Integrating Spatial Information in Single-Cell Transcriptomics Analysis

Guo-Cheng Yuan

Department of Pediatric Oncology
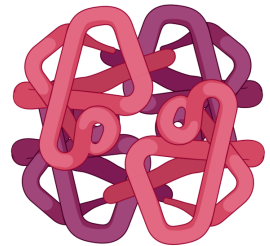
Dana-Farber Cancer Institute

Harvard Medical School
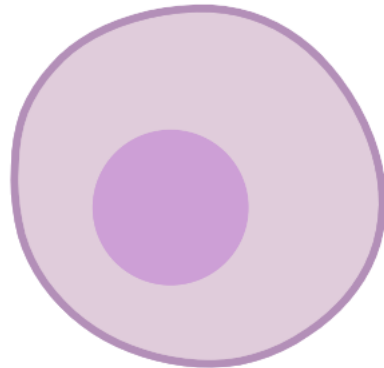
A complete cell state contains many components
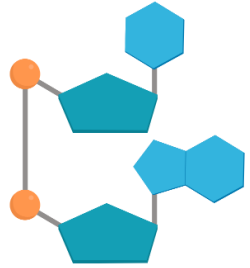
Different components are interconnected
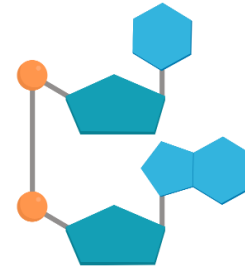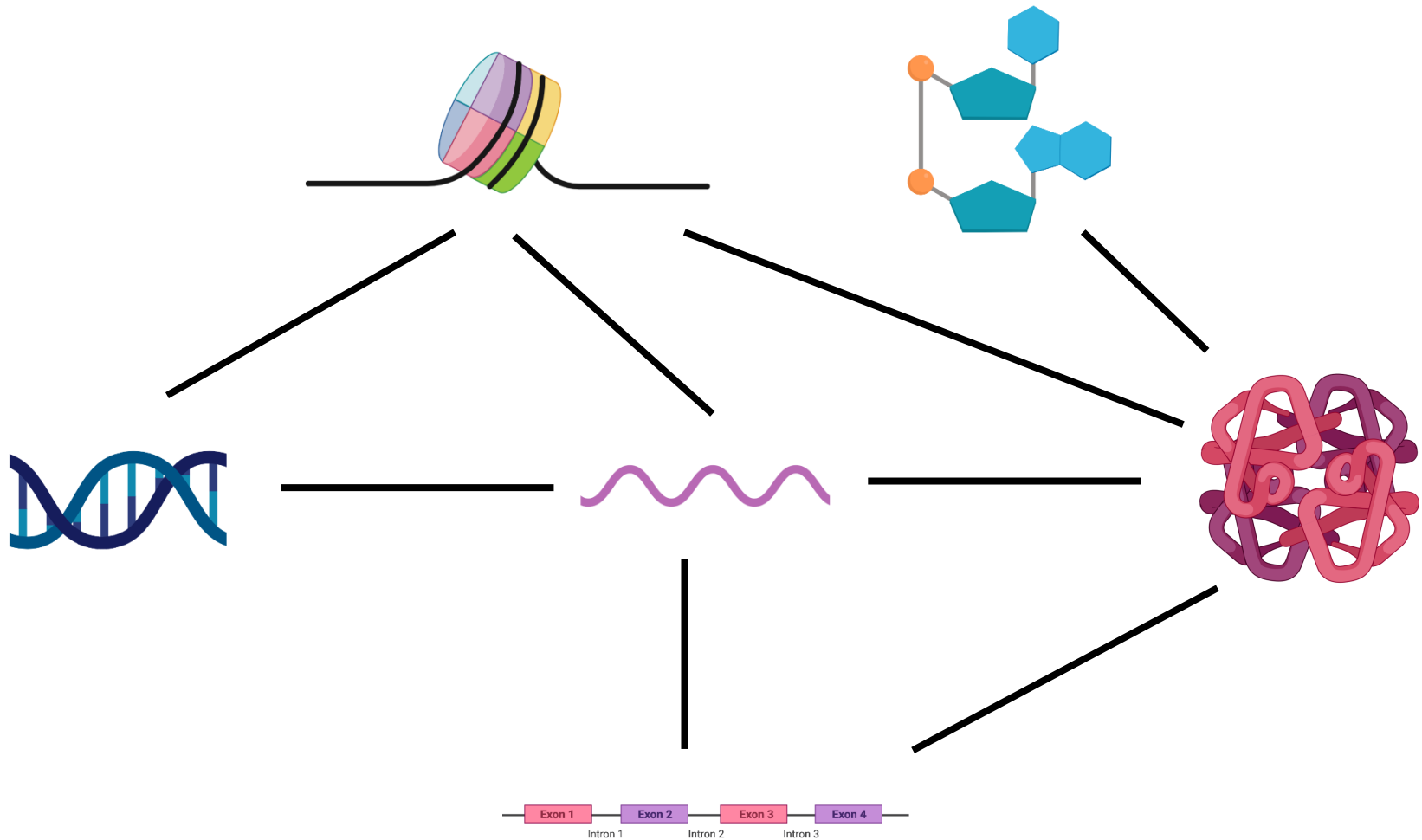

Exon 1 Exon 2 Exon 3 Exon 4

Intron 1 Intron 2 Intron 3
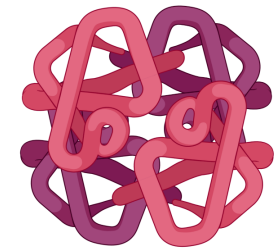
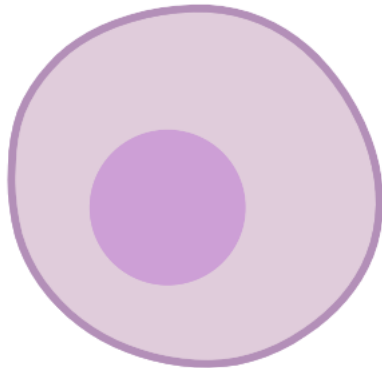# Different components are interconnected
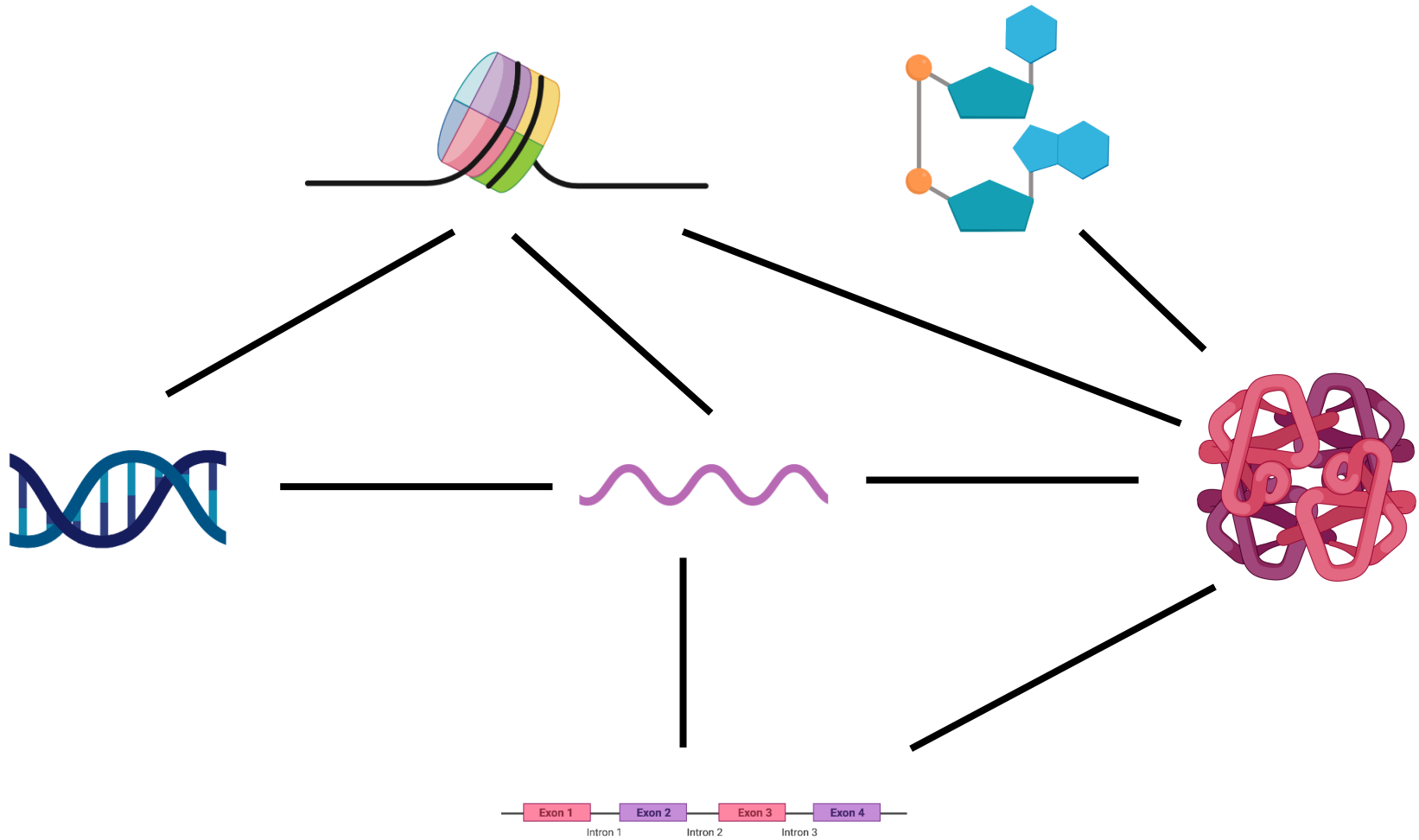
Different components are interconnected

How to represent an integrated cell state?
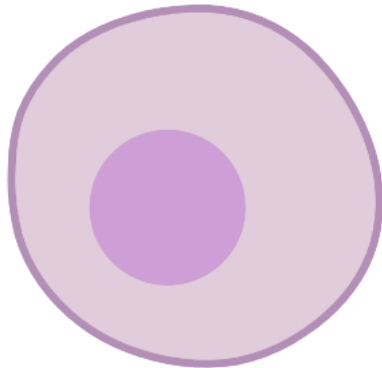
Few components can be measured in the same cell

	RNA level
Gene 1	
Gene 2	
Gene 3	
Gene 4	

	Protein level
Gene 1	
Gene 2	
Gene 3	
Gene 4	

	methylaton
CpG 1	
CpG 2	
CpG 3	
CpG 4	

	Accessibility
Bin 1	
Bin 2	
Bin 3	
Bin 4	

Few components can be measured in the same cell

	RNA level
Gene 1	
Gene 2	
Gene 3	
Gene 4	

	Protein level
Gene 1	
Gene 2	
Gene 3	
Gene 4	

	methylaton
CpG 1	
CpG 2	
CpG 3	
CpG 4	

	Accessibility
Bin 1	
Bin 2	
Bin 3	
Bin 4	

I can only measure X. Can I use it to predict Y?

Why mathematical modeling is important

- Conceptual: Modeling the system.
  - Causal inference
  - Factor analysis
  - Biological networks
  - Multi-scale modeling
  - Dynamical Systems
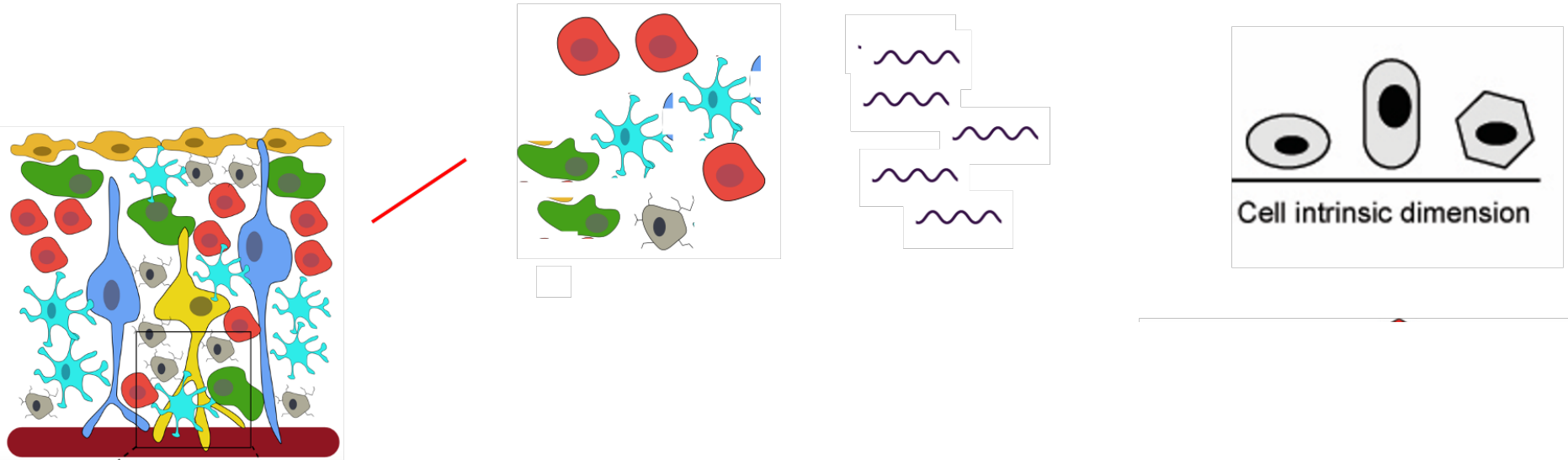
Why mathematical modeling is important

- Practical: Predicting the unknowns.
  - Supervised vs unsupervised
  - Bayesian vs frequentist
  - Model based vs data driven
  - Statistical vs machine learning
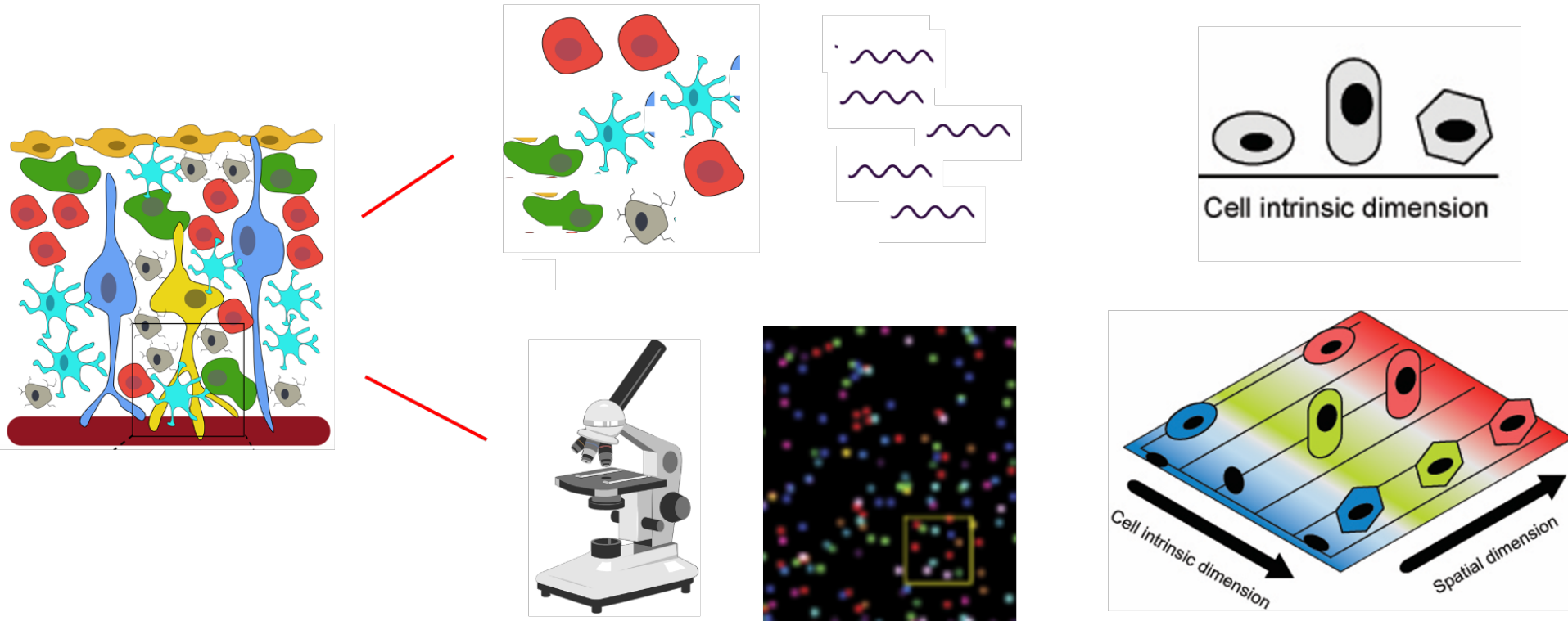
Integration of spatial information
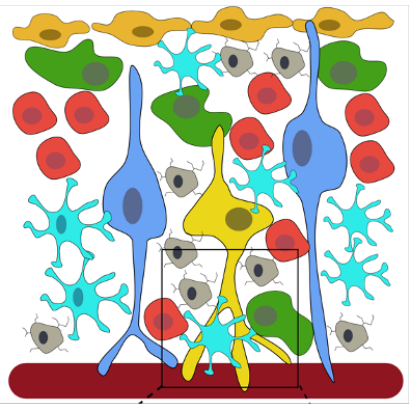
Integration of spatial information

scRNA-seq

Cell intrinsic dimension

Integration of spatial information

scRNA-seq

spatial transcriptomics

Spatial organization is important for function
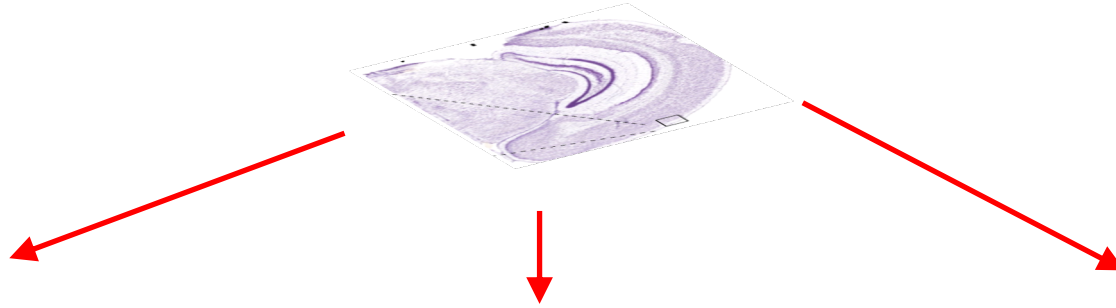
The brain has complex but structured anatomy.

Specialized functions are carried out by focal regions.

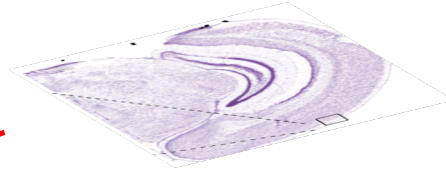Cell type composition at different regions are distinct.

Communications between different neurons are critical for carrying out brain functions.
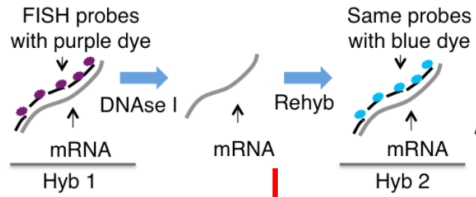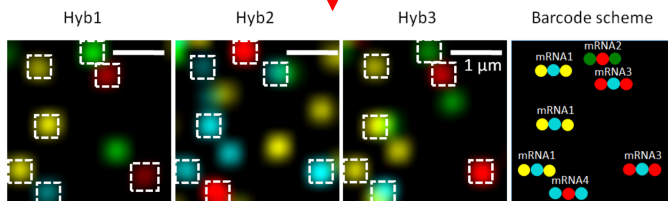

A midsagittal view showing the inner boundaries of the lobes of the cerebral cortex
(Structures outside of the cerebrum are labeled in italics.)

Precentral gyrus
Central sulcus
Postcentral gyrus
Limbic lobe
Frontal lobe
Parietal lobe
Corpus callosum
Parieto-occipital sulcus
Thalamus
Occipital lobe
Pineal gland
Hypothalamus
Corpora quadrigemina
Aqueduct of the midbrain
Optic chiasm
Fourth ventricle
Pons
Cerebellum
Temporal lobe
Mamillary body
Medulla oblongata
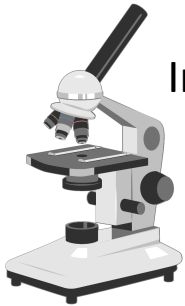© 2011 Pearson Education, Inc.

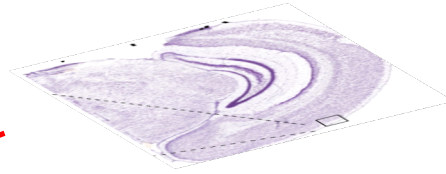Sten Linnarsson

Experimental Approaches
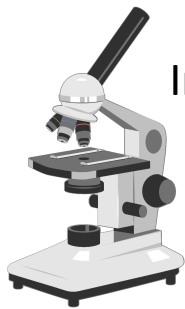
Experimental Approaches
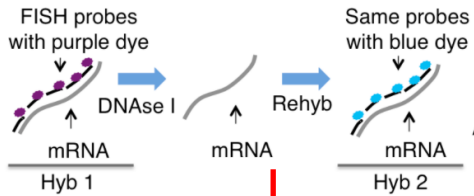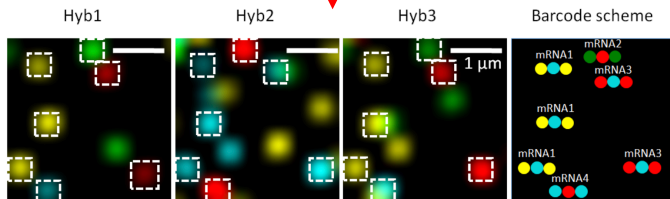

In situ hybridization
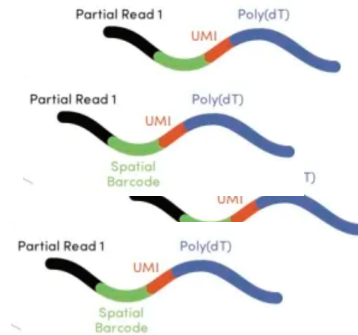
Imaging

Experimental Approaches

In situ hybridization

Spatial barcoding
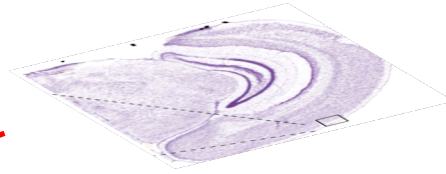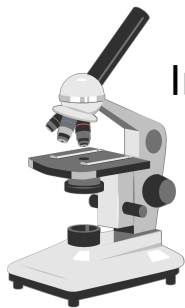
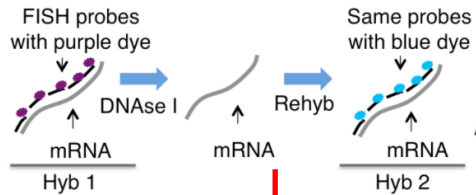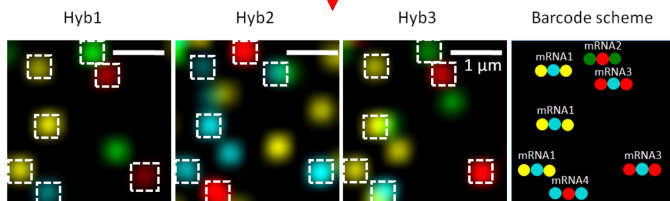Imaging
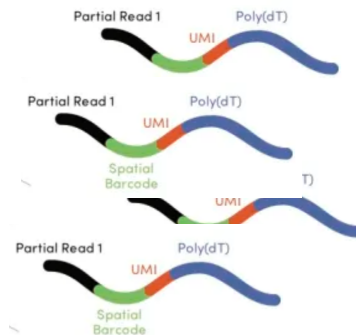
sequencing

Experimental Approaches
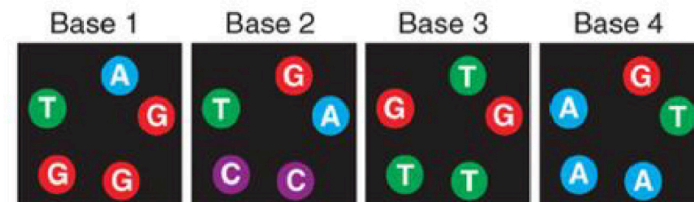

In situ hybridization

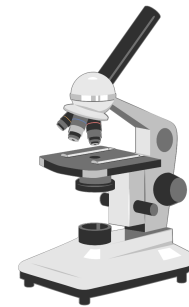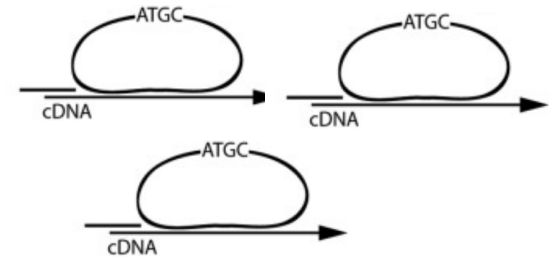Spatial barcoding

In situ sequencing

Imaging

sequencing

Types of Analysis

Cell-type mapping and spatial distribution

Spatially coherent gene detection

Spatial clustering

Spatial-temporal trajectory analysis

Cell-cell interaction
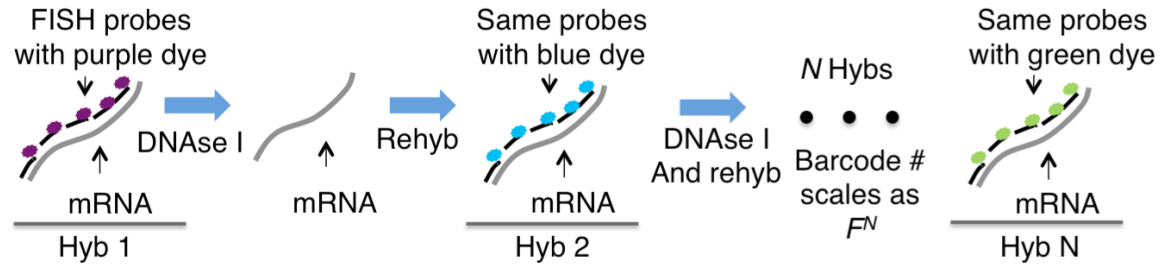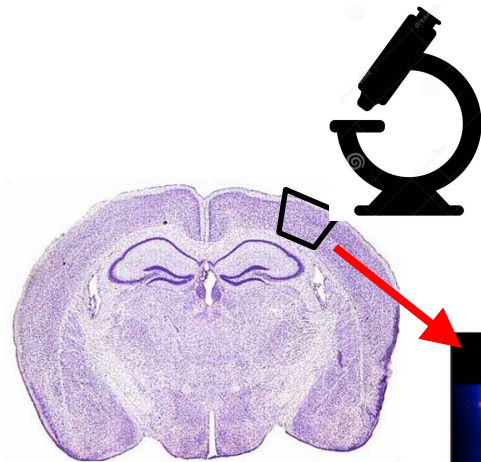
Challenges

Technological:

- Gene coverage is often limited.

- Single-cell resolution data are difficult to generate.

- 3D data are difficult to generate

- Imaging/sequencing associated artifacts

Challenges

Computational:

➢ Delineating cell boundaries.

➢ Creating spatial proximity networks.

➢ Classifying spatial patterns at cellular or subcellular scale.

➢ Mechanistic understanding of cell-cell interaction.

➢ Dissecting the contribution of spatial environment in mediating cell states

seqFISH: a case study

Mouse visual cortex
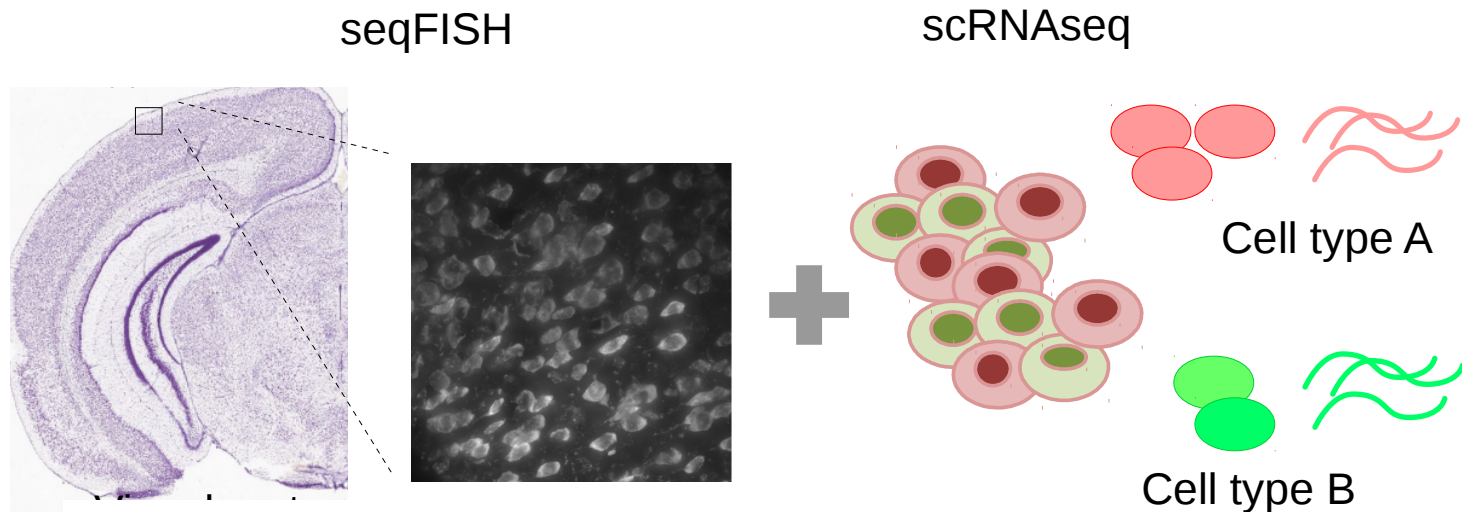125 genes
1597 cells
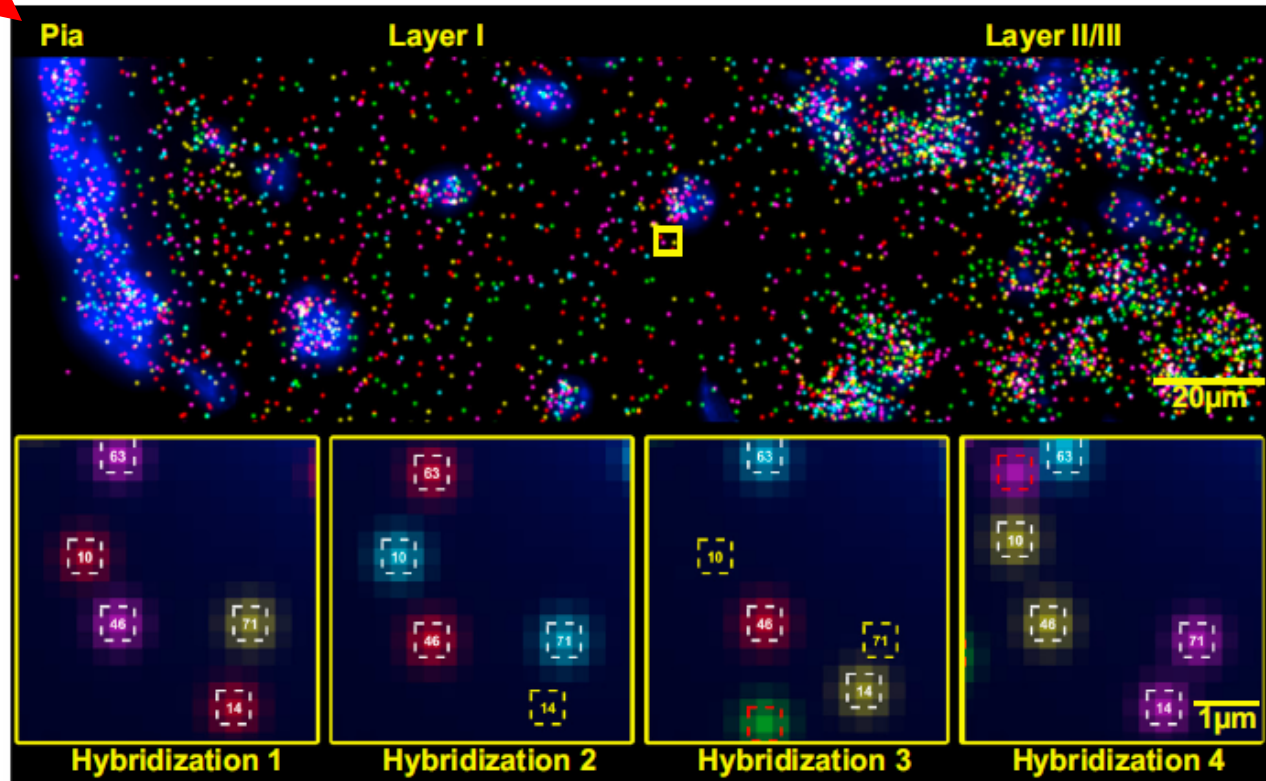
Sheel Shah, Long Cai

Question 1: How do we identify cell types from seqFISH data

Challenge: The number of profiled genes is limited.

Question 1: How do we identify cell types from seqFISH data

Challenge: The number of profiled genes is limited.

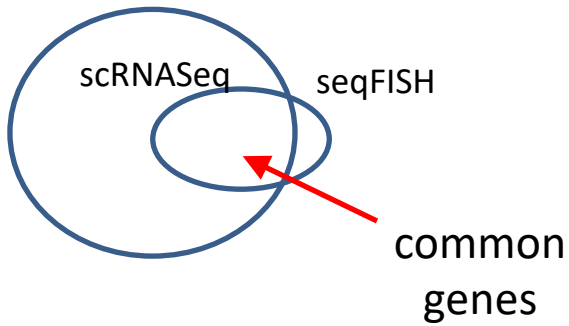Approach: Integrating seqFISH with external scRNAseq data

seqFISH

scRNAseq

Cell type A

Cell type B

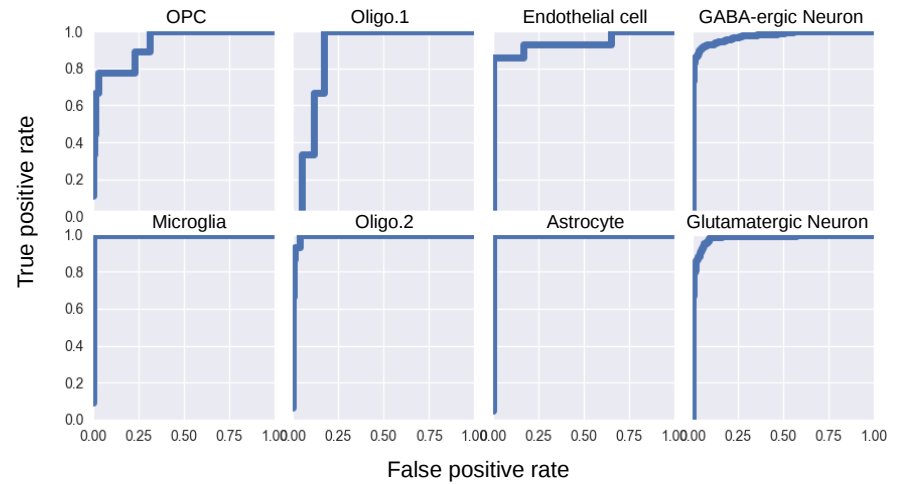# Our computational strategy

gene features

scRNASeq

seqFISH

common
genes

Our computational strategy

Our computational strategy
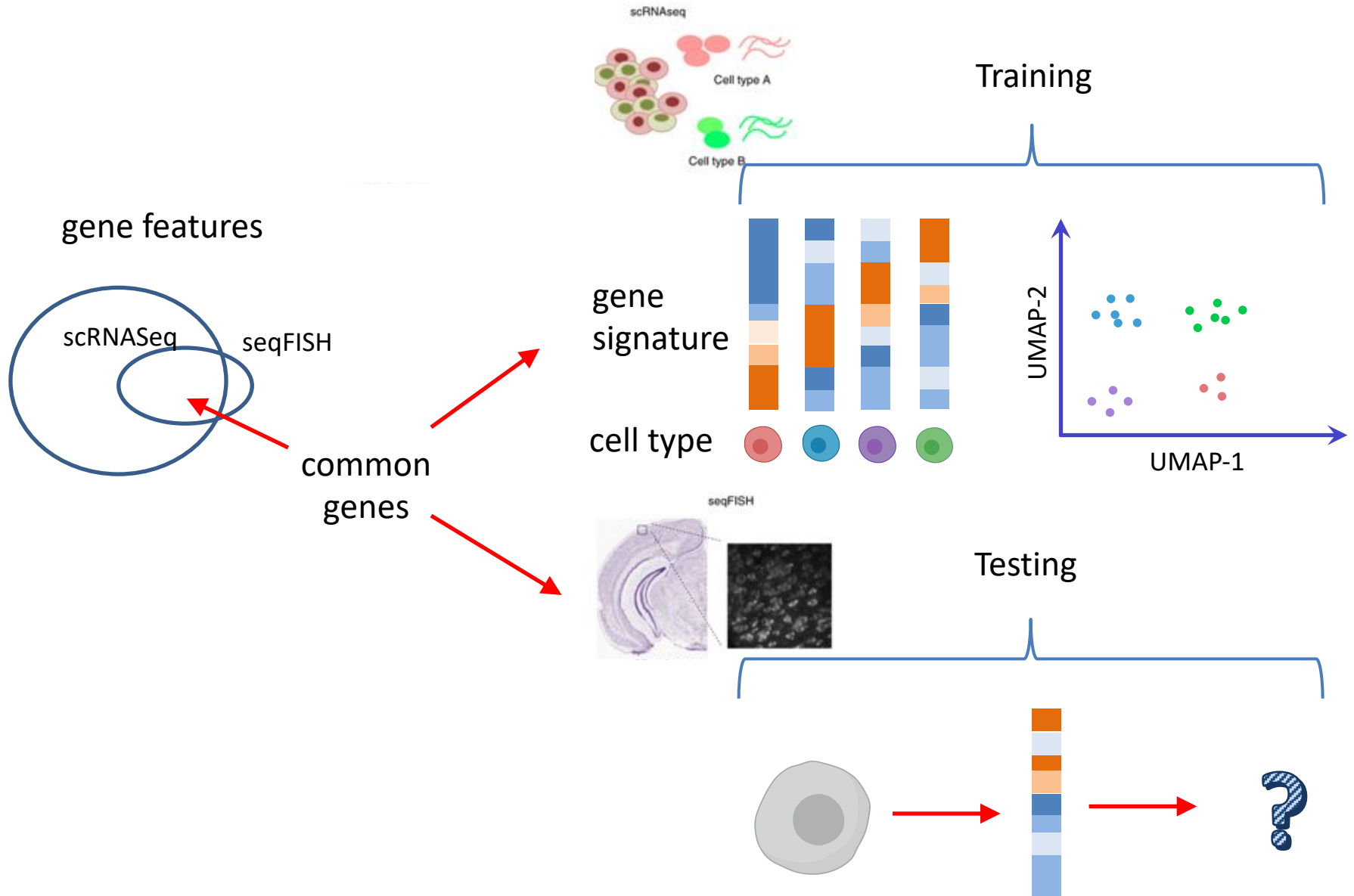
gene features

scRNASeq   seqFISH

common genes

scRNAseq
Cell type A
Cell type B

Training

gene signature

cell type

UMAP-2

UMAP-1

seqFISH
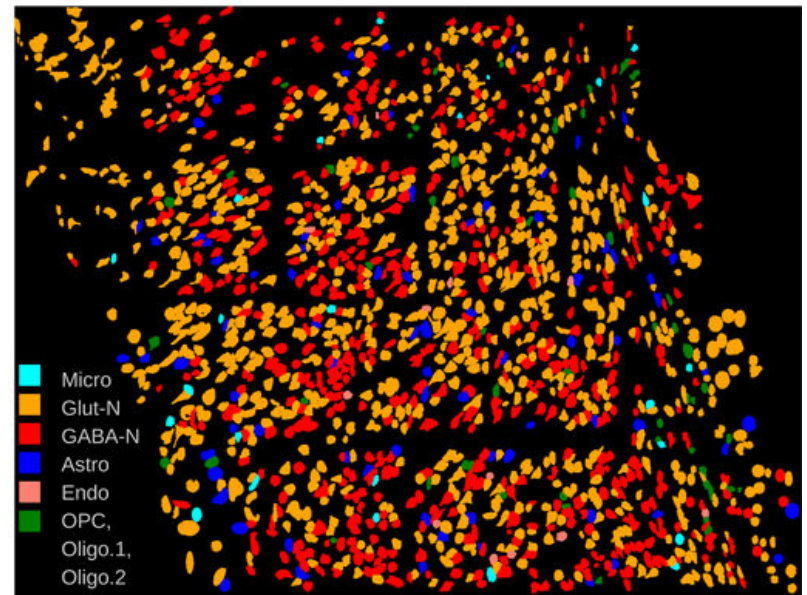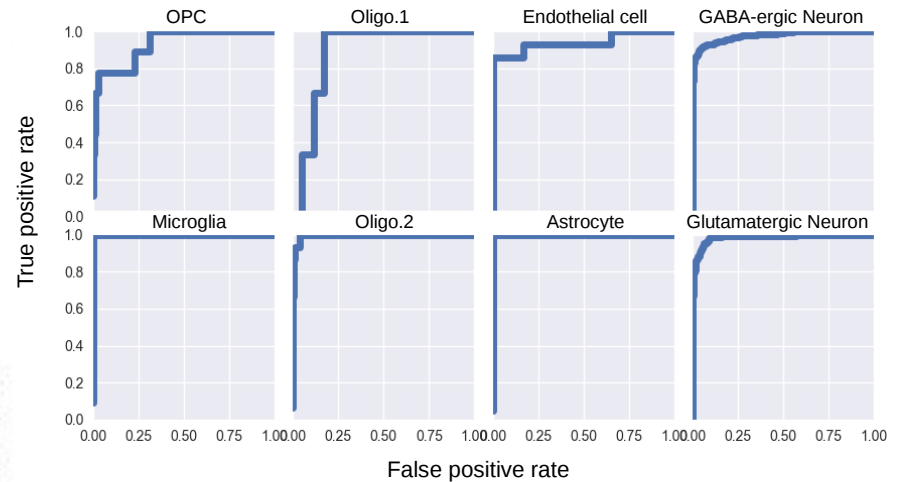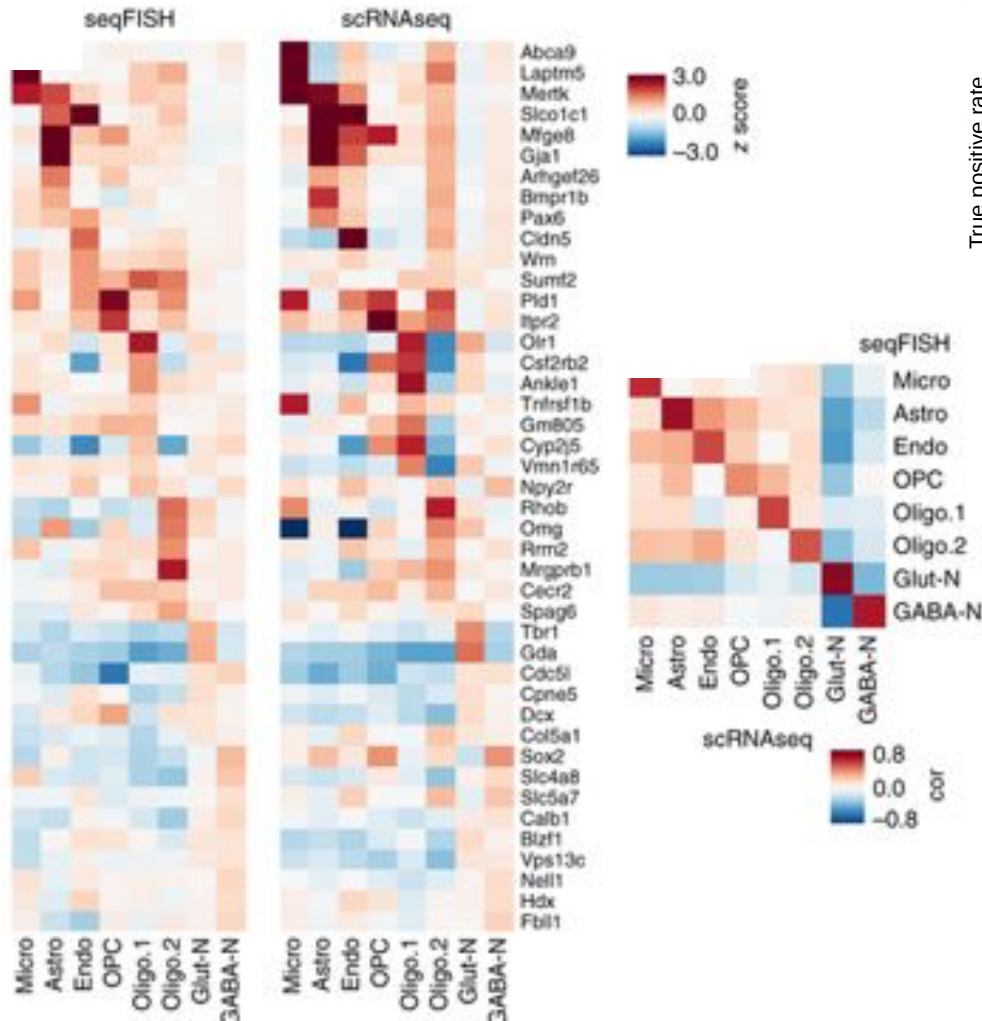
Testing

Evaluating the accuracy of cell-type mapping

Qian Zhu

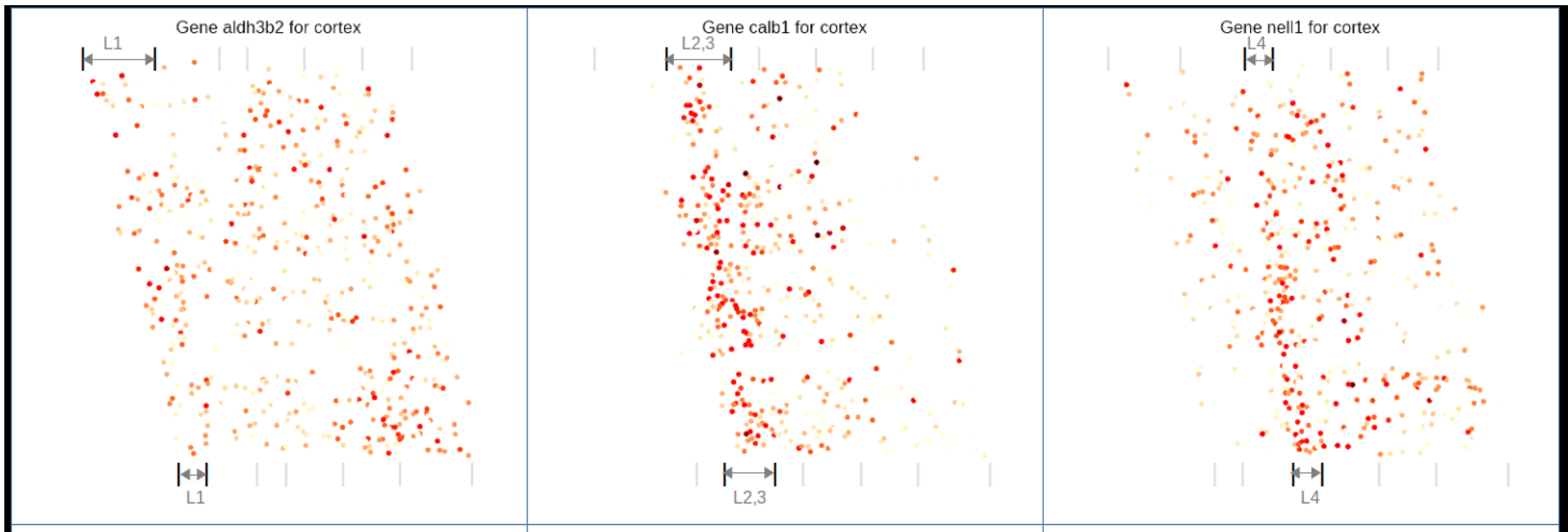# Evaluating the accuracy of cell-type mapping

Qian Zhu

Evaluating the accuracy of cell-type mapping

Qian Zhu
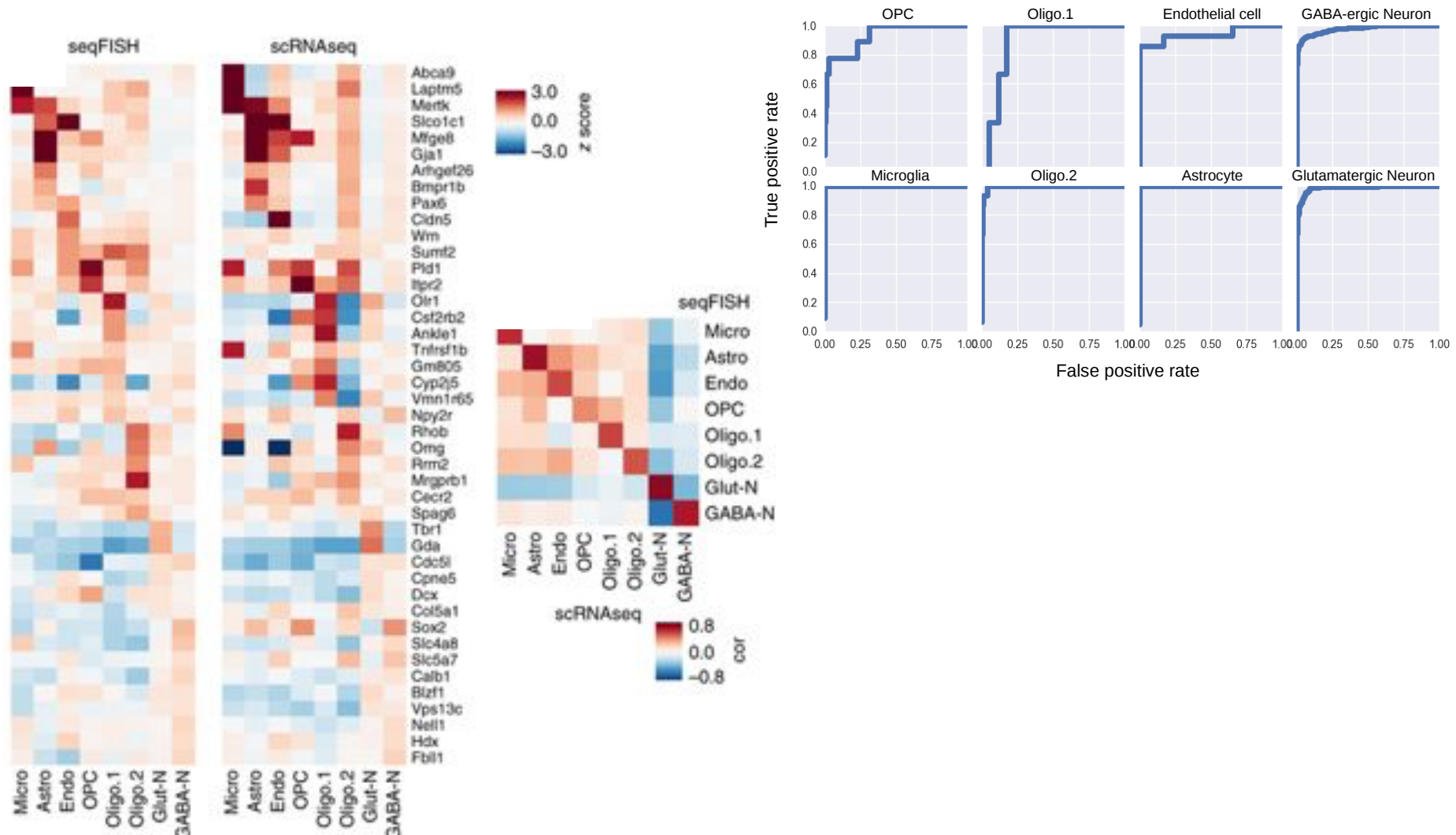
Question 2: How do we identify distinct spatial patterns?

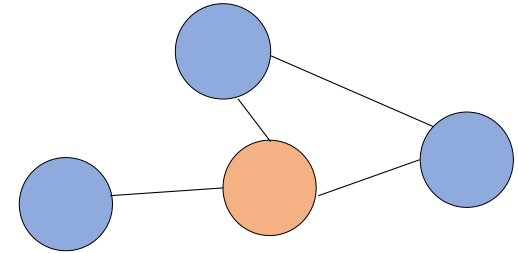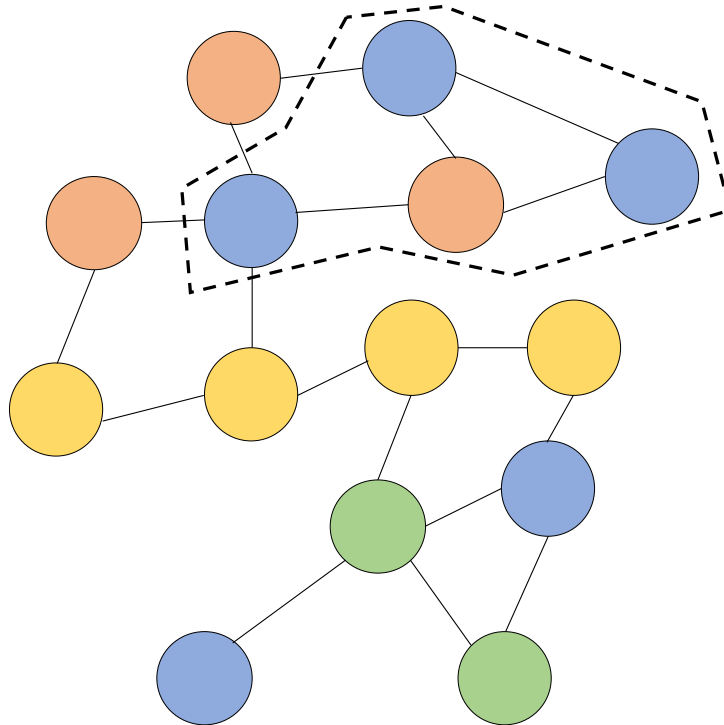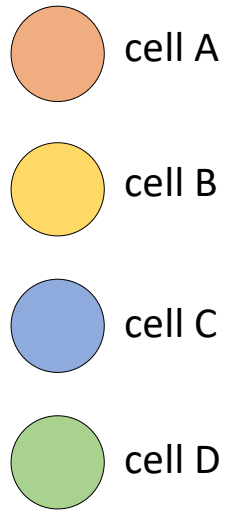# Question 2: How do we identify distinct spatial patterns?

Challenge: The number of images is too large for manual annotation.

> 100 images.

Approach: Hidden Markov Random Field (HMRF)

gene expression pattern

cell A

cell B

cell C

cell D

use HMRF to create discrete domains based on shared spatial gene expression patterns across cell types
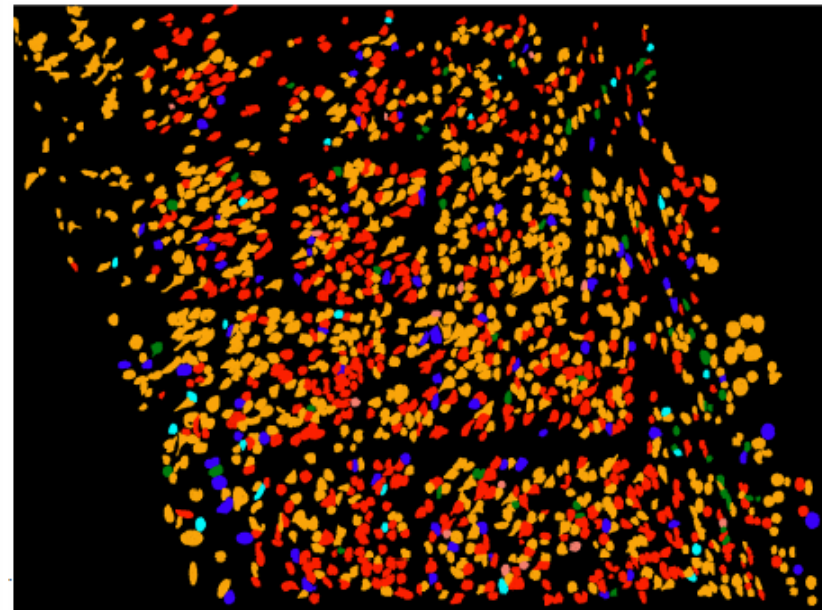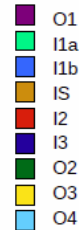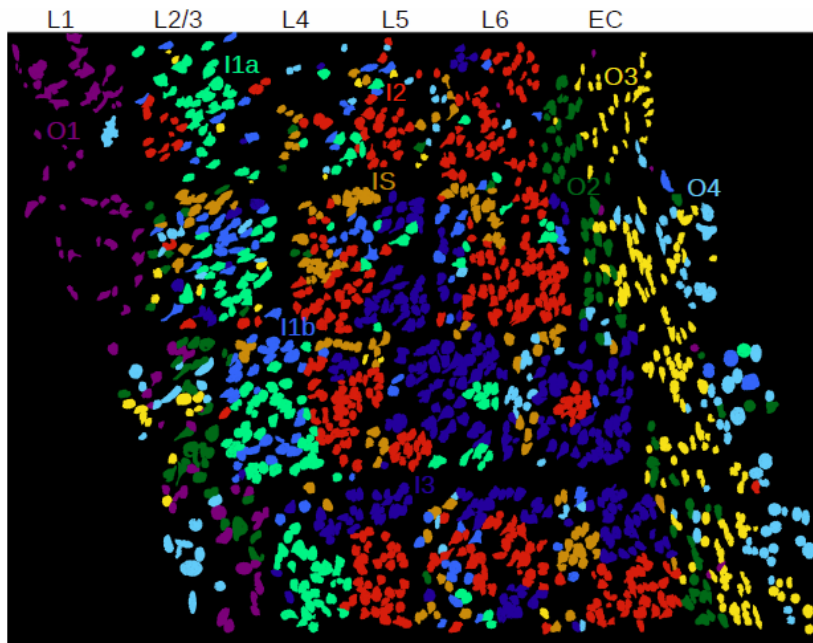
Qian Zhu

Spatial domains recapitulate layer-like structure in visual cortex

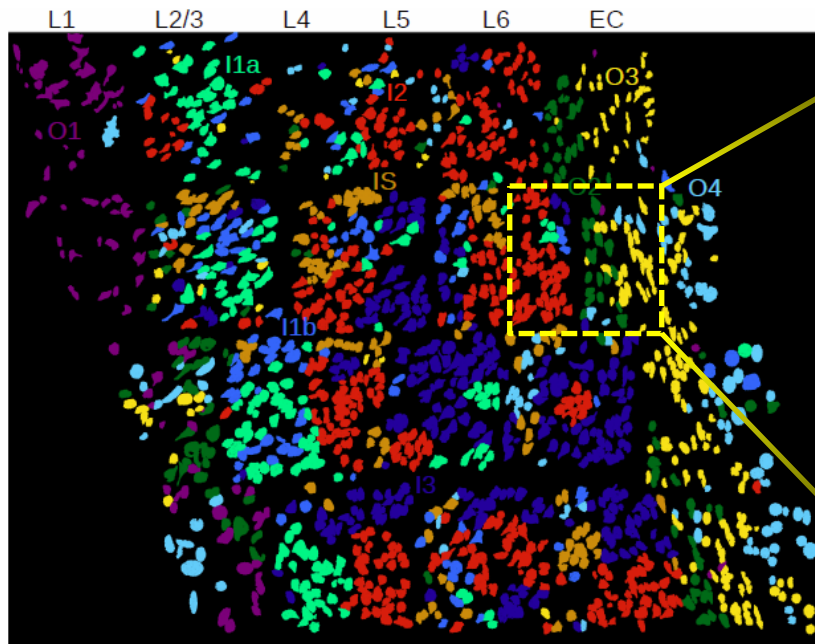# Complementary information in spatial domain and cell-type annotations
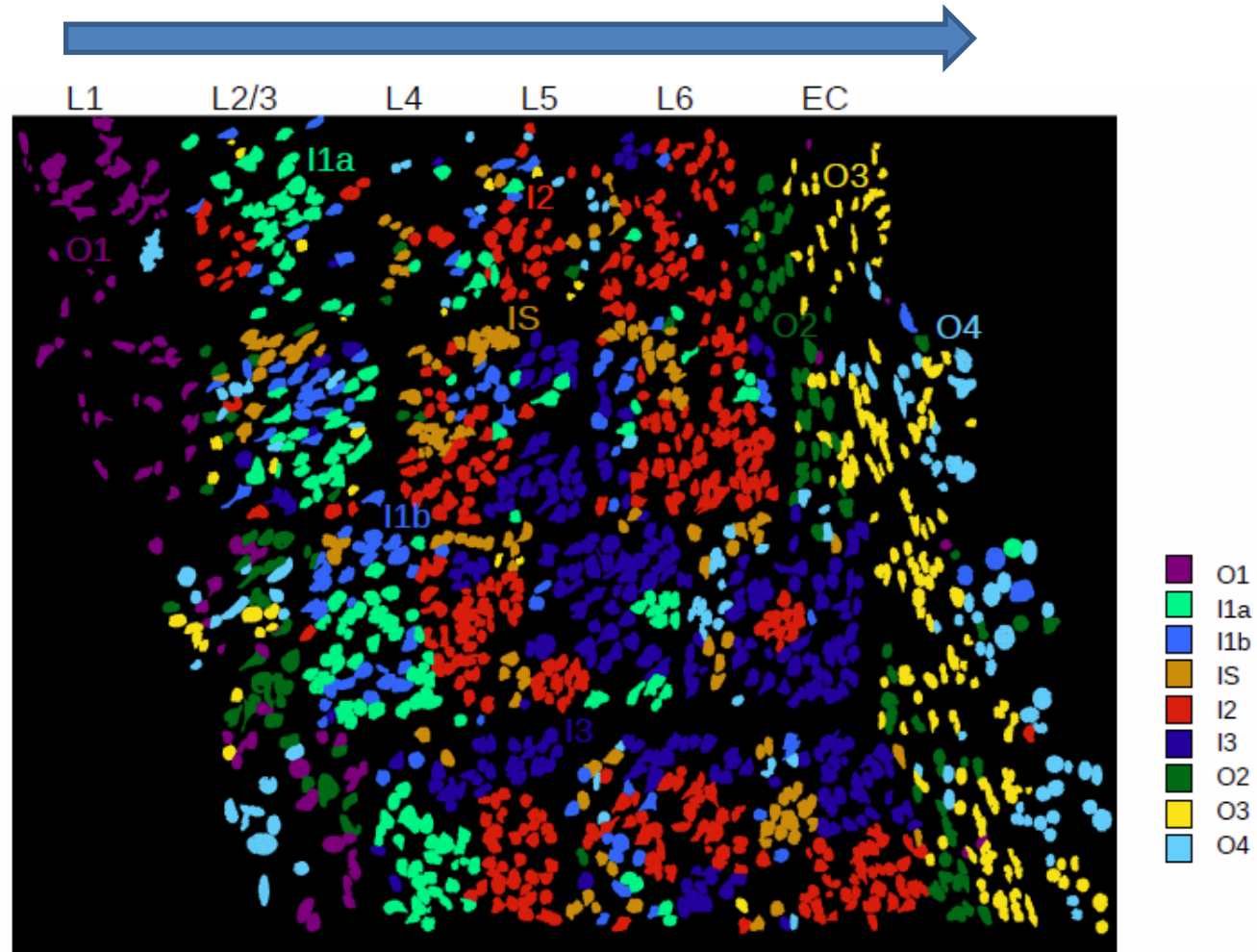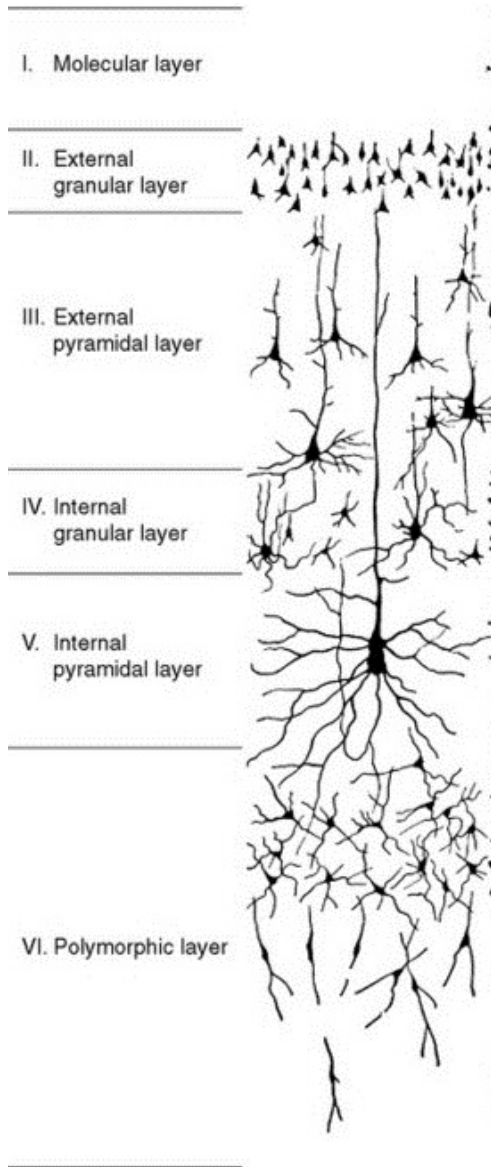
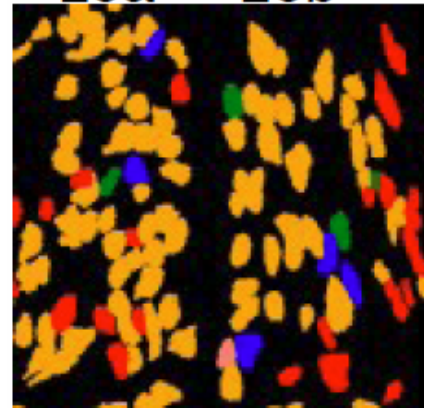# Complementary information in spatial domain and cell-type annotations
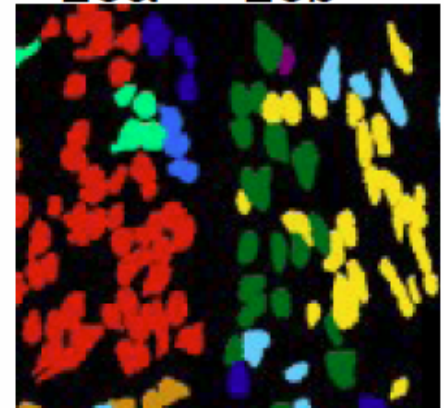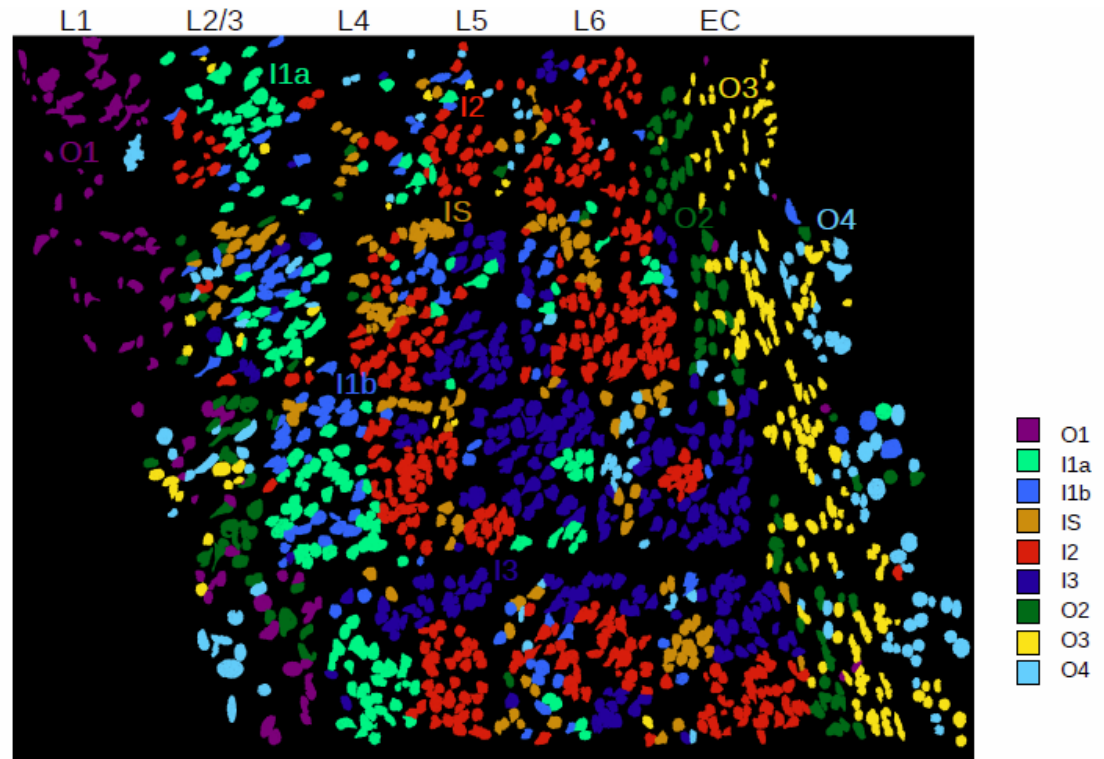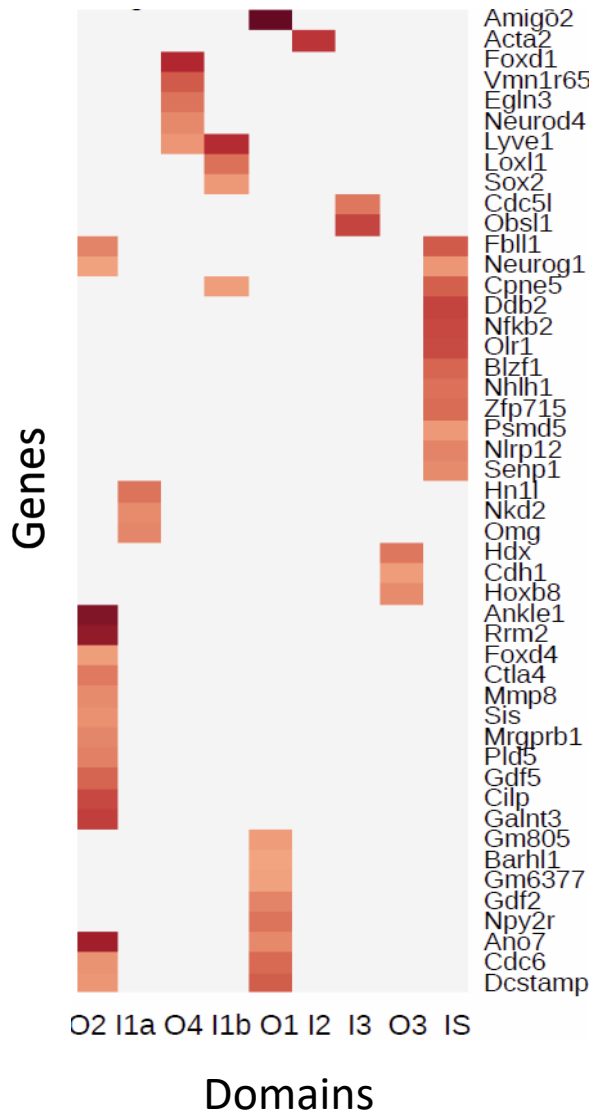

Spatial Domain

Using spatial domain annotation to study within cell-type variations associated with spatial location
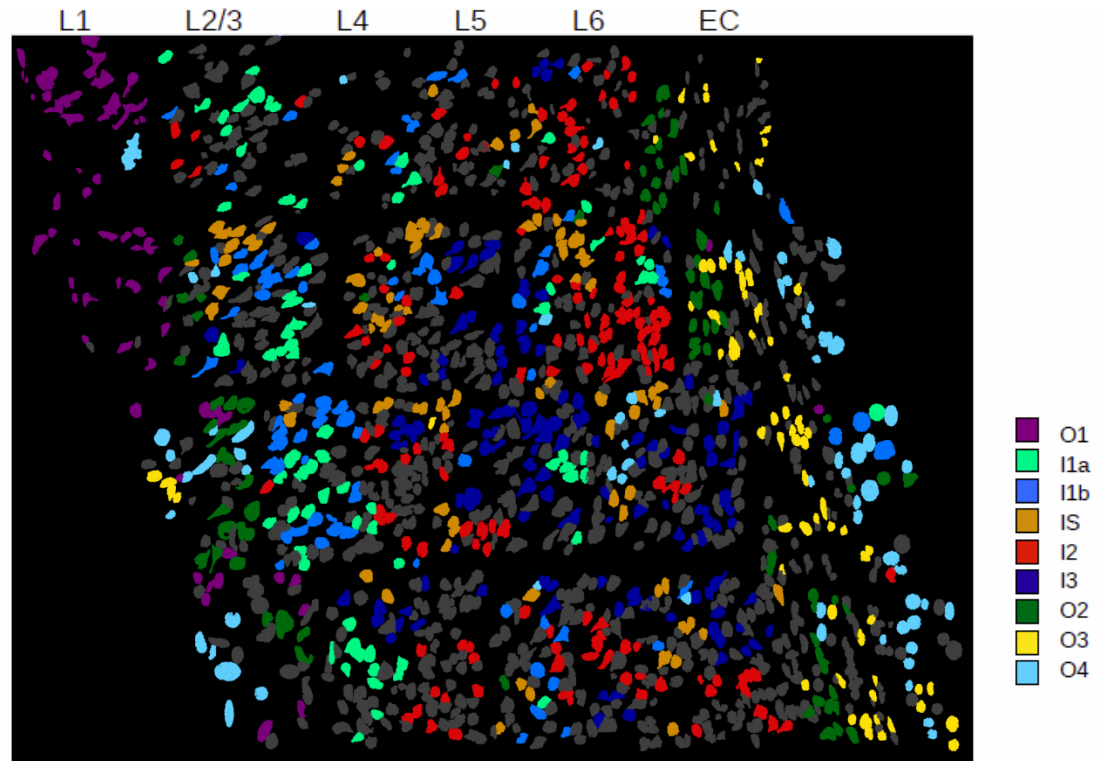
Take the image with all the cells, …

Using spatial domain annotation to study within cell-type variations associated with spatial location
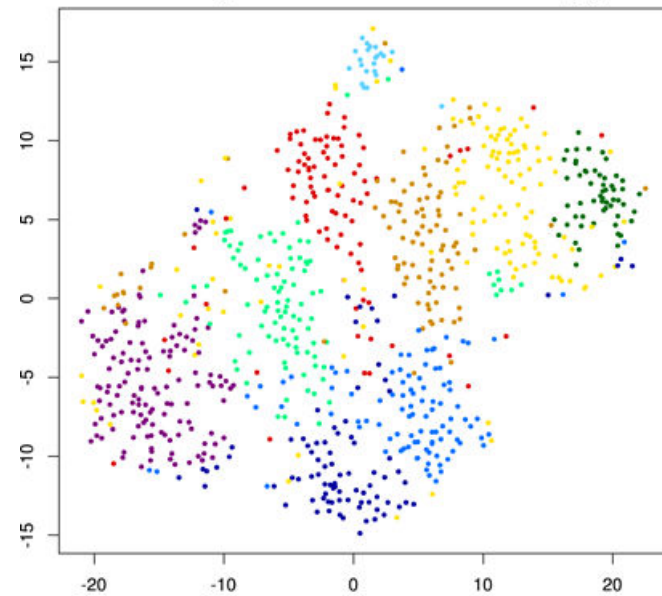

if we focus only on glutamatergic cells …
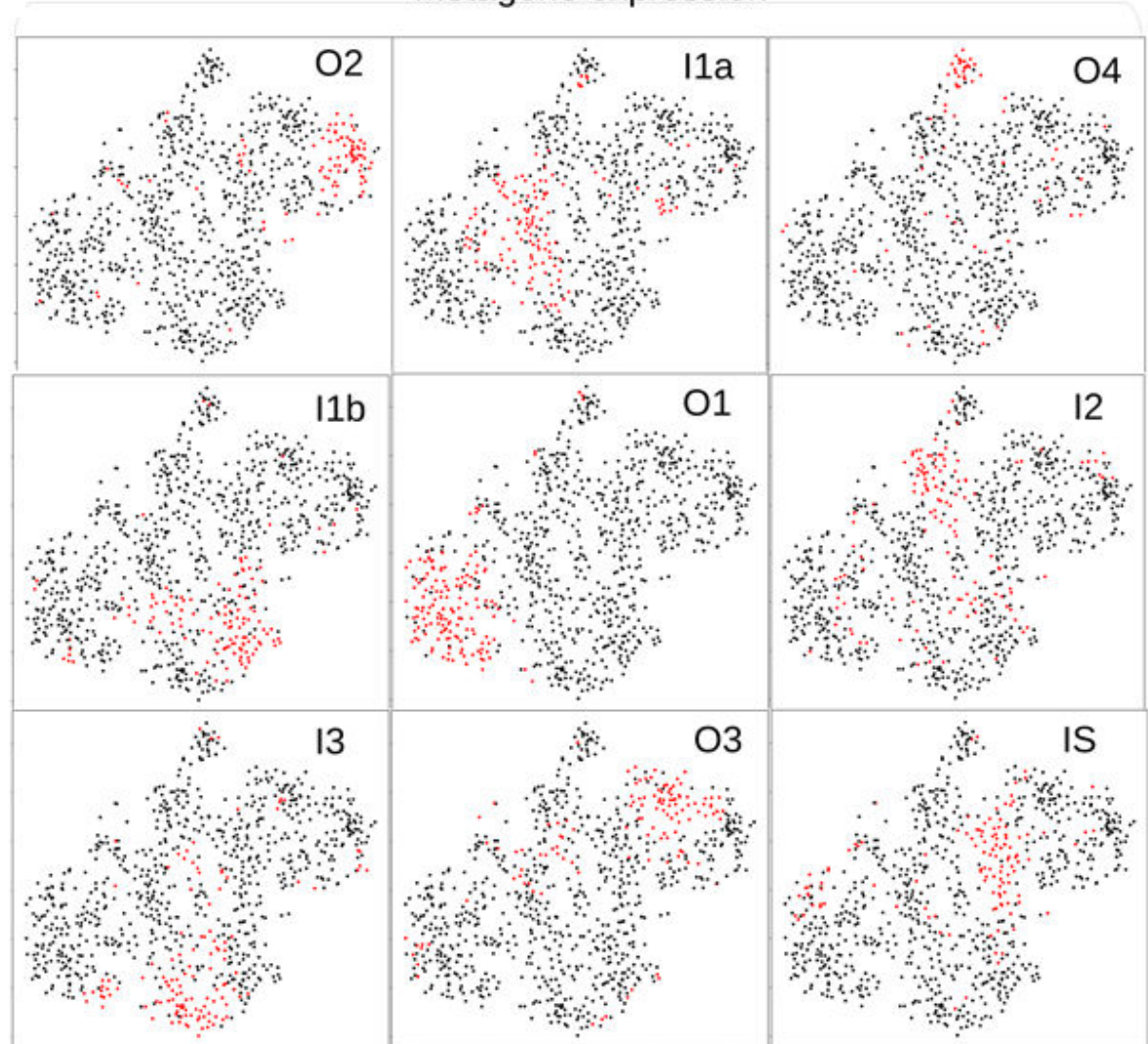
Genes

Domains

Use domain-specific signatures to reanalyze scRNAseq data

Metagene expression

Metagene-derived cell clusters(9)

Corresp. to expression

1 I1a	4 I2	7 O1
2 O2	5 IS	8 I1b
3 O4	6 O3	9 I3

Cortex-seqFISH Hackathon data

File	Description
/tasic_training_b2.txt	normalized scRNAseq data
/seqfish_cortex_b2_testing.txt	normalized seqFISH data
/seqfish_labels.tsv	spatial cluster labels and SVM learned cell types for seqFISH
/tasic_labels.tsv	cell type labels for scRNAseq
/fcortex.coordinates.txt	Spatial cell coordinates
/hmrf-usage/data/fcortex.gene.ALL.txt	z-scored matrix incorporating the spatial gene expression of 69 genes

seqFISH data source: (125 genes, 1597 cells)
Zhu Q, Shah S, Dries R, Cai L, Yuan GC. Nat Biotechnol. 2018 Oct 29:10.1038/nbt.4260.

scRNAseq data source: (24057 genes, 1723 cells)
Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, Koch C, Zeng H. Nat Neurosci. 2016 Feb;19(2):335-46.

Giotto, a toolbox for integrative analysis and visualization of spatial expression data

Ruben Dries, Qian Zhu, Rui Dong, Chee-Huat Linus Eng, Huipeng Li, Kan Liu, Yuntian Fu, Tianxiao Zhao, Arpan Sarkar, Feng Bao, Rani E George, Nico Pierson, Long Cai, Guo-Cheng Yuan

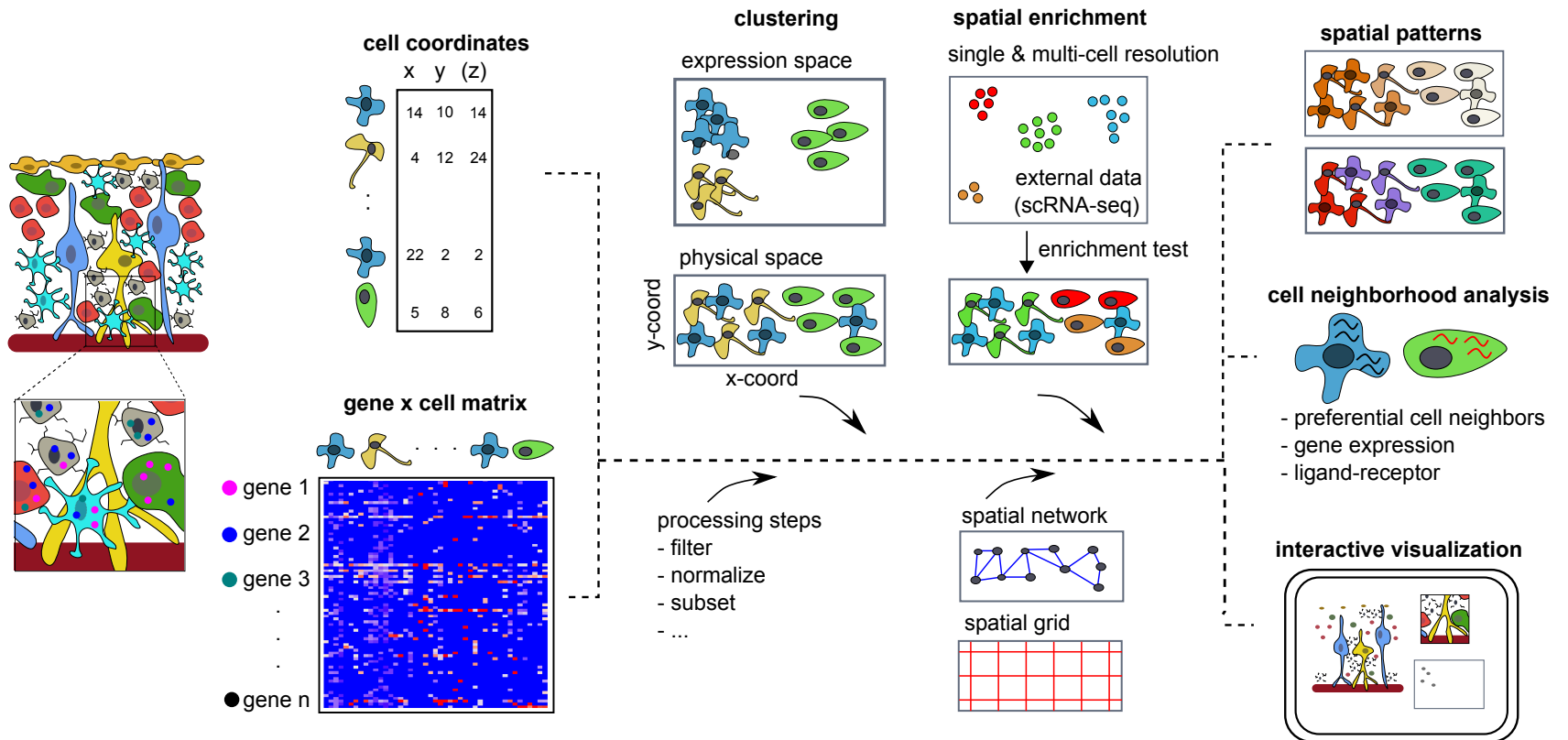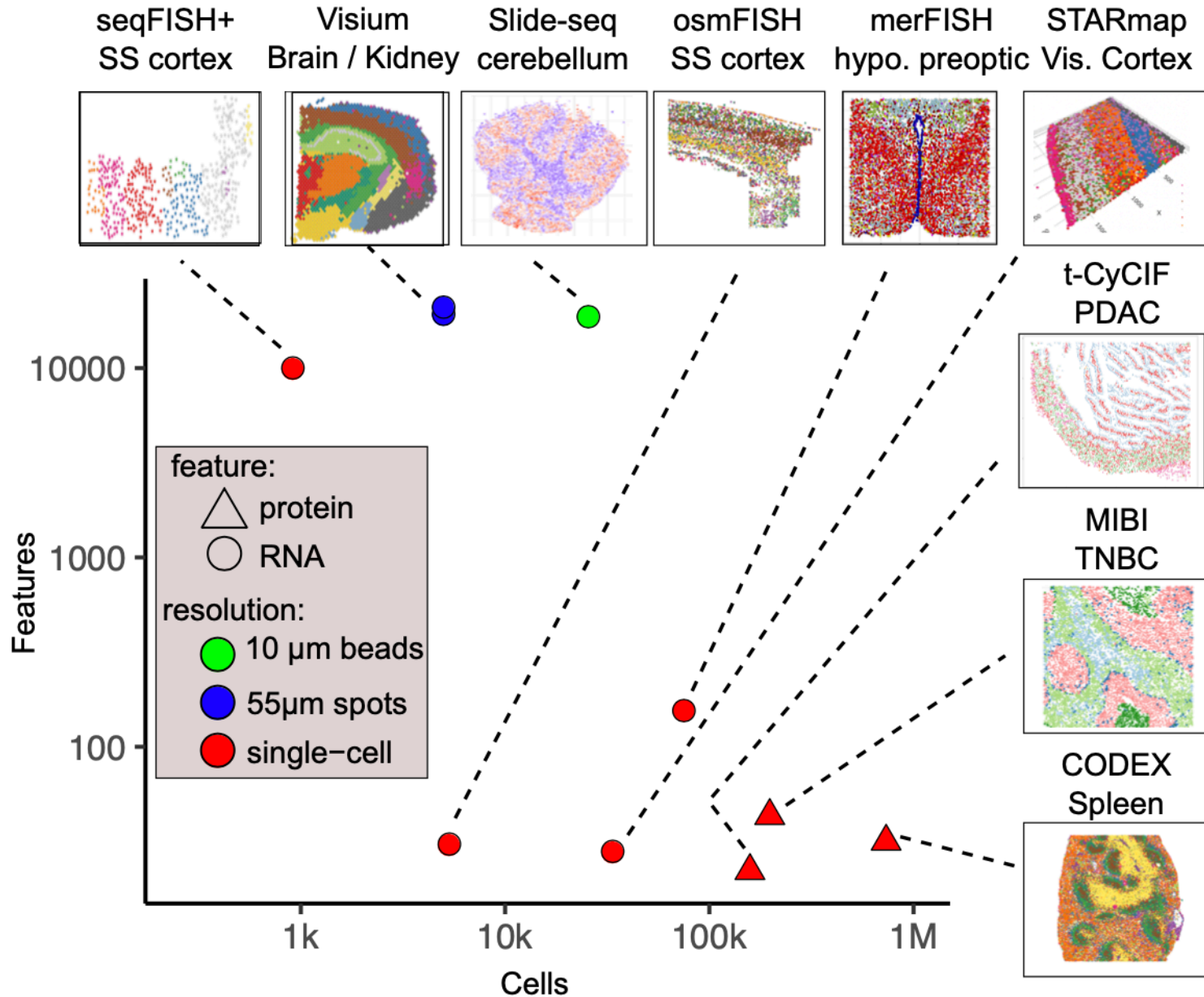Ruben Dries

Qian Zhu

Giotto in a snapshot

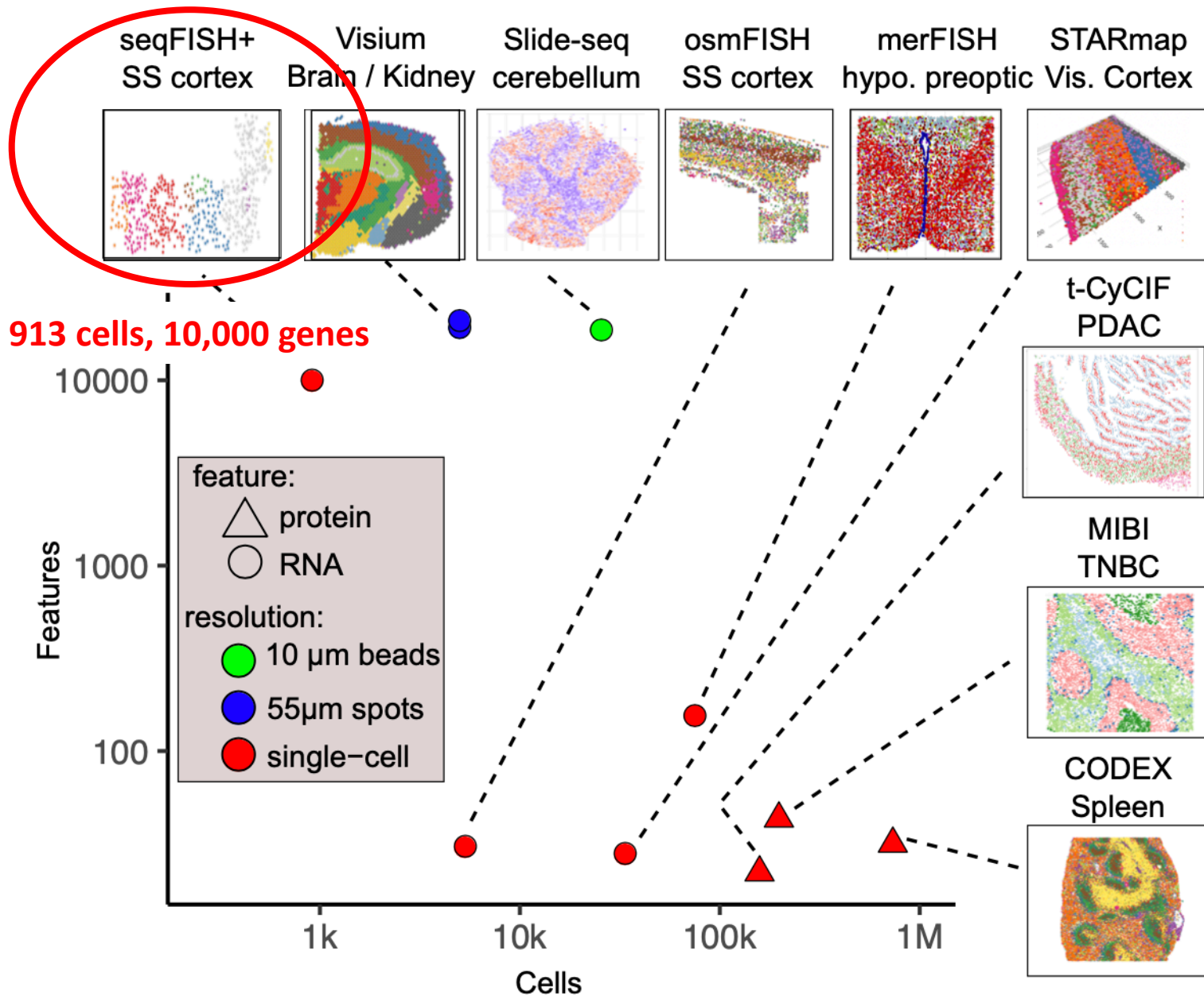**Main Website:** www.spatialgiotto.com

Development Website:

Giotto Analyzer: https://rubd.github.io/Giotto_site/

Giotto Viewer: http://spatialgiotto.rc.fas.harvard.edu/giotto-viewer/

Application of Giotto to diverse datasets

Application of Giotto to diverse datasets

Spatial gene detection
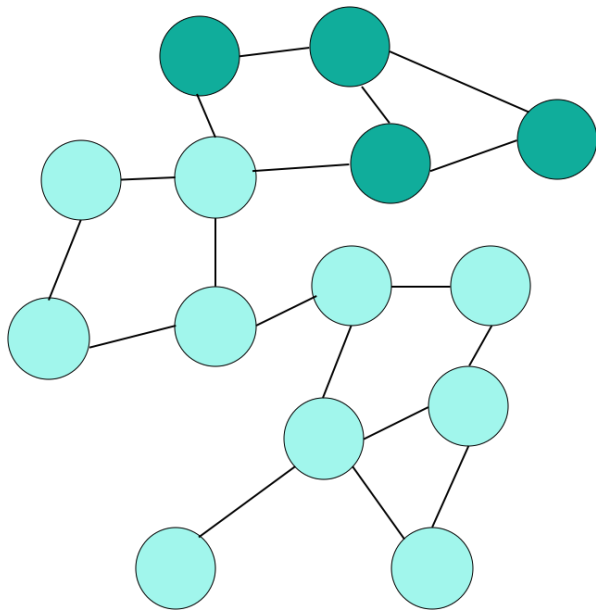
gene X high

gene X low

Ruben Dries

Spatial gene detection

gene X high

gene X low

Cux2

Grm2

Ruben Dries

Spatial Cell-Cell Interaction

Ruben Dries

Spatial Cell-Cell Interaction
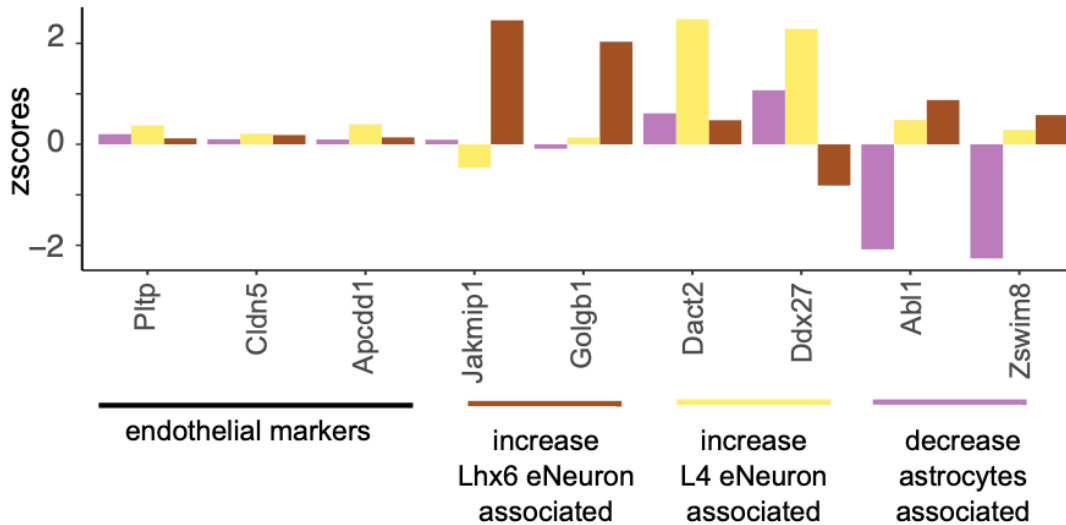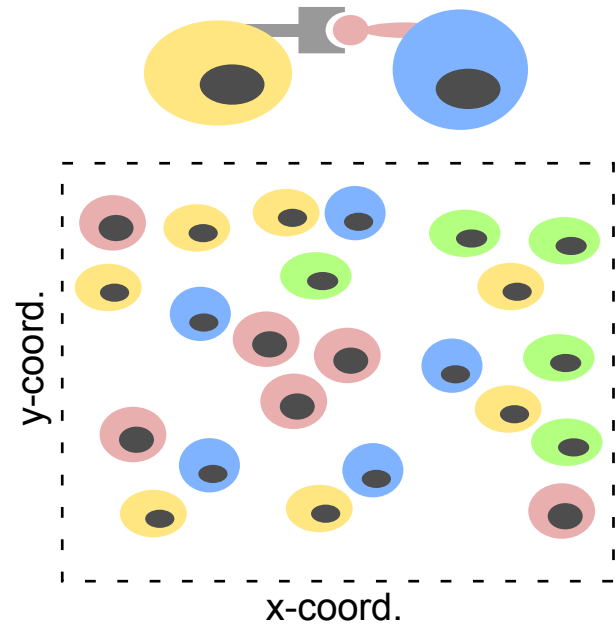

cell-cell proximity

gene (pair) enrichment
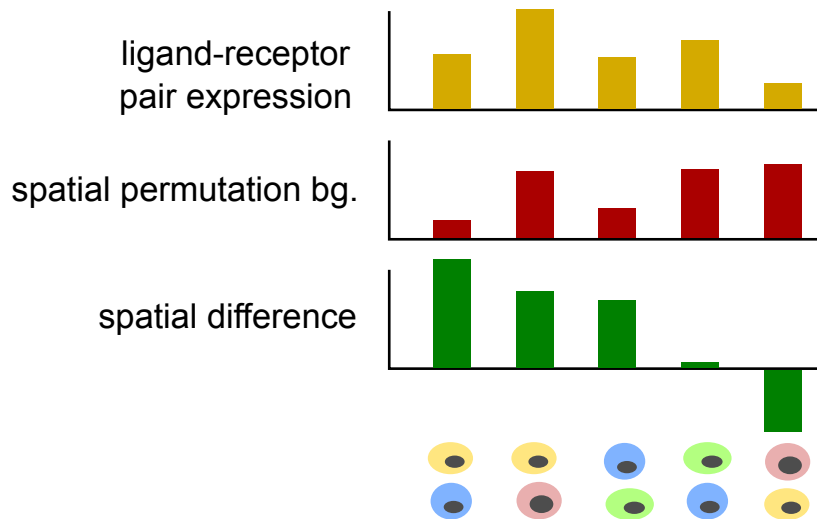
gene x gene y
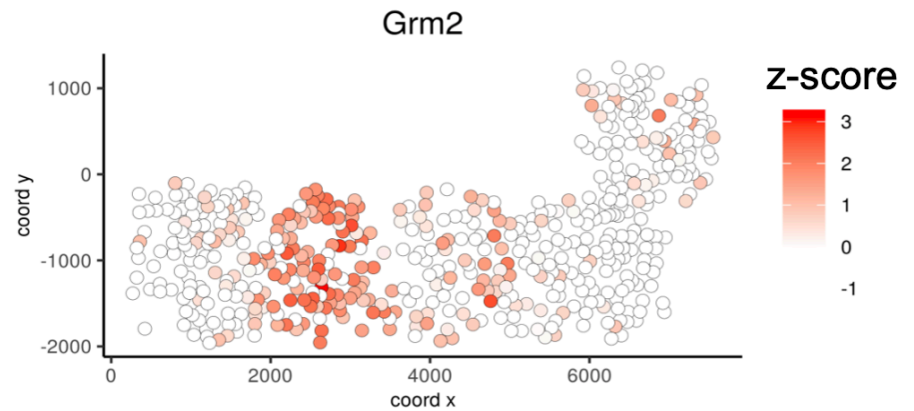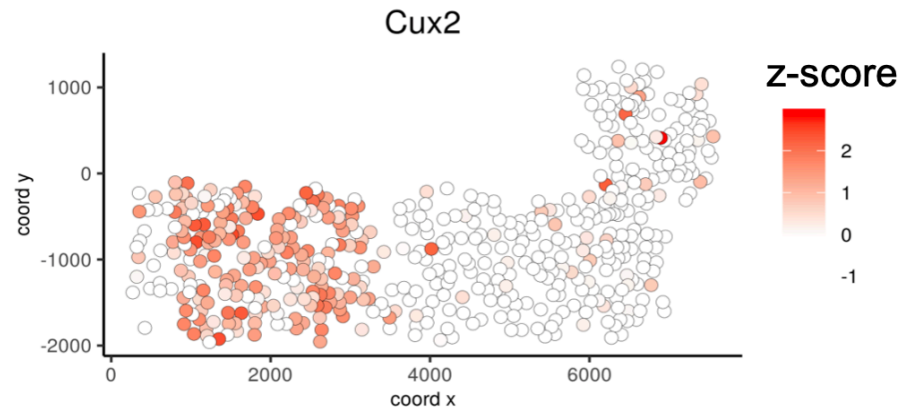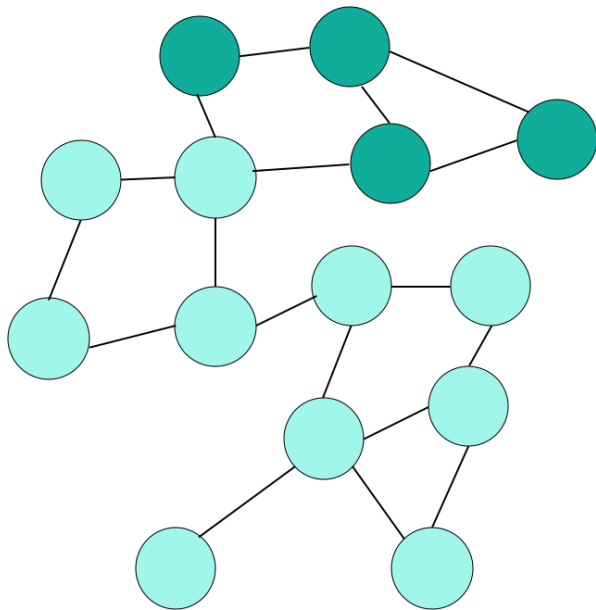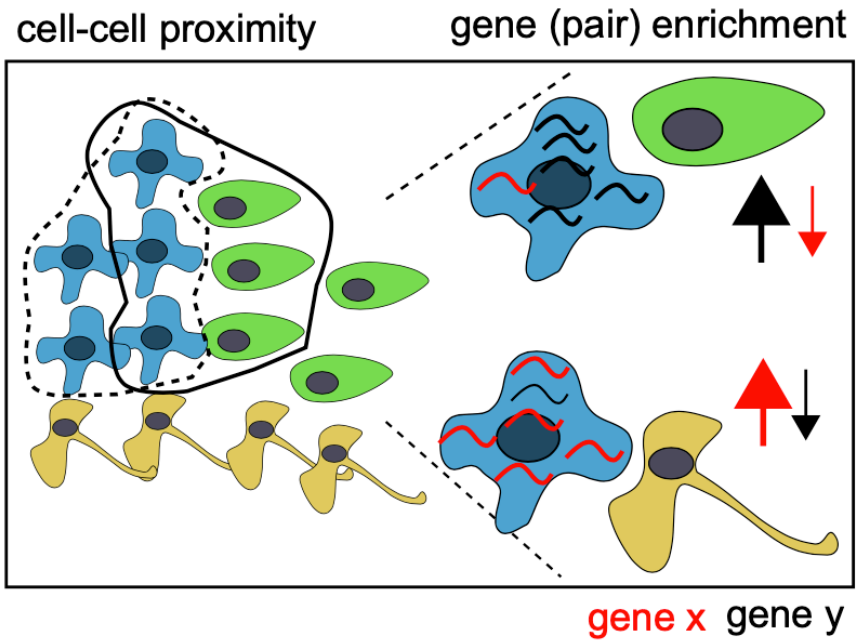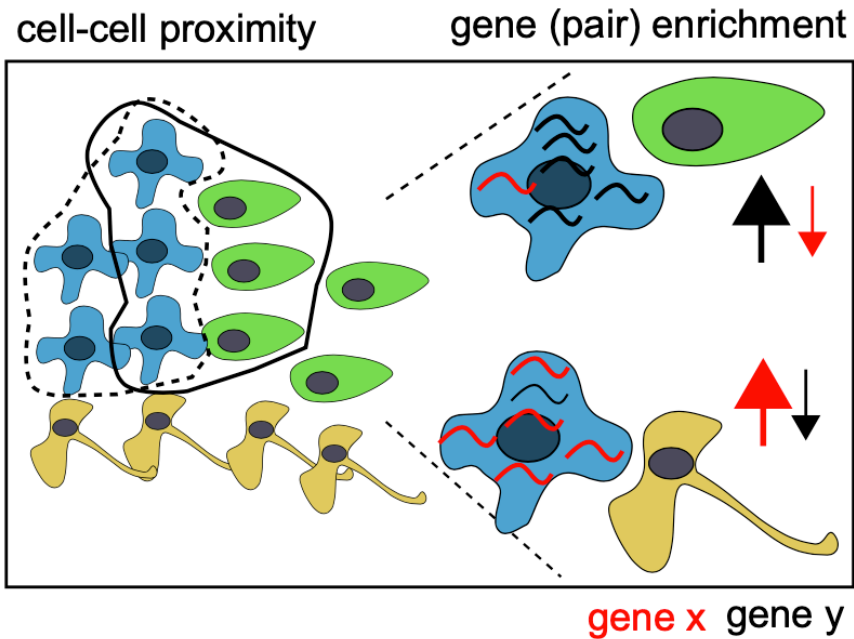
interacting cell-type

astrocytes

L4 eNeuron

Lhx6 iNeuron
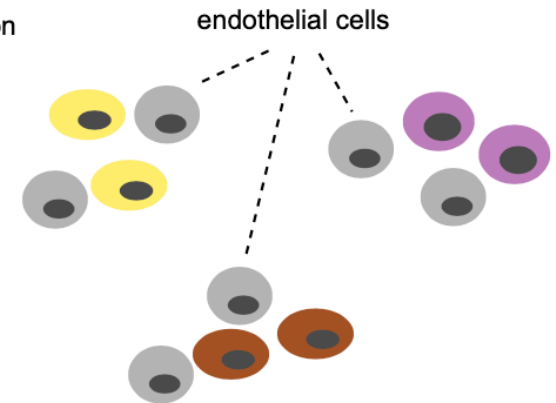
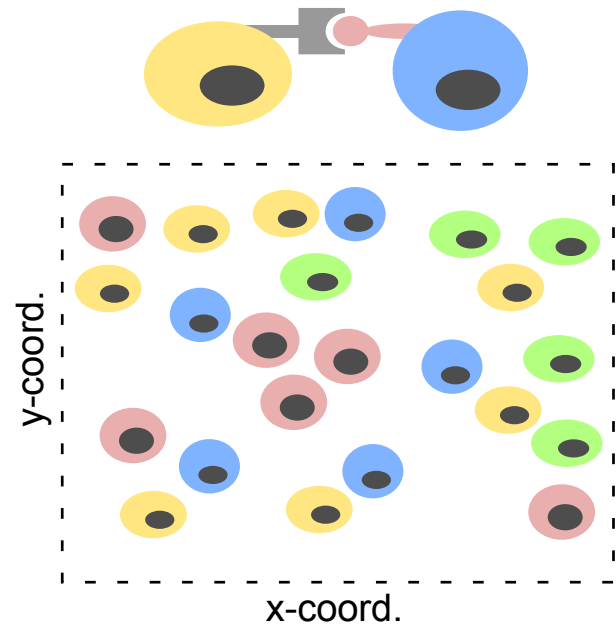endothelial cells
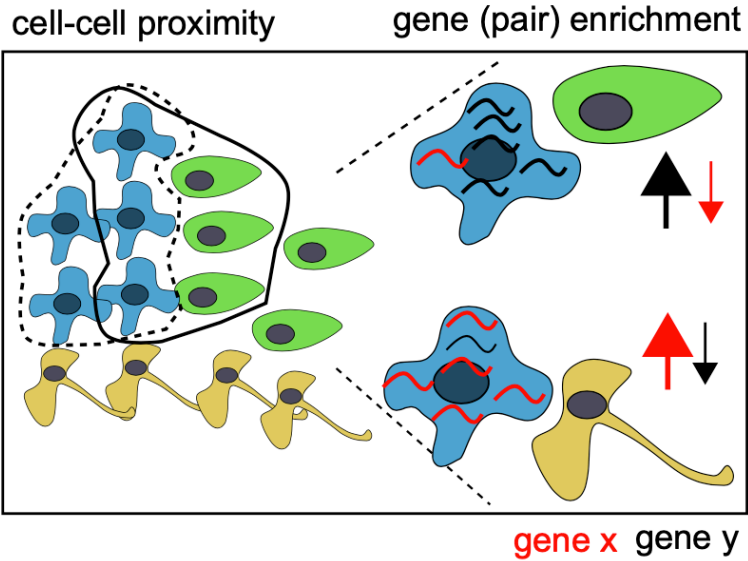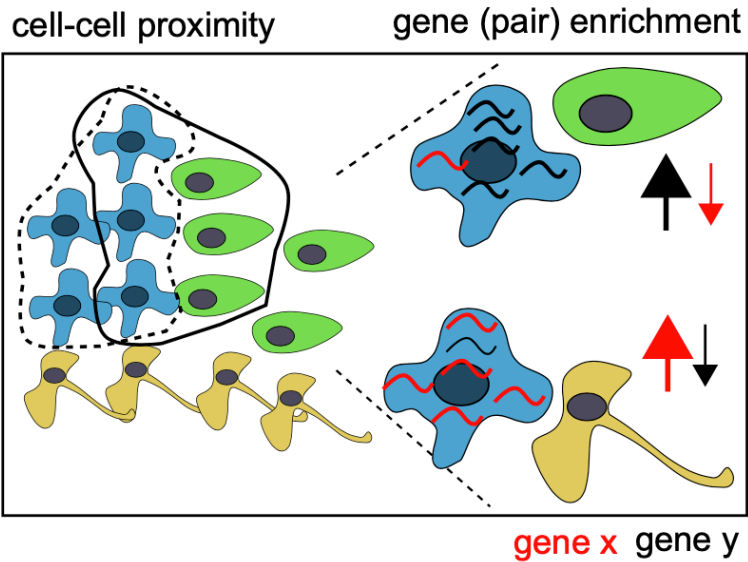
Ruben Dries

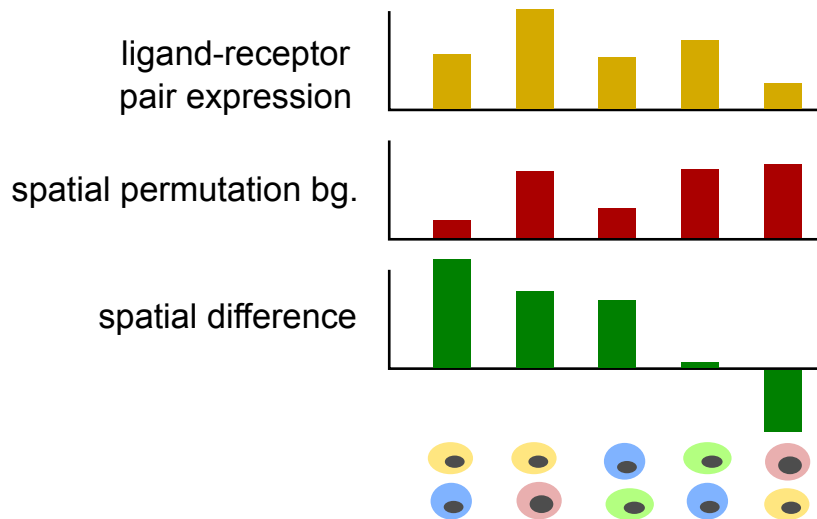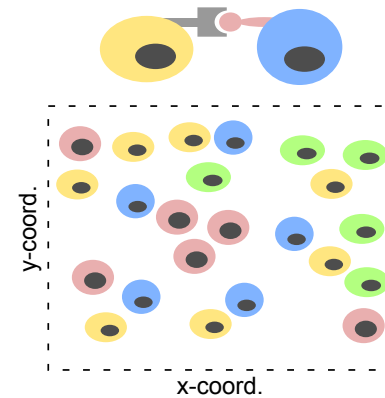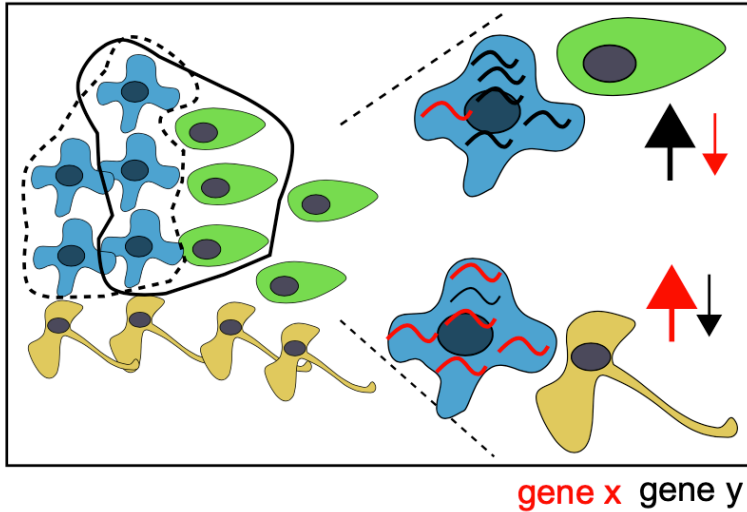# Spatial Cell-Cell Interaction

Ligand-Receptor interactions

cell-cell proximity

gene (pair) enrichment

gene x gene y

y-coord.

x-coord.

Ligand-Receptor interactions

cell-cell proximity

gene (pair) enrichment

gene x gene y

y-coord.

x-coord.

ligand-receptor pair expression

spatial permutation bg.

spatial difference

Ligand-Receptor interactions

cell-cell proximity

gene (pair) enrichment

gene x gene y

ligand-receptor pair expression

spatial permutation bg.

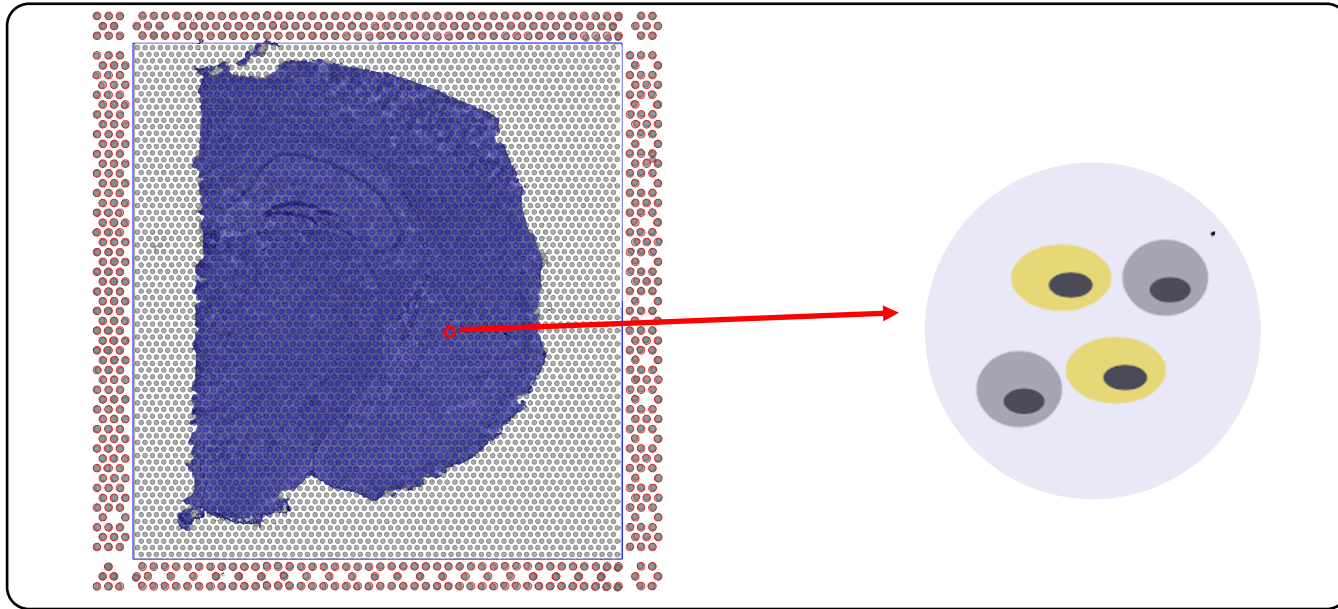spatial difference

y-coord.

x-coord.
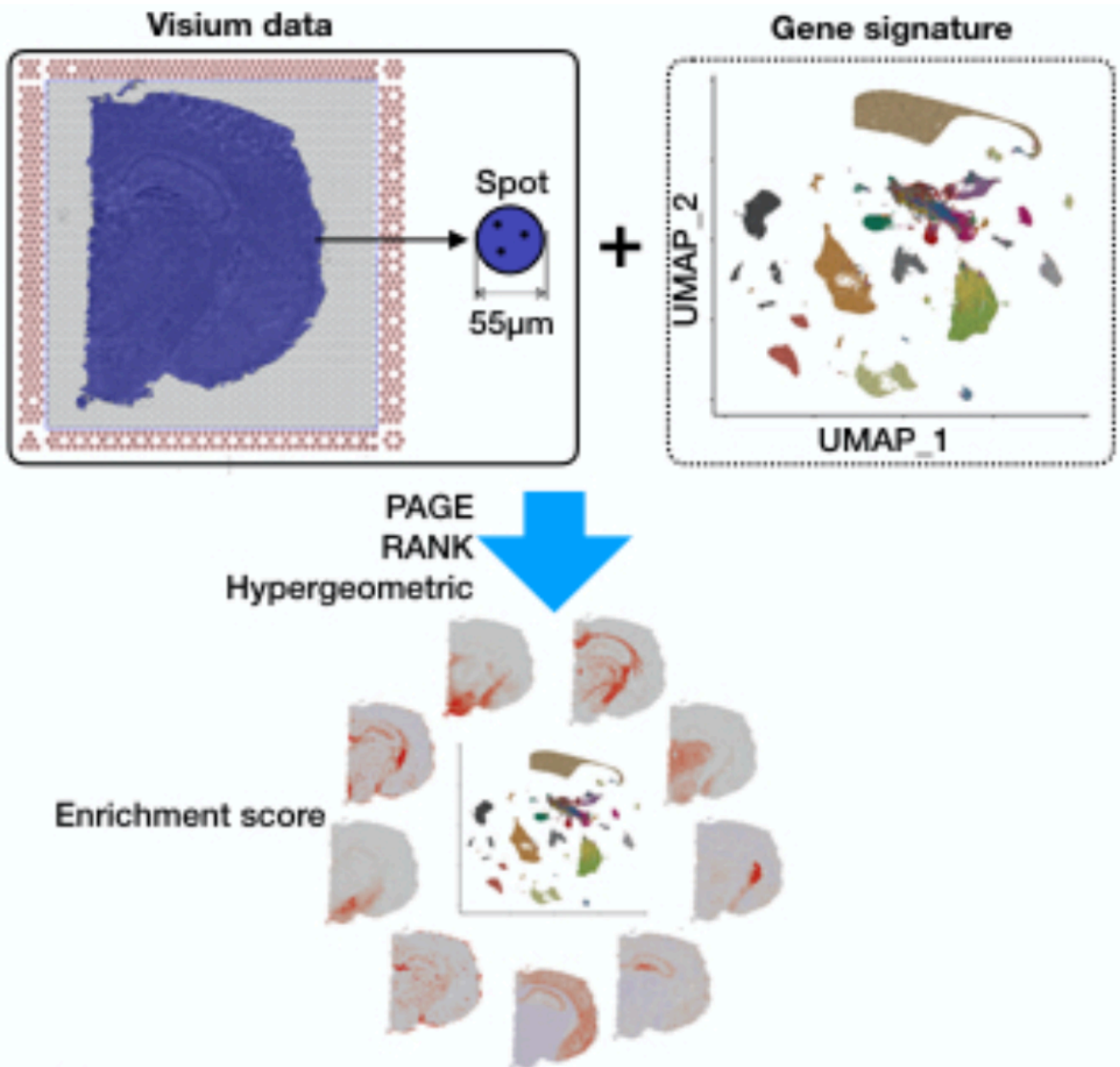
L-R pair

Cell-type pair
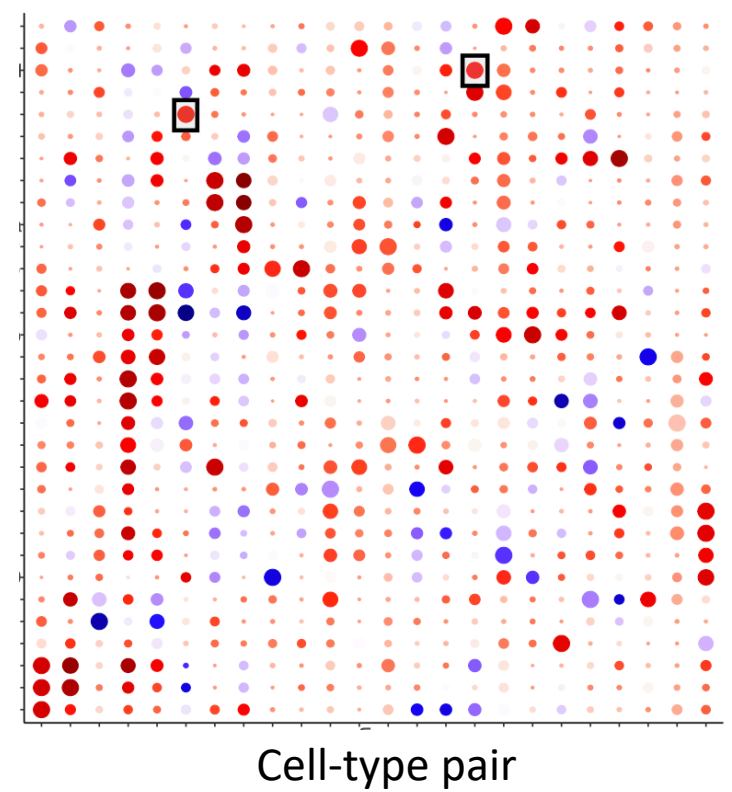
Analysis of data with lower spatial resolution
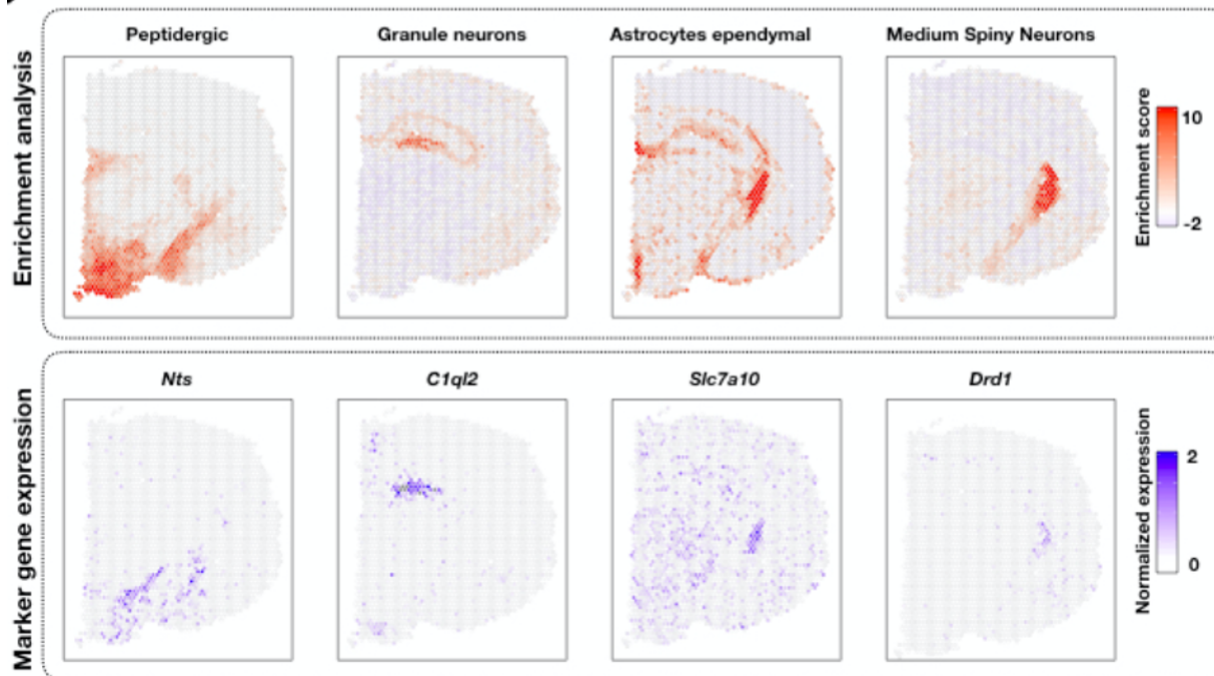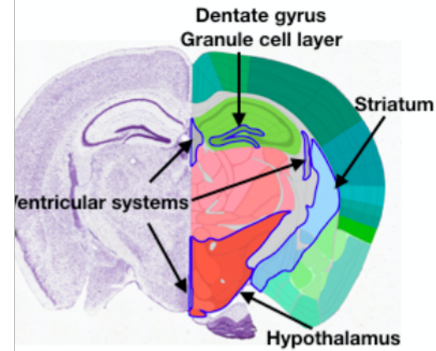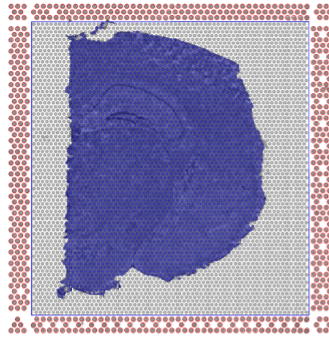
10X Genomics Visium (mouse brain)

How to estimate cell type distribution when the data does not have single-cell resolution?

Spatial enrichment of cell types
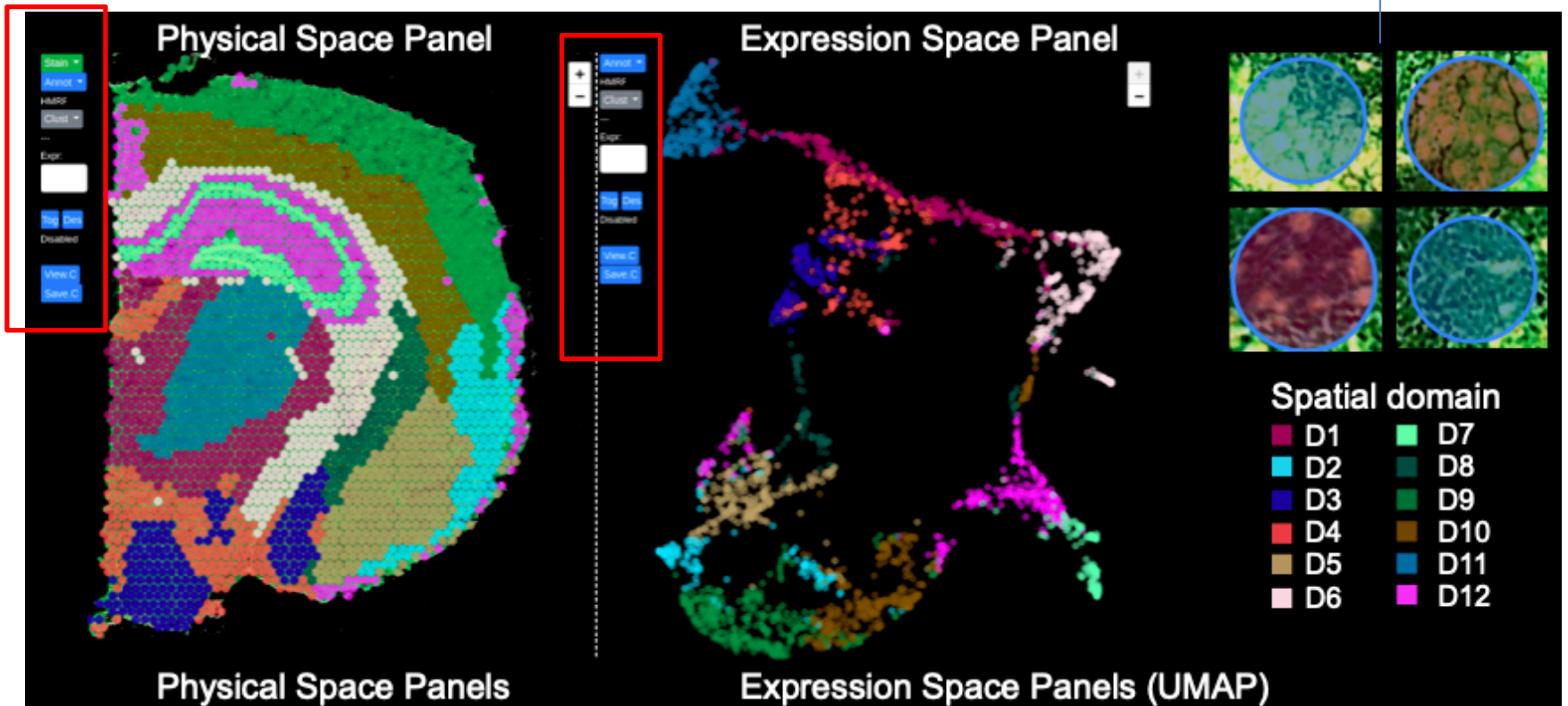

Rui Dong

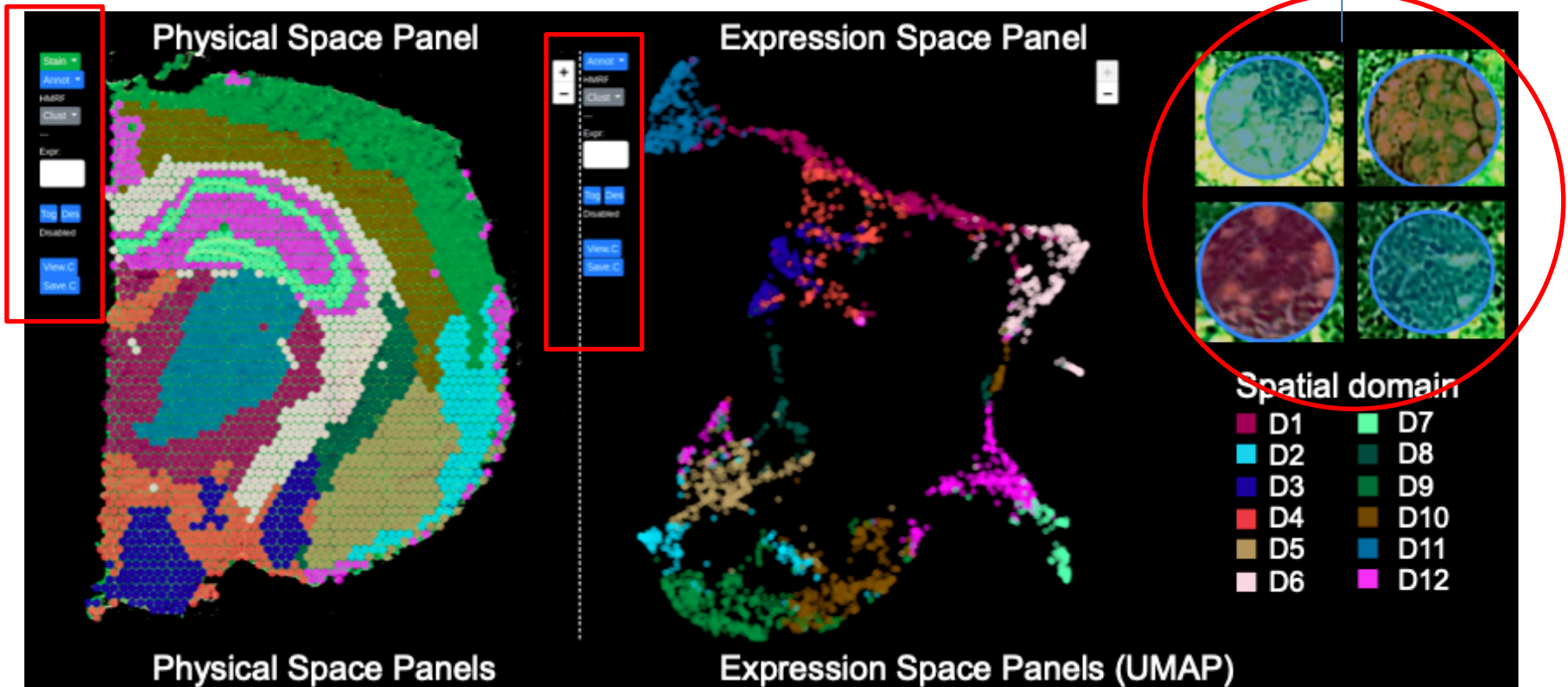# Spatial enrichment of cell types

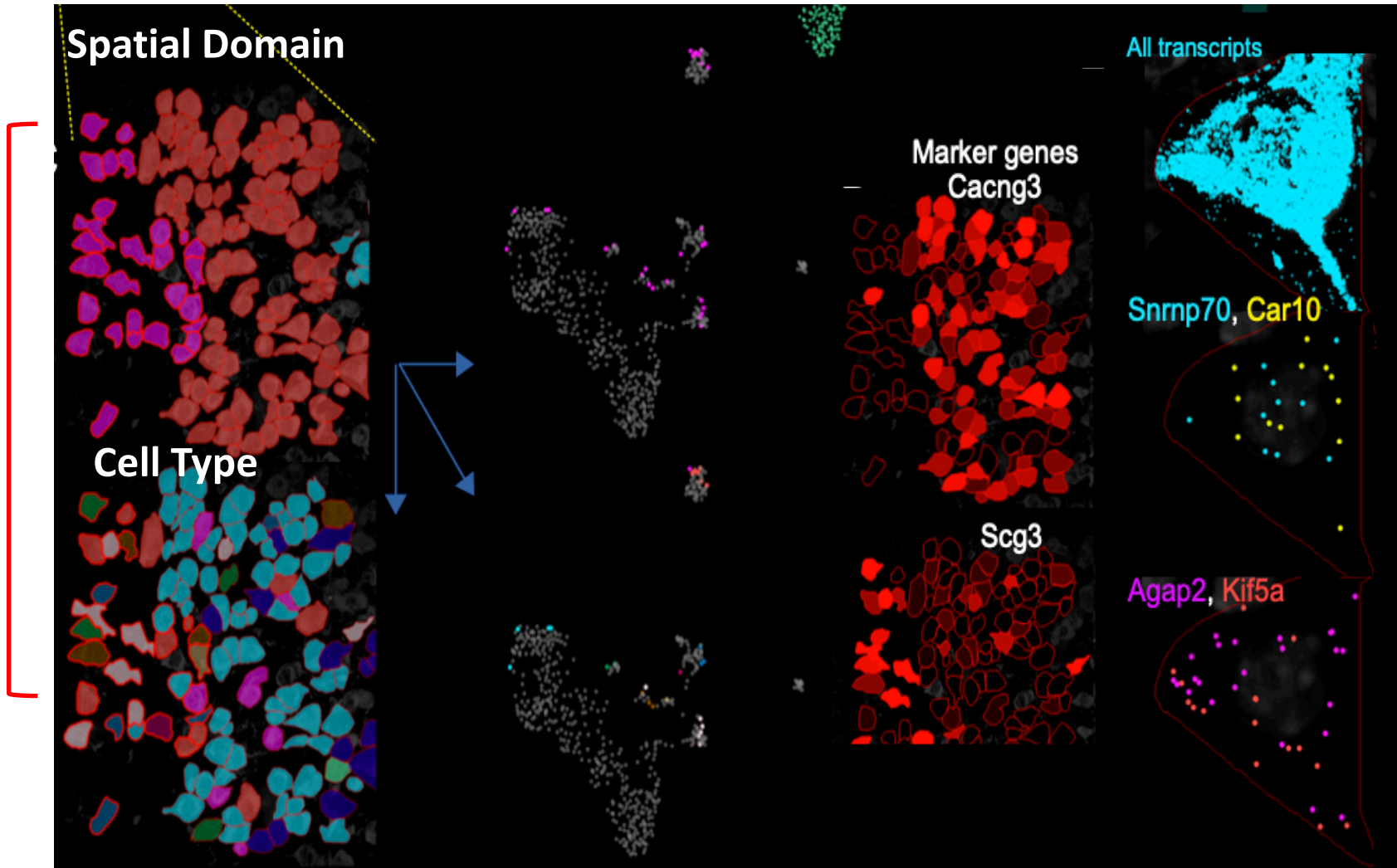10X Genomics Visium (mouse brain)

Giotto Viewer

Interactive visualization:

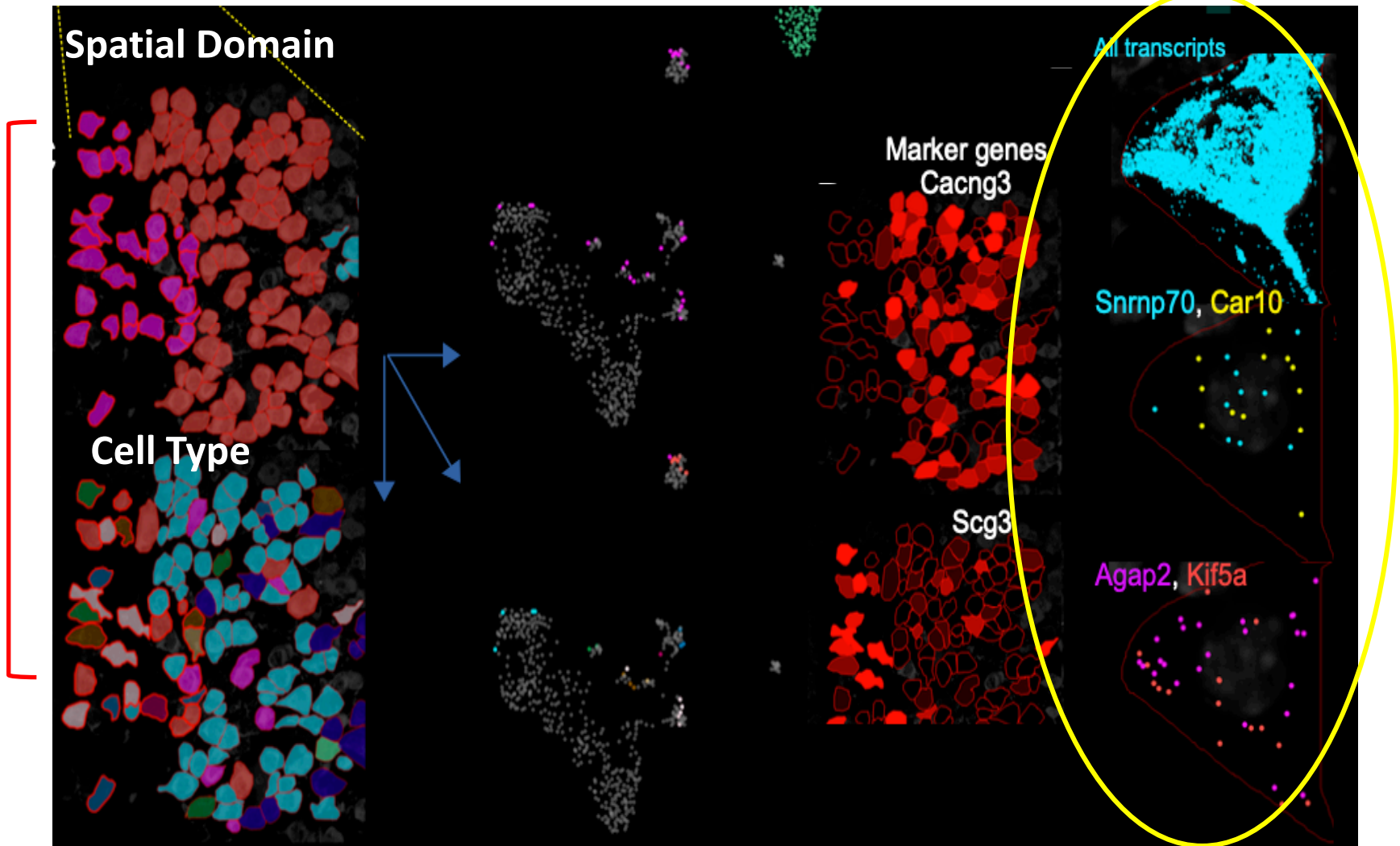Qian Zhu

Giotto Viewer

Interactive visualization:

overlaying staining images

Qian Zhu

Giotto Viewer

Compare annotations

Giotto Viewer

Compare annotations

Subcellular localization

Spatial Domain

Cell Type

Marker genes
Cacng3

Scg3

All transcripts

Snrnp70, Car10

Agap2, Kif5a

Acknowledgement

Yuan Lab

Ruben Dries
Rui Dong
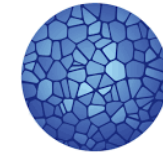Kan Liu
Tianxiao Zhao
Arpan Sarkar

Qian Zhu
Huipeng Li
Yuntian Fu
Feng Bao

Cai Lab
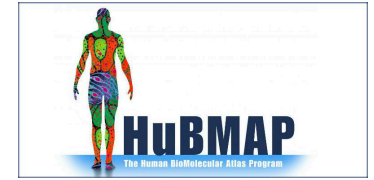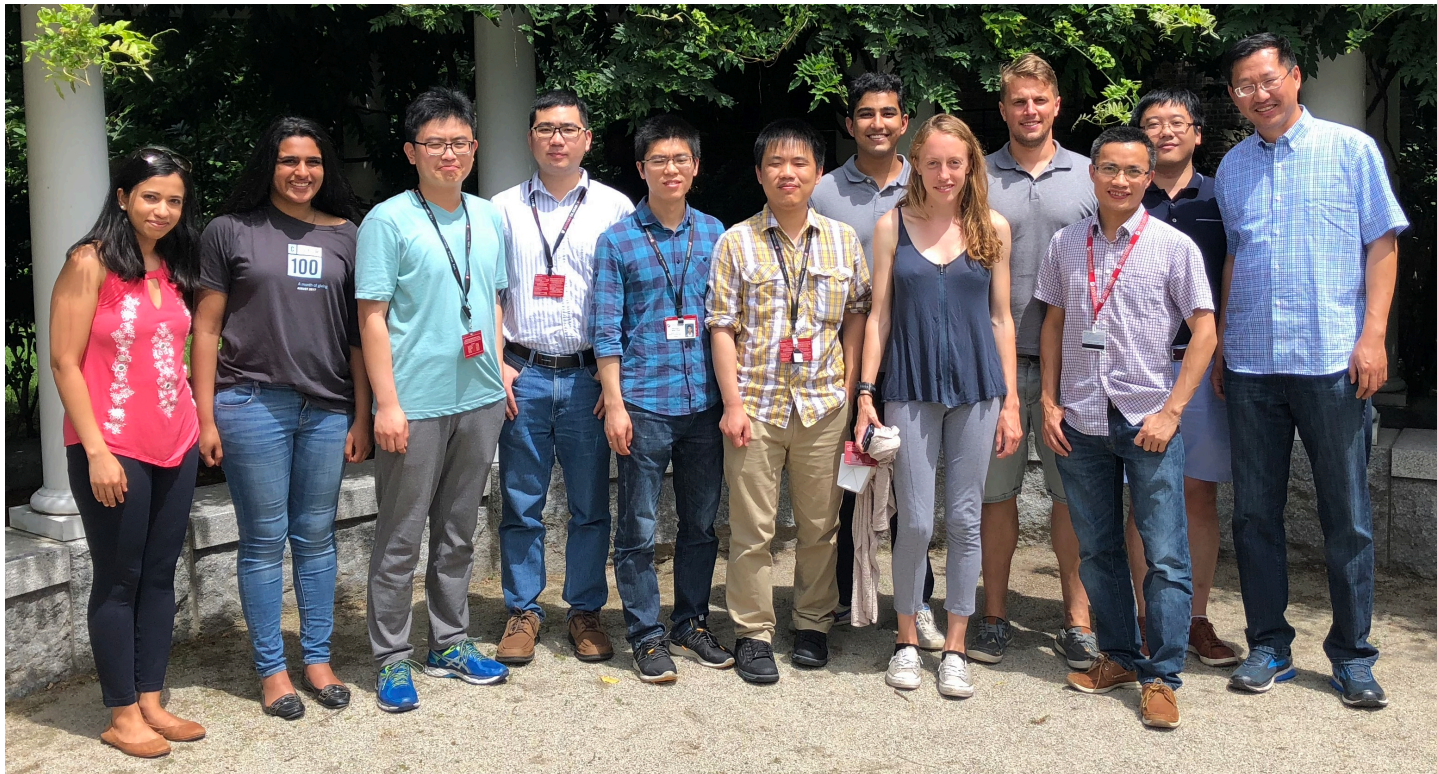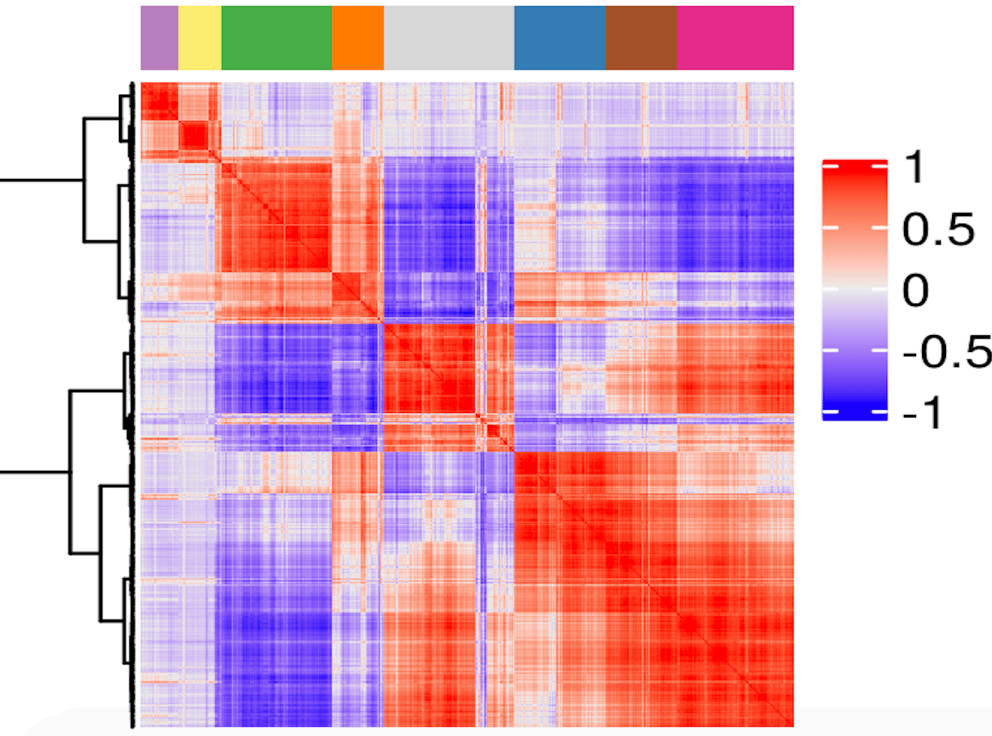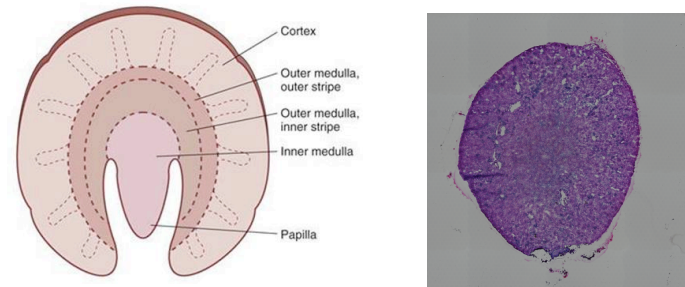
Long Cai
Sheel Shah
Linus Eng
Niko Pierson

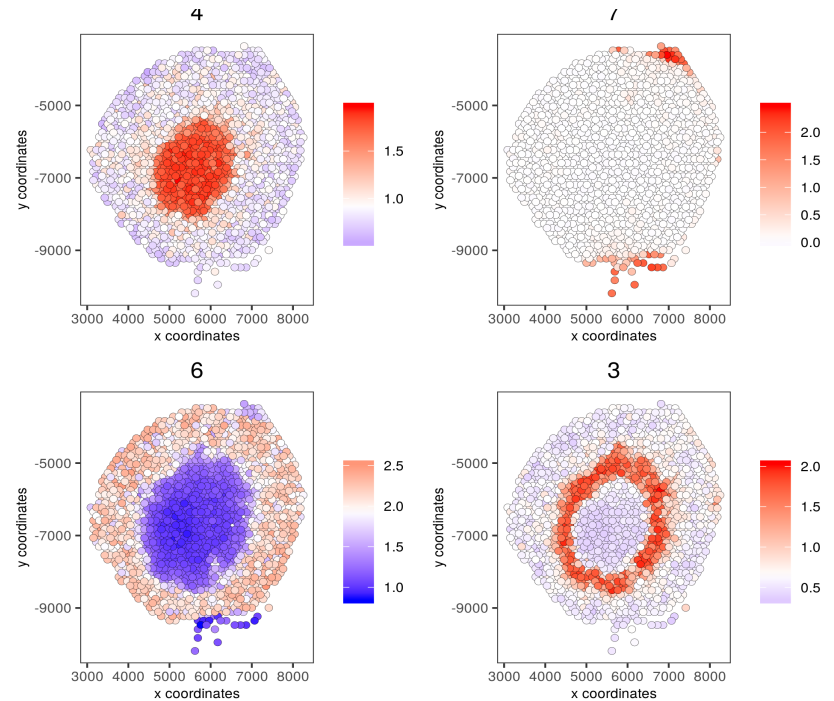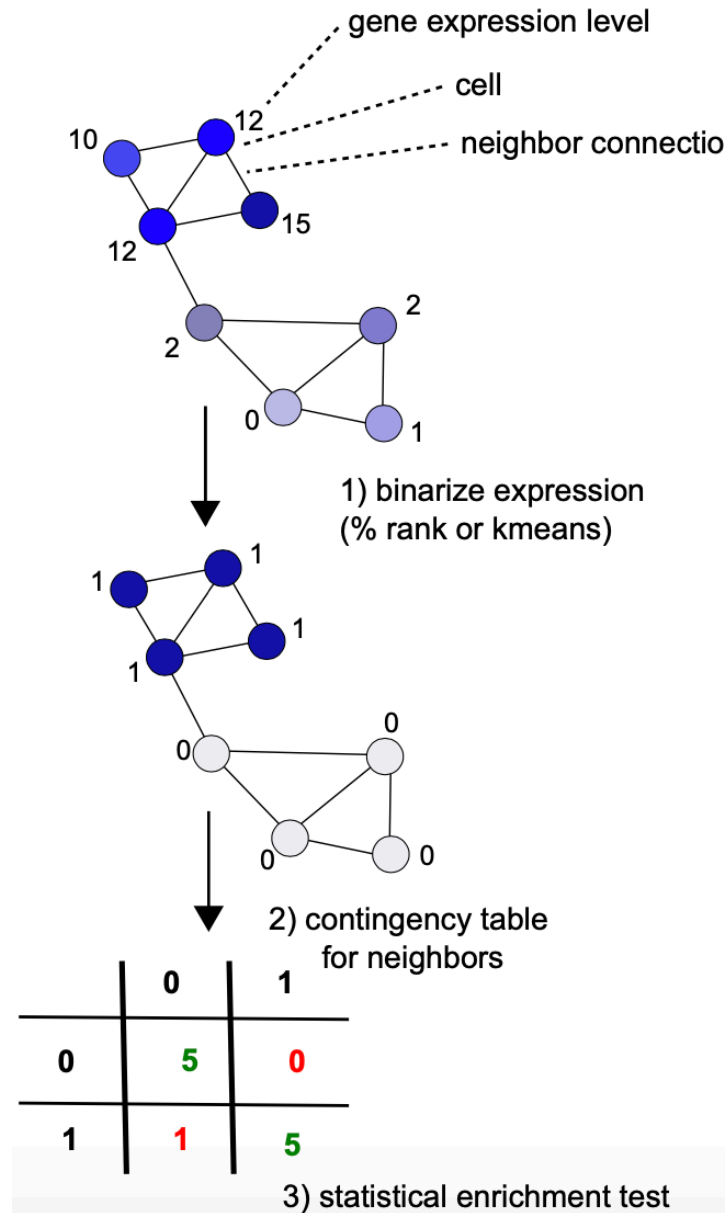# Notes

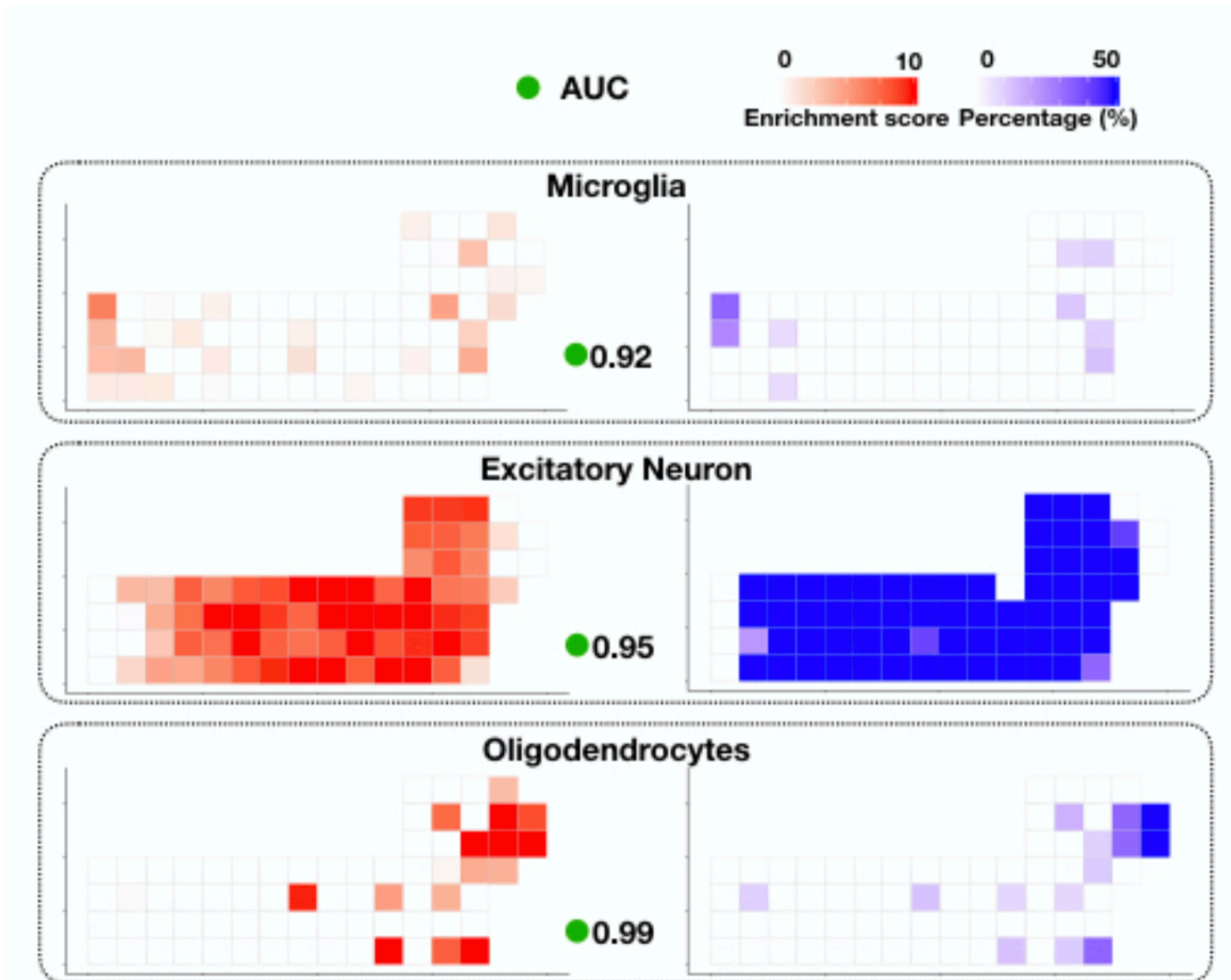# Spatial co-expression module
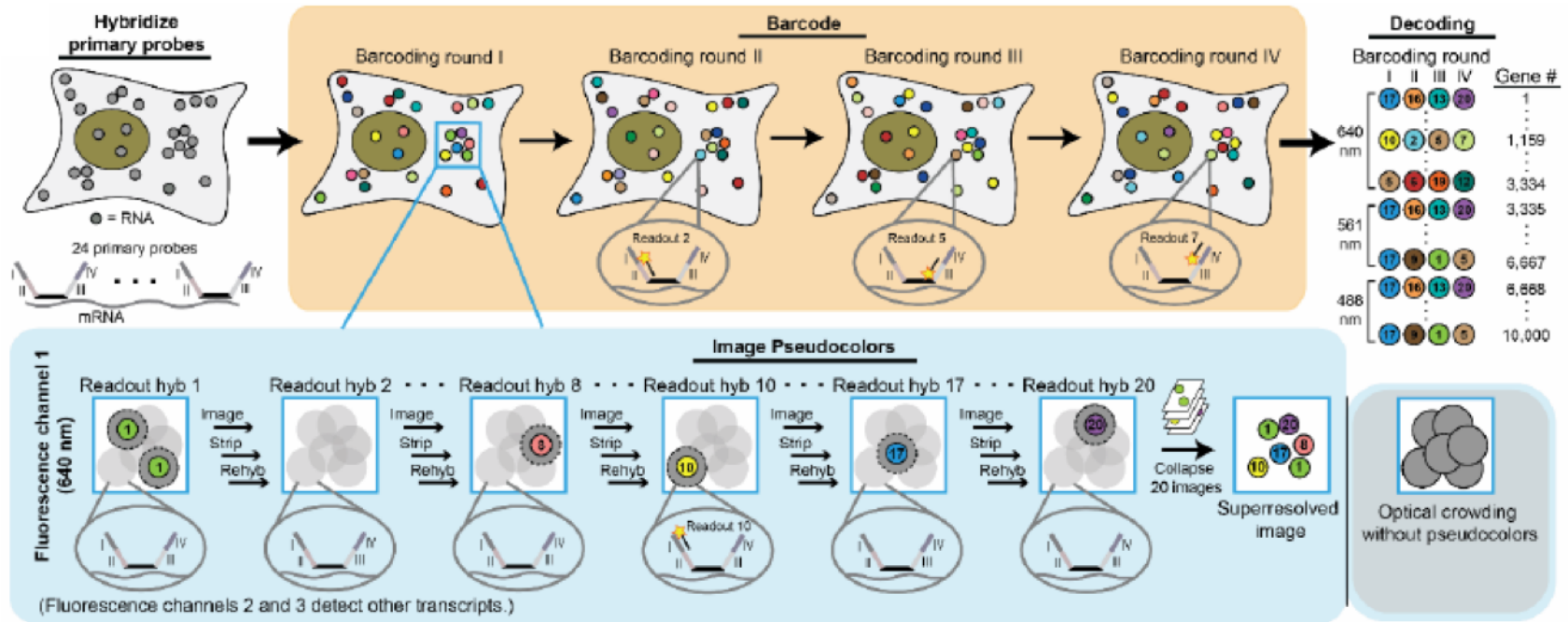

spatial co-expression modules

binSpect for spatial gene detection

Validation of spatial enrichment analysis

seqFISH+: transcriptome-scale super-resolved imaging

➤ Solving the optical crowding problem.

➤ Quantification of 10,000 genes in a single cell.

➤ Detecting subcellular spatial organization.

Eng et al. 2019, Nature