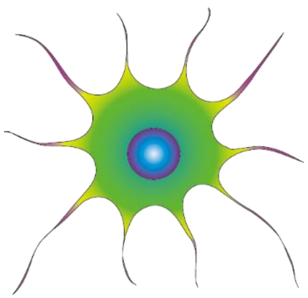
# SSAM:
# Spot-based spatial cell type analysis with multidimensional mRNA locations

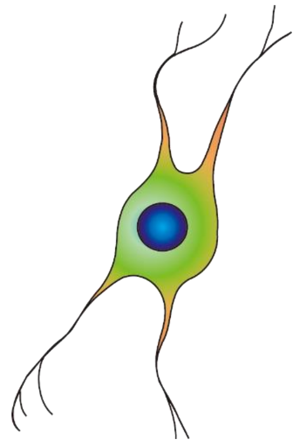
**Jeongbin Park**
Digital Health Center
Berlin Institute of Health & Charité University Hospital

# Cell types

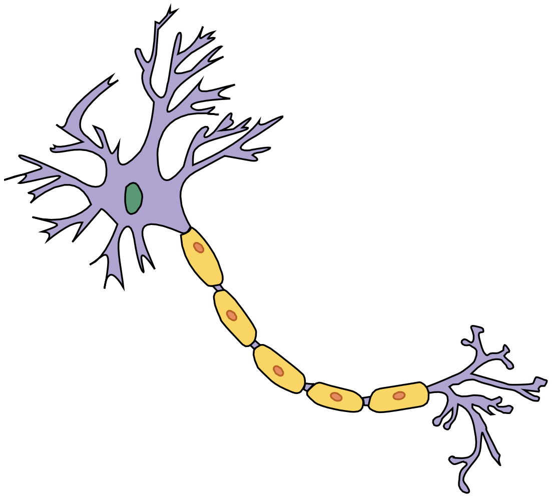
- What is cell type?



Astrocytes



Oligodendrocytes



Neurons

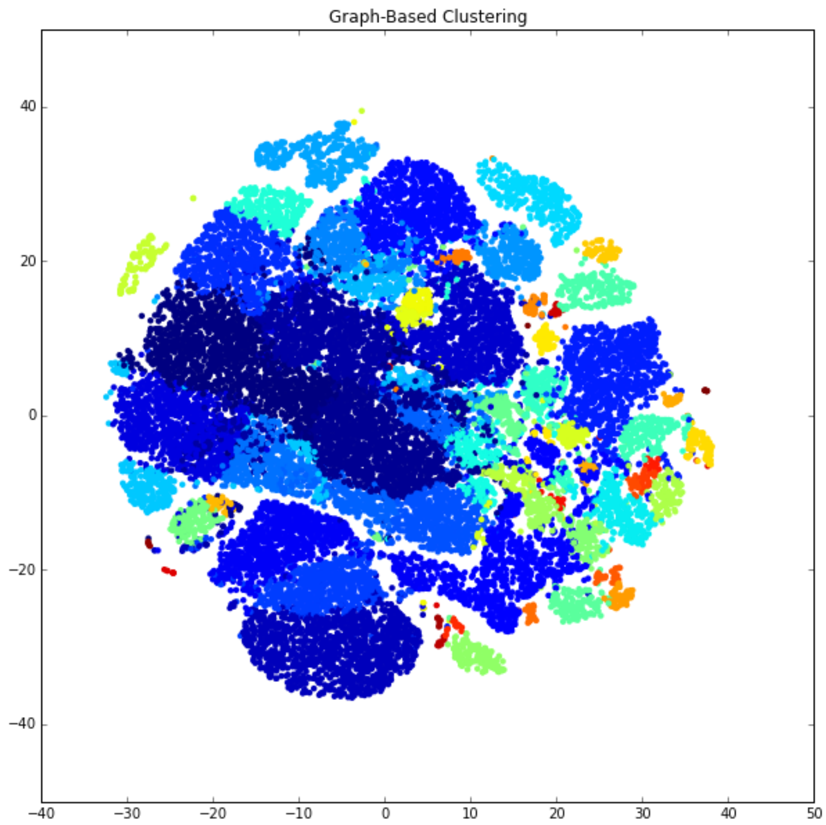
…

- Cells with different shapes and functions

  -> Cells with different gene expressions profiles?

* Cell type images from Wikimedia common

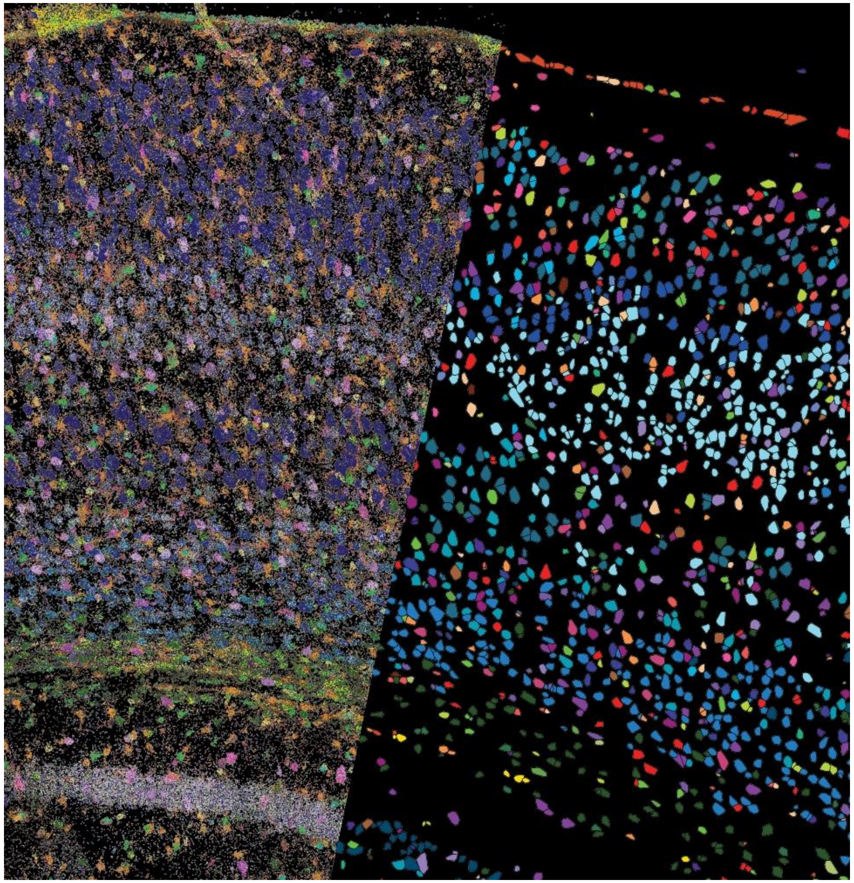
# Cell type identification in scRNAseq data




1.3 million mouse brain cells dataset
(https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/1M_neurons)

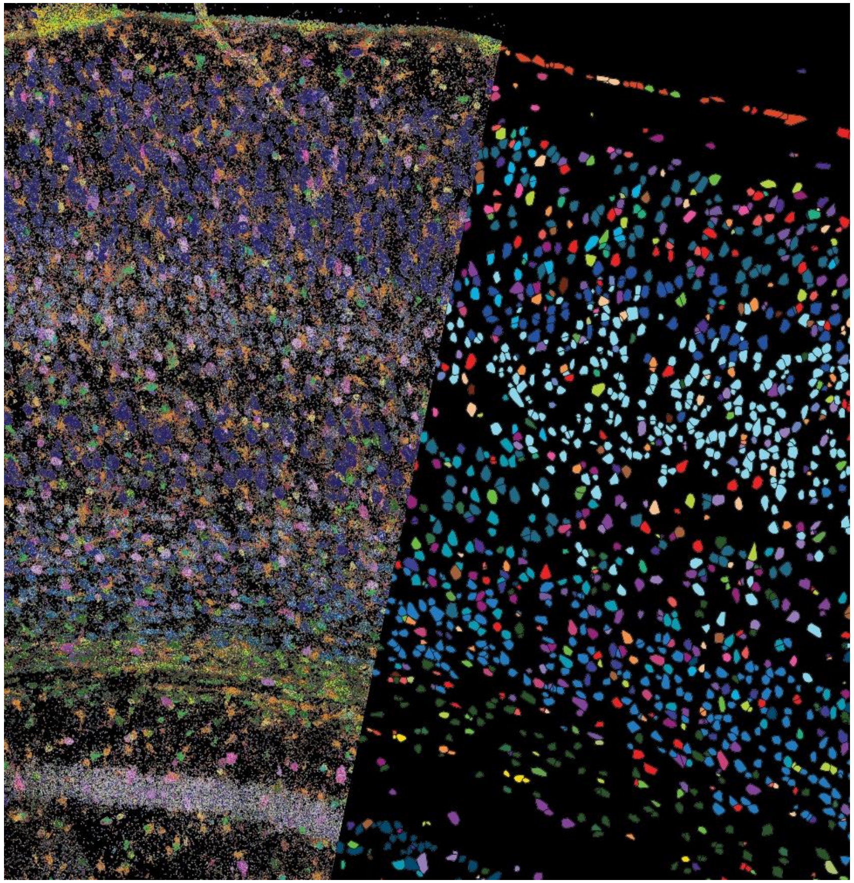
- How?
  - Sequencing individual cell
  - Clustering cells
  - Identifying cell types
- scRNAseq challenges
  - Dropouts
  - Batch effects

# Cell type identification in multiplexed FISH data


osmFISH dataset (left: mRNAs, right: cell types)

- How?
  - Cell segmentation
  - Counting mRNAs within each cell's border
  - Clustering cells
  - Identifying cell types
- Multiplexed FISH
  - Sensitive and accurate
  - **Location** of cell types

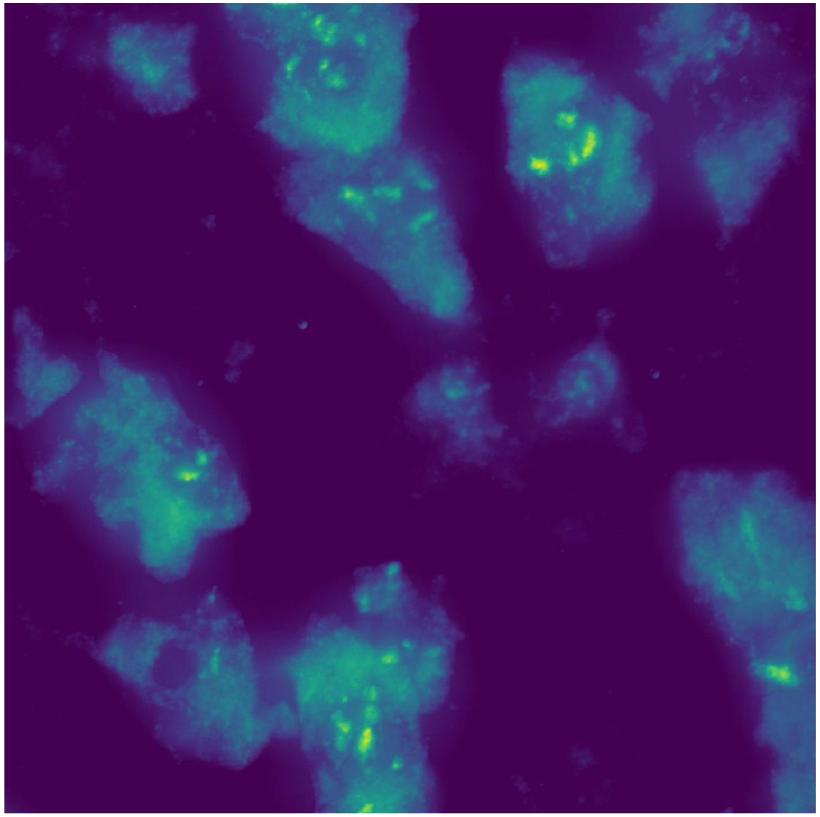
# Cell type identification in multiplexed FISH data



osmFISH dataset (left: mRNAs, right: cell types)

- How?
  - Cell segmentation
  - Counting mRNAs within each cell's border
  - Clustering cells
  - Identifying cell types
- Multiplexed FISH
  - Sensitive and accurate
  - **Location** of cell types

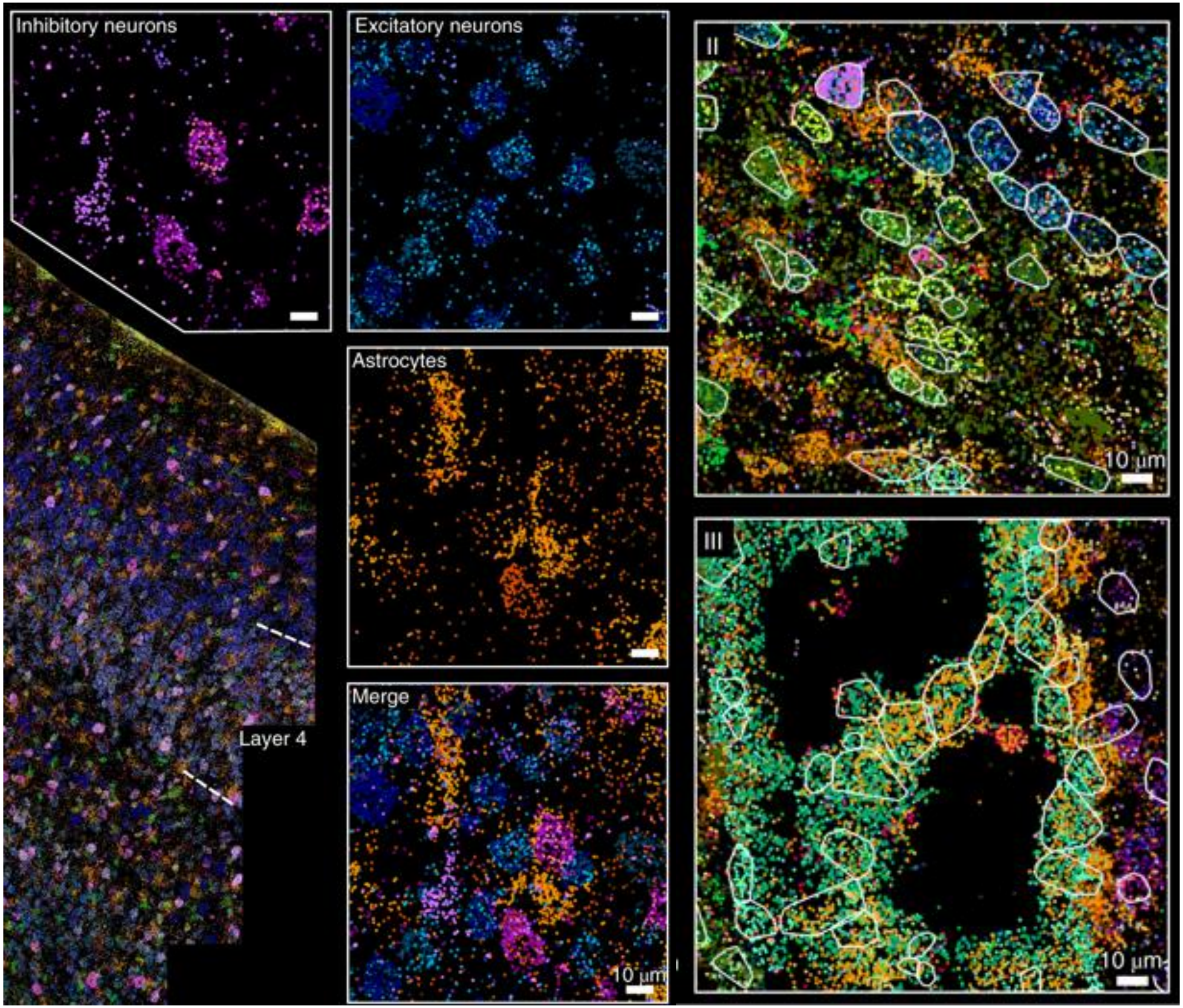
# Problem: Segmentation is not easy



Poly-A image from osmFISH dataset

- Complex cell shapes
- Requires additional images (DAPI, poly-A, etc.)
- Unclear cell borders due to imaging problem

*Nature Methods* **15**, 932–935 (2018)




mRNA distribution already looks very similar to cell shapes…

# SSAM (Spot-based Spatial cell type Analysis with Multidimensional mRNA locations)

- Spot-based spatial cell type calling
- Cell type calling without segmentation
- Works with multidimensional mRNA locations (in 2D or 3D)
- Easy to use Python package

# Requirements

- Location of mRNA (in 2D or 3D)

- Uniform distribution of mRNA within a cell
  - i.e. Density of mRNAs of a certain gene should be similar at any point in a cell

# Kernel Density Estimation

- Estimate density by summing up kernels at i-th mRNA position

$$\hat{\sigma}_h(\vec{x}) = \frac{1}{Nh}\sum_{i}^{N} K\left(\frac{\vec{x}-\vec{x}_i}{h}\right)$$

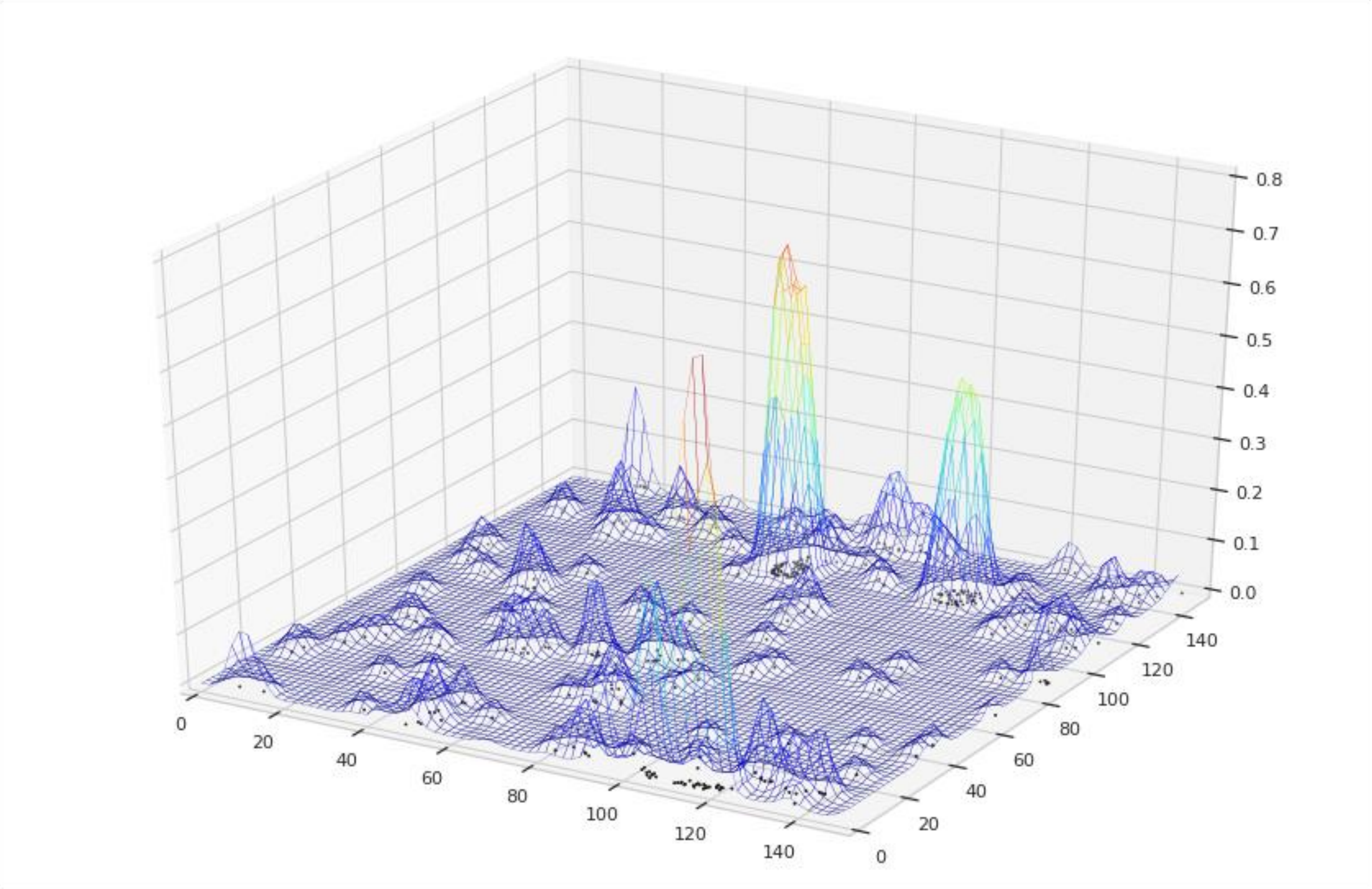
  - Where σ is density, N is number of mRNAs, K is kernel, h is bandwidth, xi is position of i-th mRNA

- We used Gaussian kernel for simplicity:

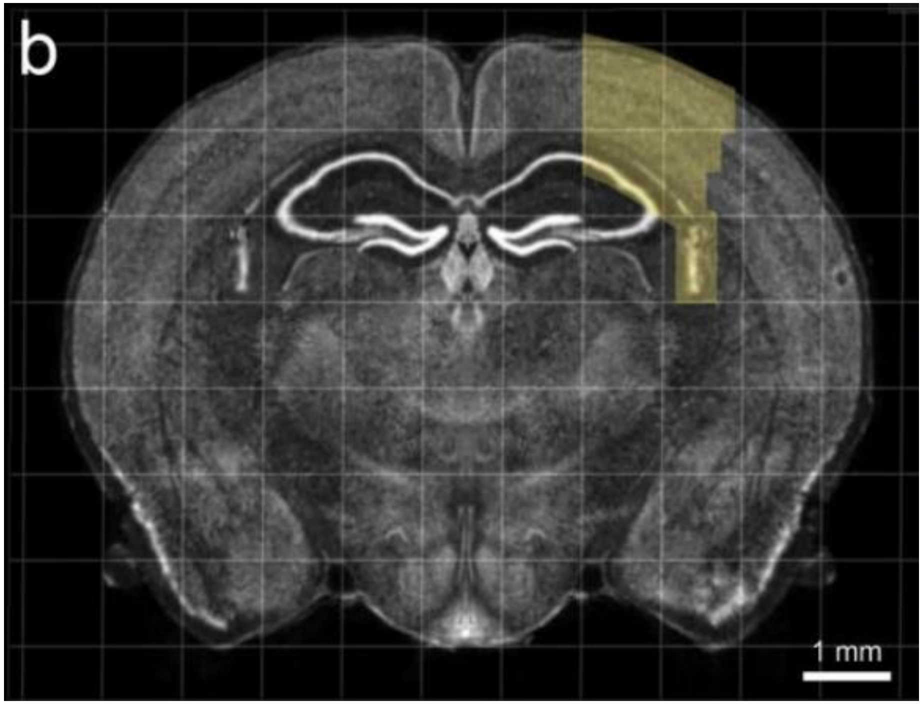
$$K_{gaussian}(\vec{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}|\vec{x}|^2\right)$$

- Select h, which makes 2*FWHM(σ) ~ cell radius (h=2, assuming that cell radius ~ 10um)

# Kernel Density Estimation



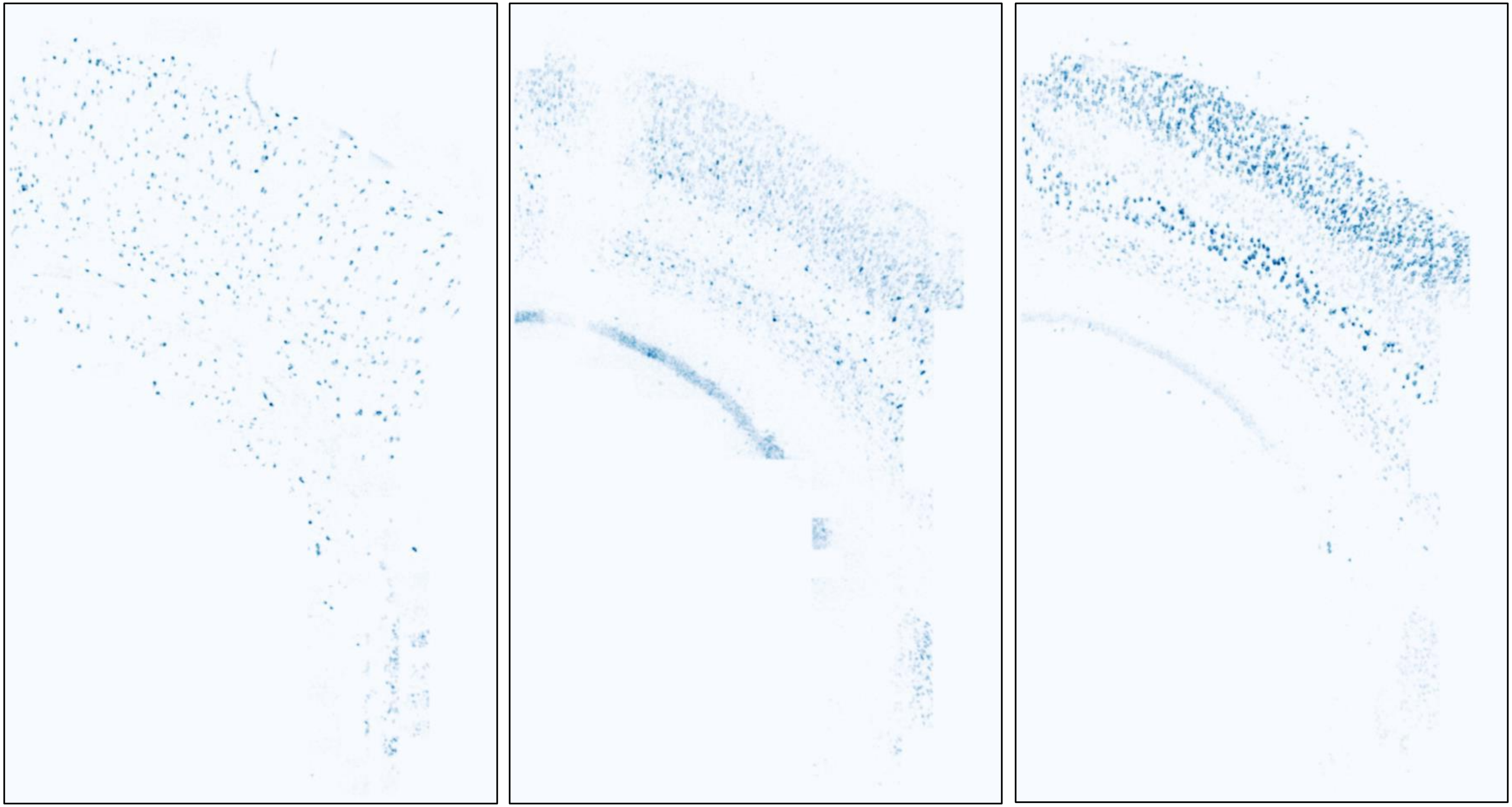
# Data analysis example - osmFISH



*Nature Methods* **15**, 932–935 (2018)

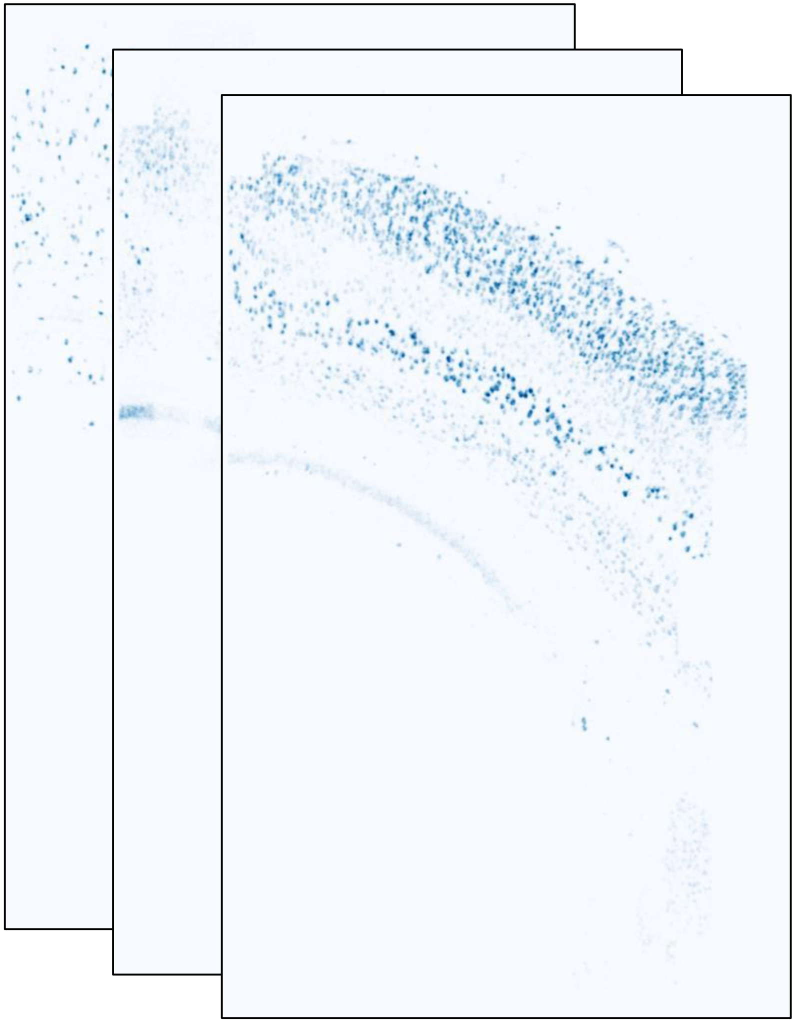
- Mouse somatosensory cortex
- 2080 x 3380 um (2D)
- 33 genes

# Kernel Density Estimation (osmFISH dataset)



Gad2 Kcnip2 Lamp5

# Kernel Density Estimation (osmFISH dataset)



- Stack all genes

$\rightarrow$ 33 dimensional vector field

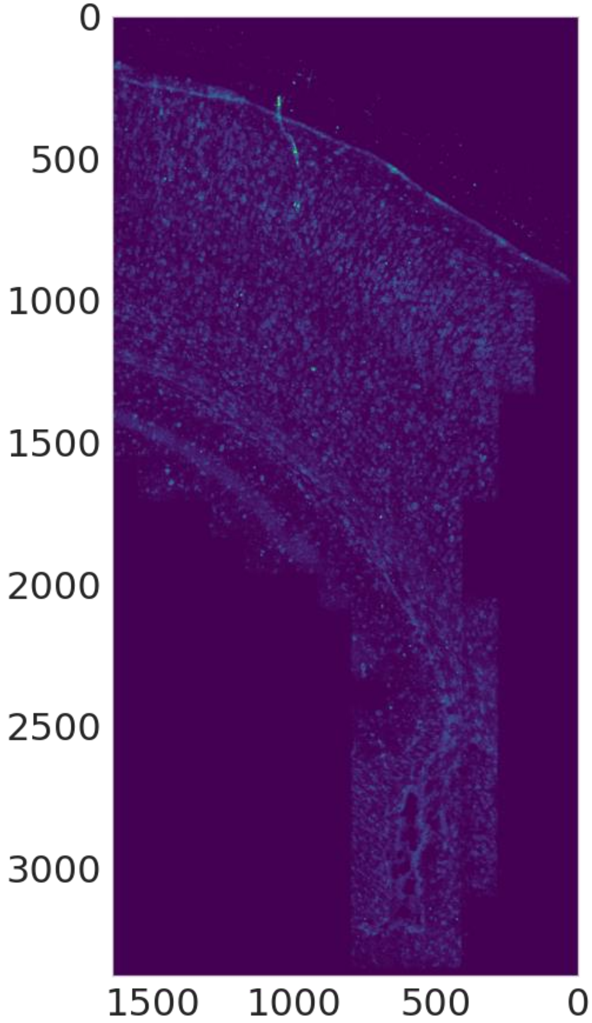
- i-th gene's expression ($E_i$):

$$\hat{E}_i(\vec{x}) = \hat{\sigma}_i(\vec{x}) N_i$$

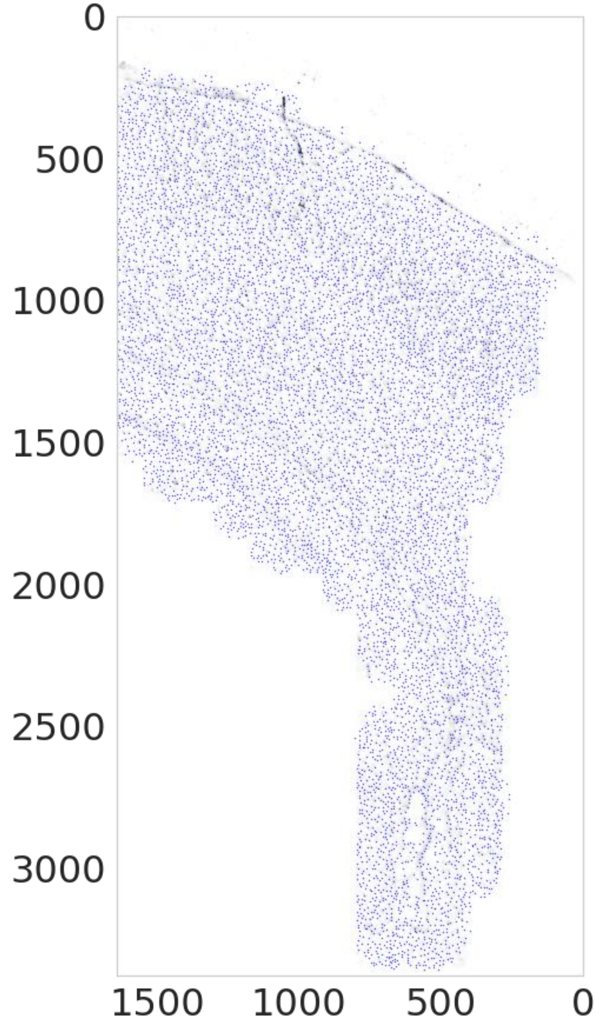
*where*

$E_i$ : expression of $i-th$ gene
$\sigma_i$ : estimated density of $i-th$ gene
$N_i$ : number of mRNAs of $i-th$ gene

# Selection of representative vectors

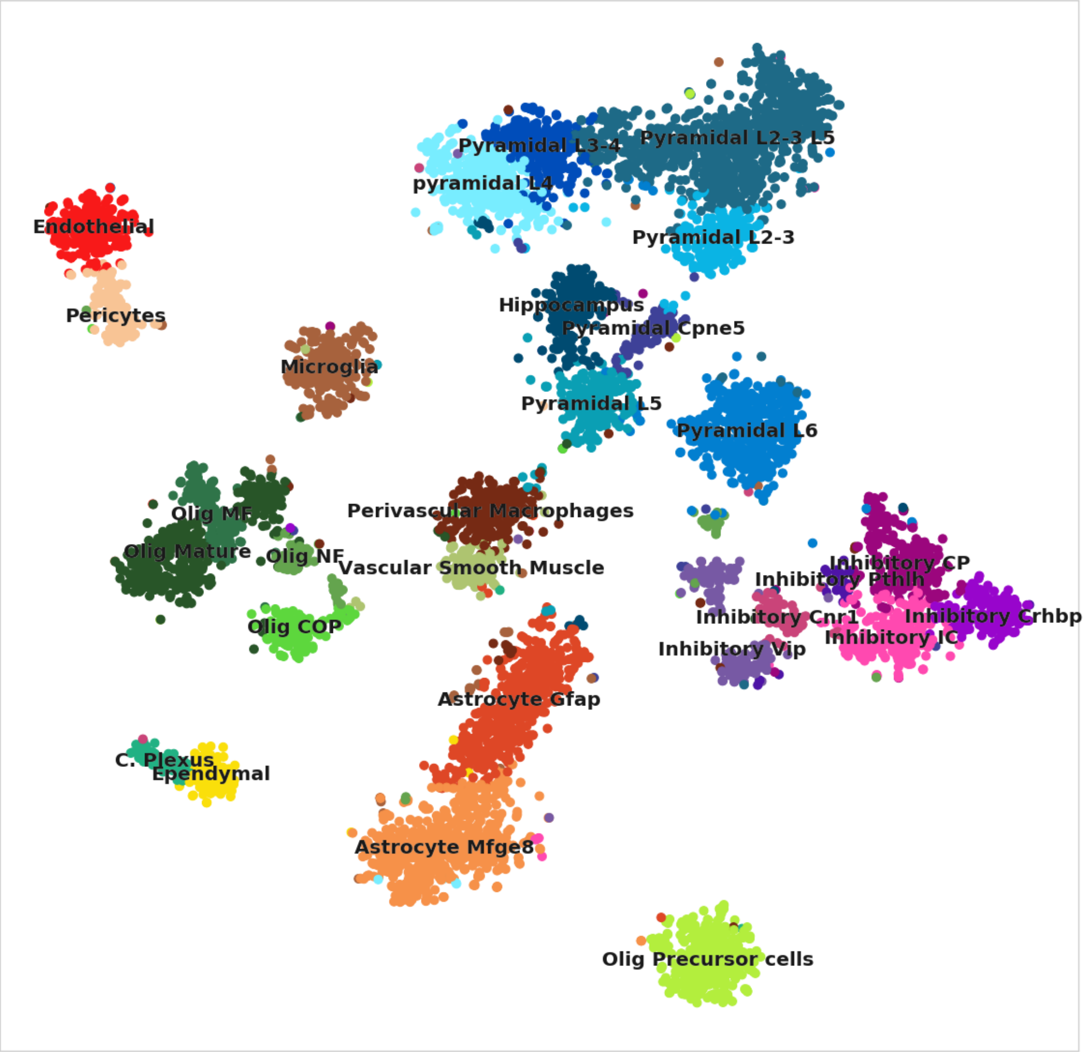



3,380 x 2,080 = 7,030,400 vectors




L1 maxima

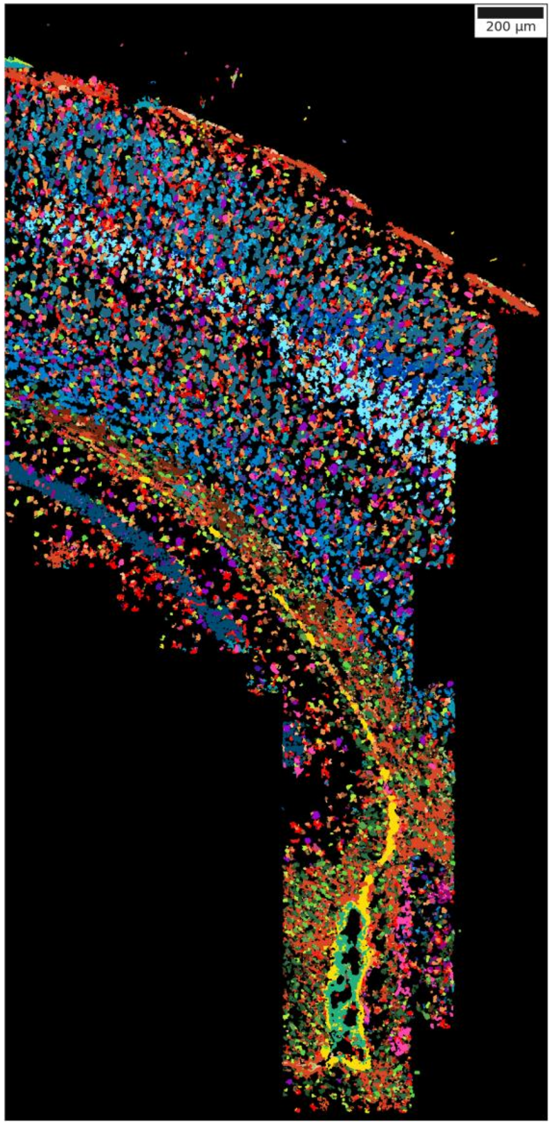
# Clustering

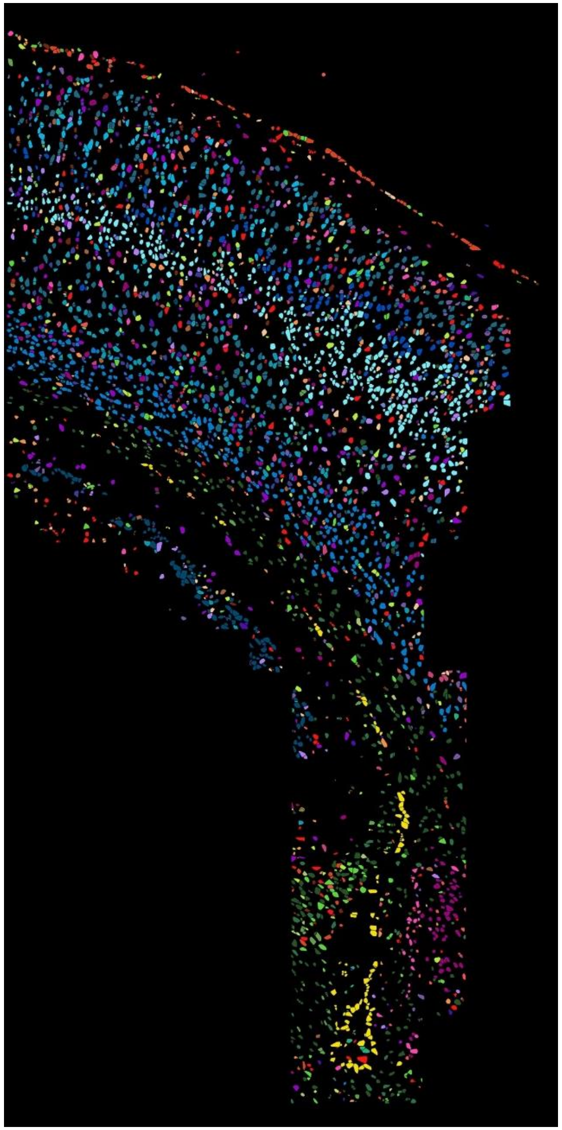



SNN (weighted by Jaccard index) + Louvain algorithm, 28 clusters

# Side-by-side comparison of cell type map

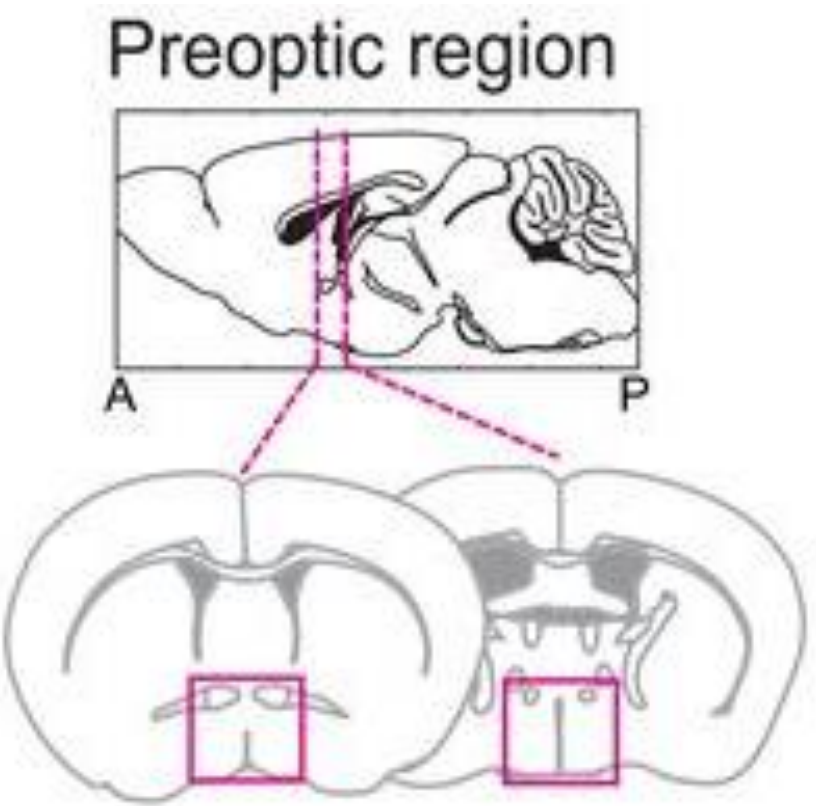
SSAM generated cell type map



Segmentation based cell type map

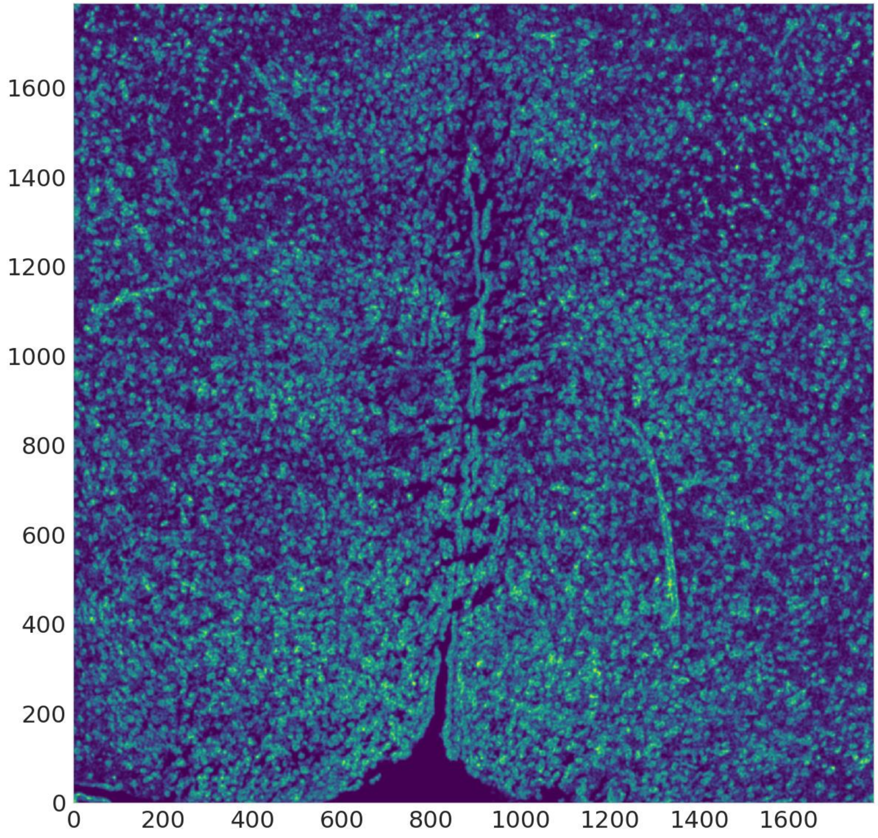
# MERFISH data



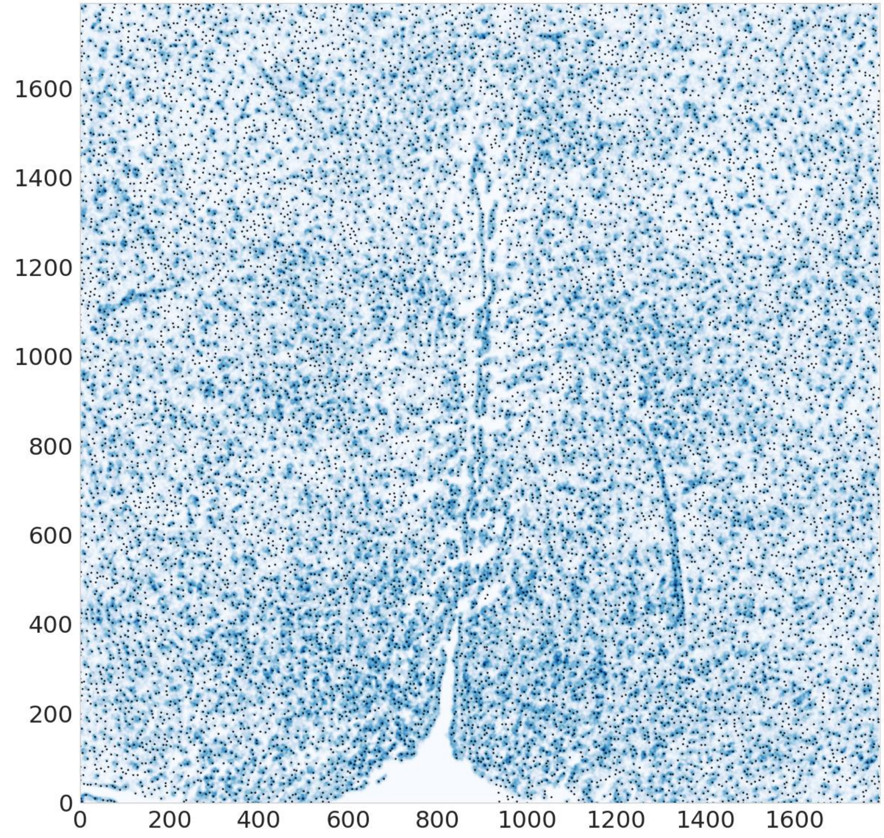

Moffit *et al. Science* 16 Nov 2018: Vol. 362, Issue 6416, eaau5324

- Mouse hypothalamic preoptic region
- 1790 x 1790 x 9 um (3D)
- 135 genes

# Selection of representative vectors

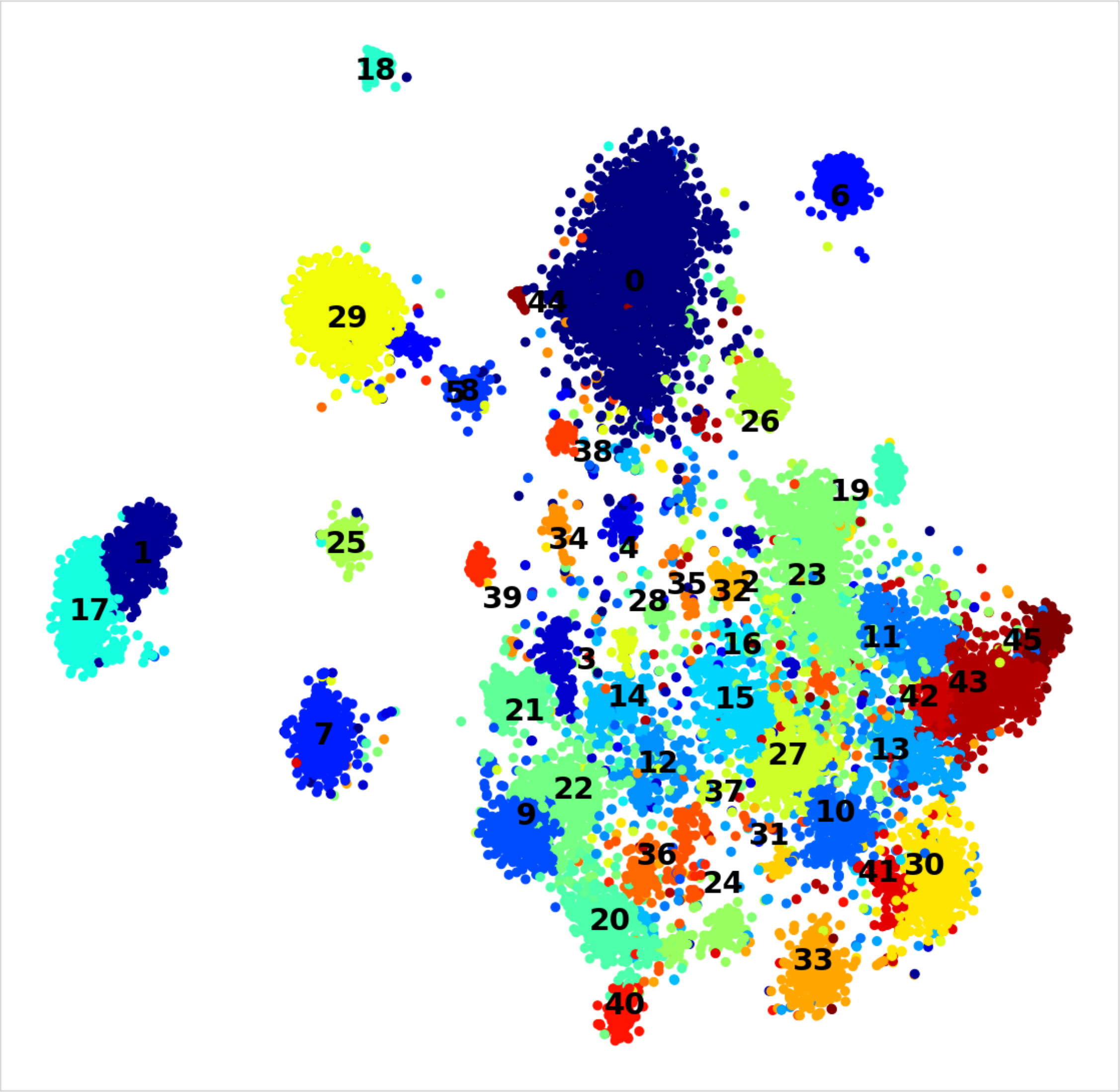



1,790 x 1,790 x 9 = 28,836,900 vectors
(image is at z=4um)

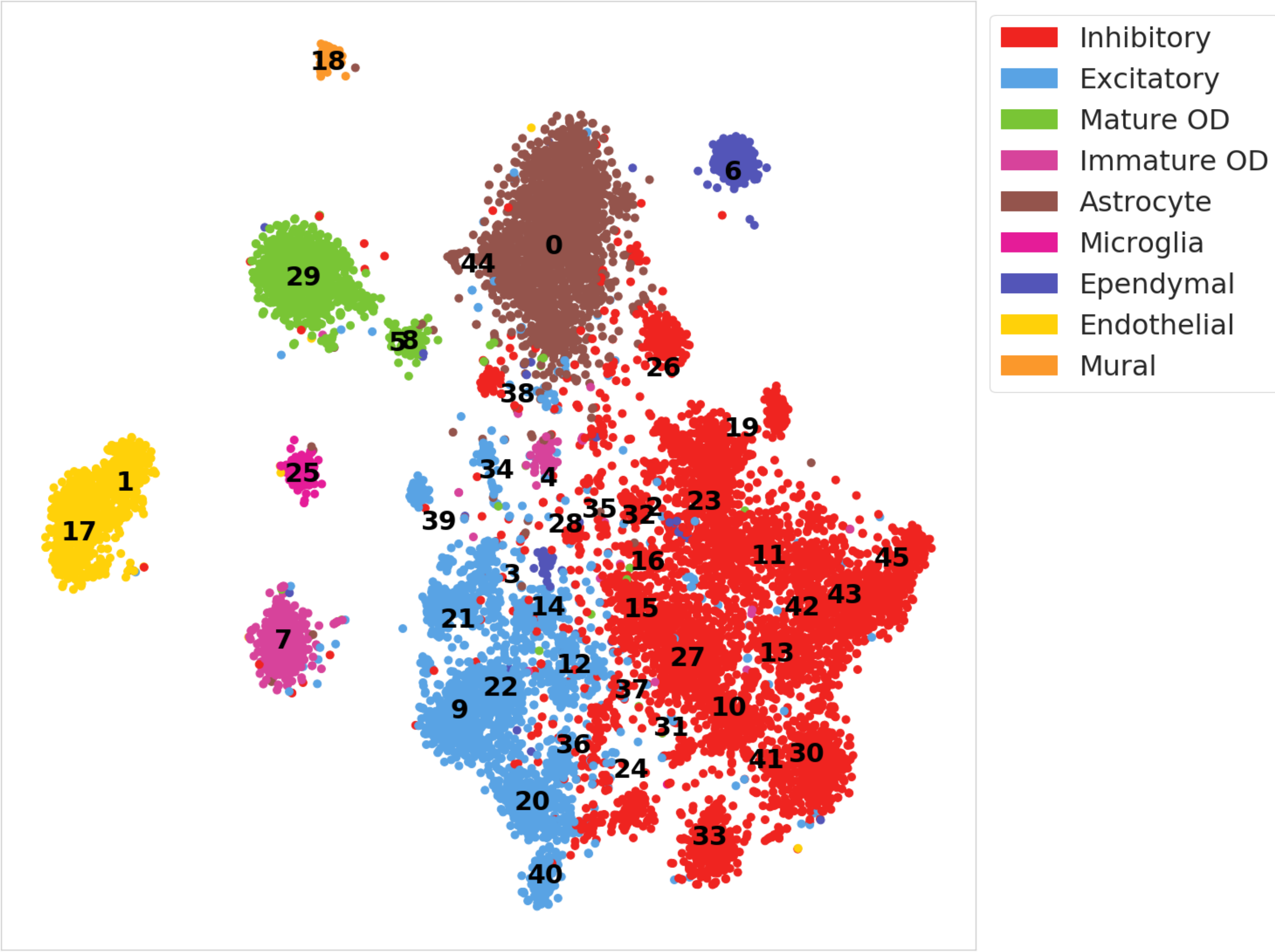



L1 maxima

# Clustering


18
6
29
0
44
58
26
38
19
1
25
34
4
39
35
32
2
23
17
28
11
45
16
3
43
42
14
15
21
7
13
27
12
22
37
10
9
31
36
41
30
24
20
33
40

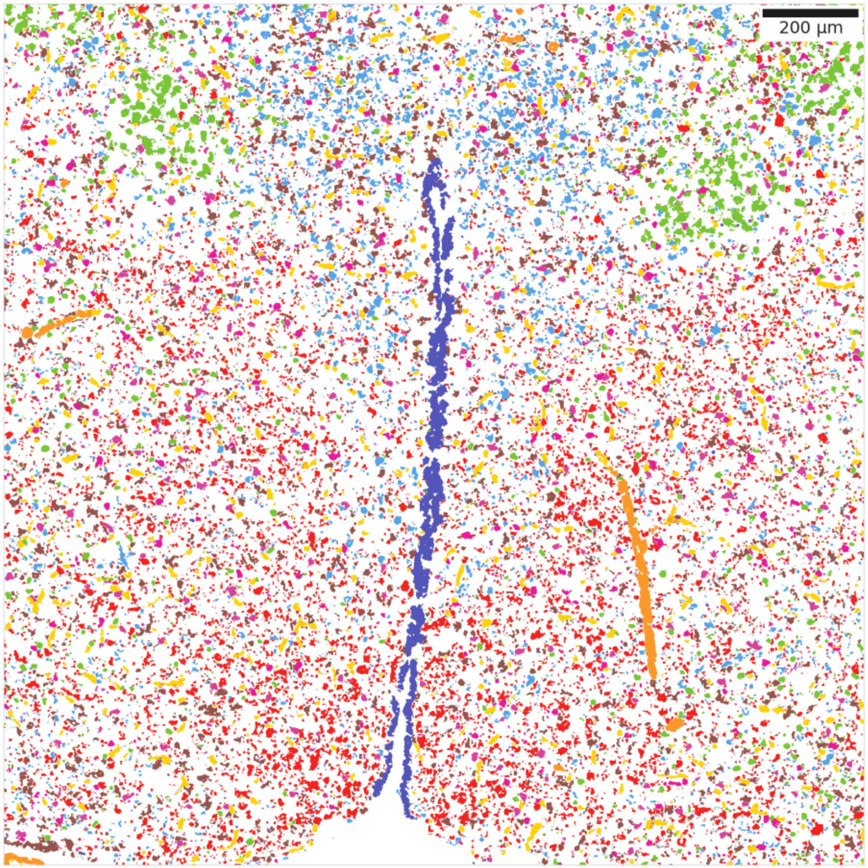
# Clustering



Inhibitory
Excitatory
Mature OD
Immature OD
Astrocyte
Microglia
Ependymal
Endothelial
Mural
18
6
0
44
29
58
26
38
19
25
1
34
4
17
39
35
32
2
23
28
16
11
45
3
14
15
42
43
21
7
12
27
13
22
37
9
10
31
36
41
30
24
20
33
40

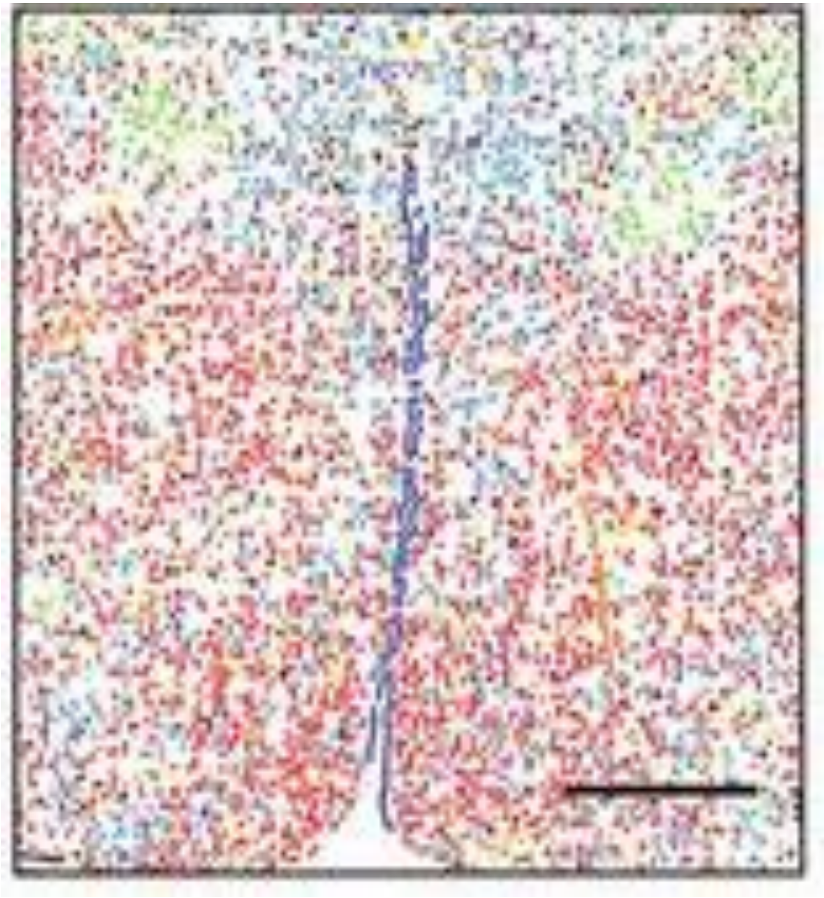
# Cell type map

SSAM generated
cell type map



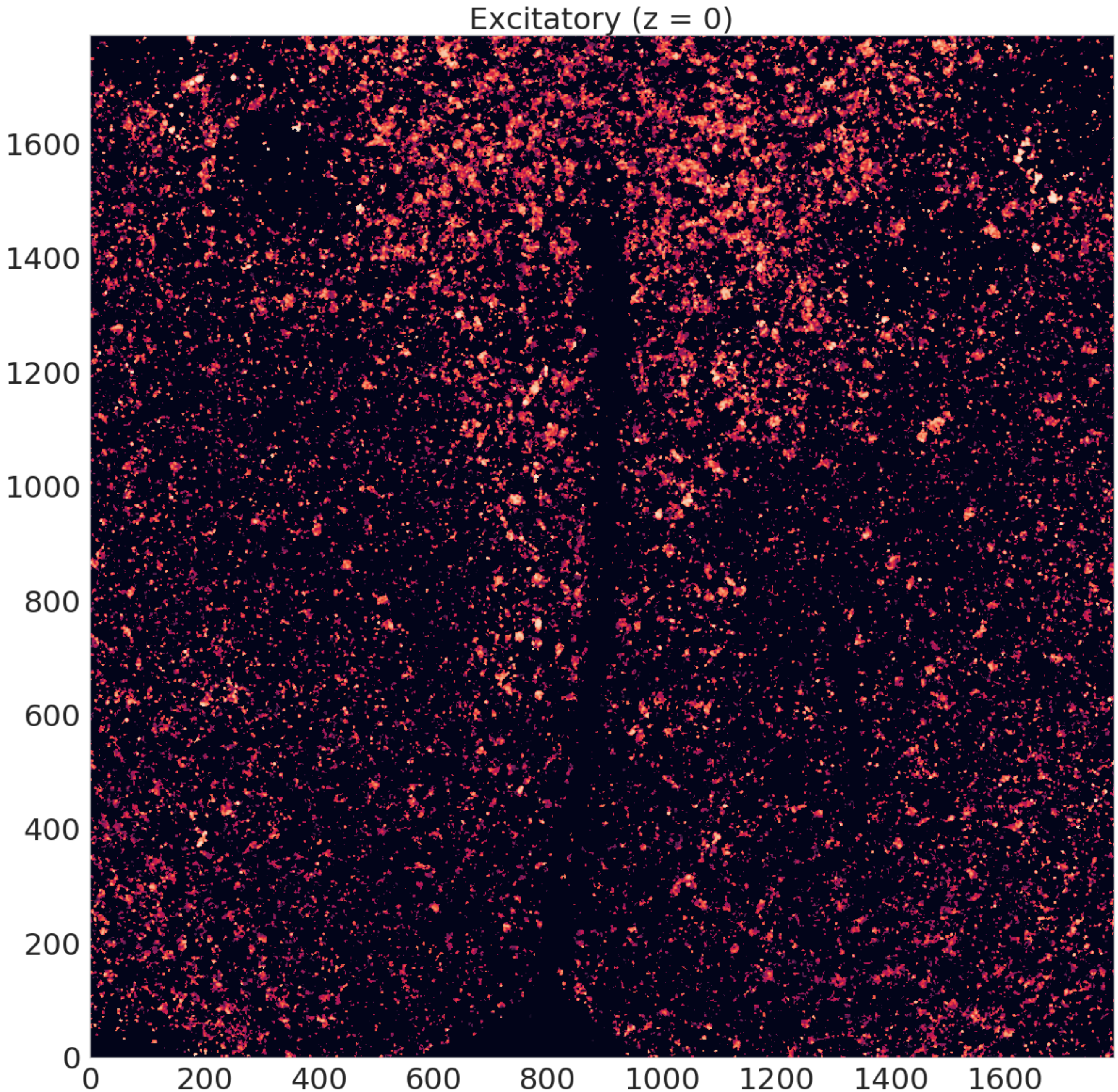

At z = 4um

Segmentation-based
cell type map (Moffit, *et al*.)

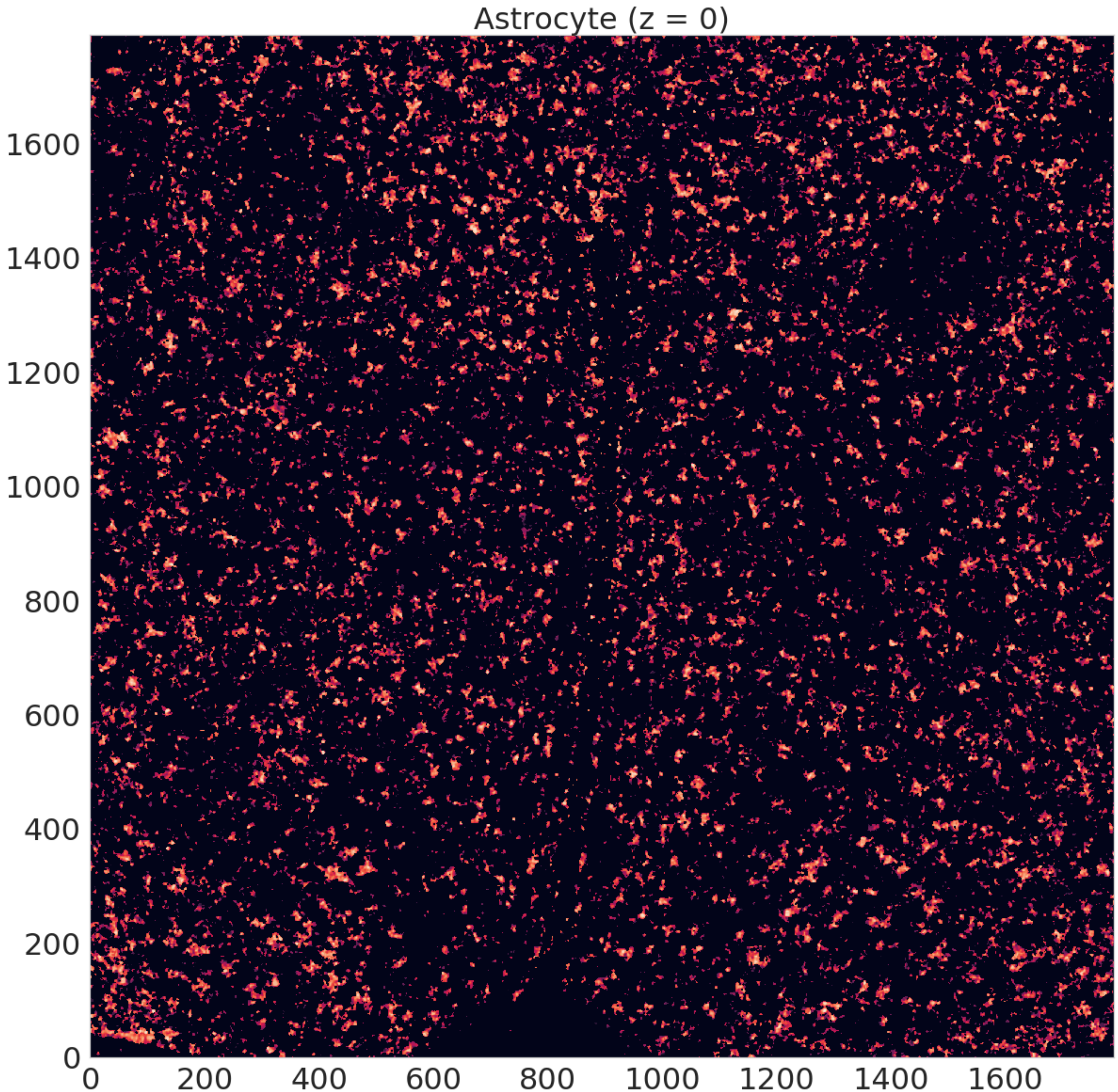


# 3D cell type map

- Movies, sweeping z directions
  - Excitatory neurons
  - Astrocytes



Excitatory (z = 0)
1600
1400
1200
1000
800
600
400
200
0
0
200
400
600
800
1000
1200
1400
1600


Astrocyte (z = 0)
1600
1400
1200
1000
800
600
400
200
0
0
200
400
600
800
1000
1200
1400
1600

# Summary

- We developed SSAM, a segmentation-free method to call cell types

- SSAM can reproduce prior results, also provides more detailed structure of cell types in tissue

# Acknowledgements


- **Wonyl Choi (Boston U.)**

**BIH/Charite, Berlin**

- Naveed Ishaque
- Roland Eils
- Luca Tosti
- Christian Conrad

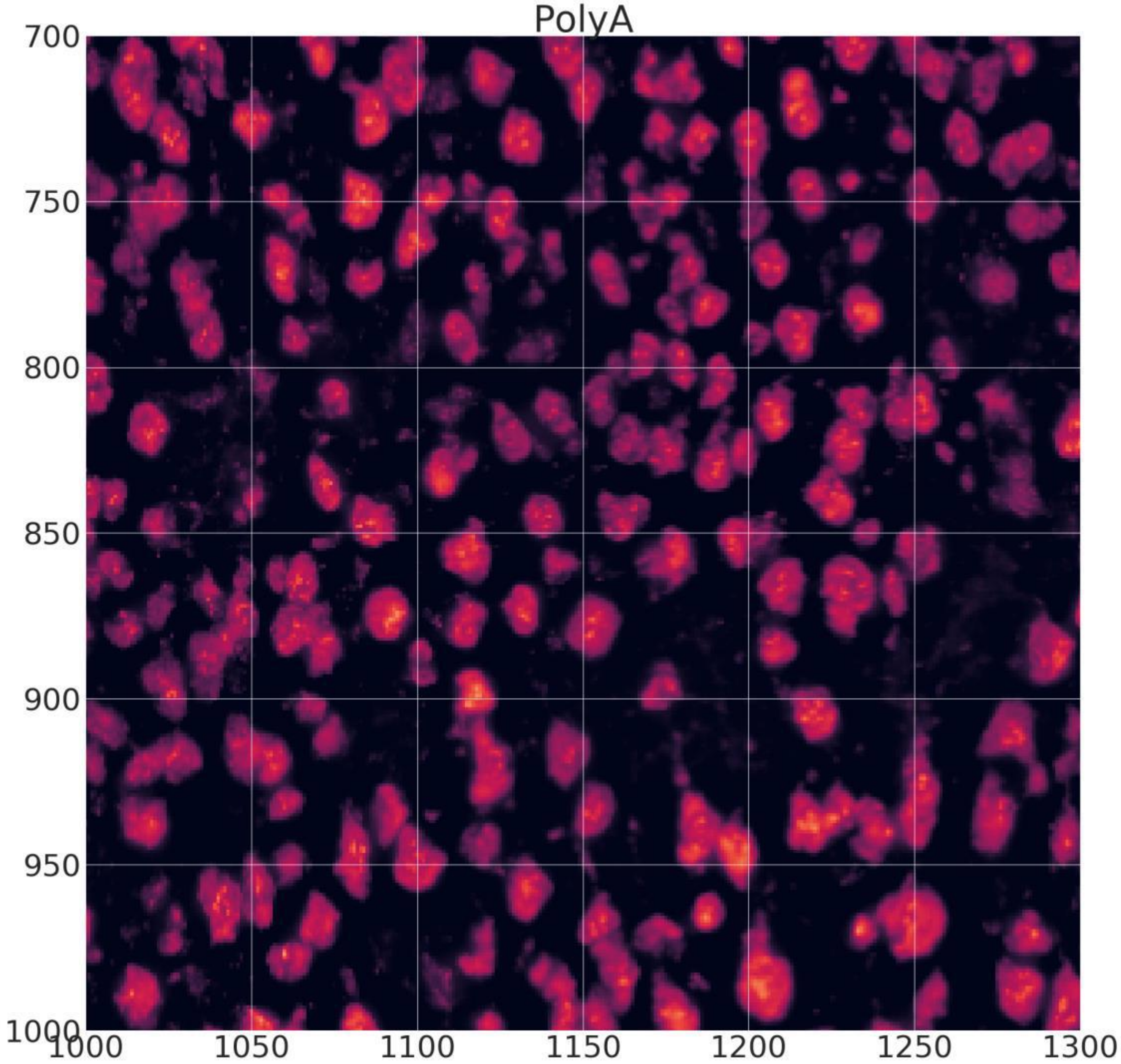
**DKFZ, Heidelberg**

- Daniel Hübschmann
- Matthias Schlesner
- Zuguang Gu
- Nagarajan Paramasivam
- Stephen Krämer

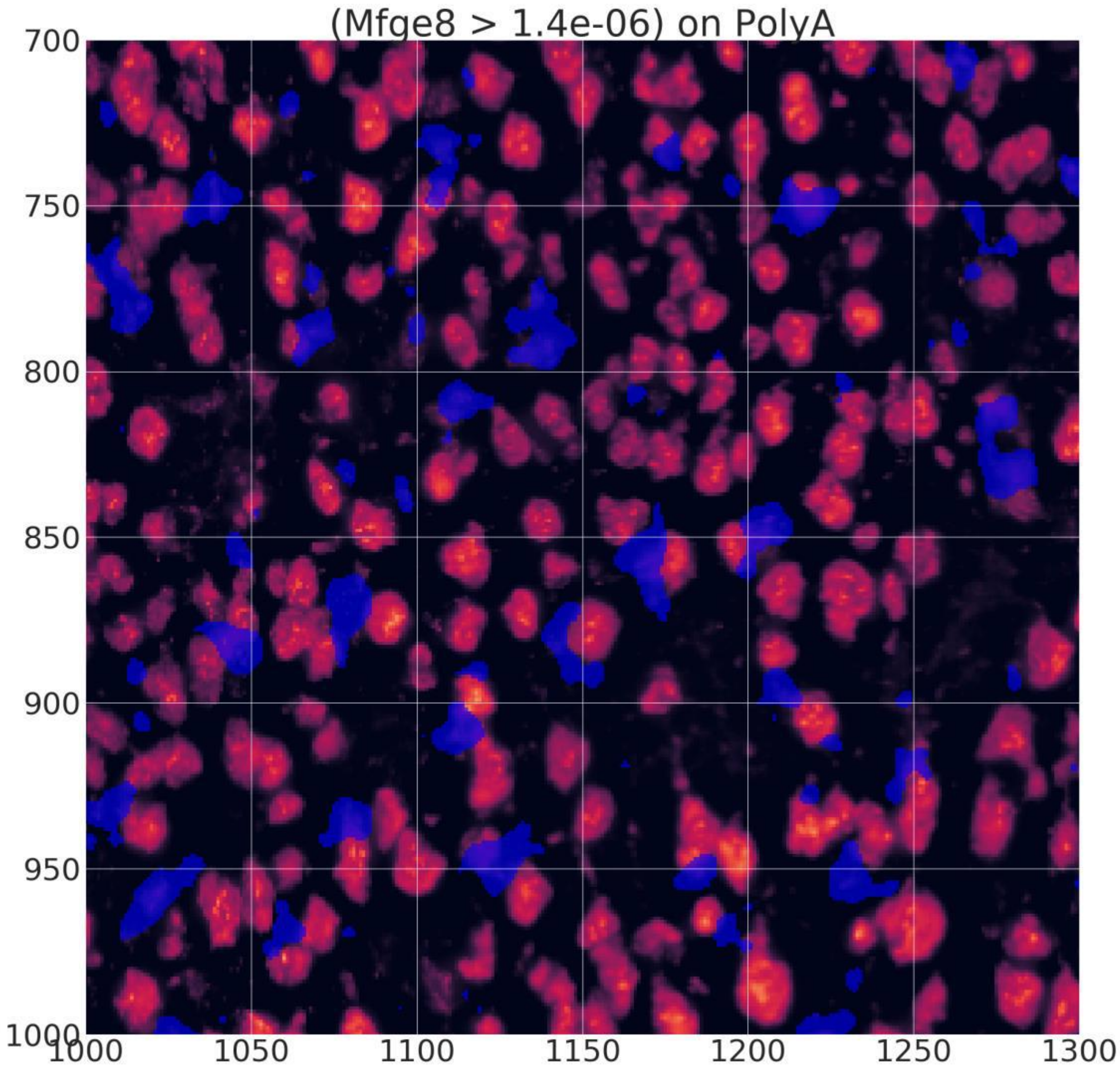
**Karolinska (Linnarsson Lab), Stockholm**

- Lars Borm
- Simone Codeluppi

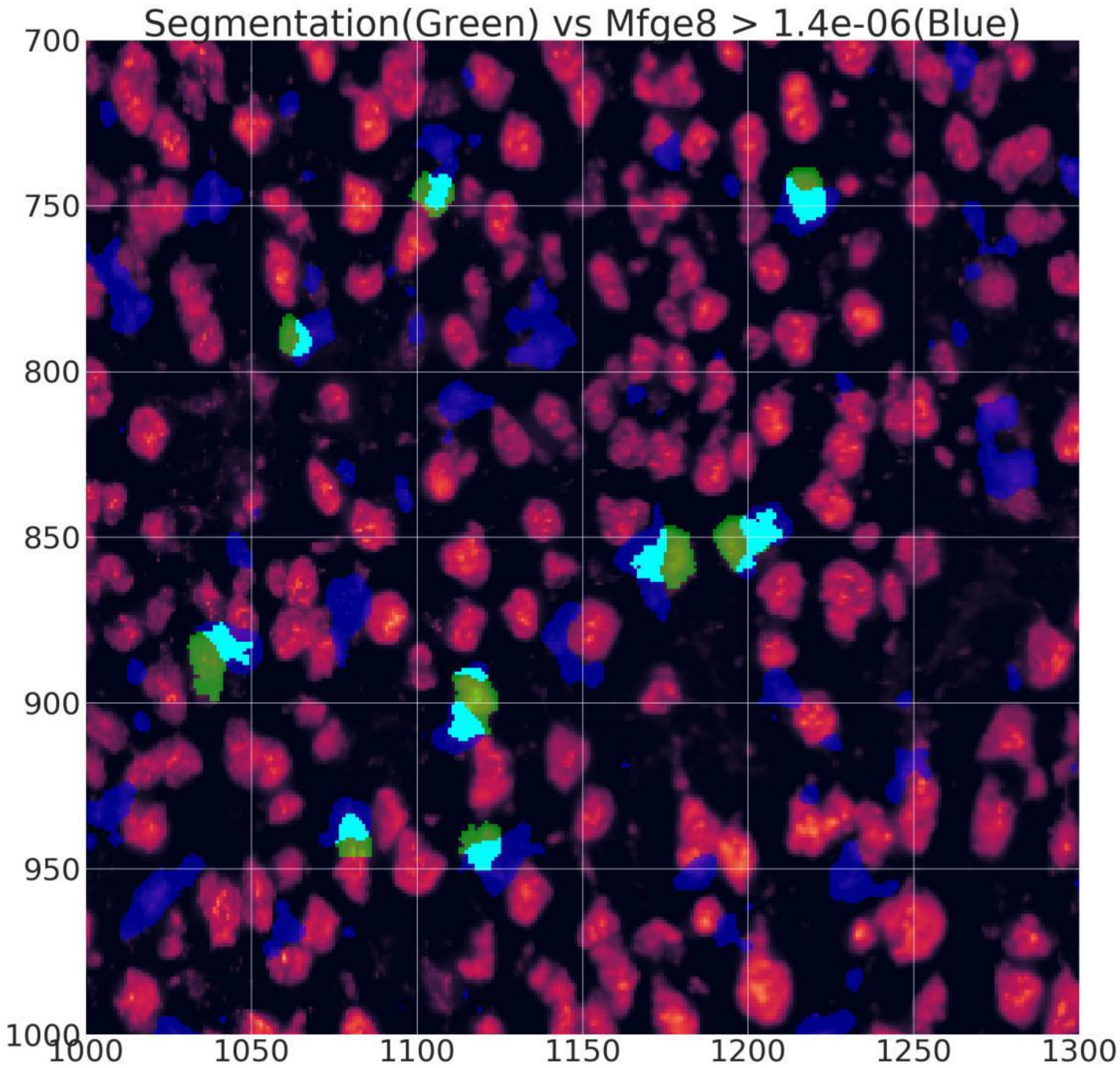
# Astrocyte Mfge8



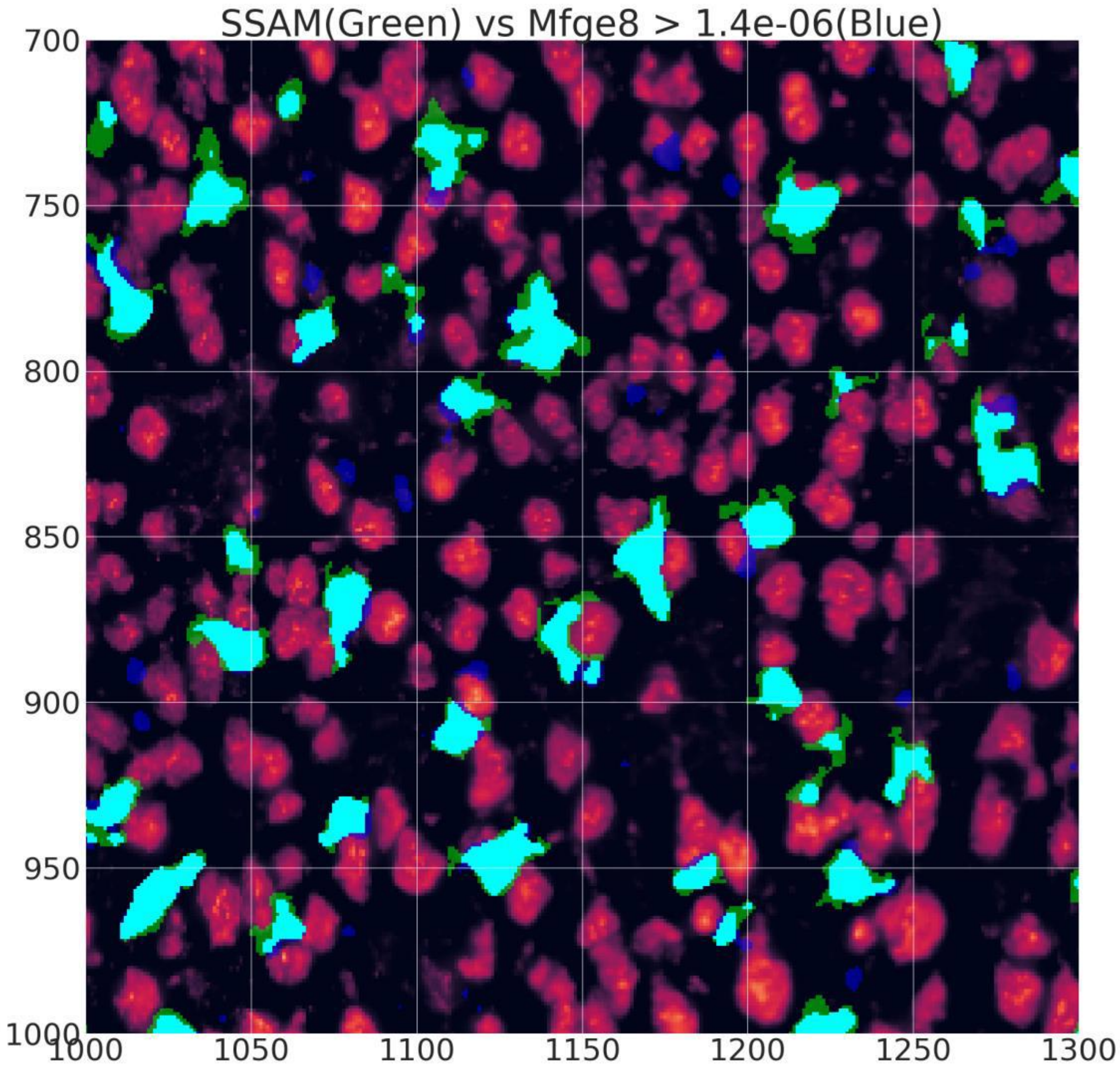
# Astrocyte Mfge8



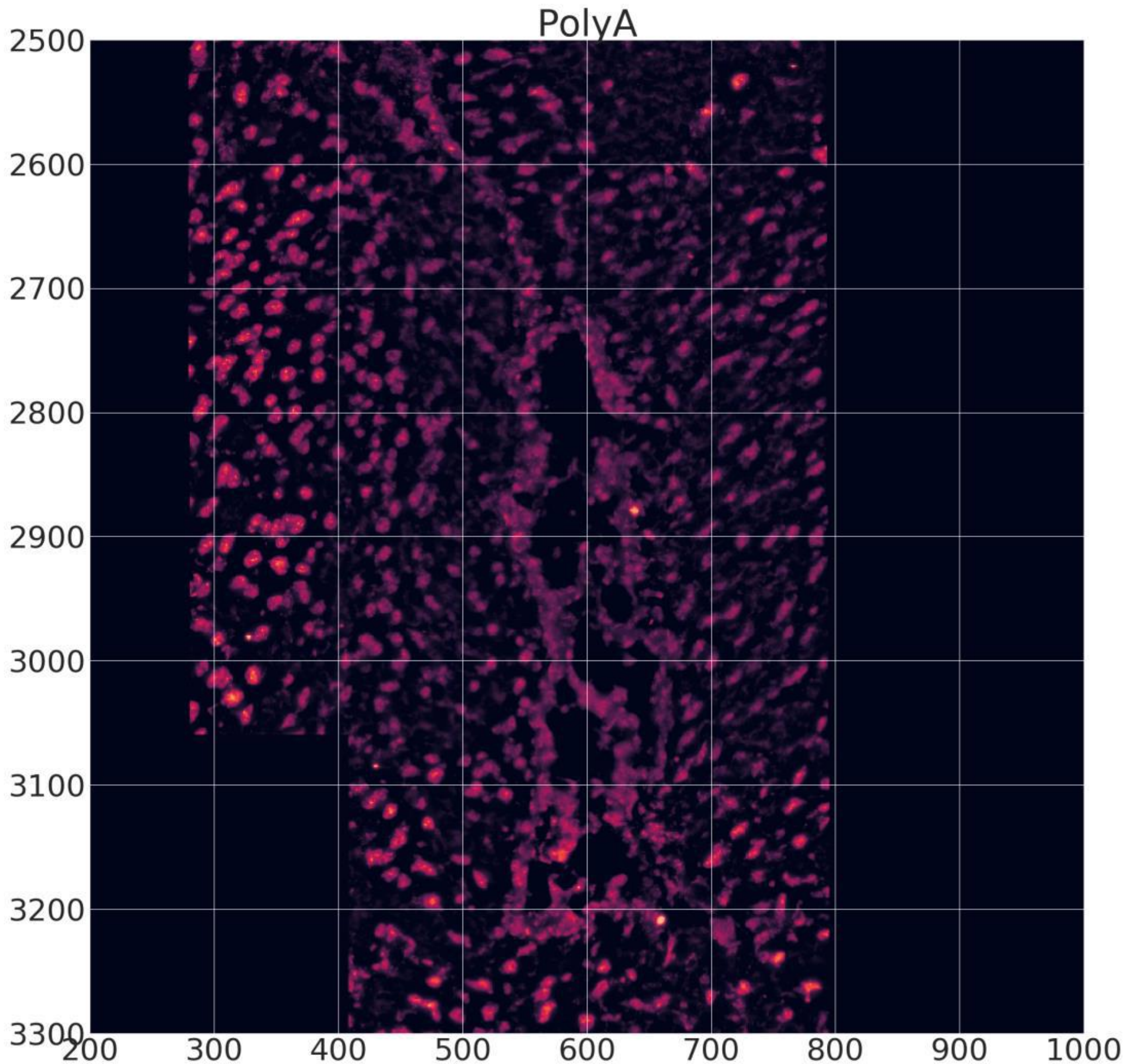
# Astrocyte Mfge8



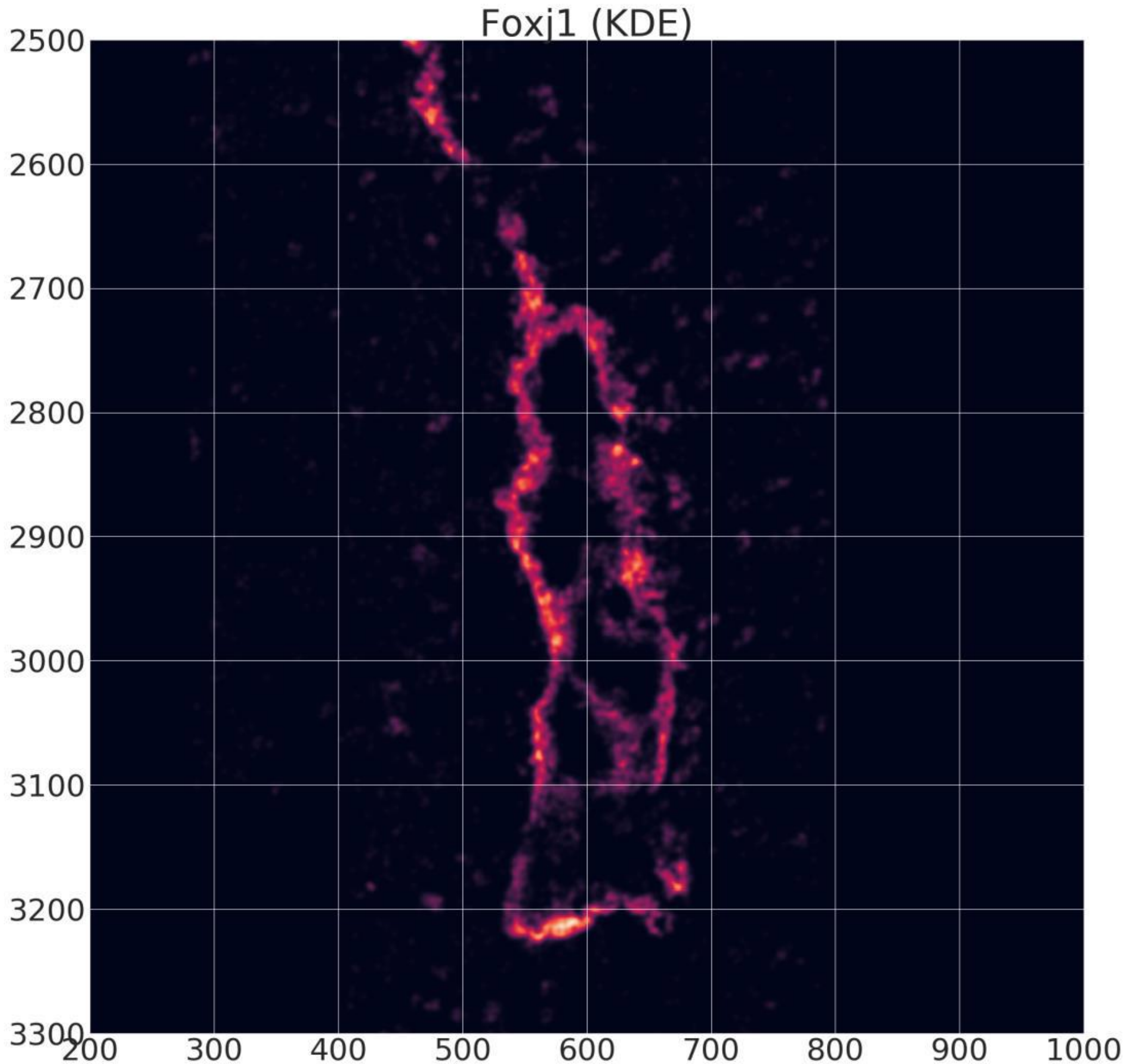
# Astrocyte Mfge8



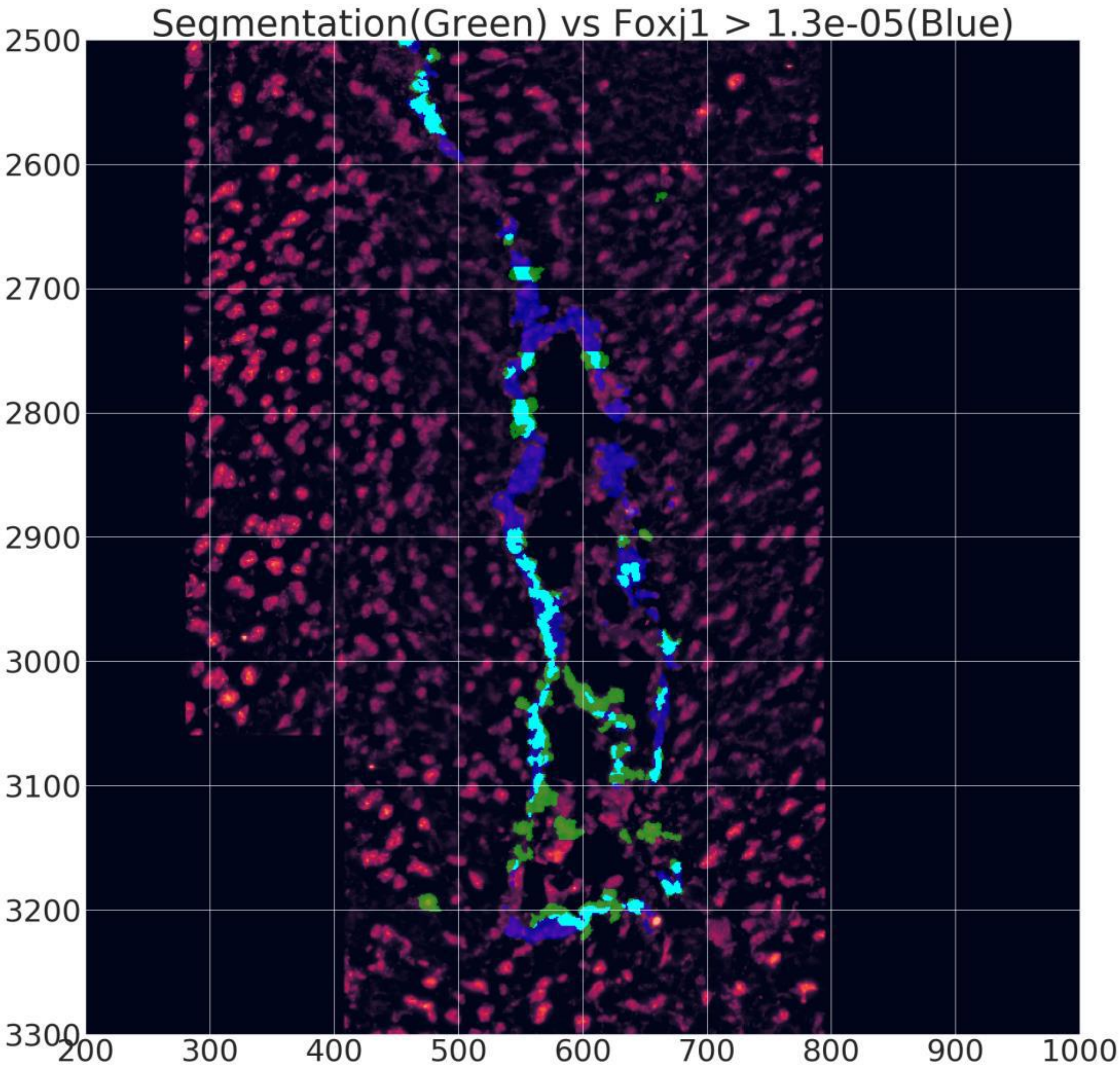
# Ependymal cells



# Ependymal cells



# Ependymal cells



# Ependymal cells