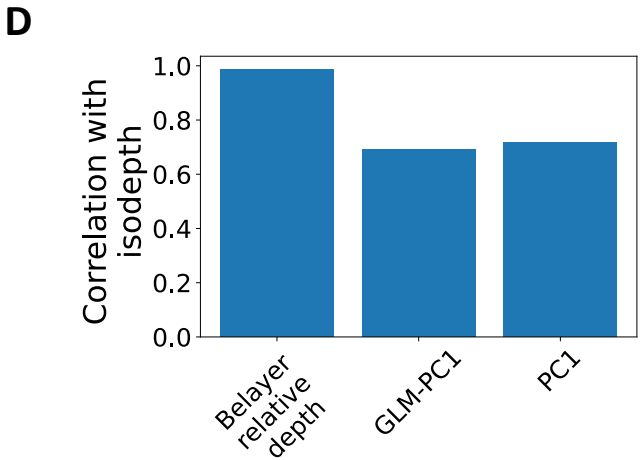


GASTON – DLPFC

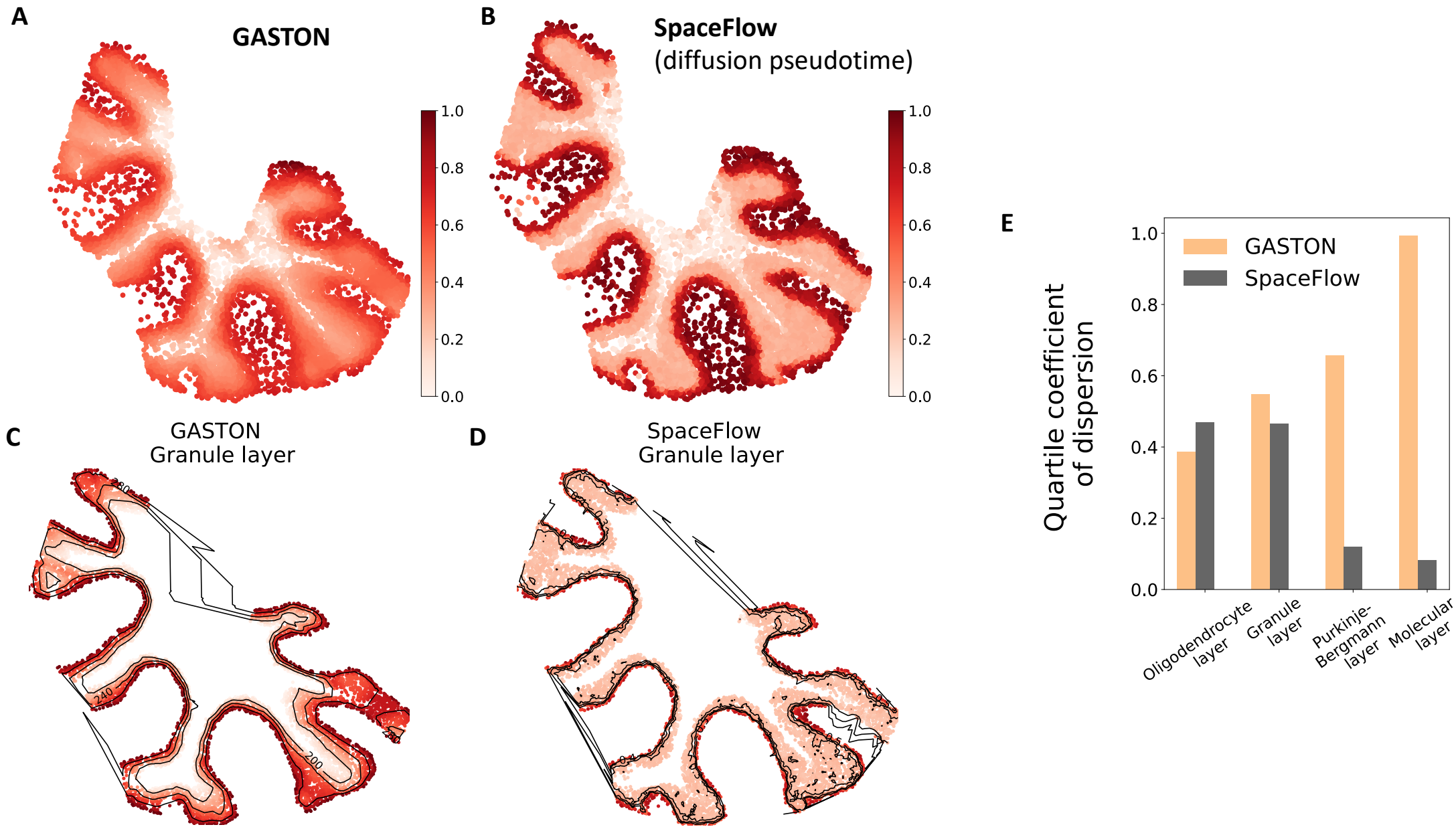




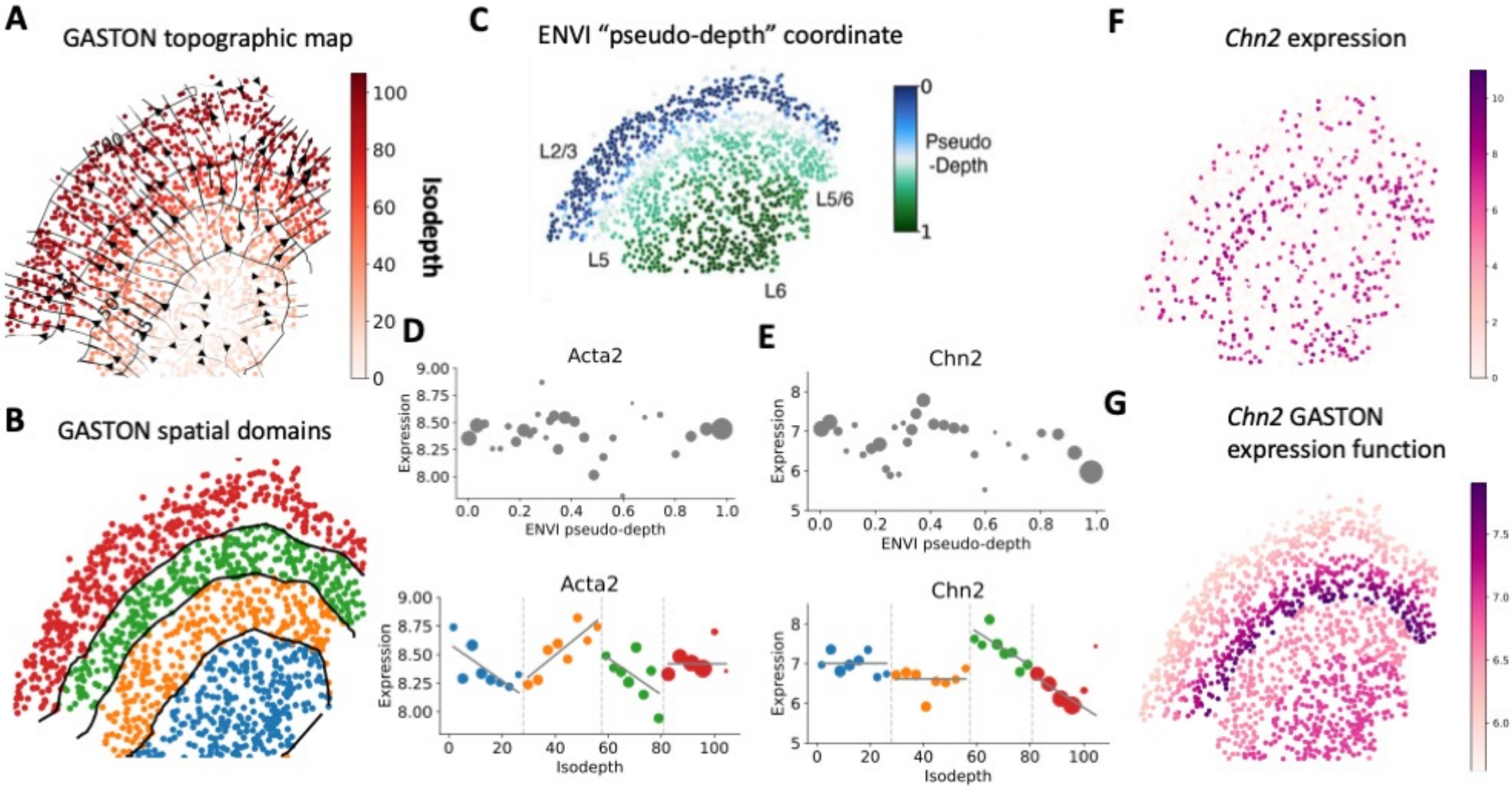




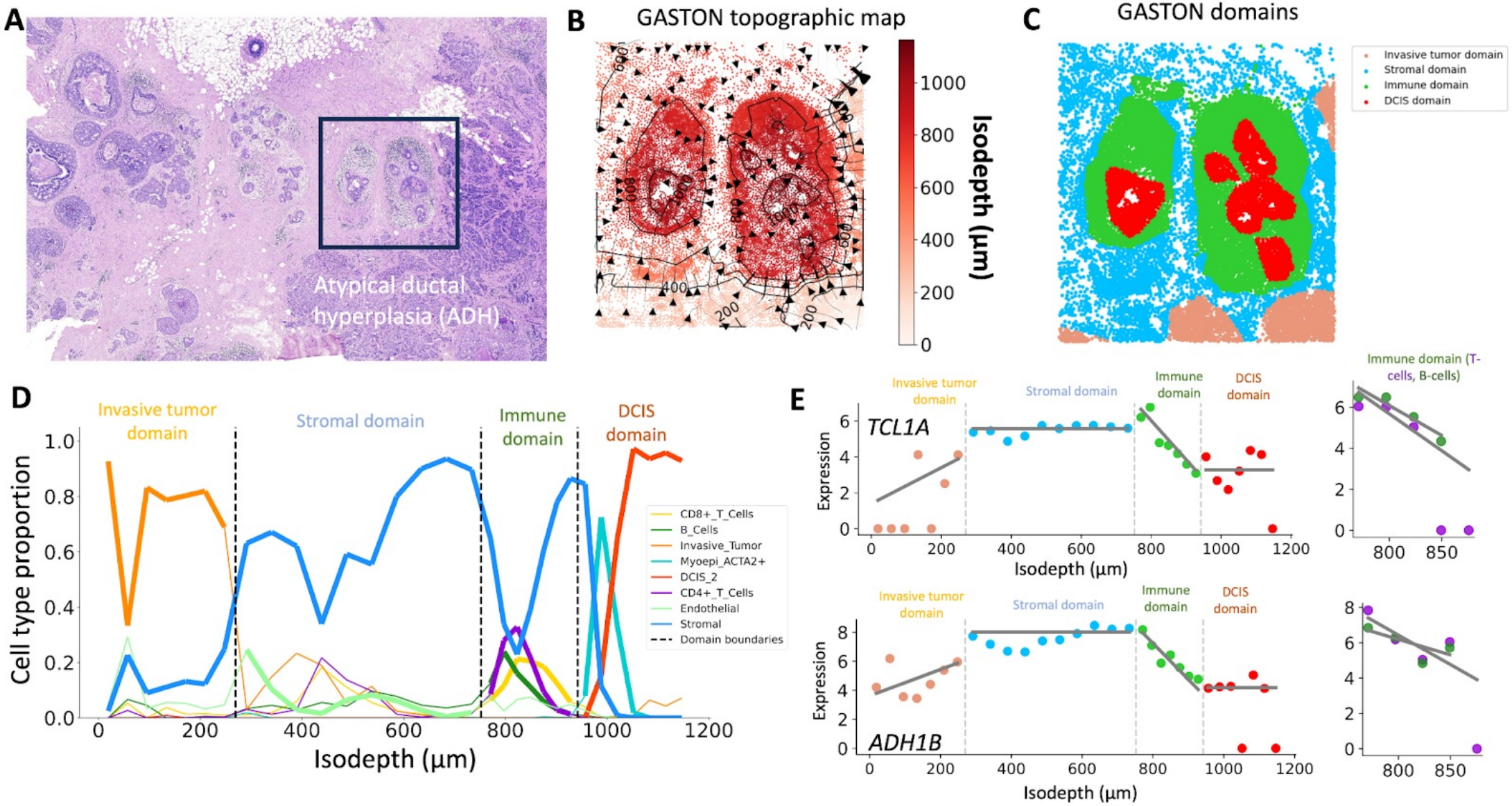
GASTON – SpaceFlow comparison (cerebellum)

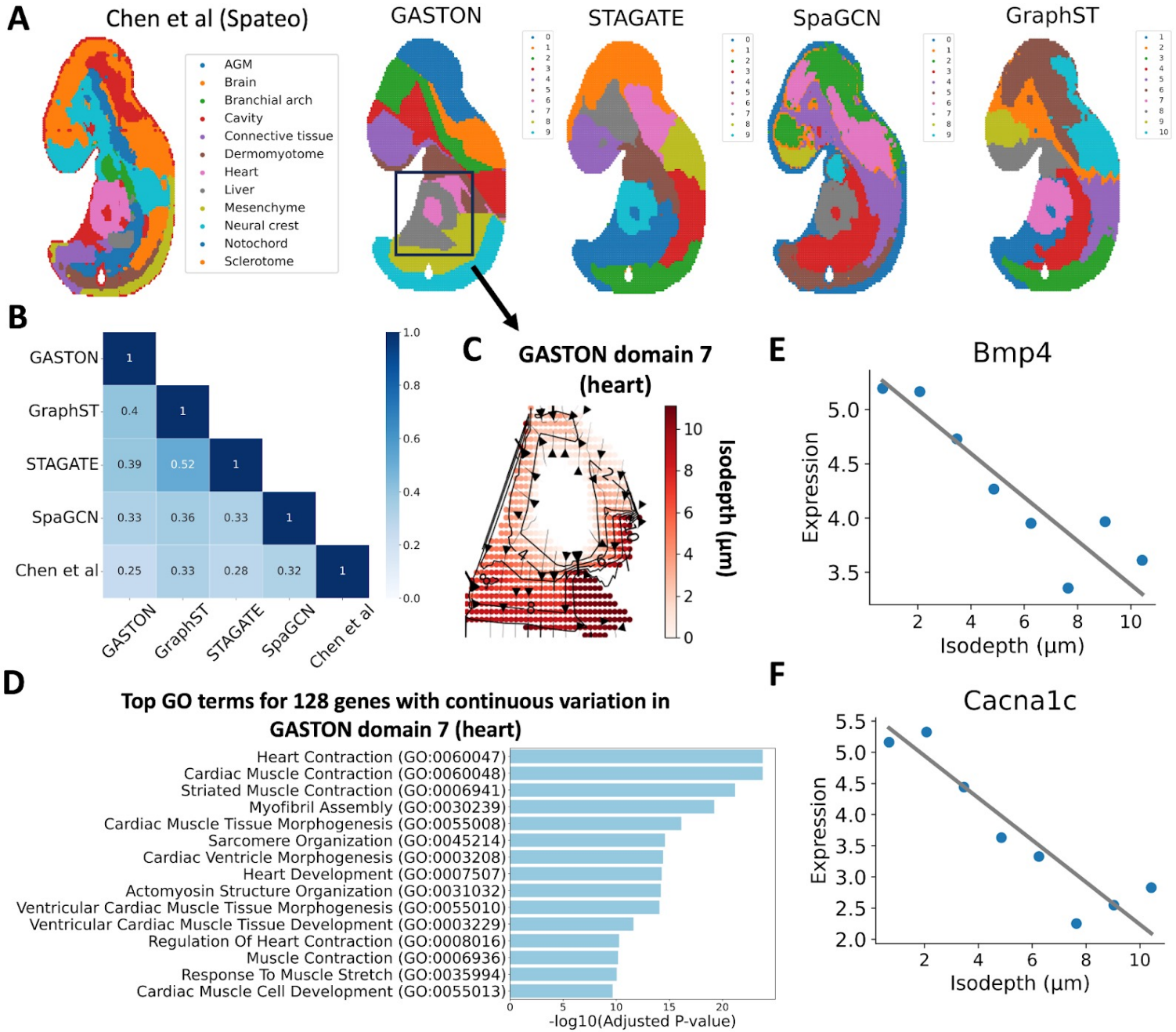


GASTON – mouse primary motor cortex (MERFISH)



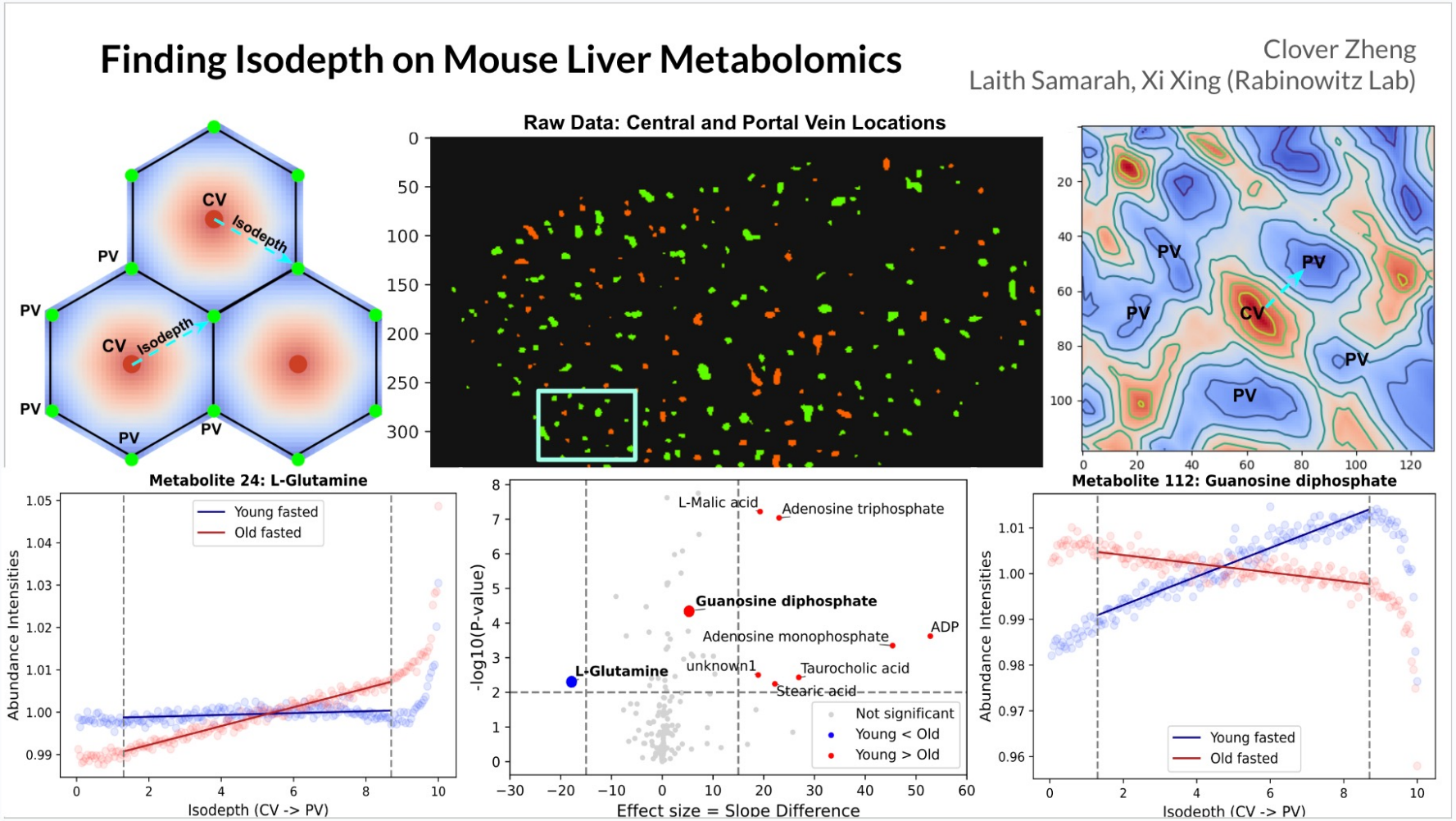
GASTON – breast cancer (10X Genomics Xenium)





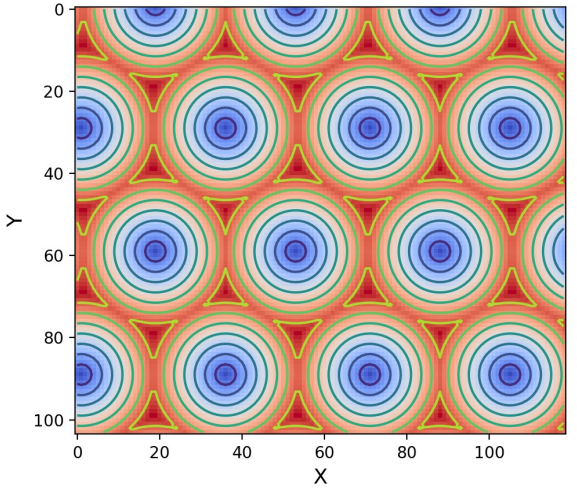
GASTON – mouse embryo day 9.5 (Stereo-seq)

Application of GASTON to metabolomics (Clover Zheng)

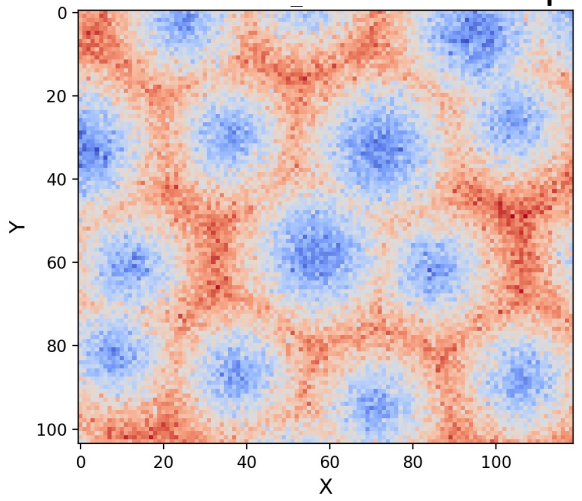


Testing GASTON w/ simulated hexagonal geometries:

True isodepth d(x,y)



GASTON-estimated isodepth



GASTON – model selection (elbow)