

SPArrOW, a tool for subcellular resolution Spatial Transcriptomics

What can you do with it, and why would you?

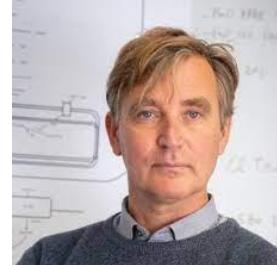
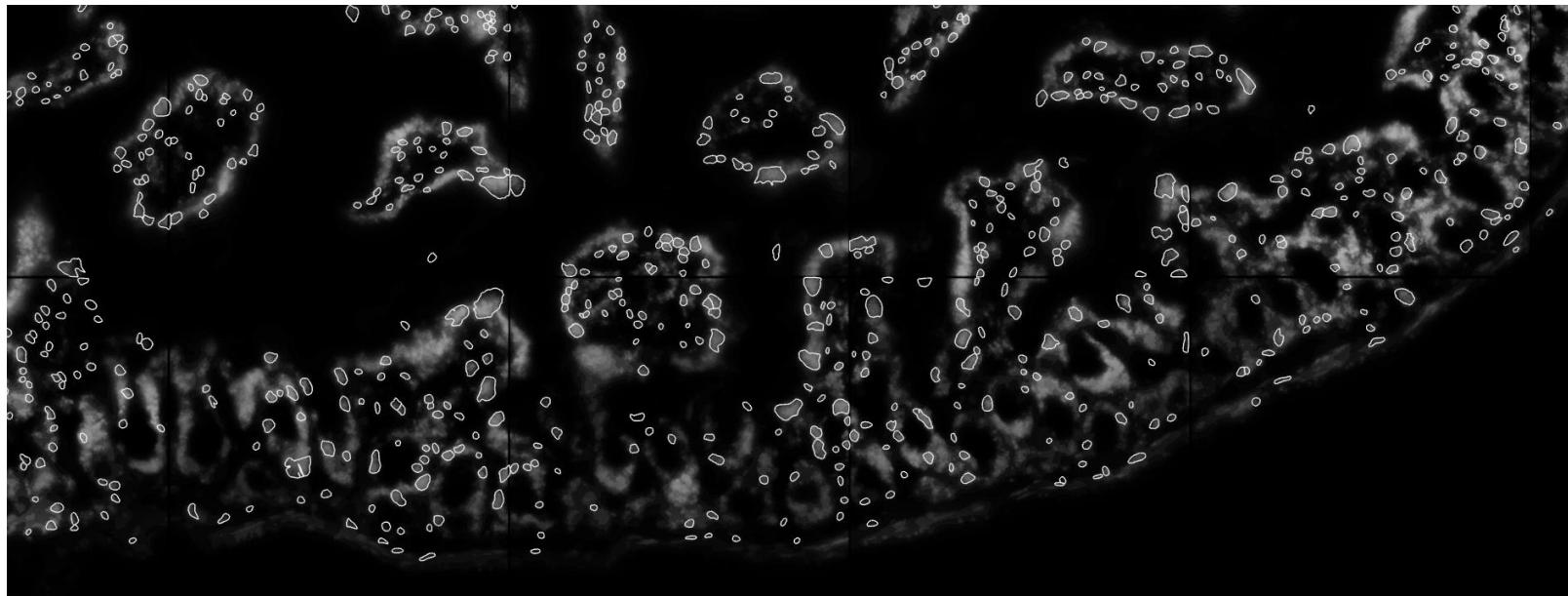
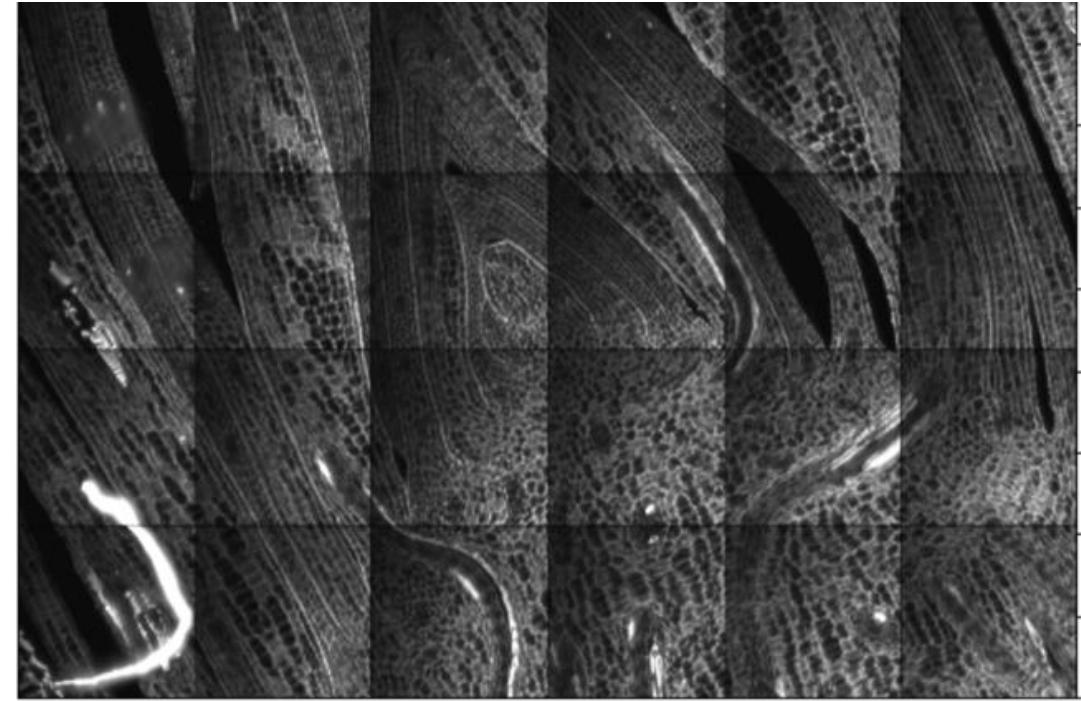
Overview

- Why go through the pain on preprocessing the data yourself?
- A detour: SpatialData
- How does SPArrOW work?
- QC with SPArrOW
- How to annotate Spatial Transcriptomics?
- SPArrOW: the effect of different membrane stainings
- How to use SPArrOW?
- If time permits: a usecase

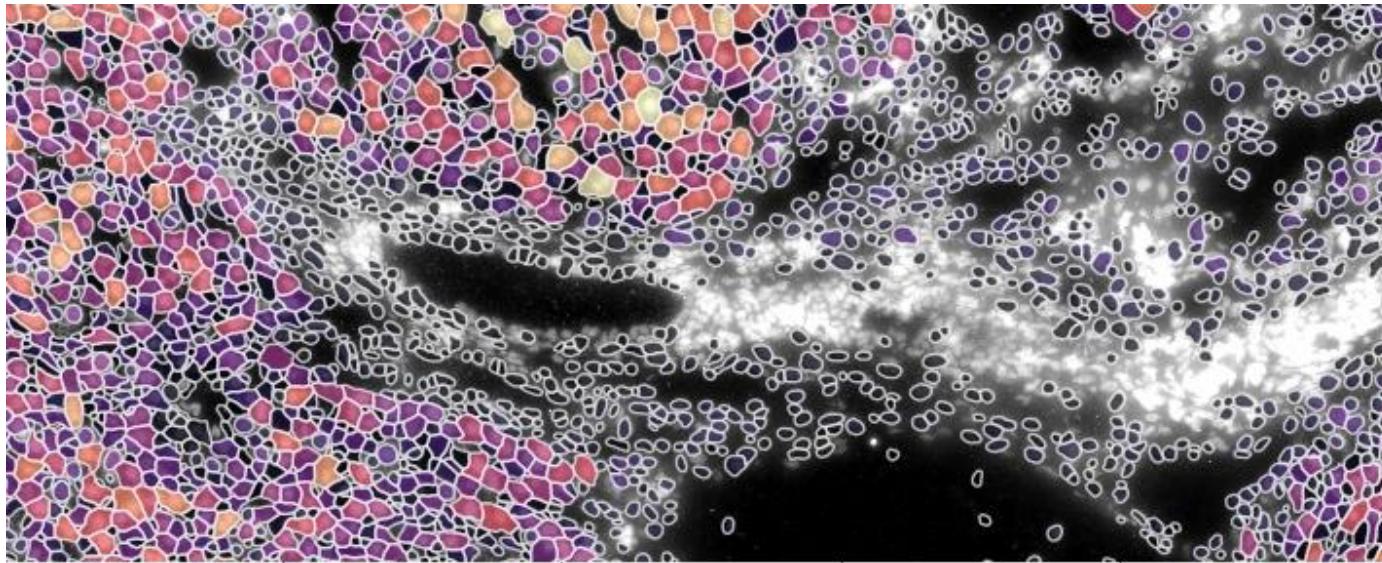
Why go through the pain on preprocessing the data yourself?

RESOLVE

- No proper segmentation provided
- Bad image quality
 - Lots of artefacts

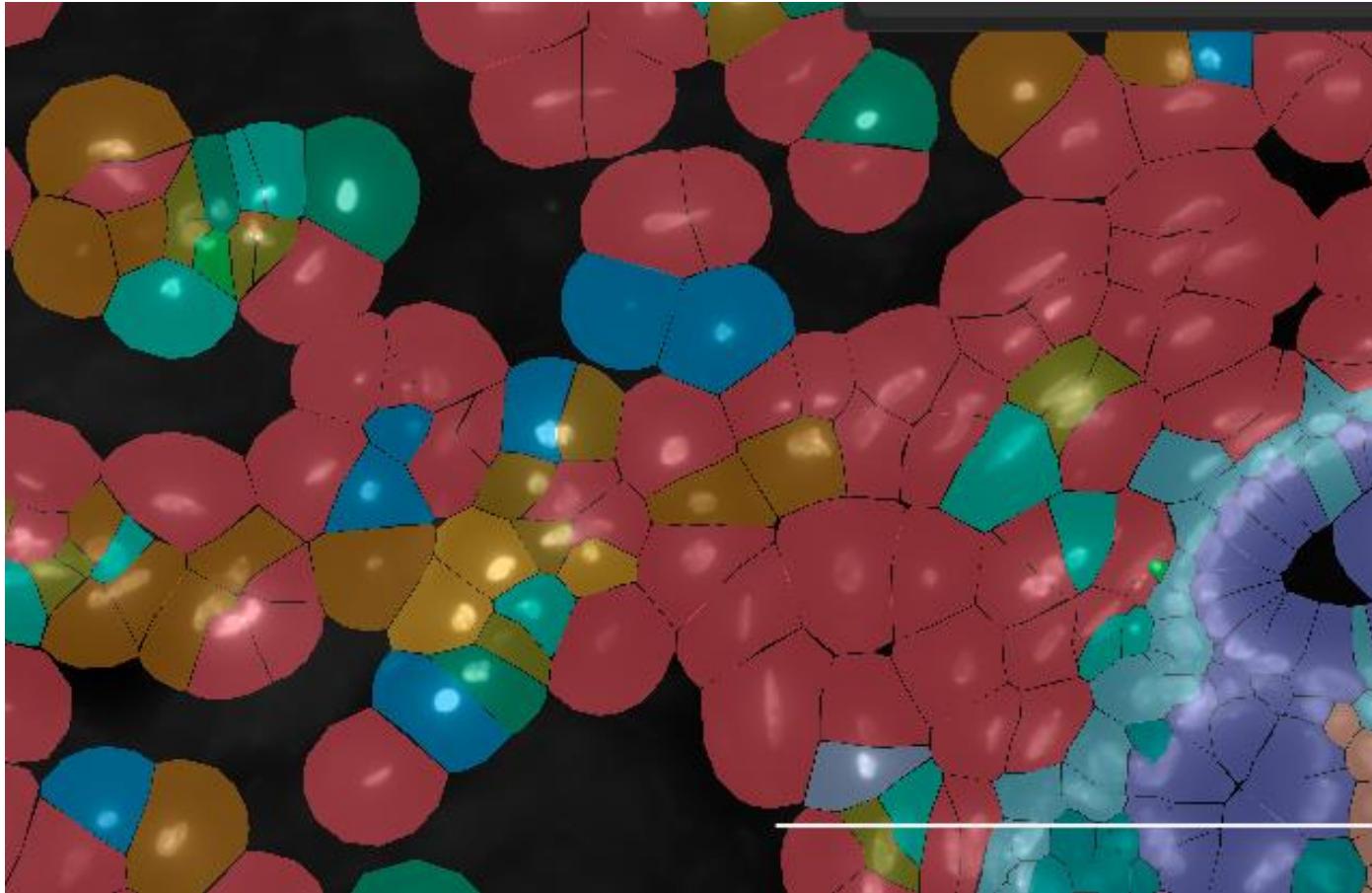


Vizgen MERSCOPE

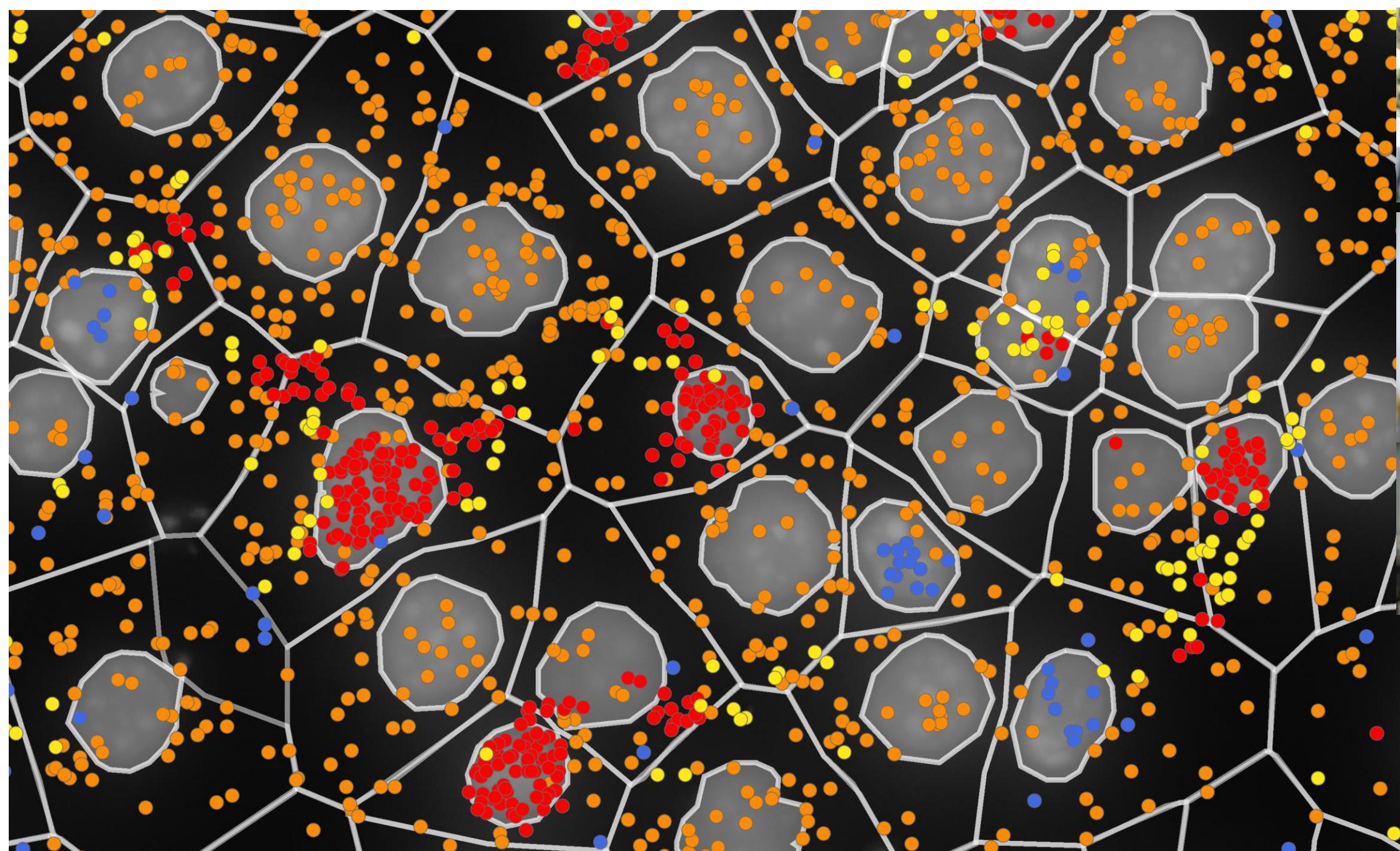


Cells are missed, not segmented

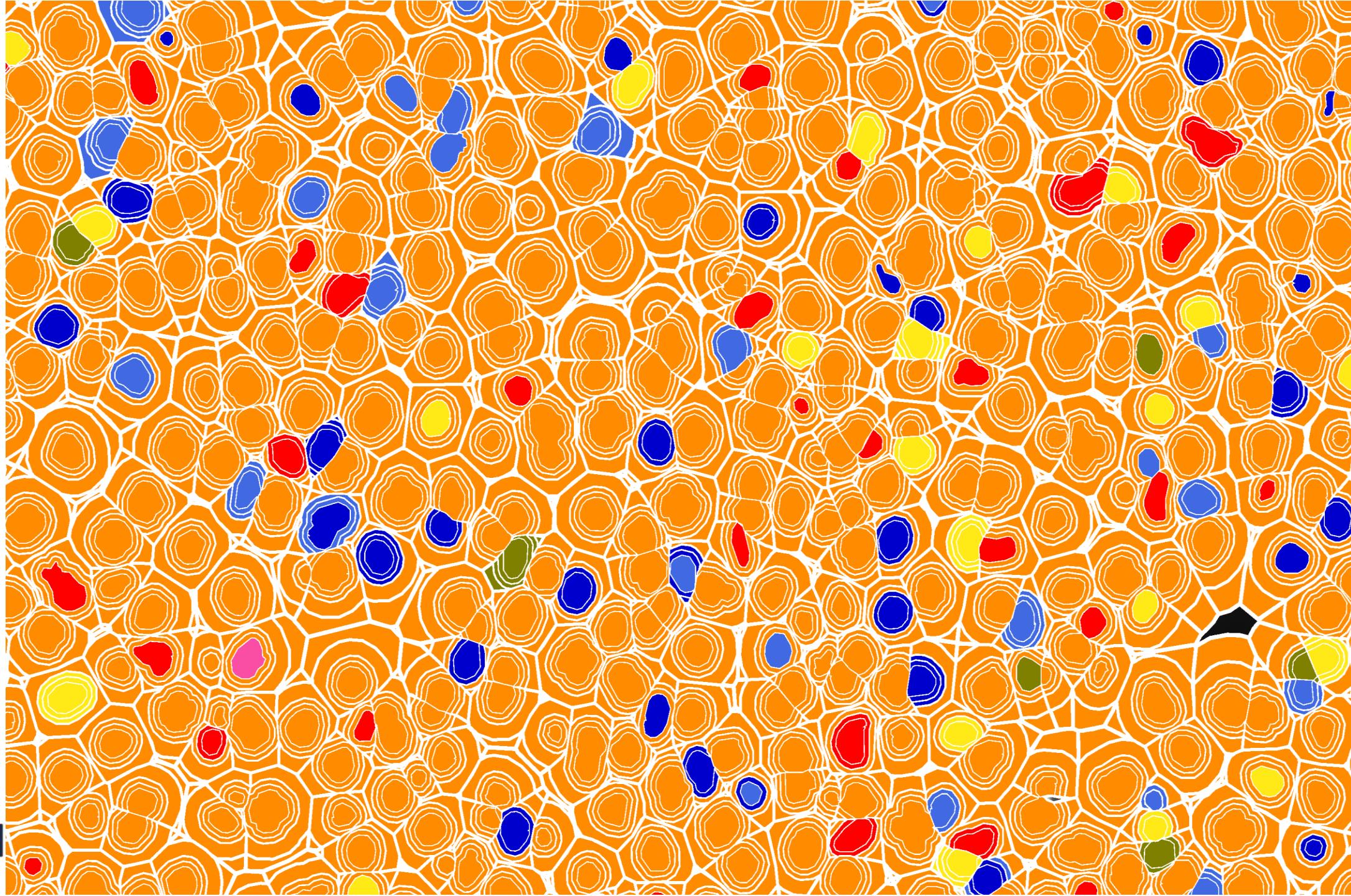
10x Xenium

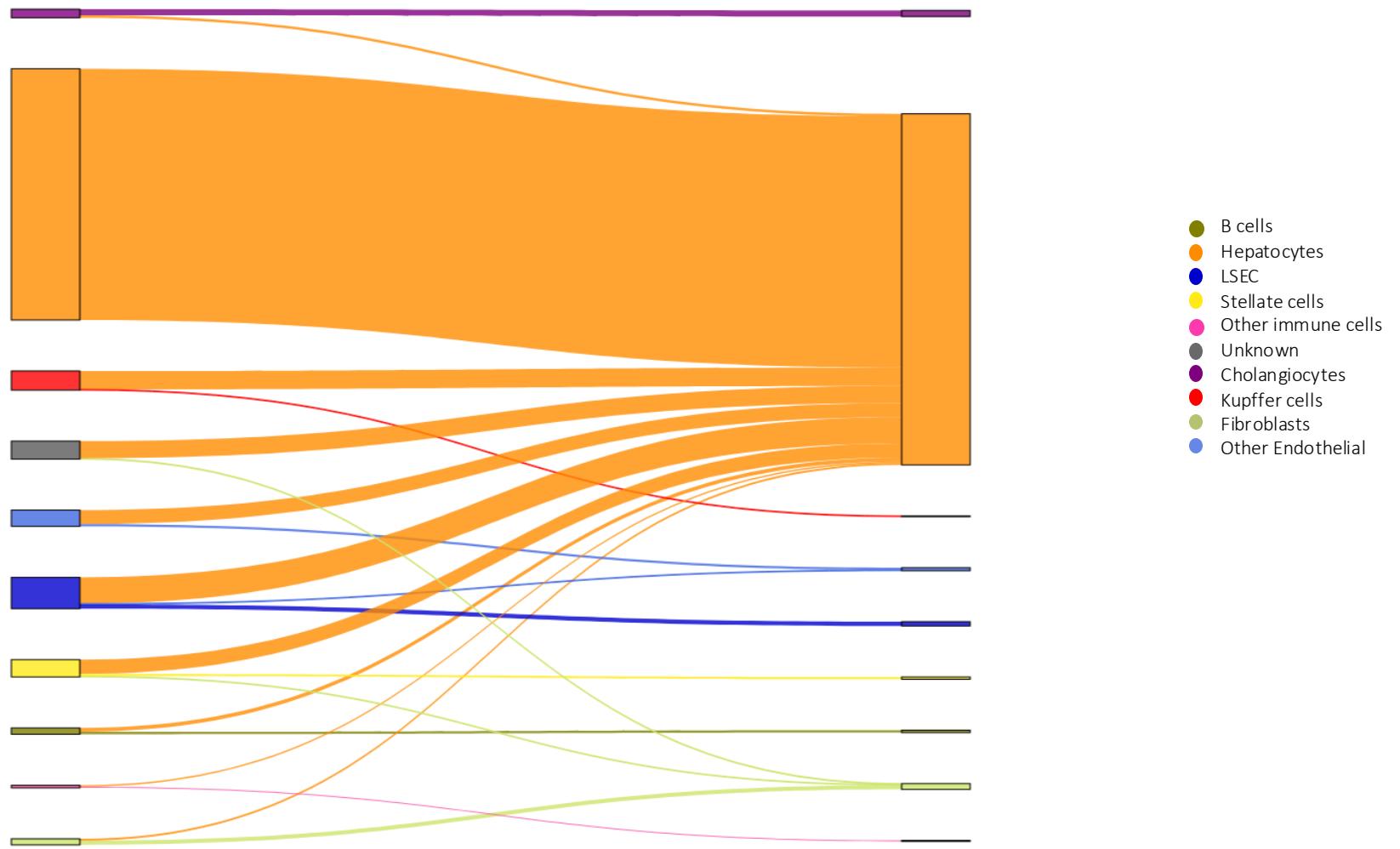


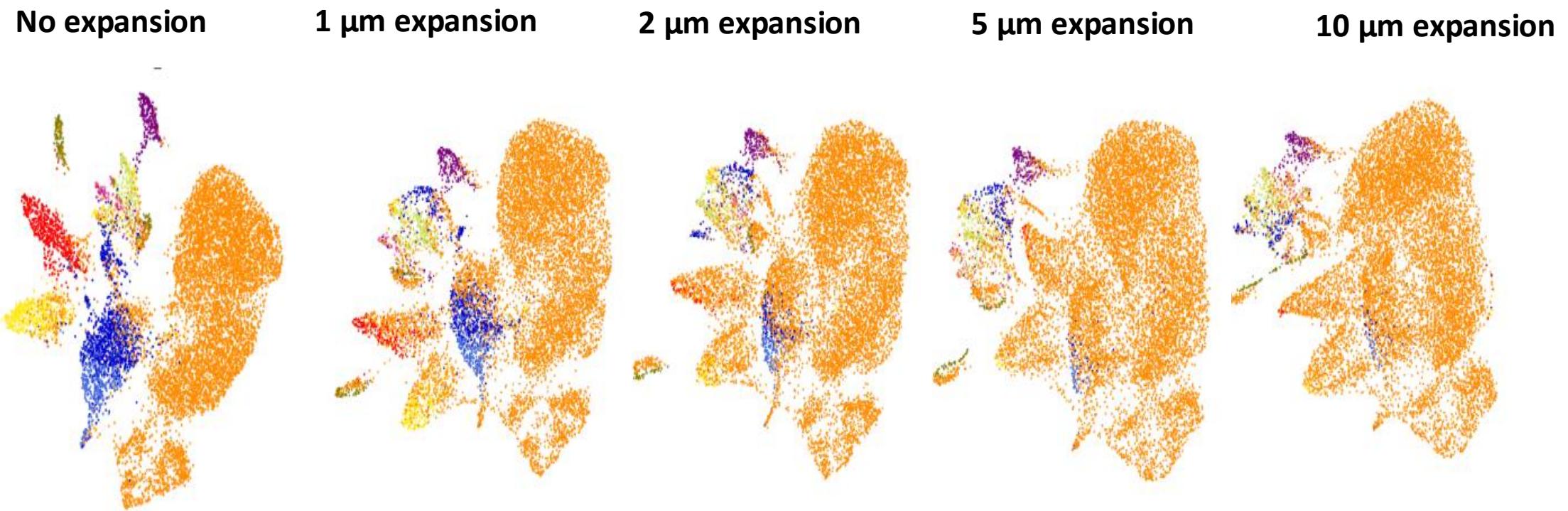
Cells are expanded beyond the nucleus,
leading to cells with mixed fingerprints



SCIENCE MEETS LIFE



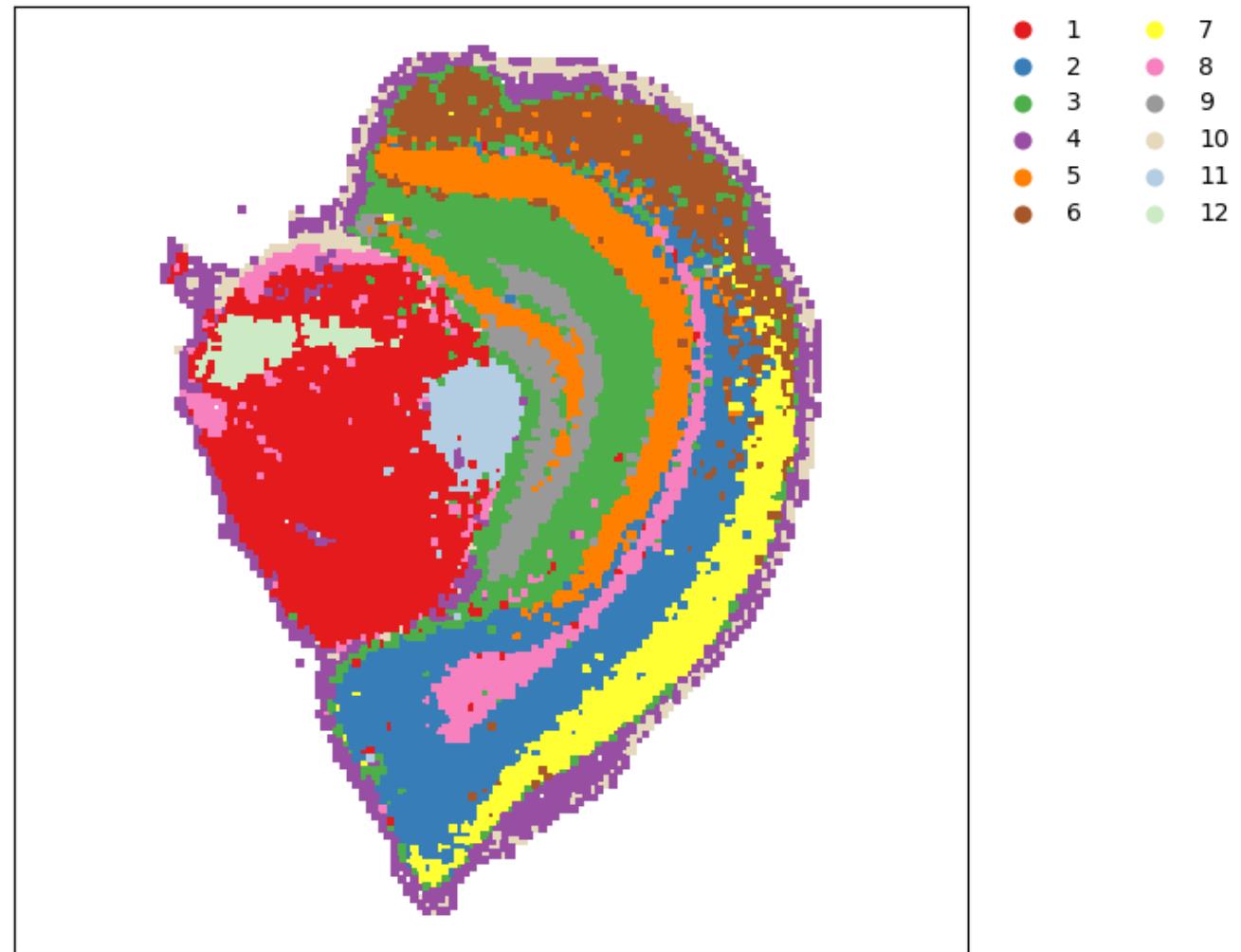




Stereo-seq

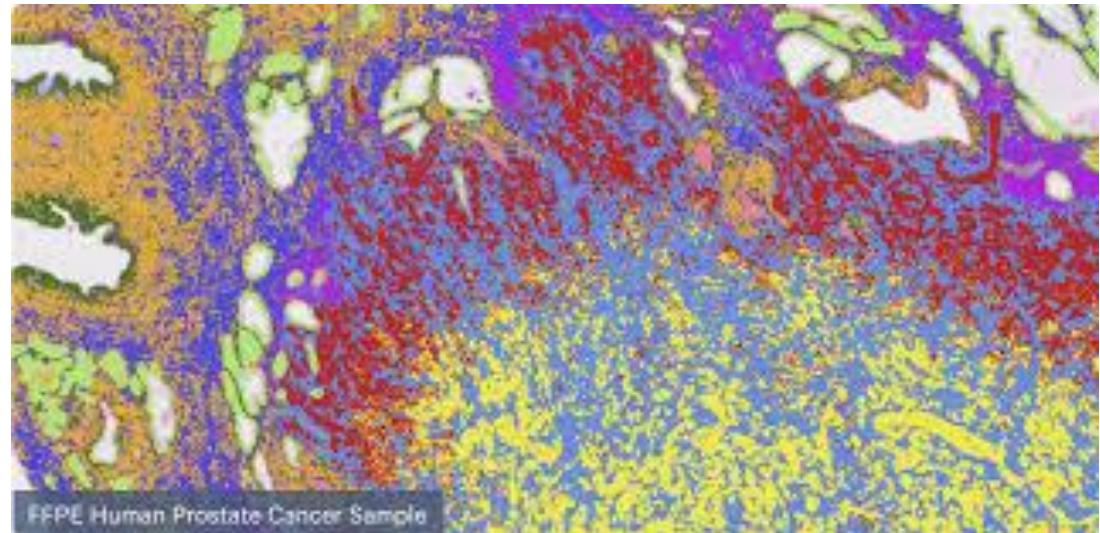
Binning makes it not at single cell resolution

They stated diffusion of 5 nanoballs (2.5 micrometer) is acceptable, but this makes this data quite messy.



VisiumHD

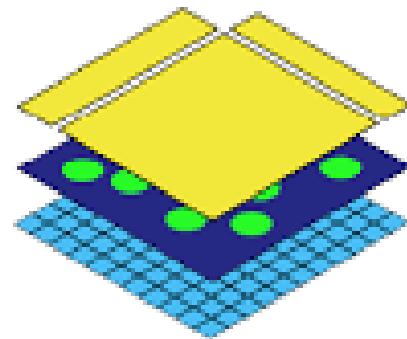
- 8 micron bin don't provide single cell resolution, but mixed identities
- Note: registration issues make using the images difficult



Conclusion

- All datasets providers have their flaws when processing the raw data
- Starting from scratch is recommendable
 - ▶ Same analysis for different platforms
 - ▶ Processing can be adapted to your tissue
 - ▶ More and better quality cells
 - ▶ More aware of artefacts in the data (plus can be filtered)

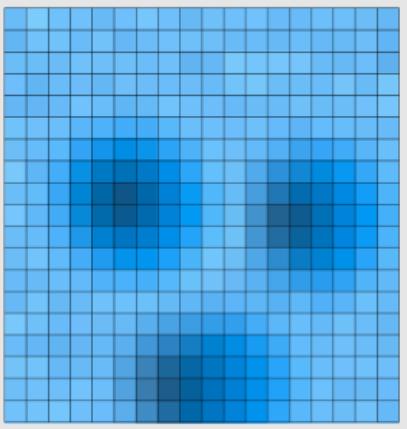
A Detour to



SpatialData

The dataformat

Images



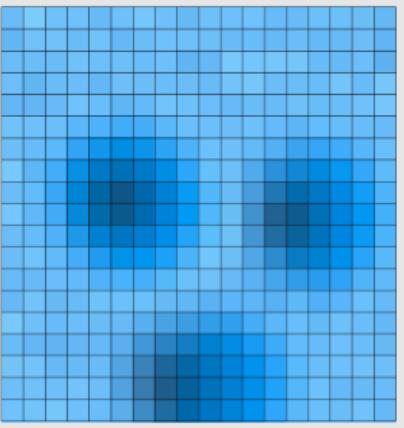
H&E image

DAPI

Protein signals

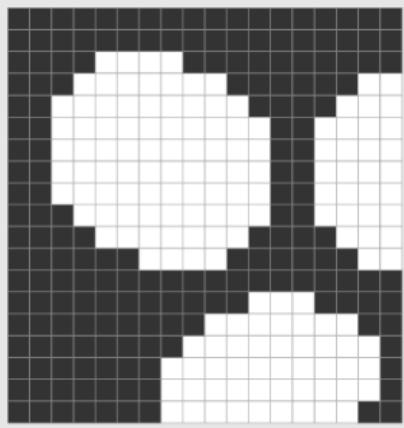
Membrane stainings

Images



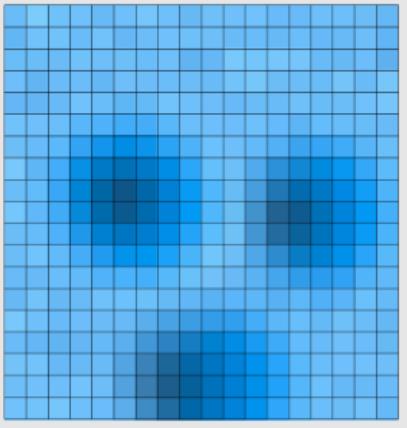
H&E image
DAPI
Protein signals
Membrane stainings

Labels



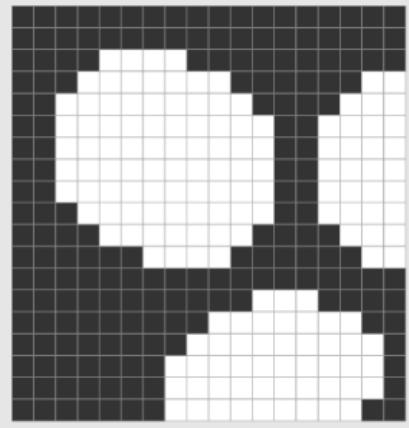
Pixel annotations of
images:
Segmentation mask
ROI's

Images



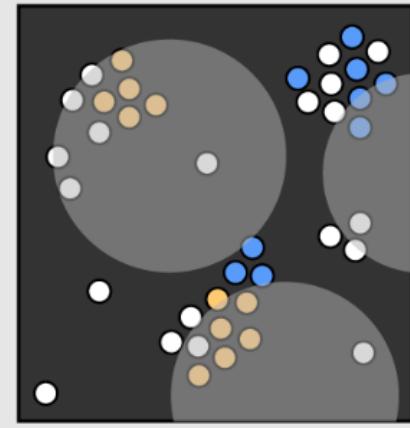
H&E image
DAPI
Protein signals
Membrane stainings

Labels



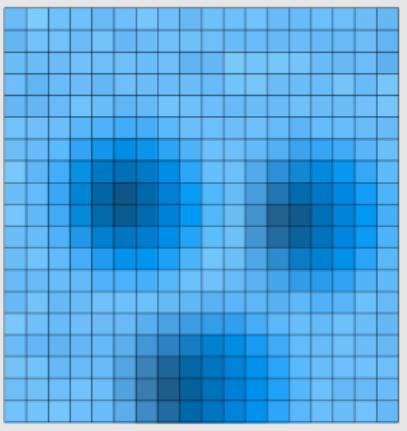
Pixel annotations of
images:
Segmentation mask
ROI's

Points



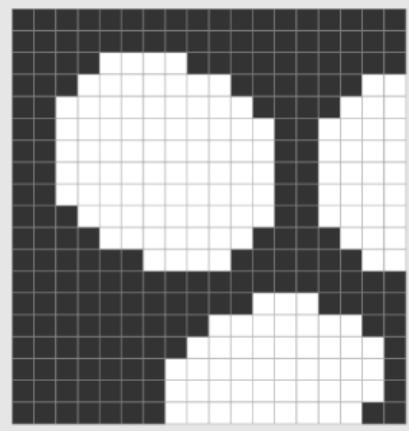
Transcript locations
Visium spots

Images



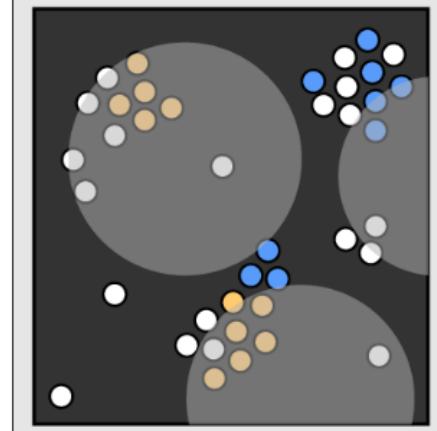
H&E image
DAPI
Protein signals
Membrane stainings

Labels



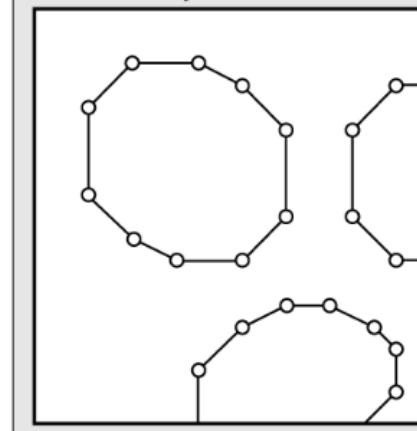
Pixel annotations of
images:
Segmentation mask
ROI's

Points



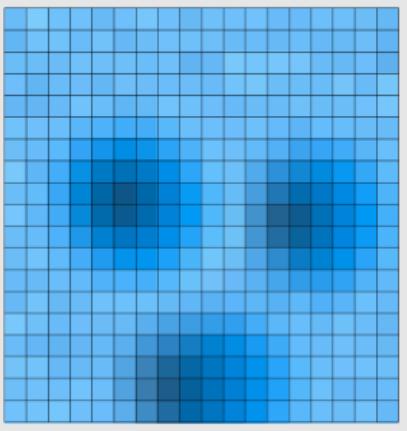
Transcript locations
Visium spots

Polygons



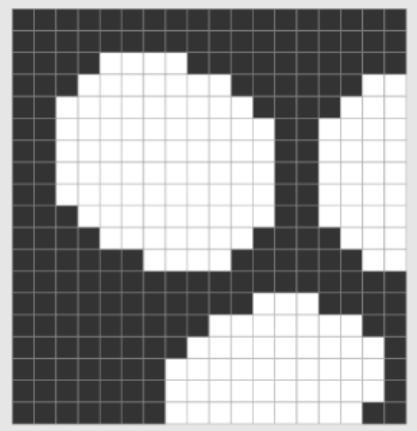
Cell shapes
ROI's
Image annotations of
doctors

Images



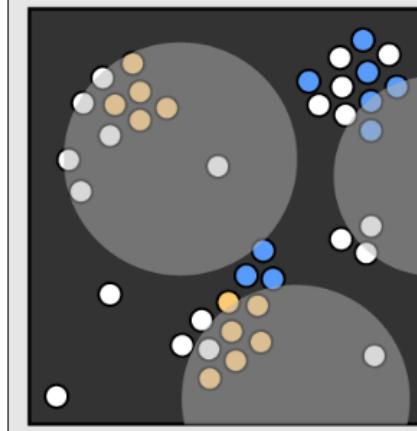
H&E image
DAPI
Protein signals
Membrane stainings

Labels



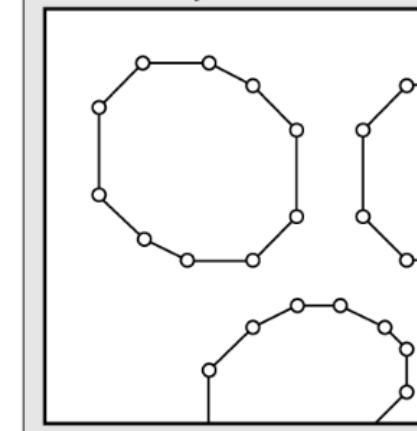
Pixel annotations of images:
Segmentation mask
ROI's

Points



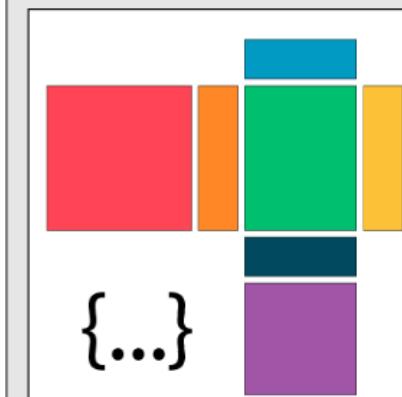
Transcript locations
Visium spots

Polygons

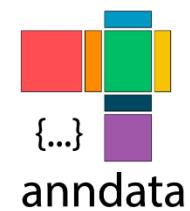
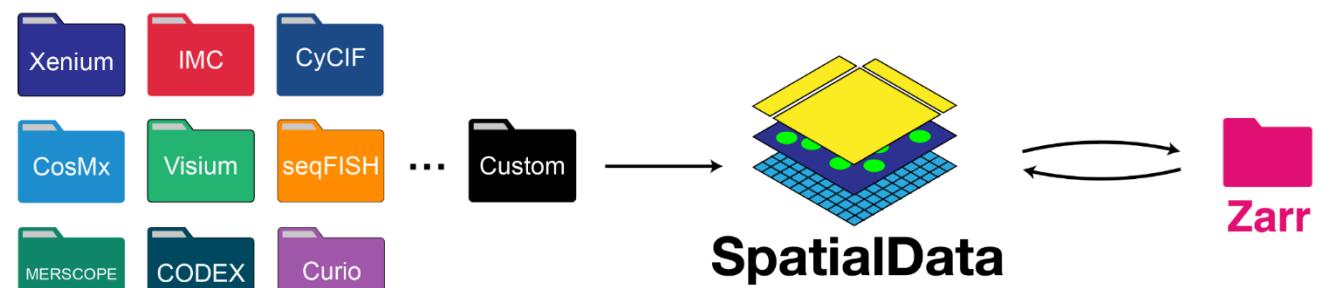
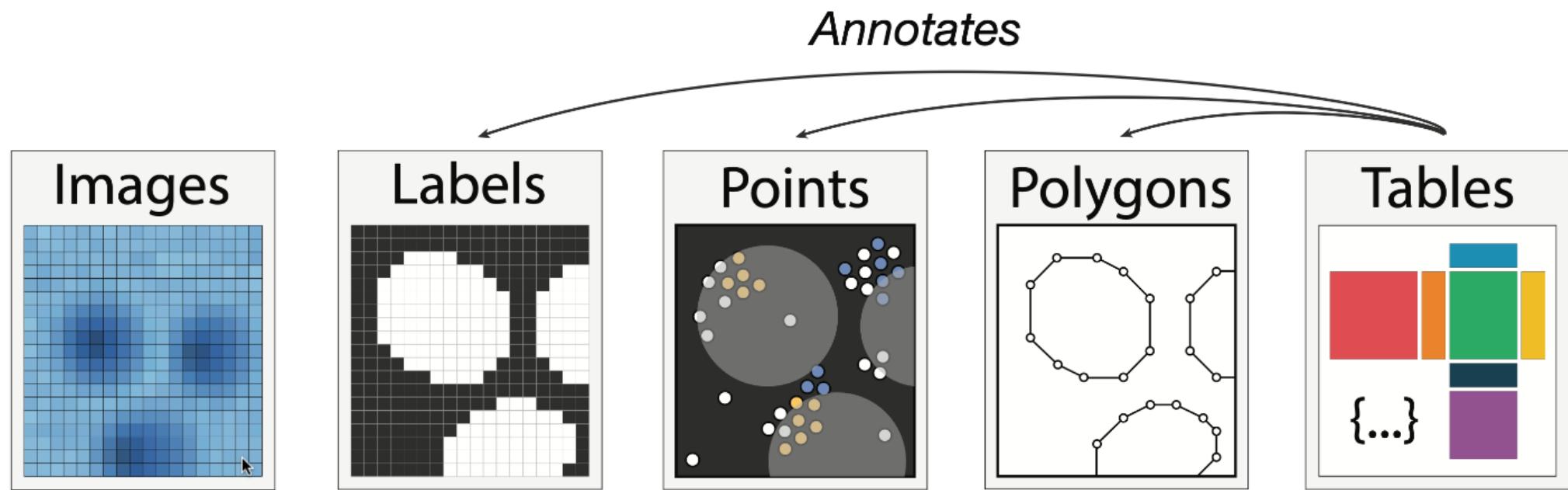


Cell shapes
ROI's
Image annotations of doctors

Tables



Anndata
Gene expression
Protein expression
Deep annotation of ROI's



```

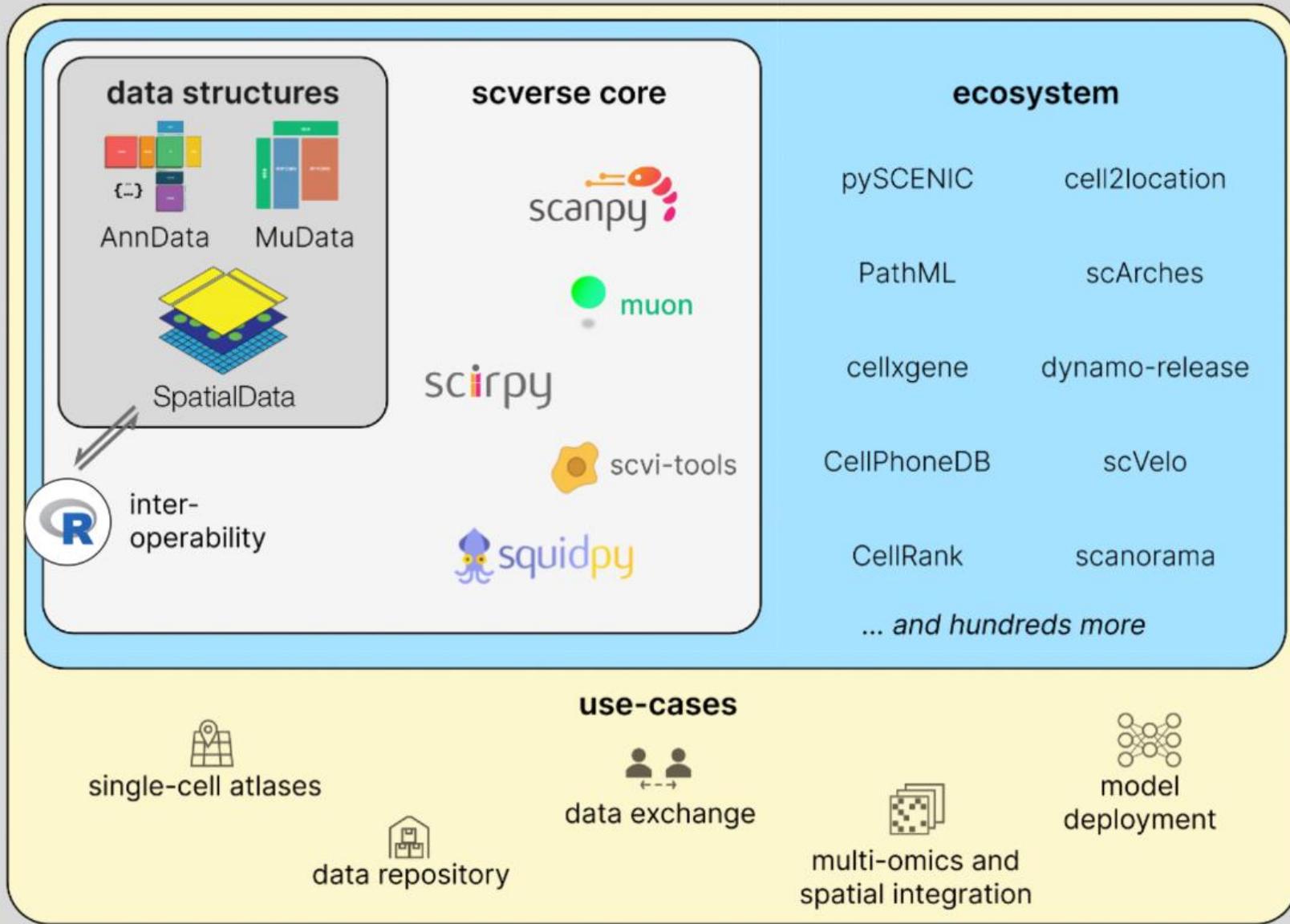
SpatialData object
├── Images
│   ├── 'blobs_image': SpatialImage[cyx] (3, 512, 512)
│   └── 'blobs_multiscale_image': MultiscaleSpatialImage[cyx] (3, 512, 512), (3, 256, 256), (3, 128, 128)
└── Labels
    ├── 'blobs_labels': SpatialImage[yx] (512, 512)
    └── 'blobs_multiscale_labels': MultiscaleSpatialImage[yx] (512, 512), (256, 256), (128, 128)
├── Points
│   └── 'blobs_points': DataFrame with shape: (<Delayed>, 4) (2D points)
└── Shapes
    ├── 'blobs_circles': GeoDataFrame shape: (5, 2) (2D shapes)
    ├── 'blobs_multipolygons': GeoDataFrame shape: (2, 1) (2D shapes)
    └── 'blobs_polygons': GeoDataFrame shape: (5, 1) (2D shapes)
└── Tables
    └── 'table': AnnData (26, 3)
with coordinate systems:
    ▶ 'global', with elements:
        blobs_image (Images), blobs_multiscale_image (Images), blobs_labels (Labels), blobs_multiscale_labels (Labels)

```

```

sdataA1-2_optimized
└── sdata.zarr
    ├── images
    │   ├── clahe
    │   ├── raw_image
    │   ├── tiling_correction
    │   └── tophat_filtered
    └── .zgroup
    └── labels
        ├── segmentation_mask
        └── .zattrs
        └── .zgroup
    └── points
        ├── transcripts
        └── .zgroup
    └── shapes
        ├── filtered_low_counts_segmentation_mask_boundaries
        ├── filtered_segmentation_segmentation_mask_boundaries
        ├── filtered_size_segmentation_mask_boundaries
        ├── segmentation_mask_boundaries
        ├── to_plot_napari
        └── .zgroup
    └── table
        └── table
            ├── layers
            ├── obs
            ├── obsm
            ├── obsp
            ├── raw
            ├── uns
            ├── var
            ├── varm
            ├── varp
            └── X
            └── .zattrs
            └── .zgroup
            └── .zgroup
            └── .zgroup

```





- established interoperable format for spatial omics based on OME-NGFF

- in-memory multimodal representation
- processing, visualization
- scales to large datasets



Luca
Marconato



Giovanni
Palla



Kevin
Yamauchi



Isaac
Virshup



Tim Treis



Wouter-Michiel
Vierdag



Josh
Moore



Sonja
Stockhaus



Elyas
Heidari



Marcela
Toth



Quentin
Blampey



Laurens
Lehner



Rahul B.
Shrestha



Benjamin
Rombaut



Lotte
Pollaris



Harald
Vöhringer



Oliver
Stegle



Fabian
Theis



Moritz
Gerstung



Sinem
Saka



Yvan
Saeys



Wolfgang
Huber



Nature Methods, 2024

First authors are underlined

EMBL

Oliver Stegle
Luca Marconato
Sinem Saka
Wouter-Michiel Vierdag
Wolfgang Huber
Harald Vöhringer
Constantin Ahlmann-Eltze
Mike Smith

Helmholtz Munich

Fabian Theis
Giovanni Palla
Isaac Virshup
Tim Treis
Sonja Stockhaus
Laurens Lehner
Marcella Toth
Rahul Shrestha

DKFZ

Ilia Kats
Tobias Graf
Moritz Gerstung
Elyas Heidari
Josh Moore (OME, GerBi)
Kevin Yamauchi (ETH)
Yvan Saeys (UGhent)

Lotte Pollaris (UGhent)
Benjamin Rombaut (UGhent)
Christian Tischer (EMBL)
Andreas S. Eisenbarth (EMBL)
Omer Bayraktar (Sanger)
Tong Li (Sanger)
Ilan Gold (Helmholtz)
Helena Crowell (CNAG)

Nils Eiling (UZH)
Will Moore (OME, UDundee)
Quentin Blampey (UParis Saclay)
Florian Wünnemann (UKHD)
Mark Keller (HMS)
10x Genomics team
CZI cellxgene team
...



spatialdata.scverse.org



Funded by

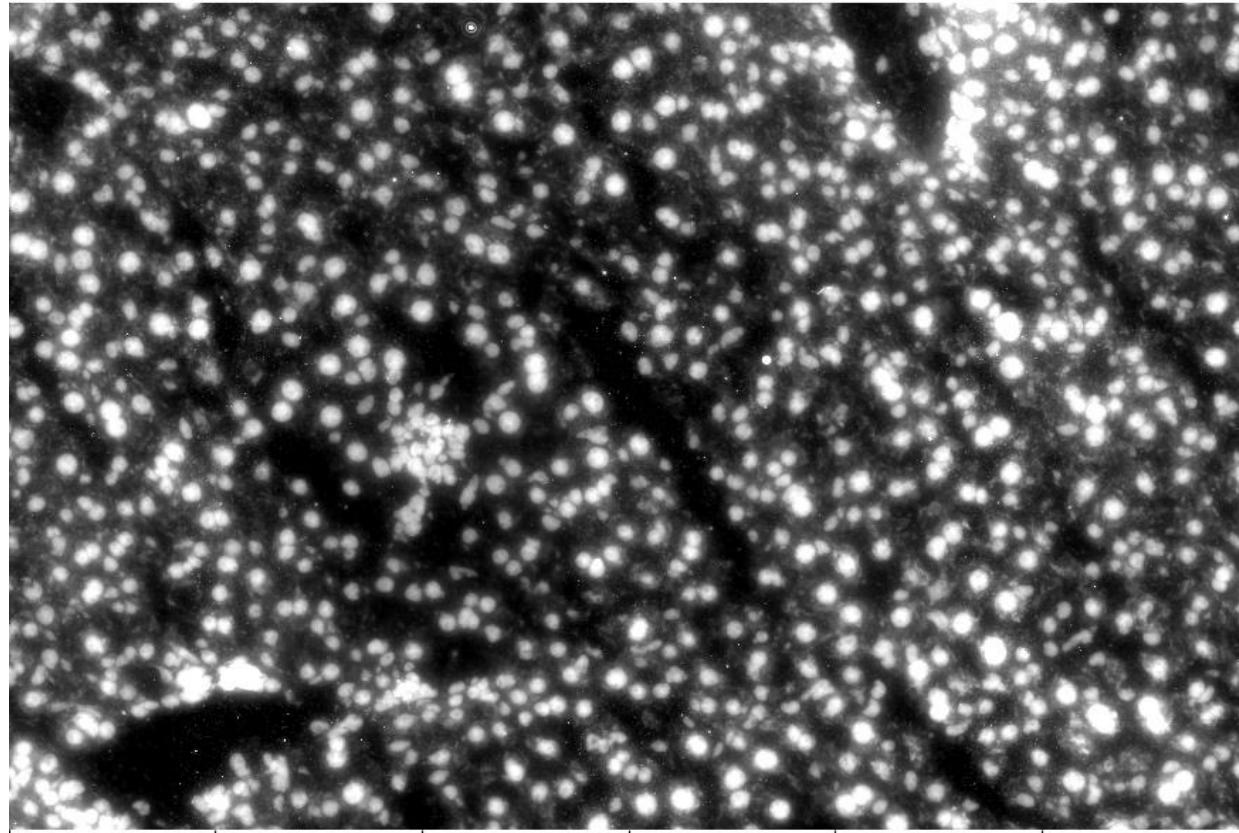


SCIENCE MEETS LIFE

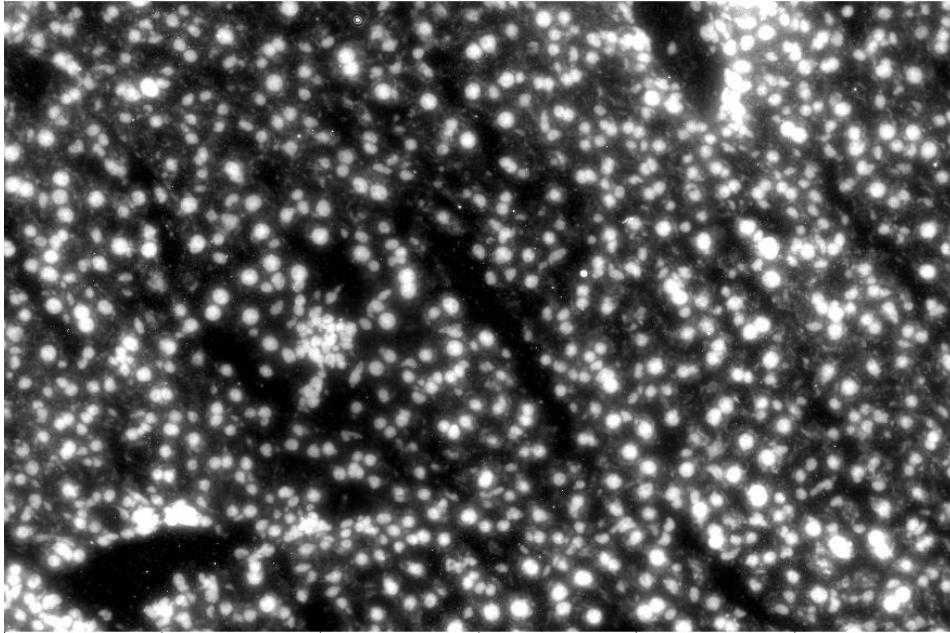
How does



Arrow Work?

**Raw image** (tiff, zarr,...)**Transcript coordinates** (csv,txt,...)

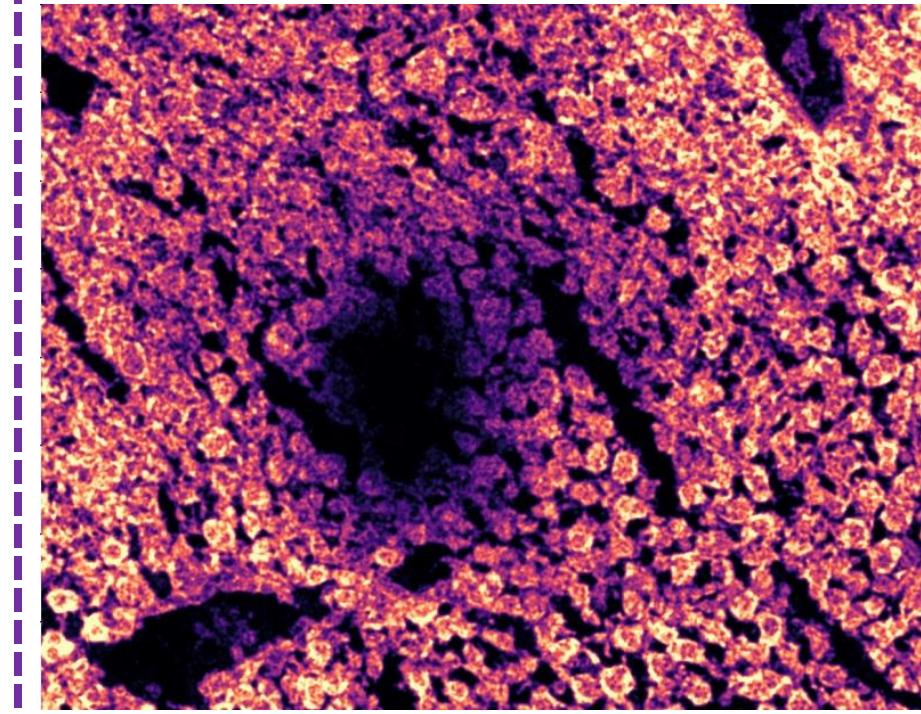
Input



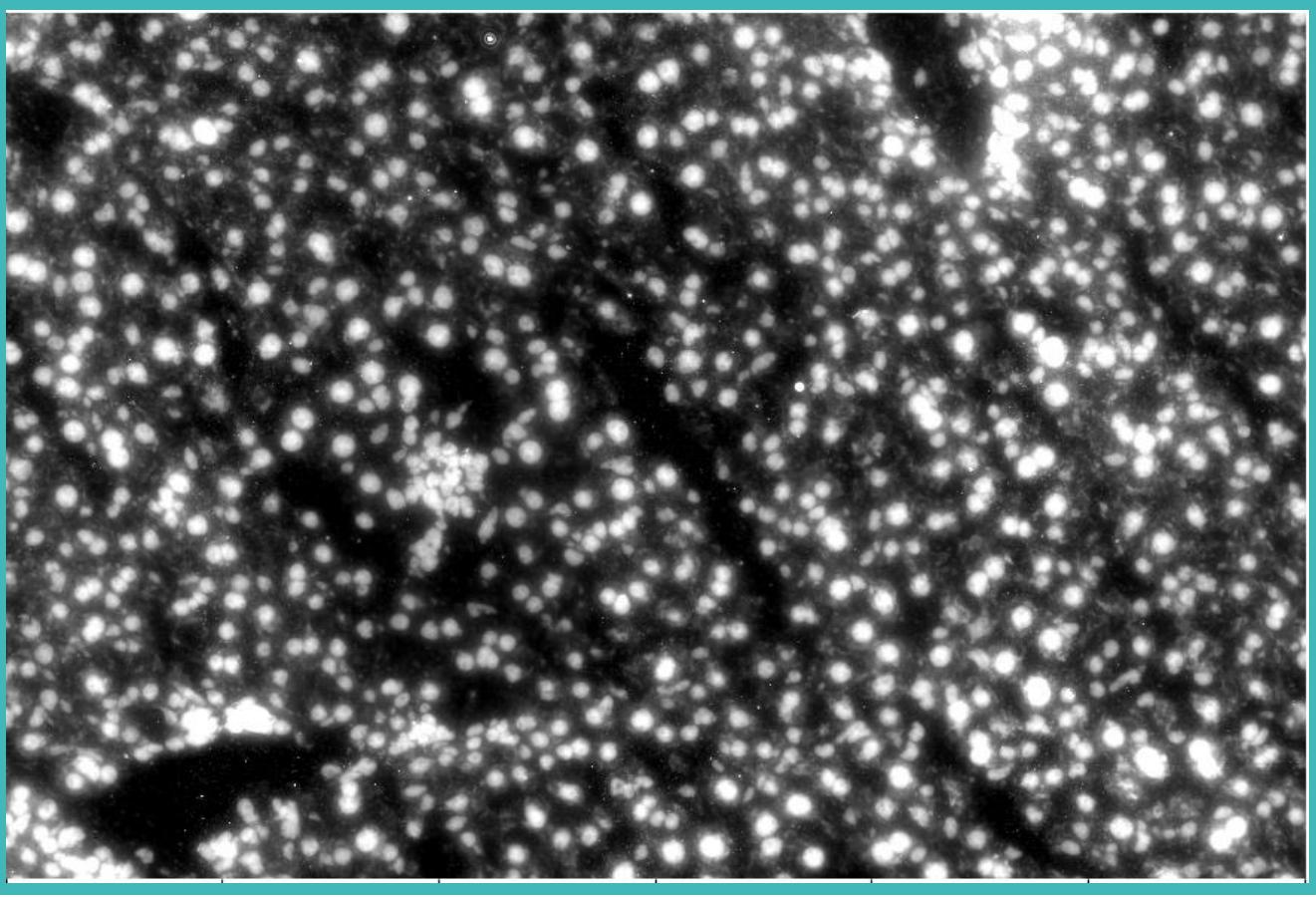
Raw image (tiff, zarr,...)

X	Y	Gene
23	1018	Glul
4567	12347	Hgf
...
7048	9345	Hal

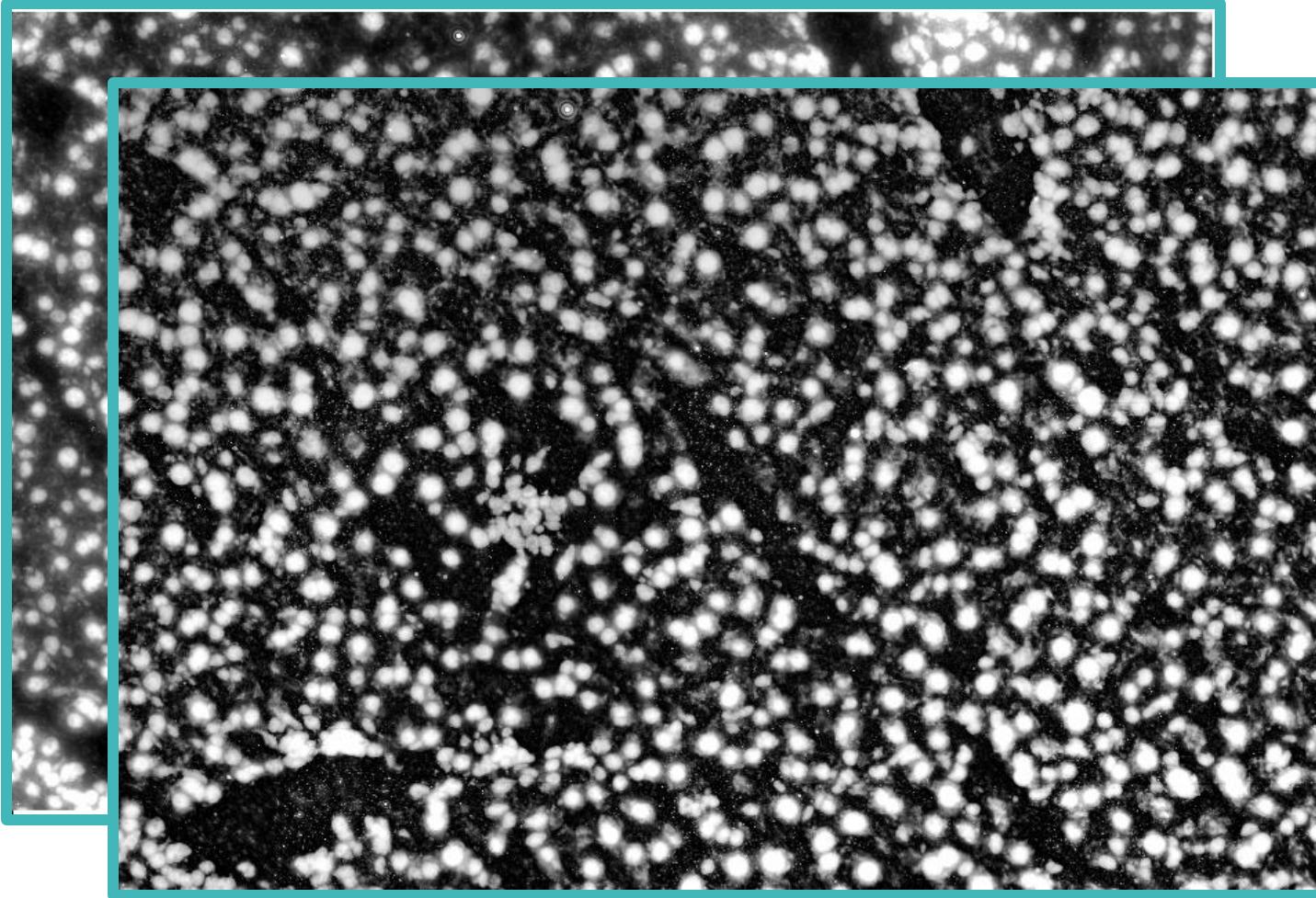
Control



Transcript density



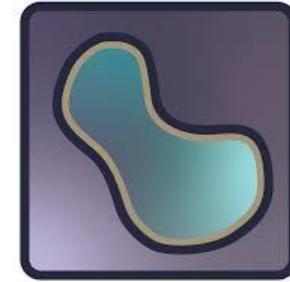
Raw image (tiff, zarr,...)



Processed image

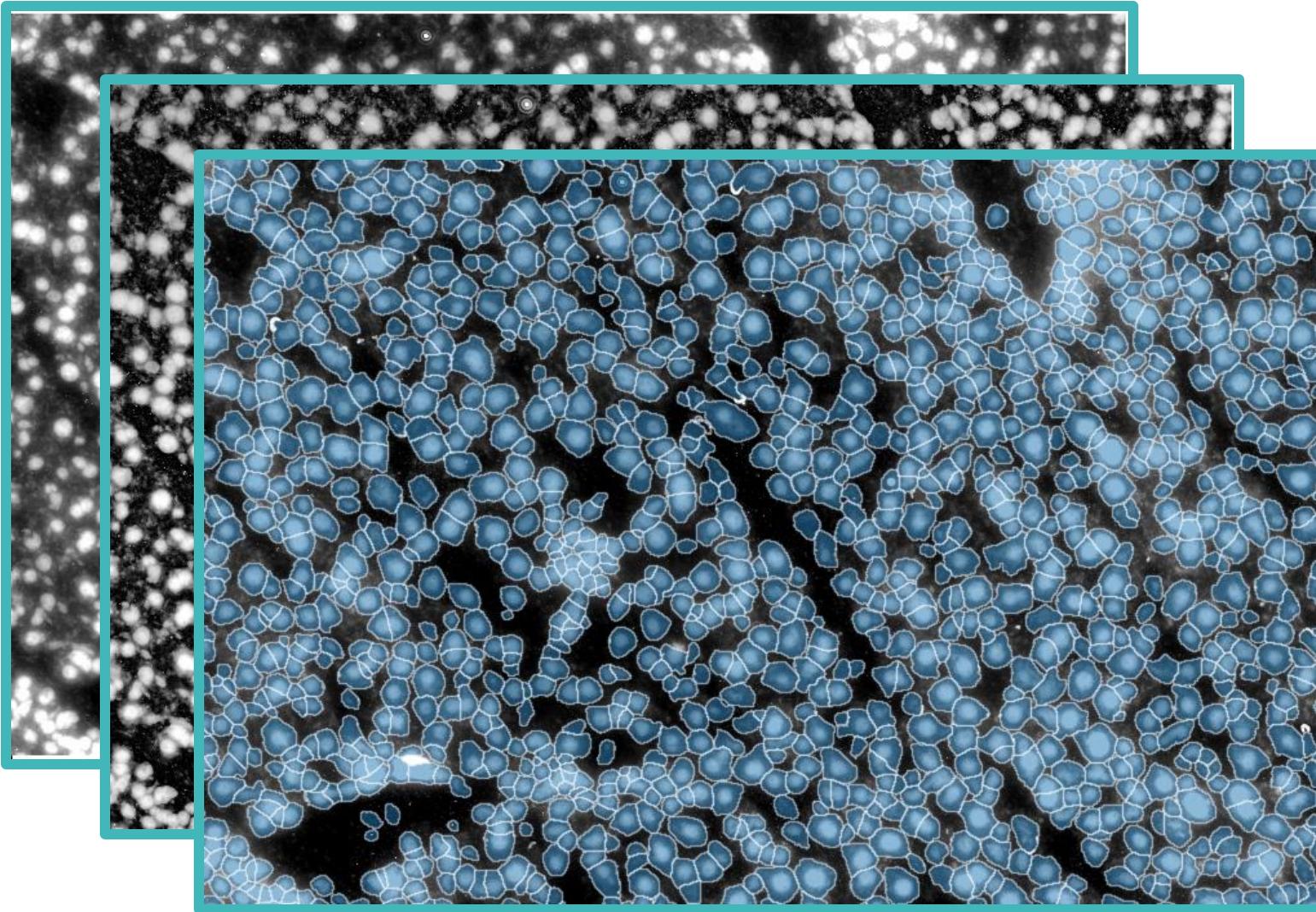


Coding



napari

Non-Coding



Segmented image

Raw image

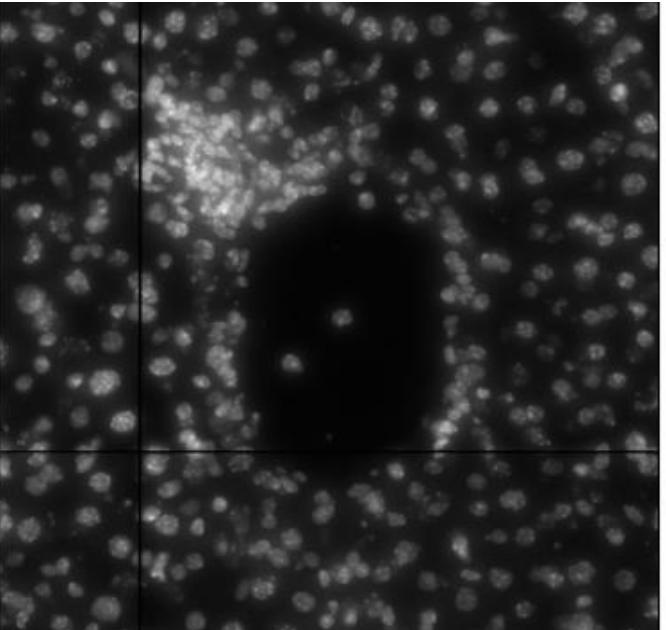
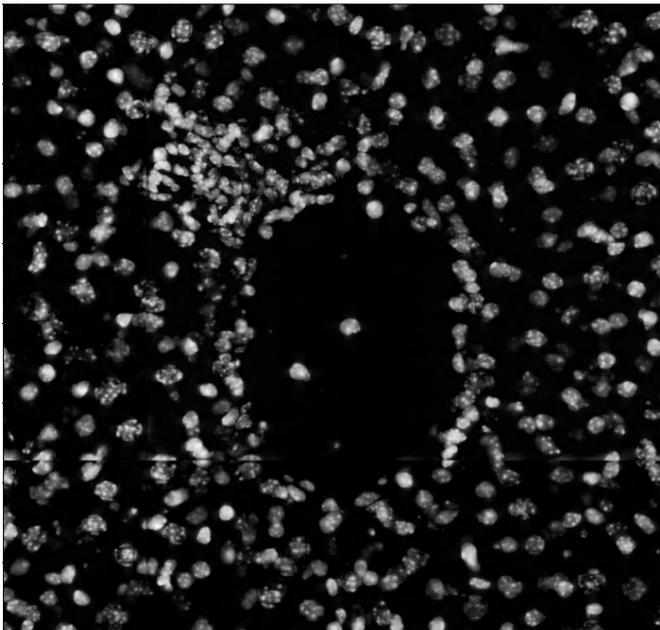


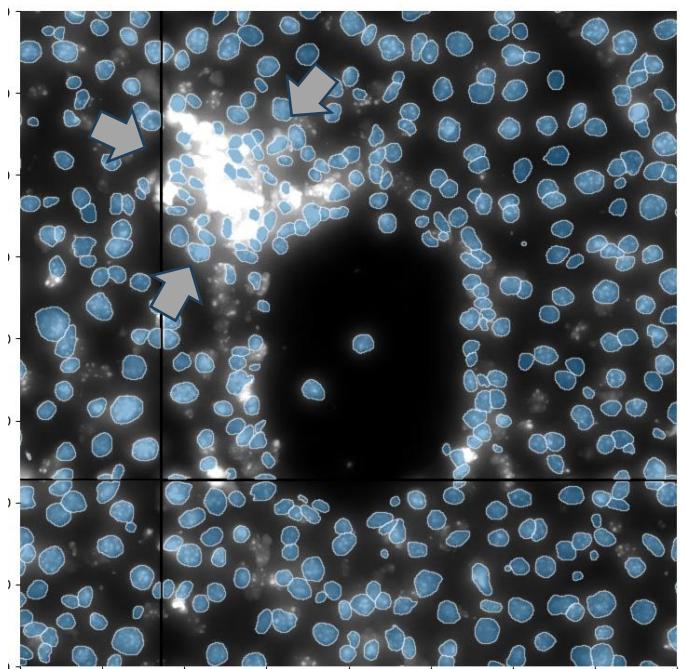
Image correction



Cleaned image



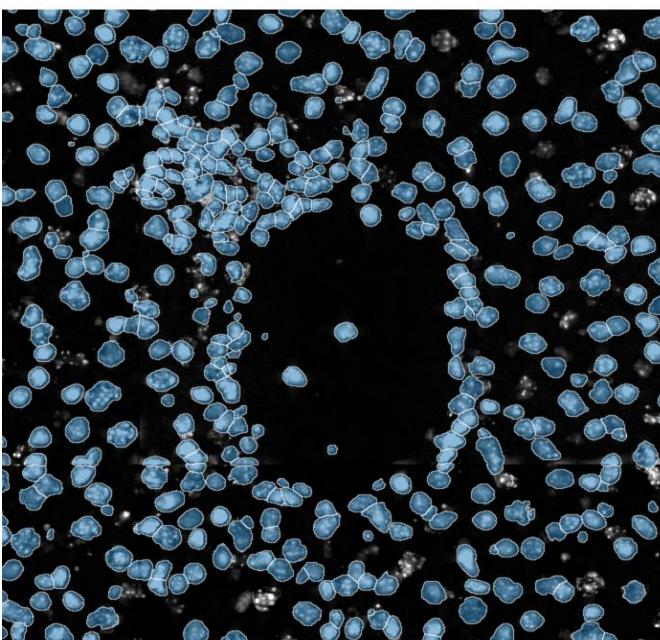
Cellpose



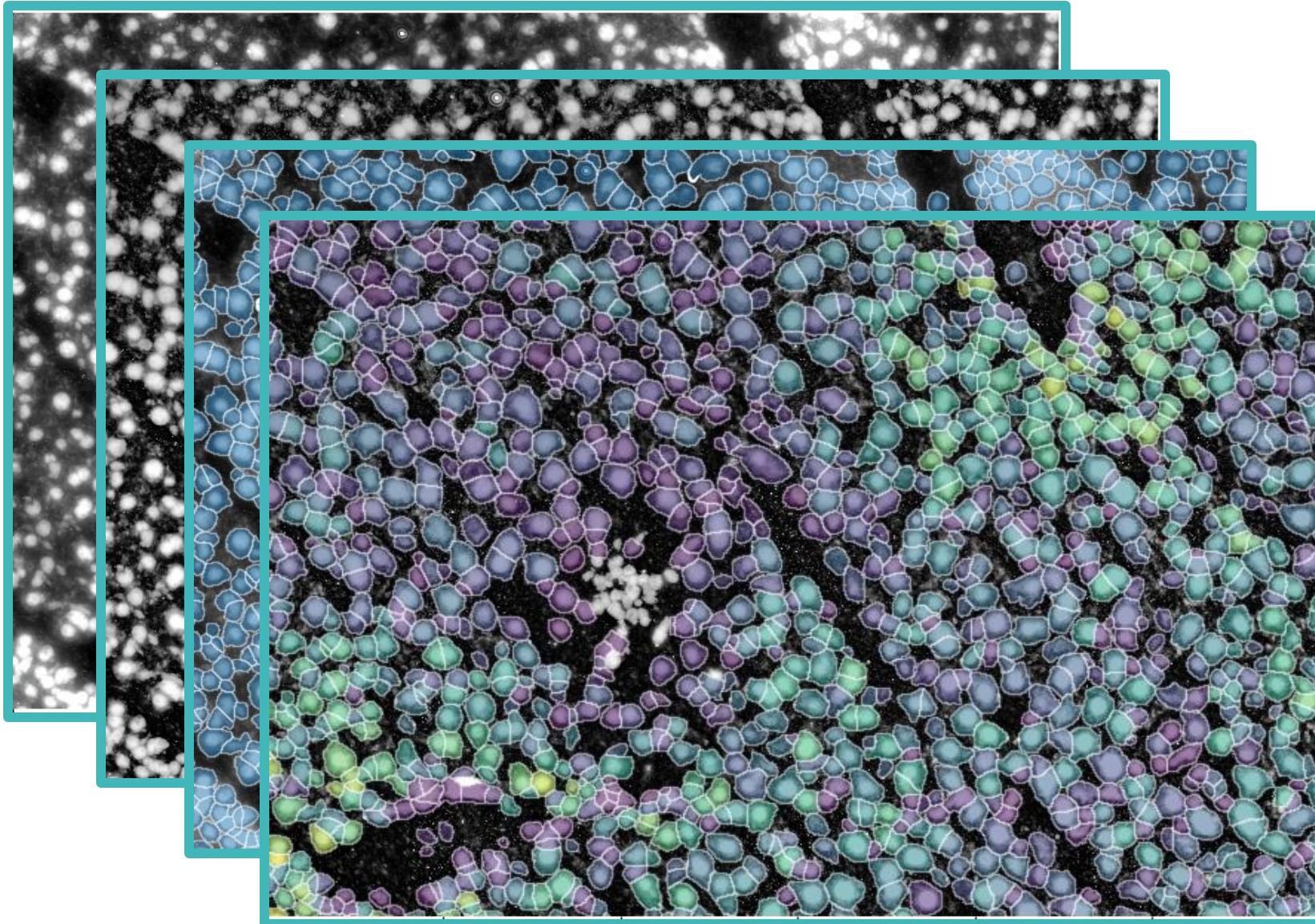
Low quality
segmentation



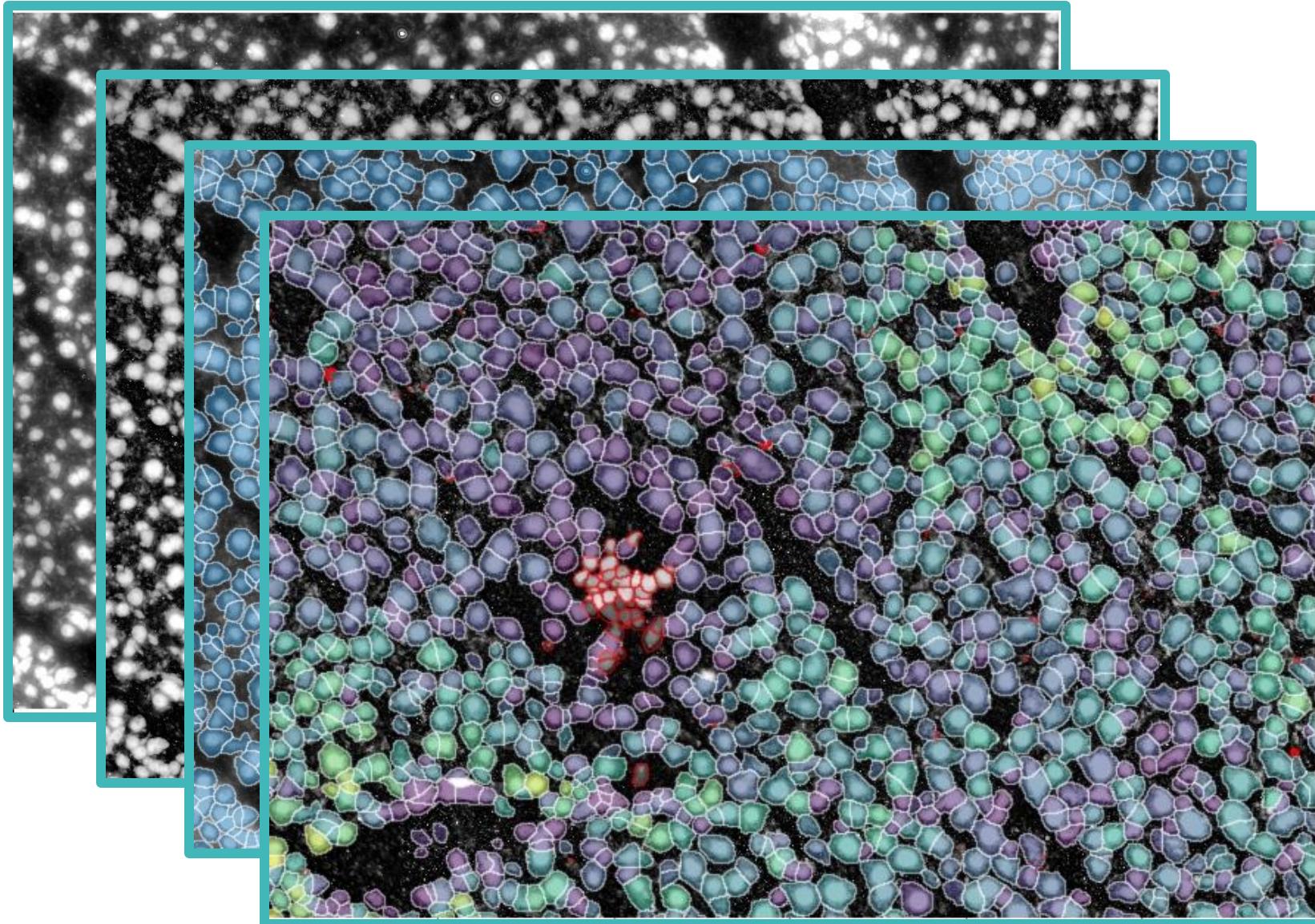
Cellpose



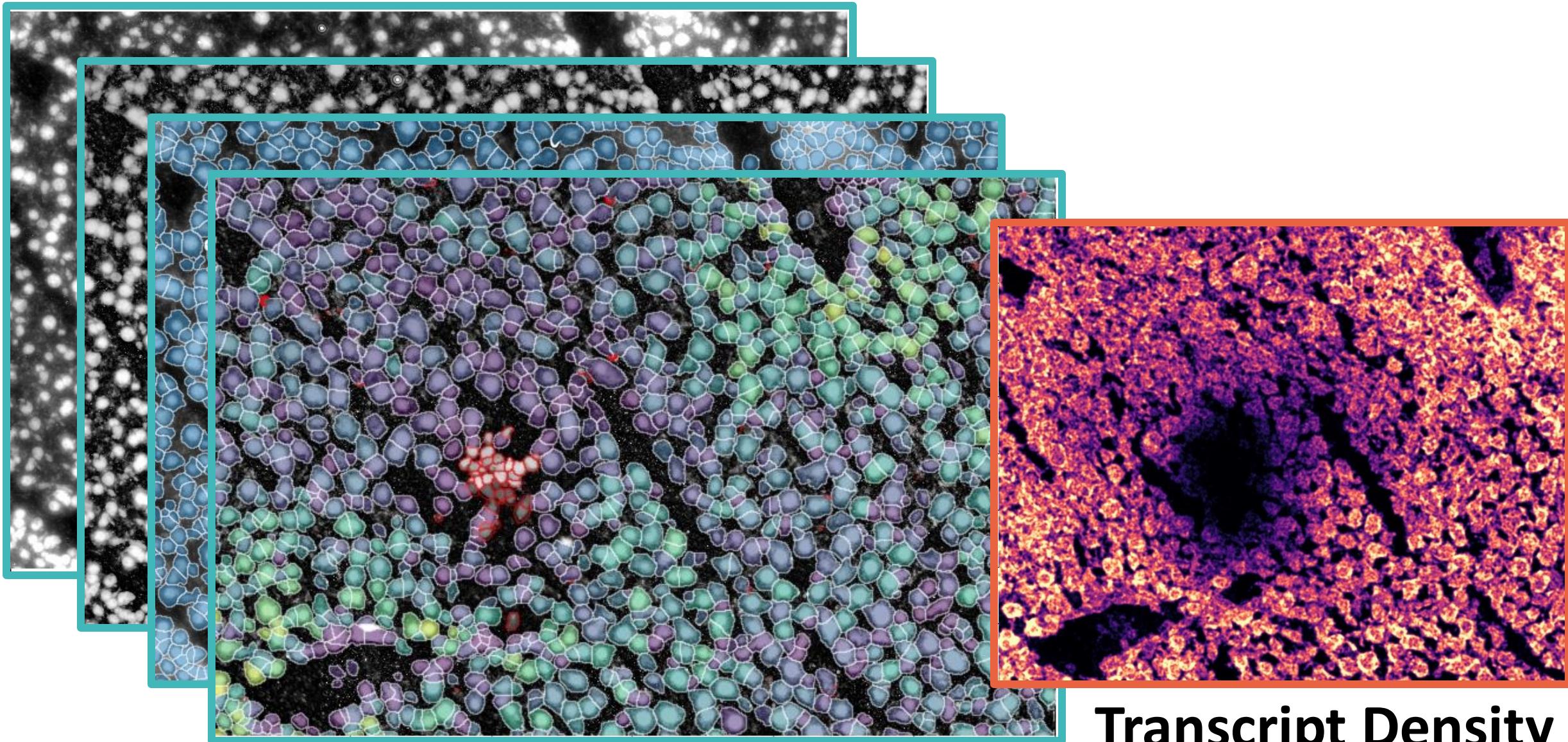
High quality
segmentation

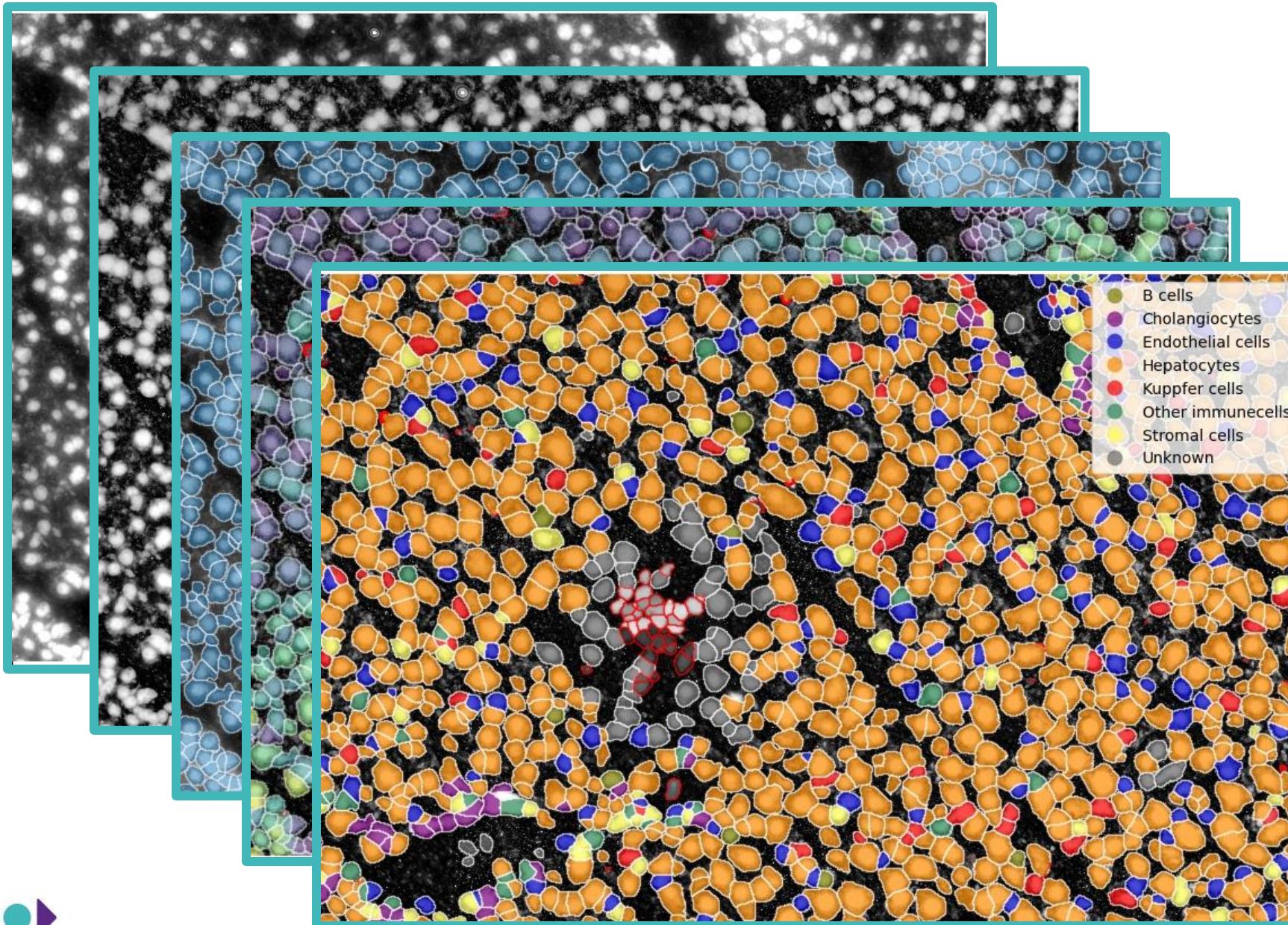


Pck1 expression

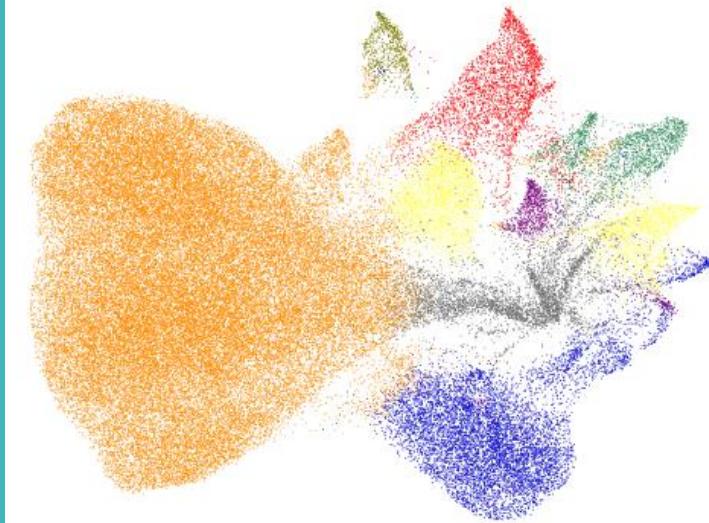


Pck1 expression

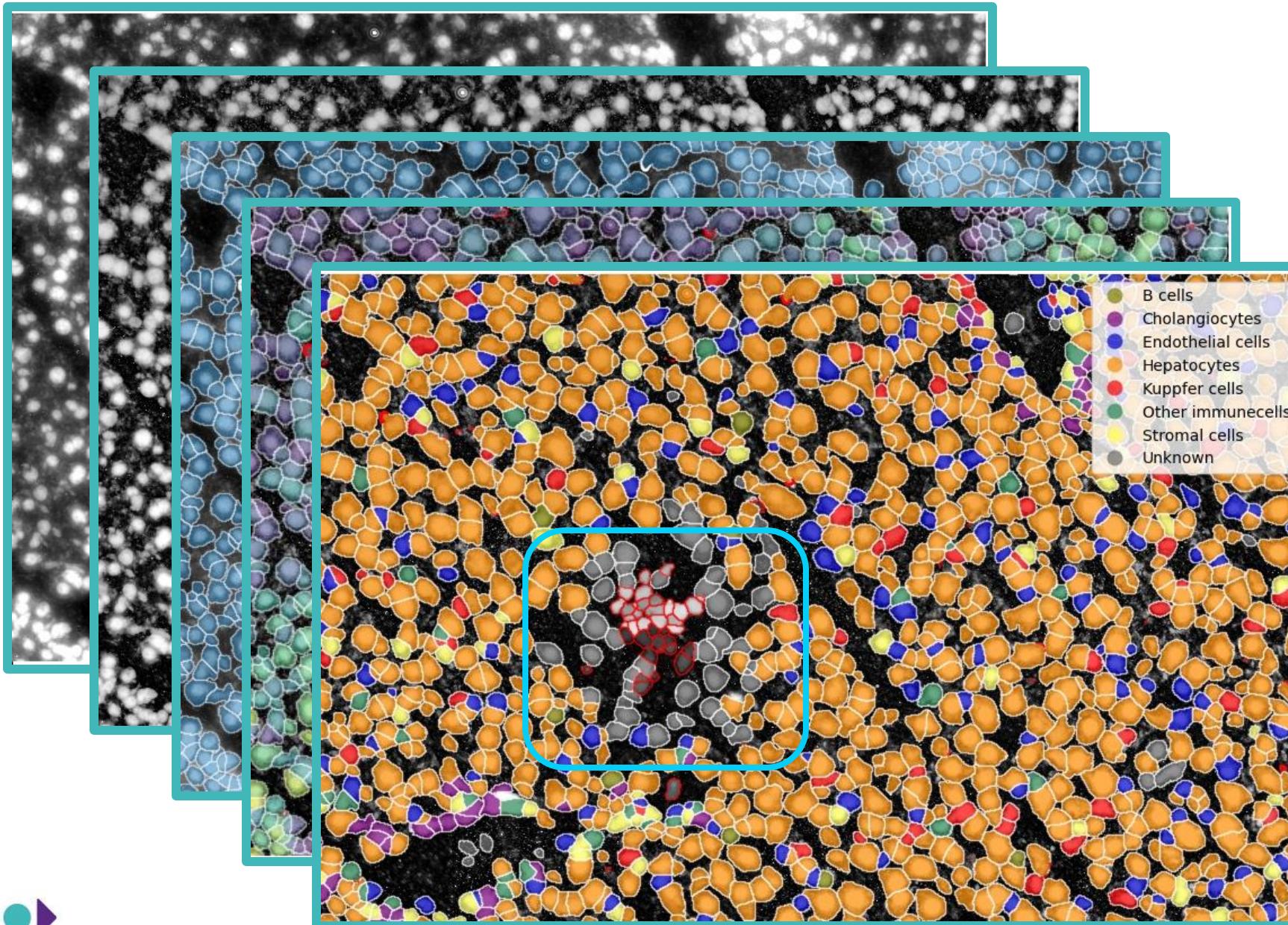


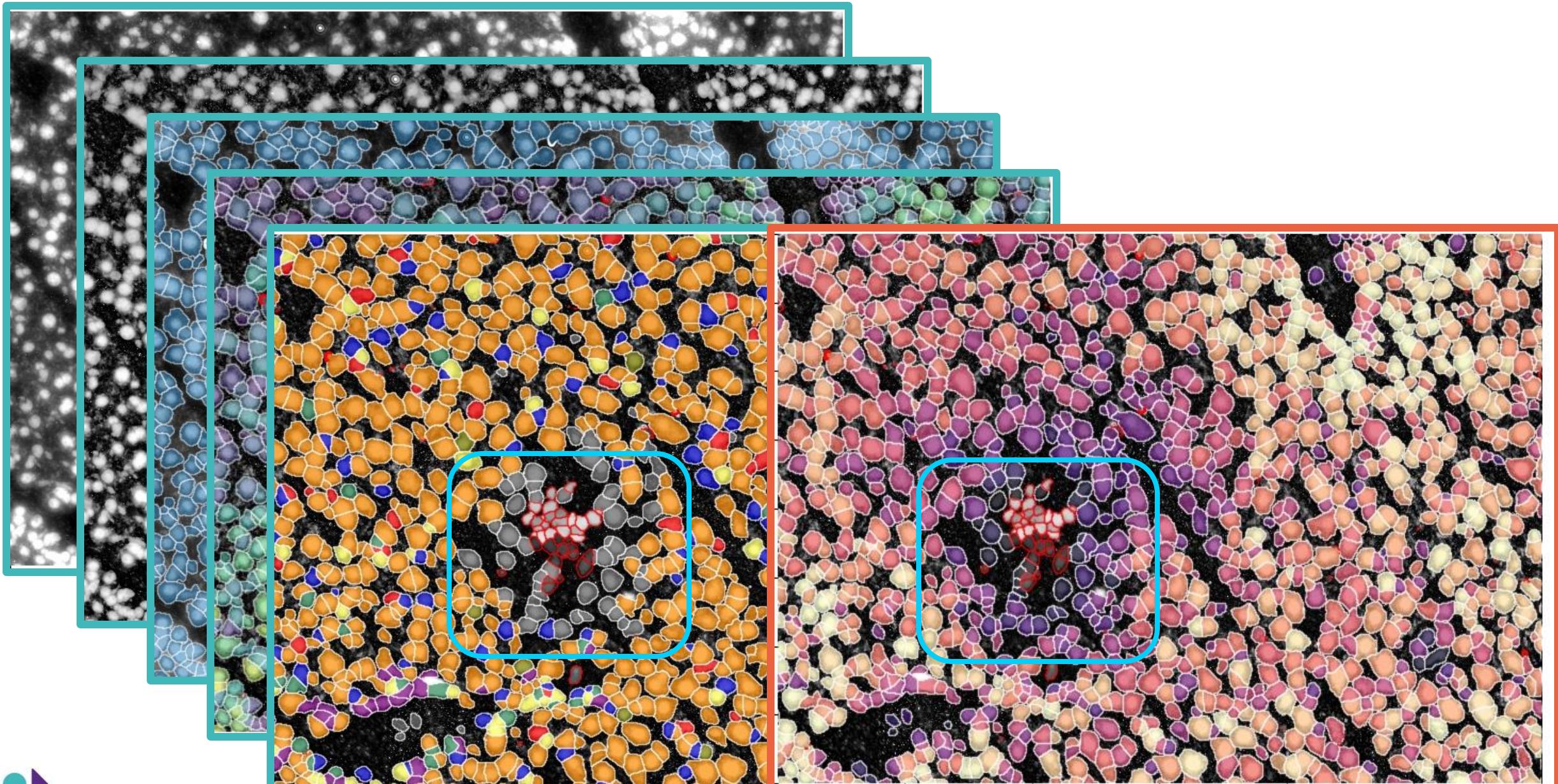


Cell annotation

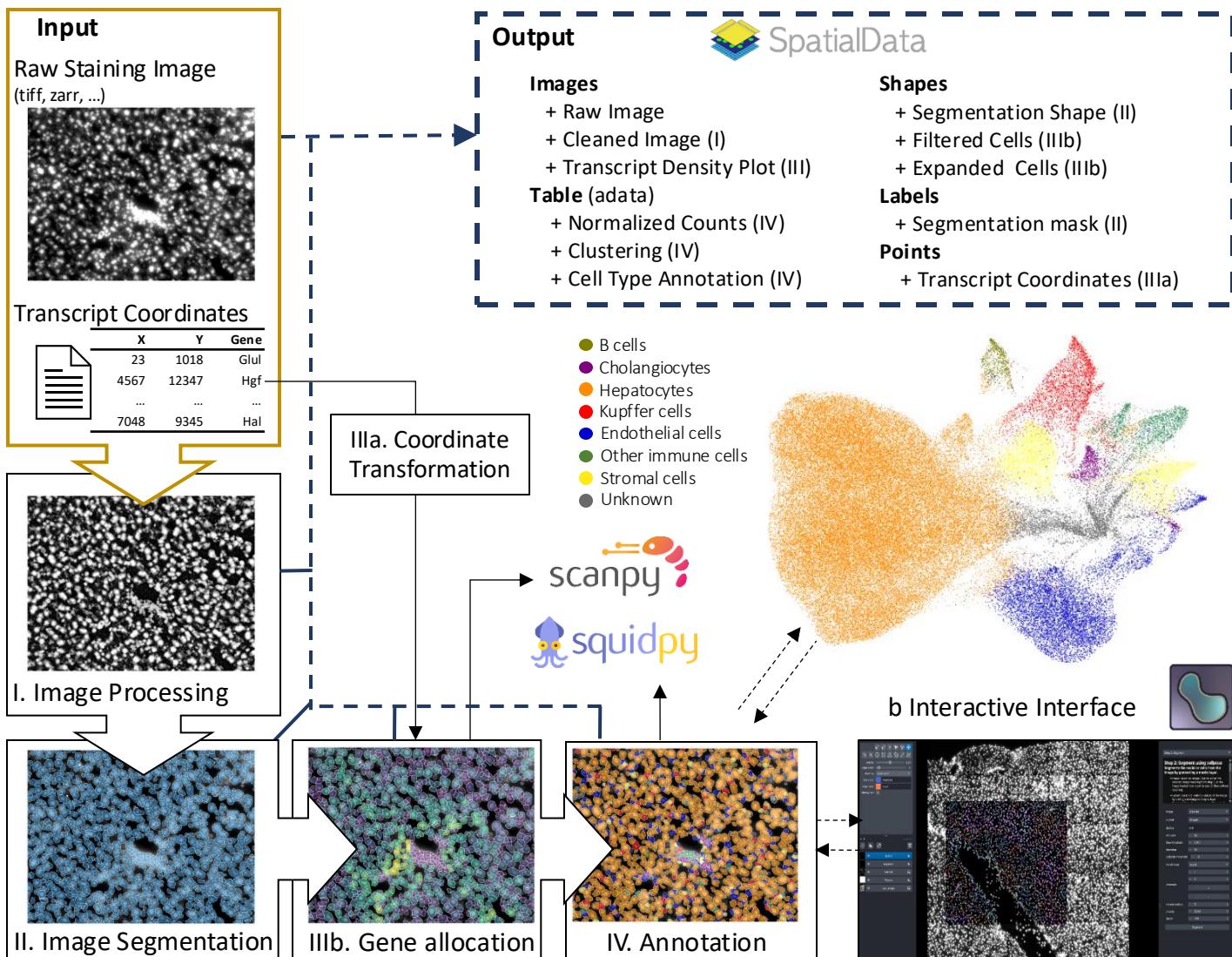


- B cells
- Cholangiocytes
- Endothelial cells
- Hepatocytes
- Kupffer cells
- Other immunecells
- Stromal cells
- Unknown





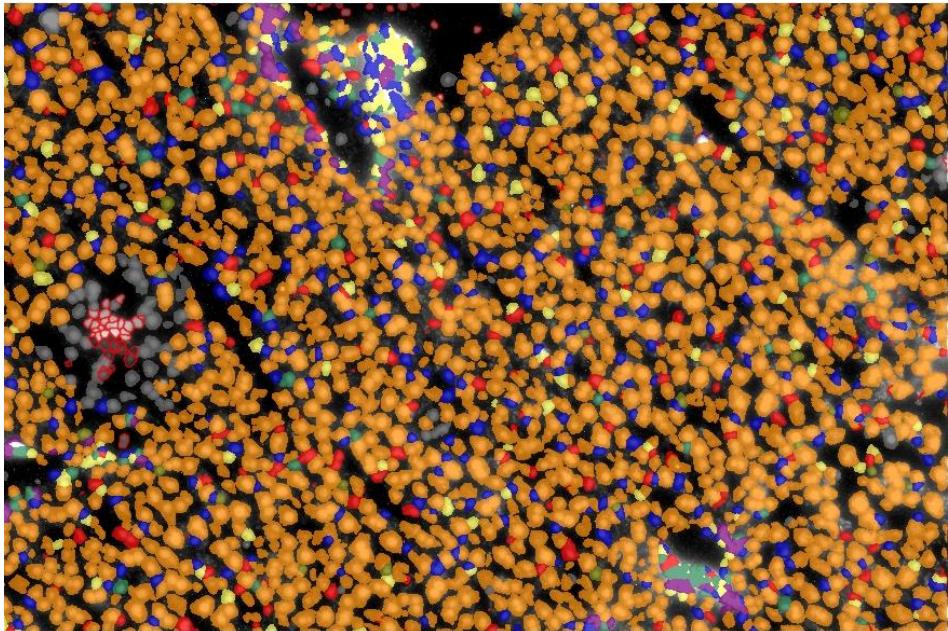
a Pipeline Overview



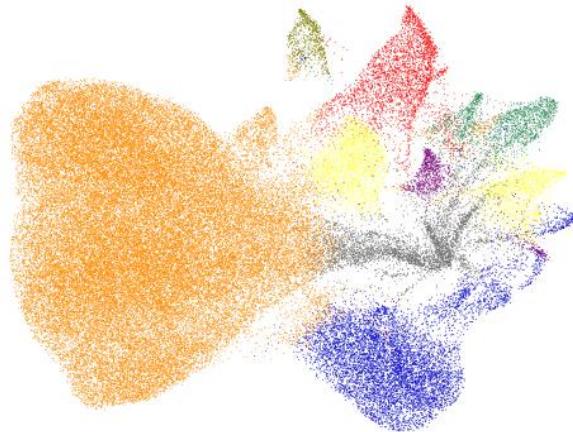


**Works better than standard
MERSCOPE analysis**

SPARROW



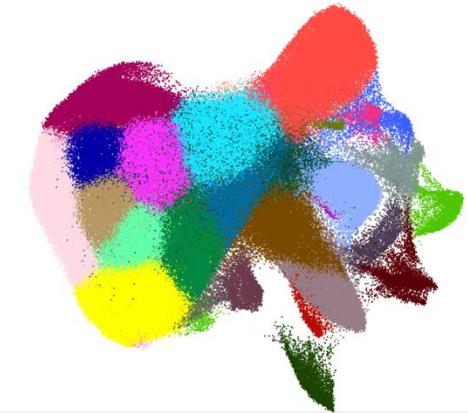
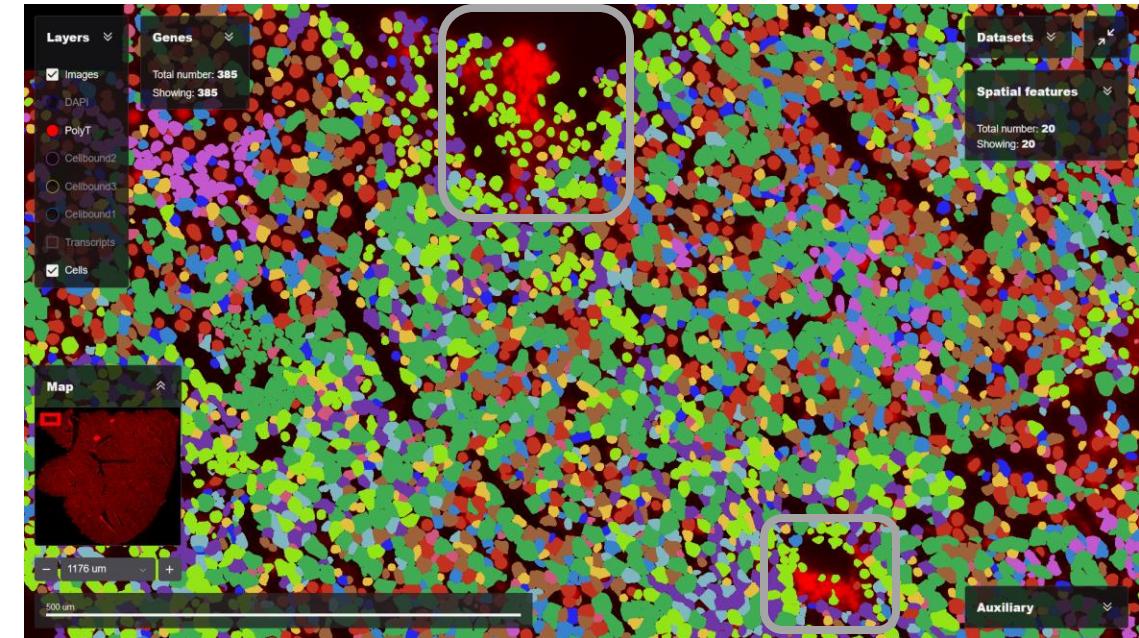
- B cells
- Cholangiocytes
- Endothelial cells
- Hepatocytes
- Kupffer cells
- Other immune cells
- Stromal cells
- Unknown



Vizgen MERFISH mouse liver data

Vs.

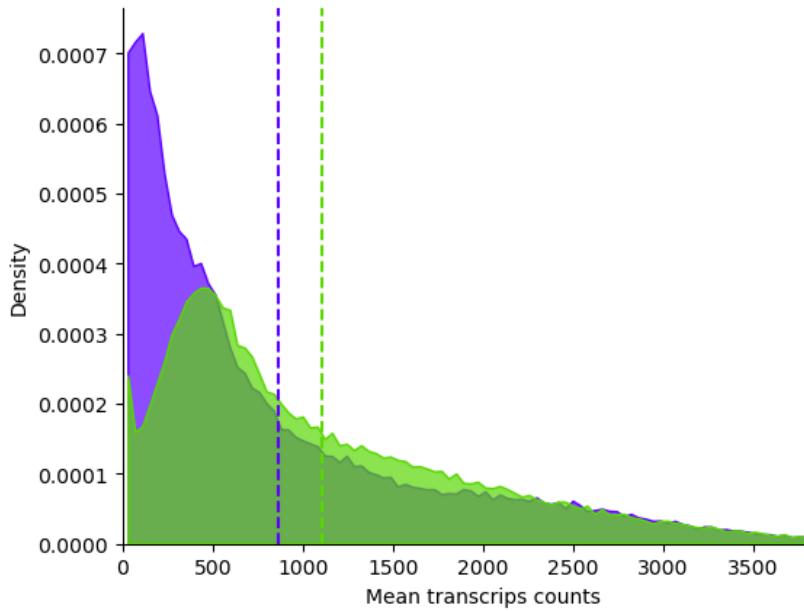
MERSCOPE



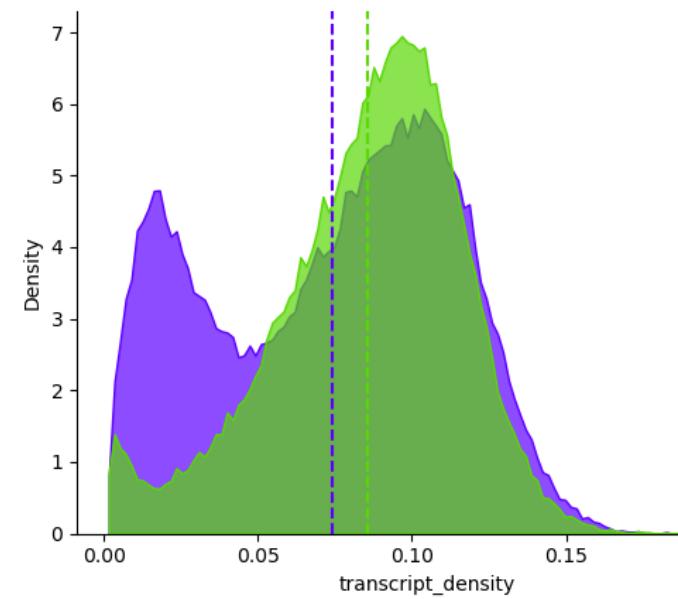
SCIENCE MEETS LIFE

When using SPArrOW more high quality cells and less debris is picked up

Mean transcript count



Transcript density

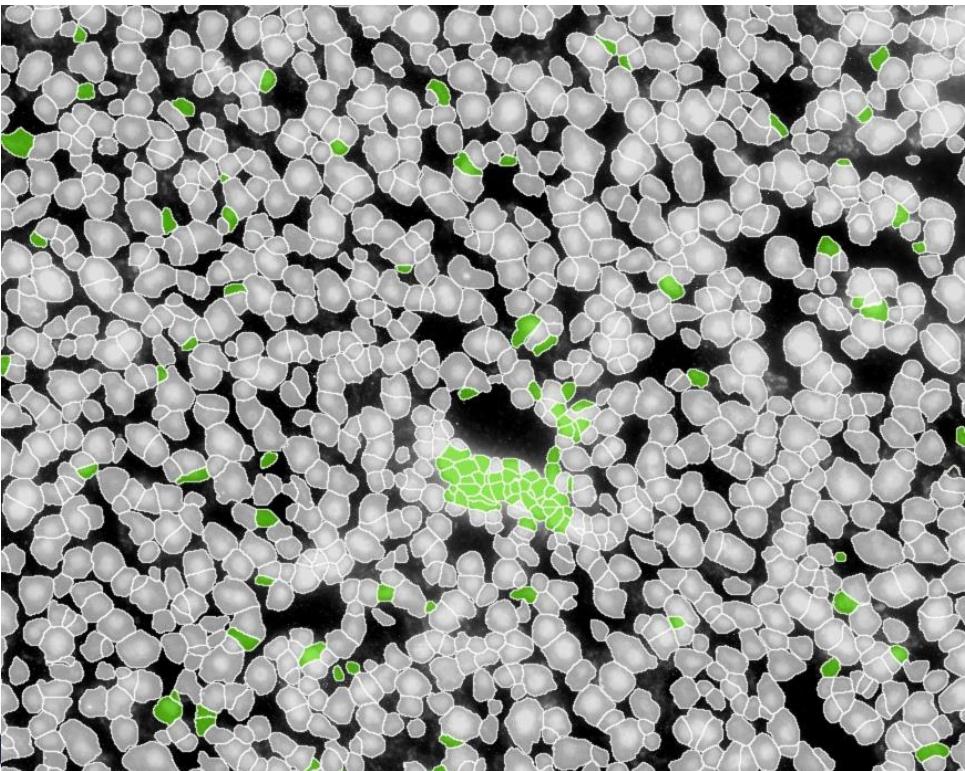
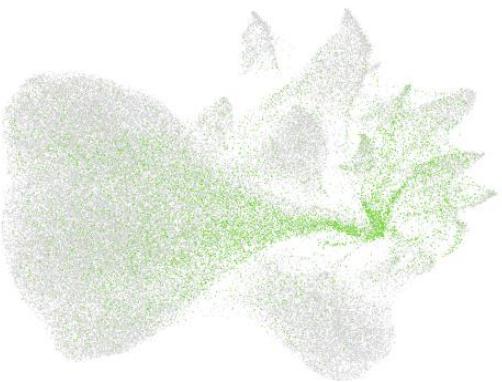


SPArrOW

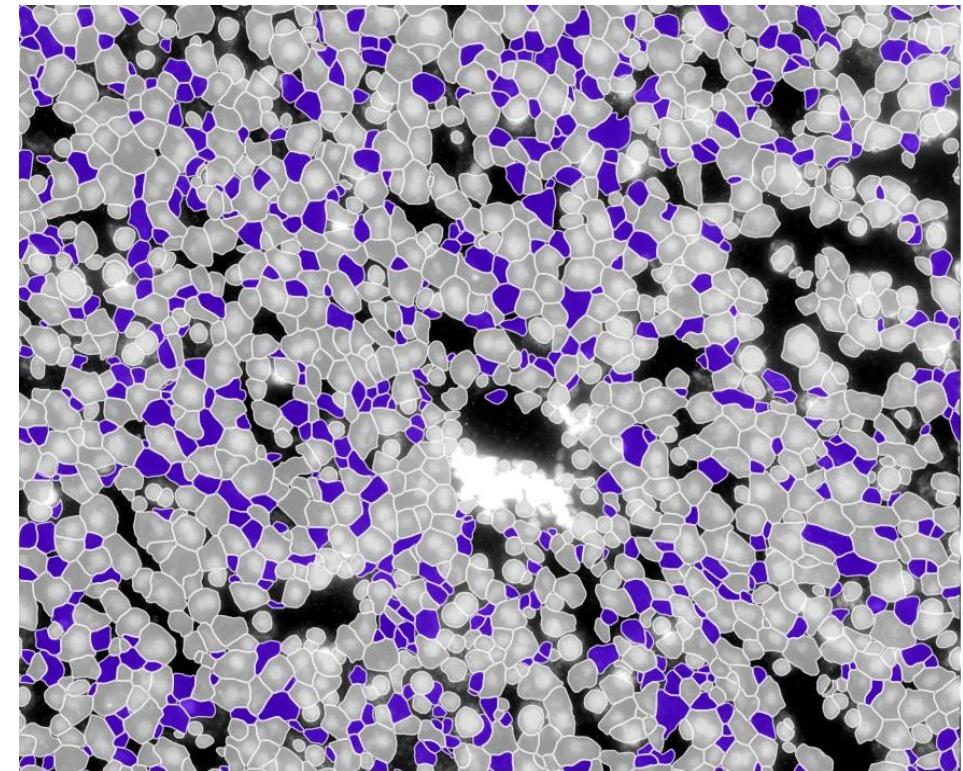
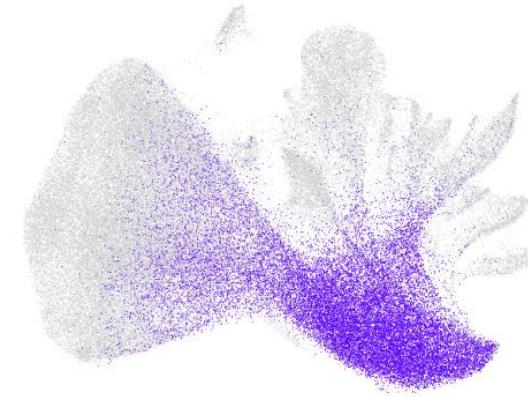


MERSCOPE output

SPArrOW

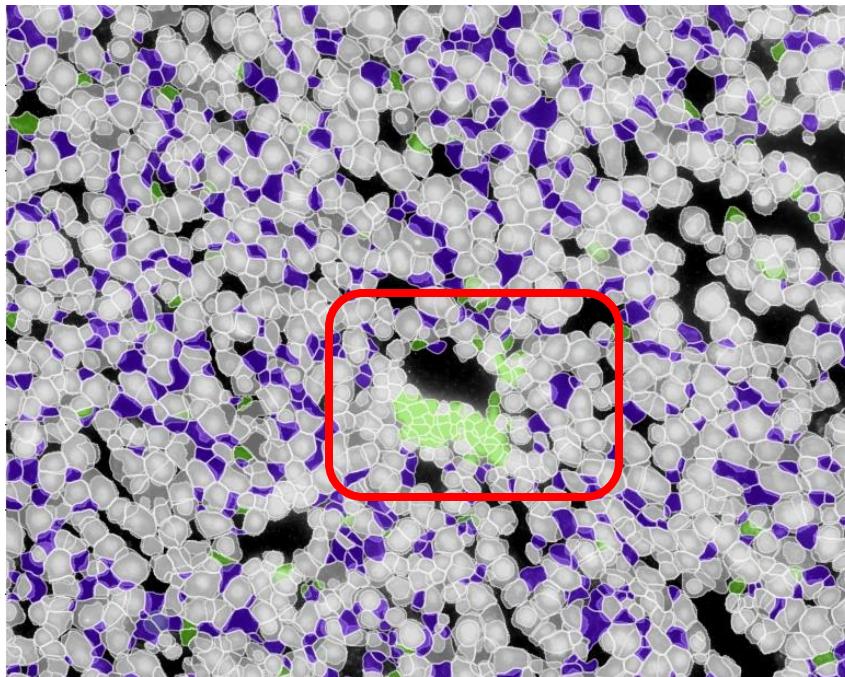


MERSCOPE

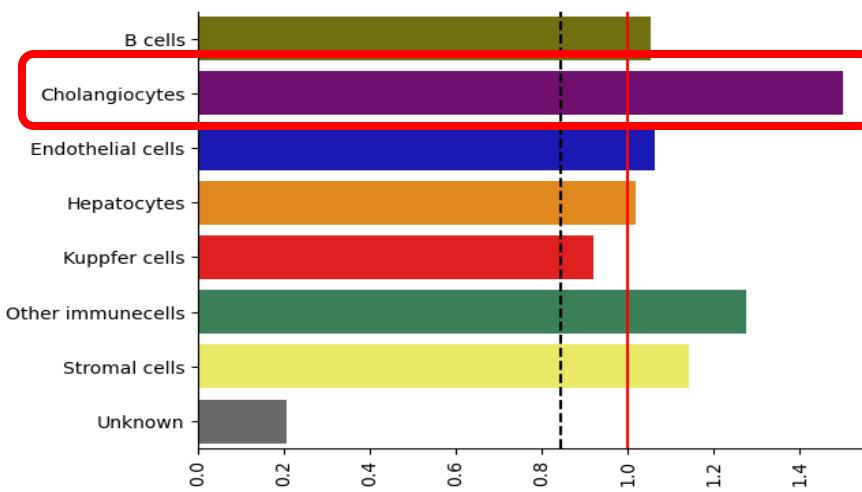
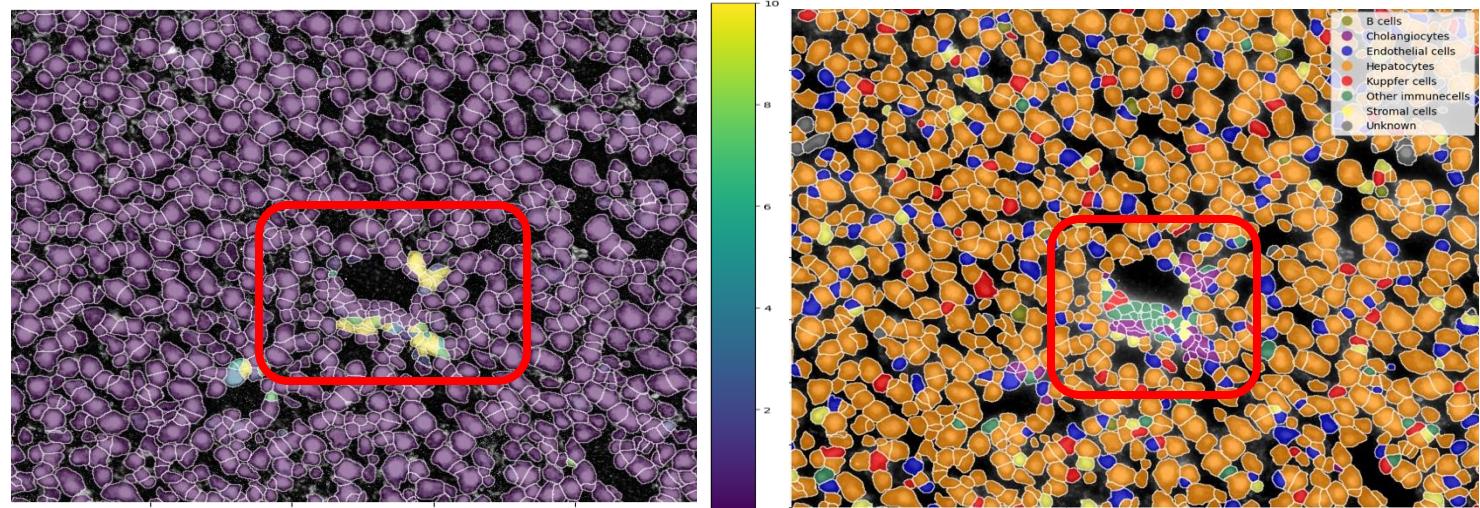


SCIENCE MEETS LIFE

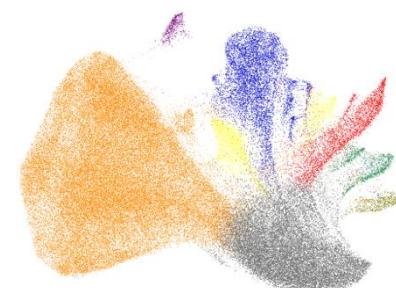
The extra cells are celltype specific



Epcam gene expr.



UMAP MERSCOPE+ SPArrOW



- B cells
- Cholangiocytes
- Endothelial cells
- Hepatocytes
- Kupffer cells
- Other immune cells
- Stromal cells
- Unknown

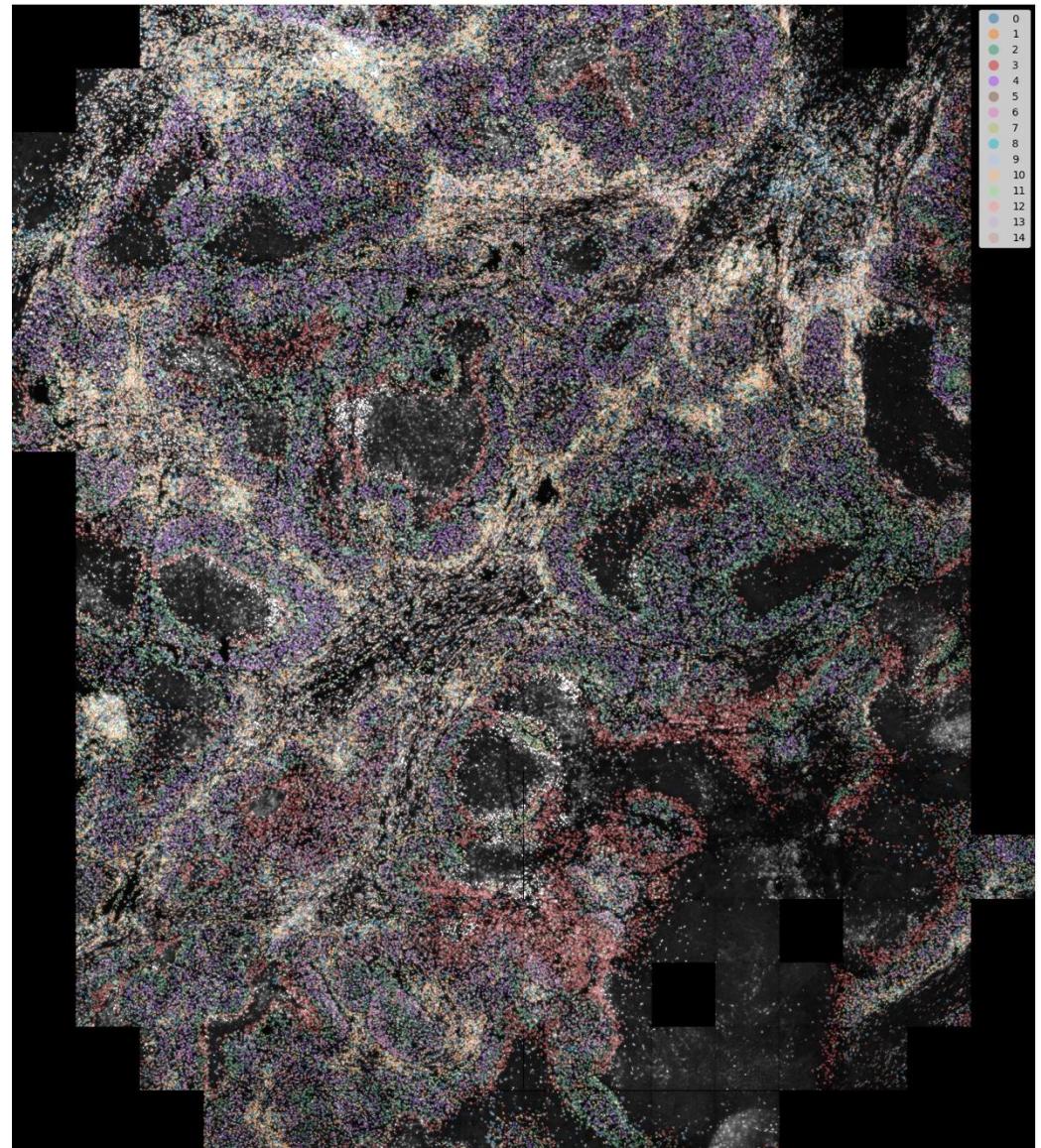
UMAP SPArrOW

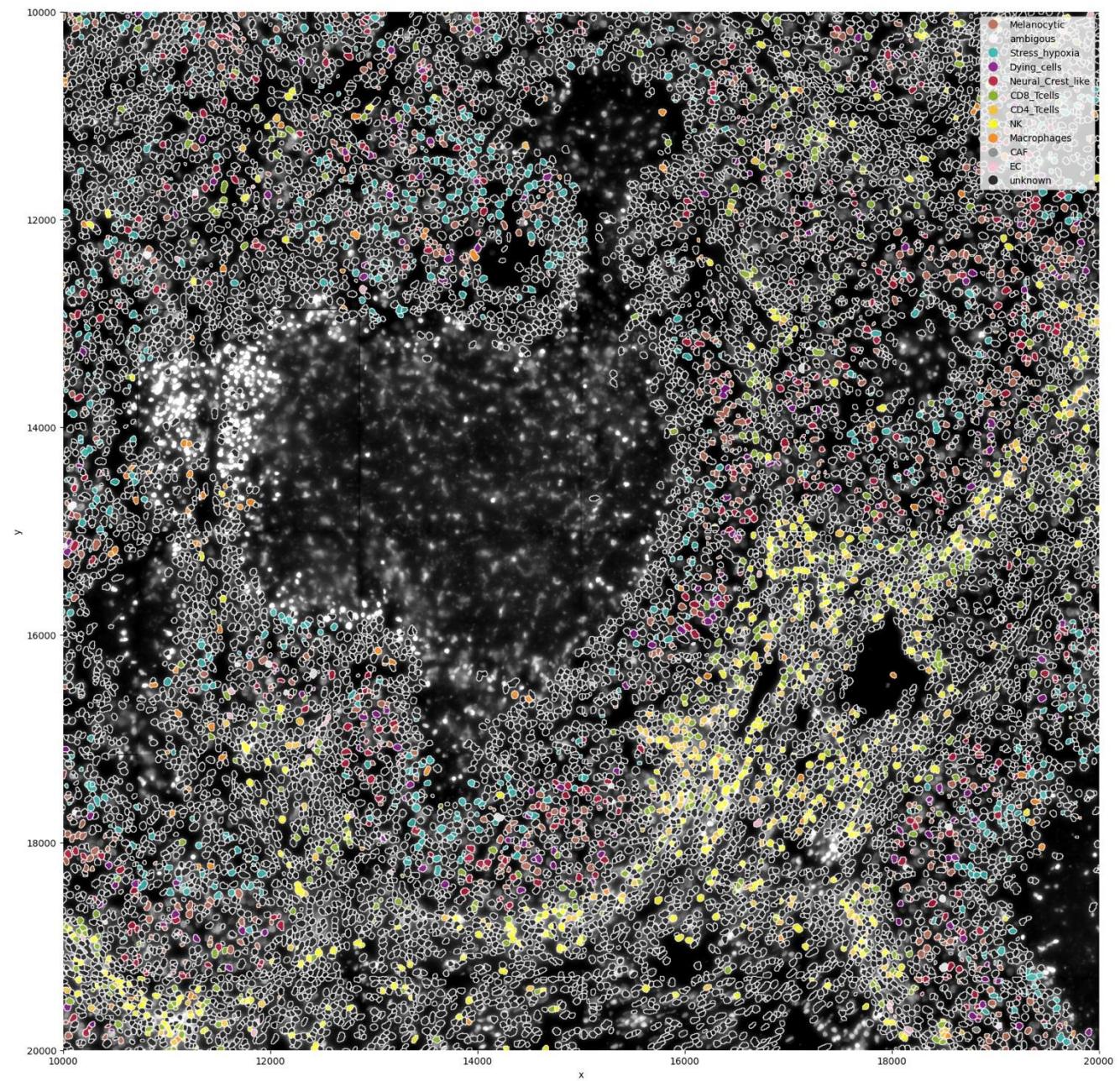


SPArrOW quality control

Human Melanoma dataset

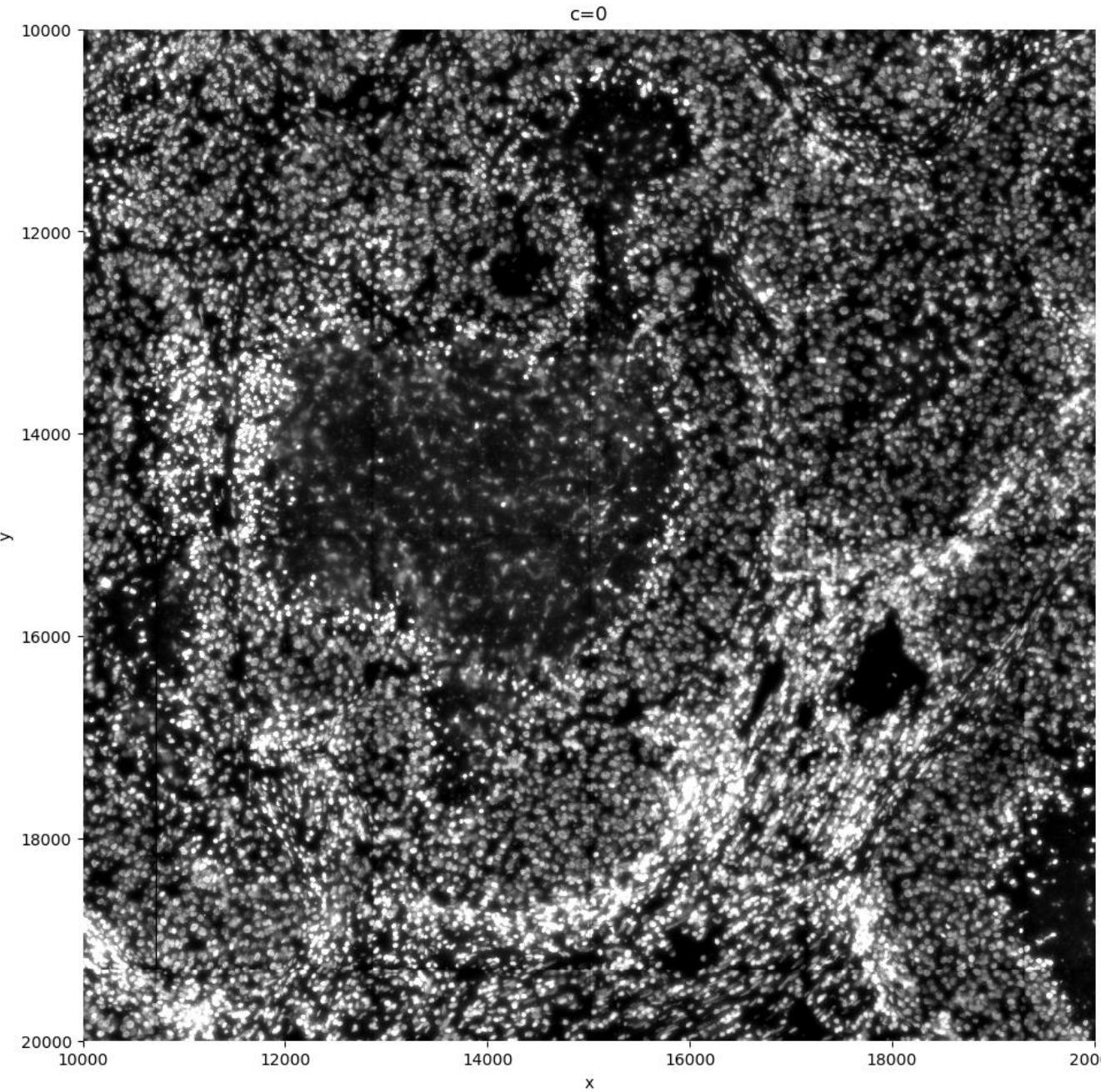
Lots of cells get lost in the analysis: why and where?





Problems with the raw image?

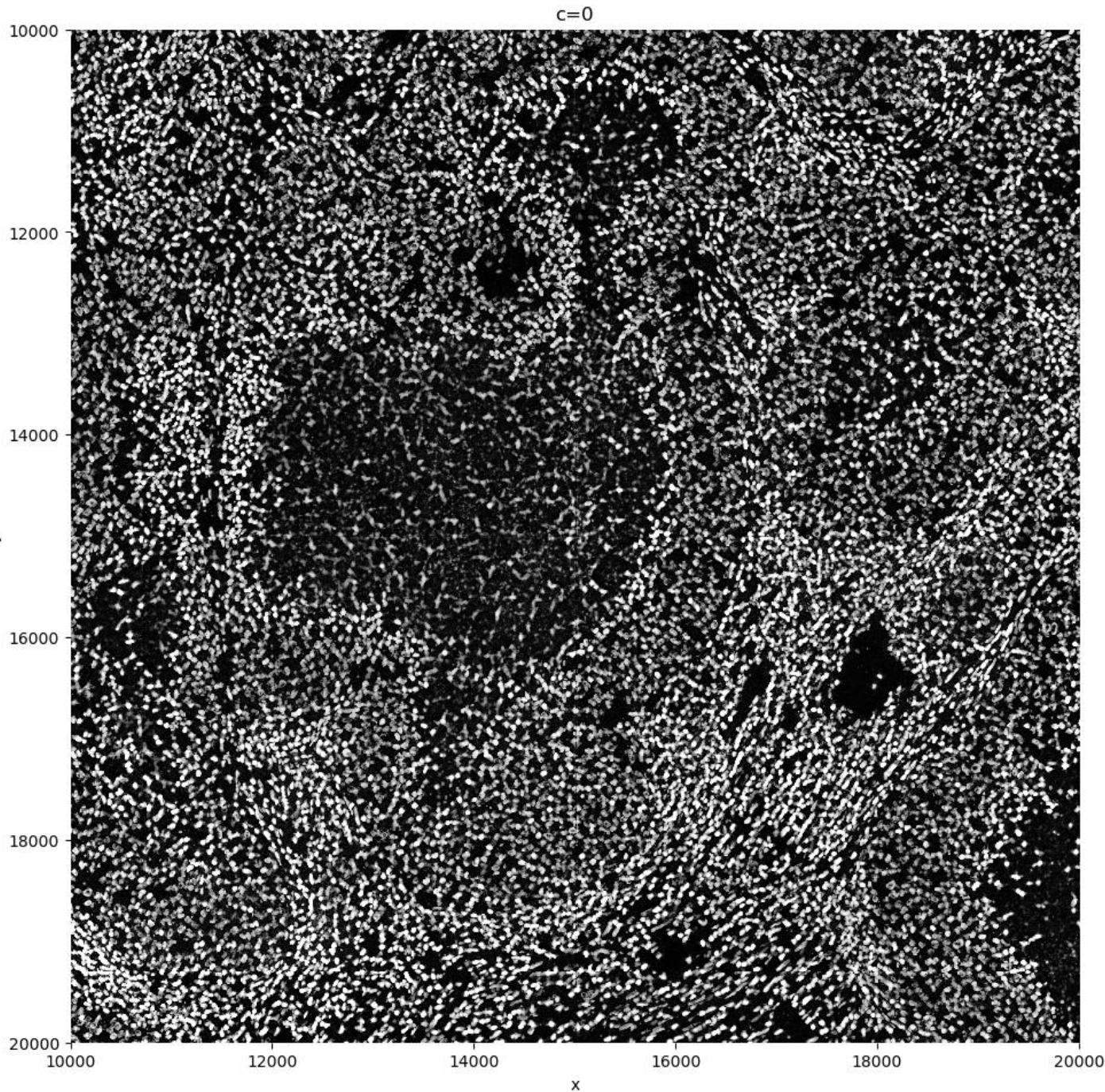
- Brightness differences
- Enlonged cells



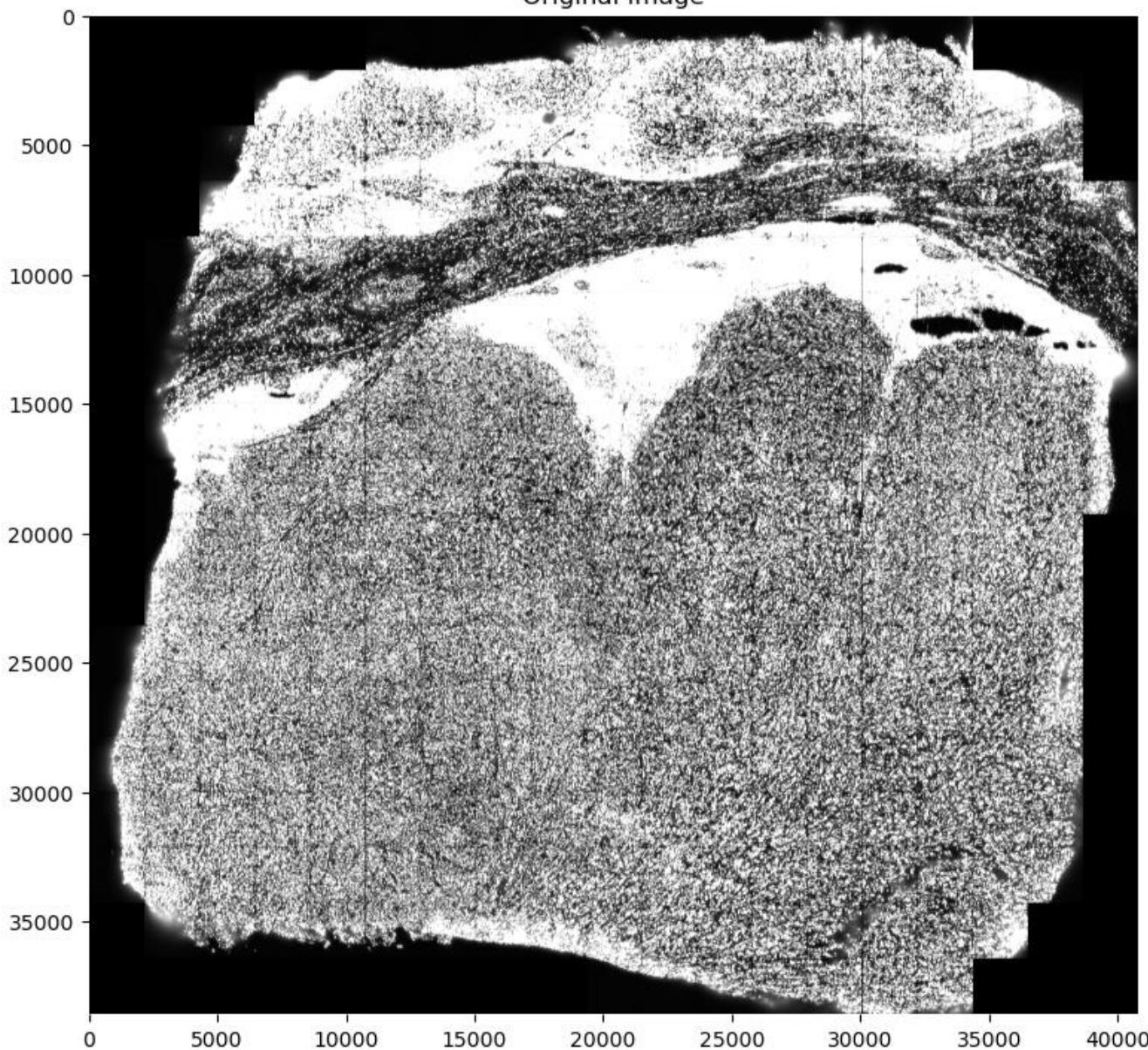
Processing the image:

Brightness issues are resolved

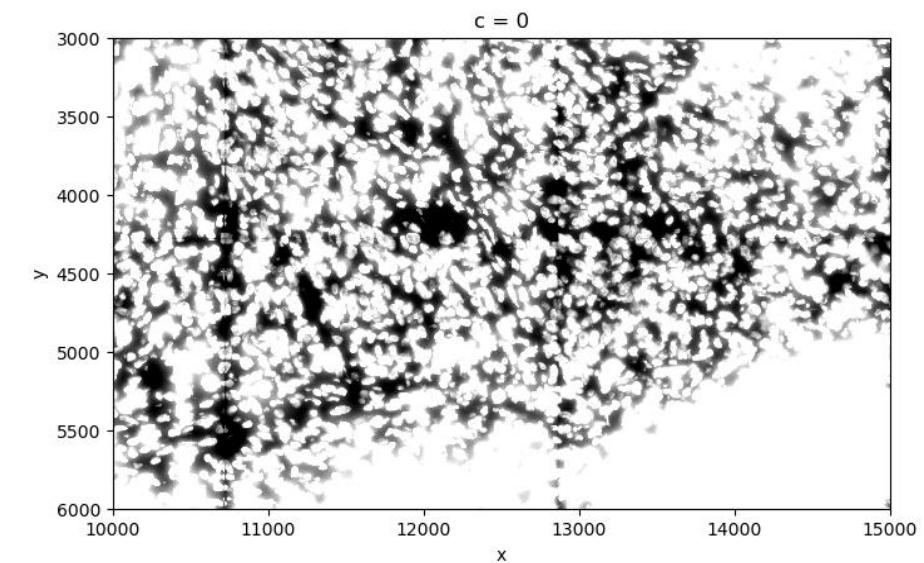
Quality control 1



Original image



Not always possible...

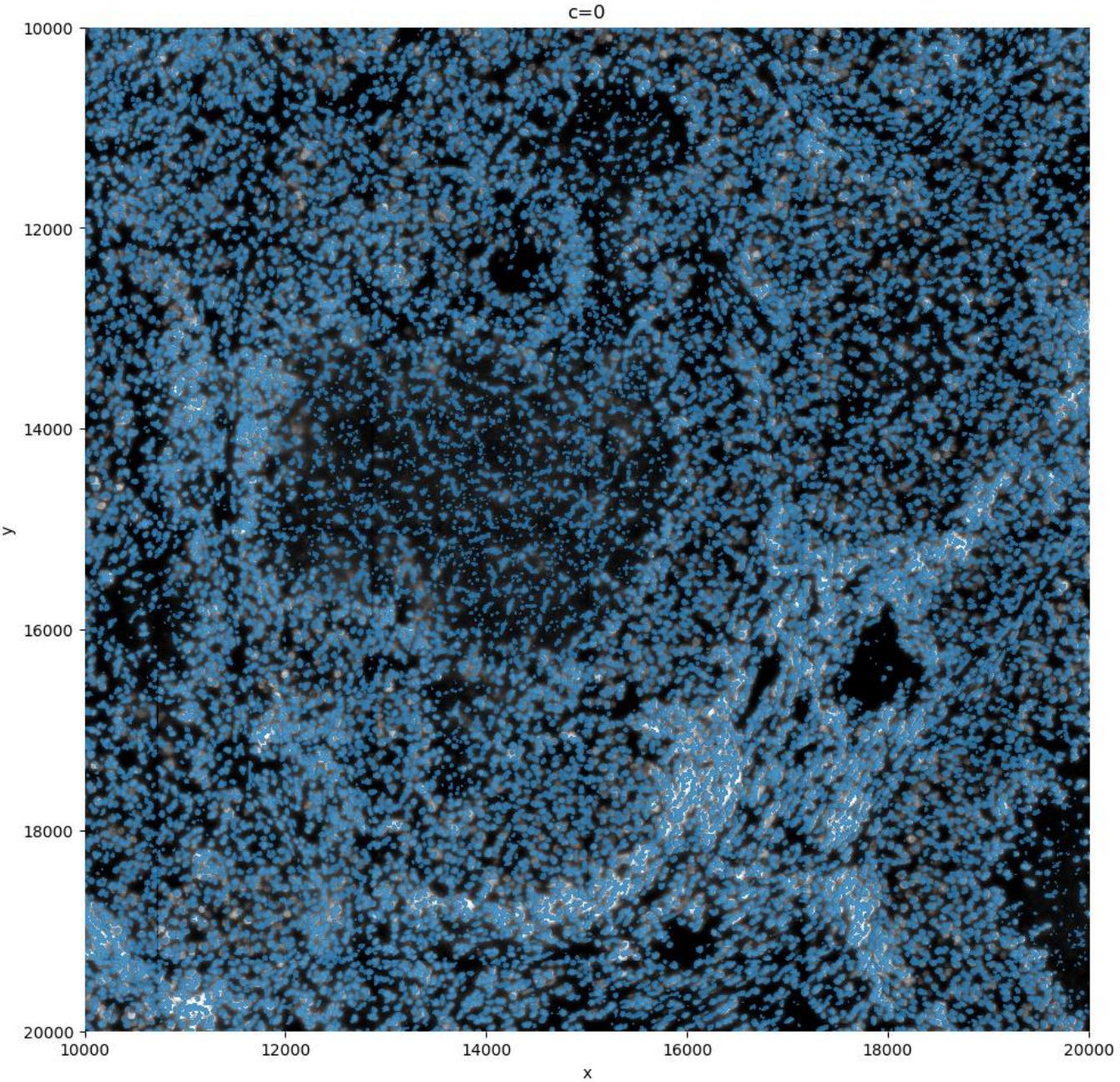


Segmenting the image:

Almost all cells are segmented, not a big issue.

If not: you can optimize the segmentation

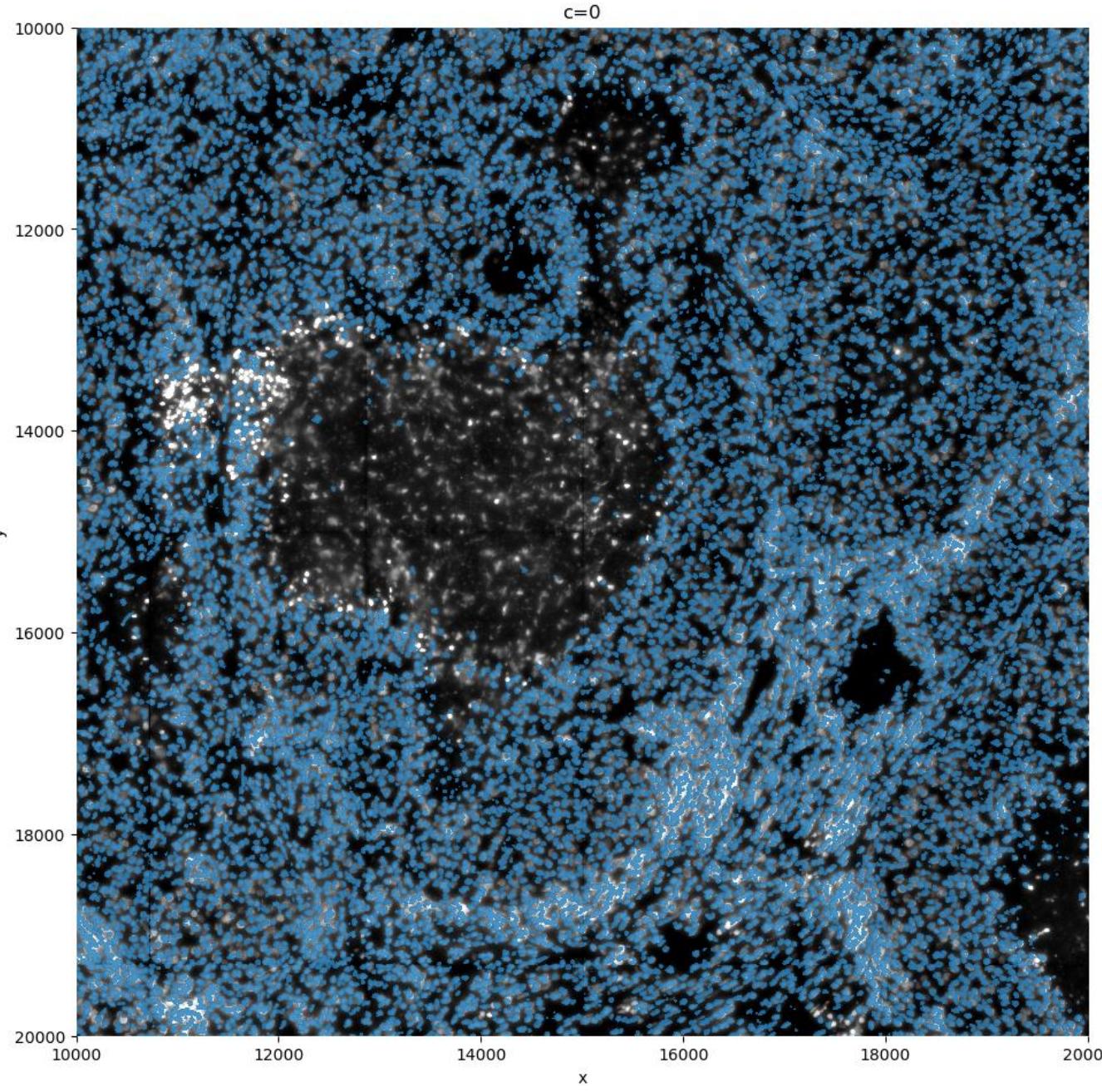
Quality control 2

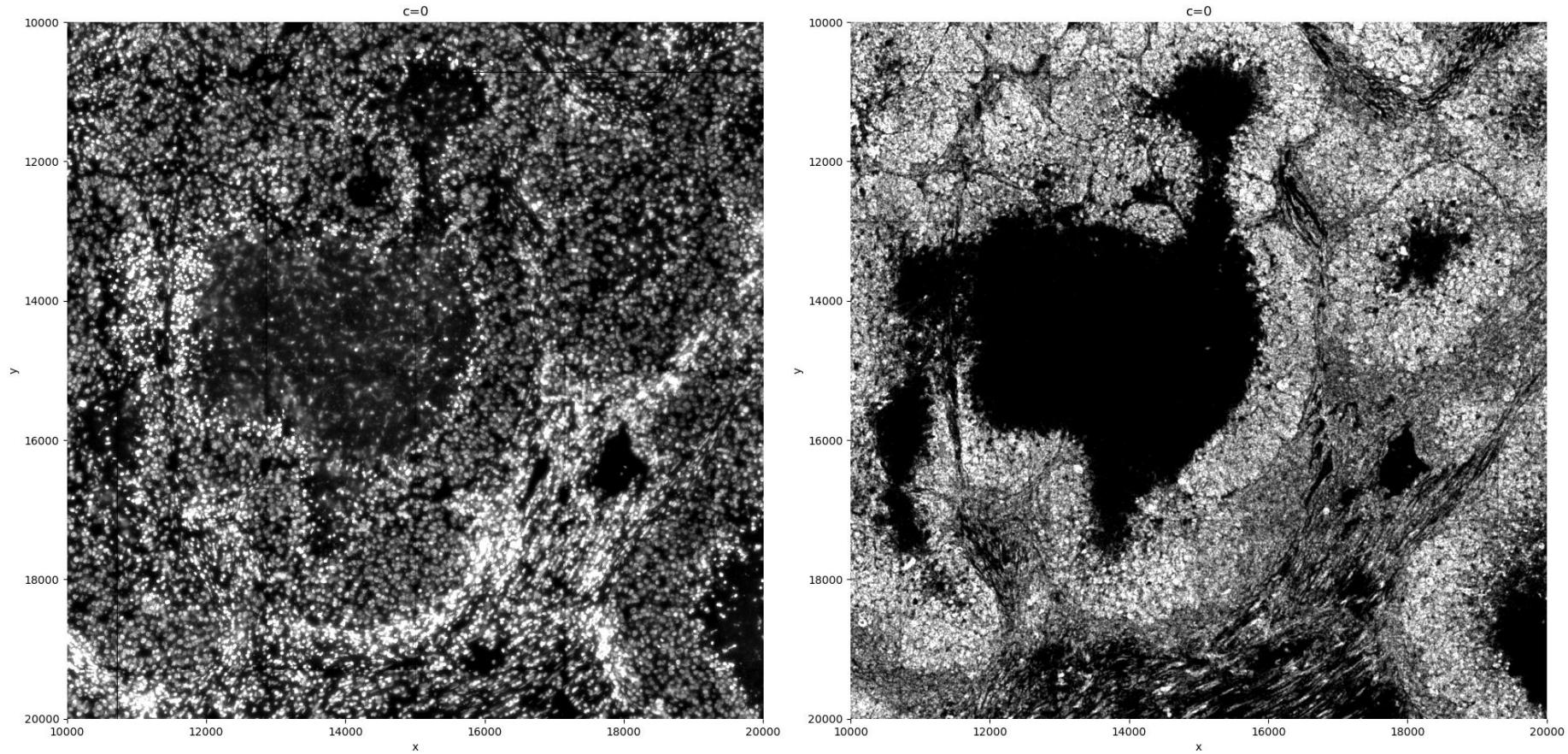


Allocating the genes:

Quite some cells disappear:
no transcripts are belonging
to these cells

How is this possible?





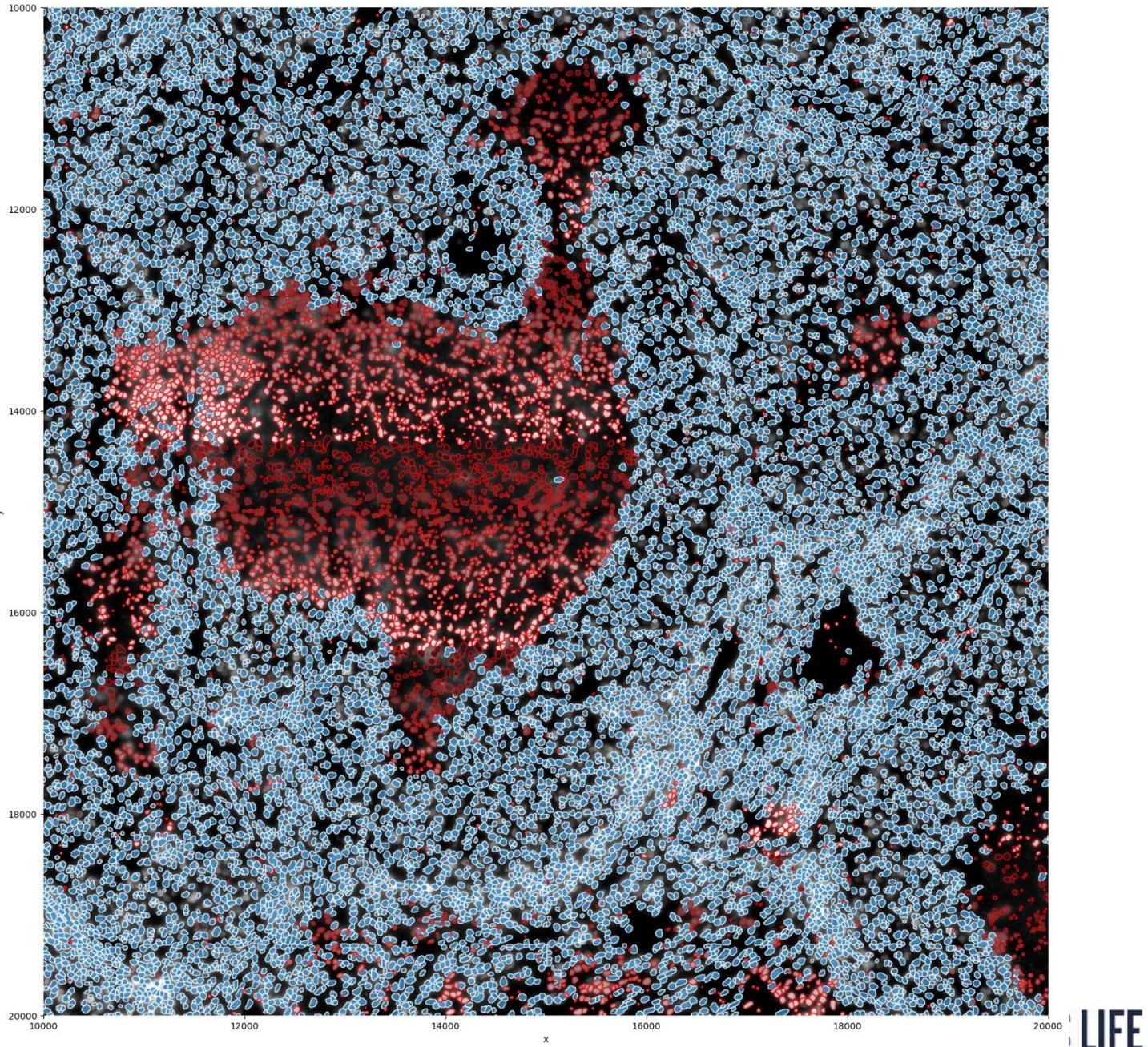
Transcript density plots based on locations of transcripts

Some regions appear to be without transcripts: biological or technical?

Necrosis, Autofluorescence, other reasons?

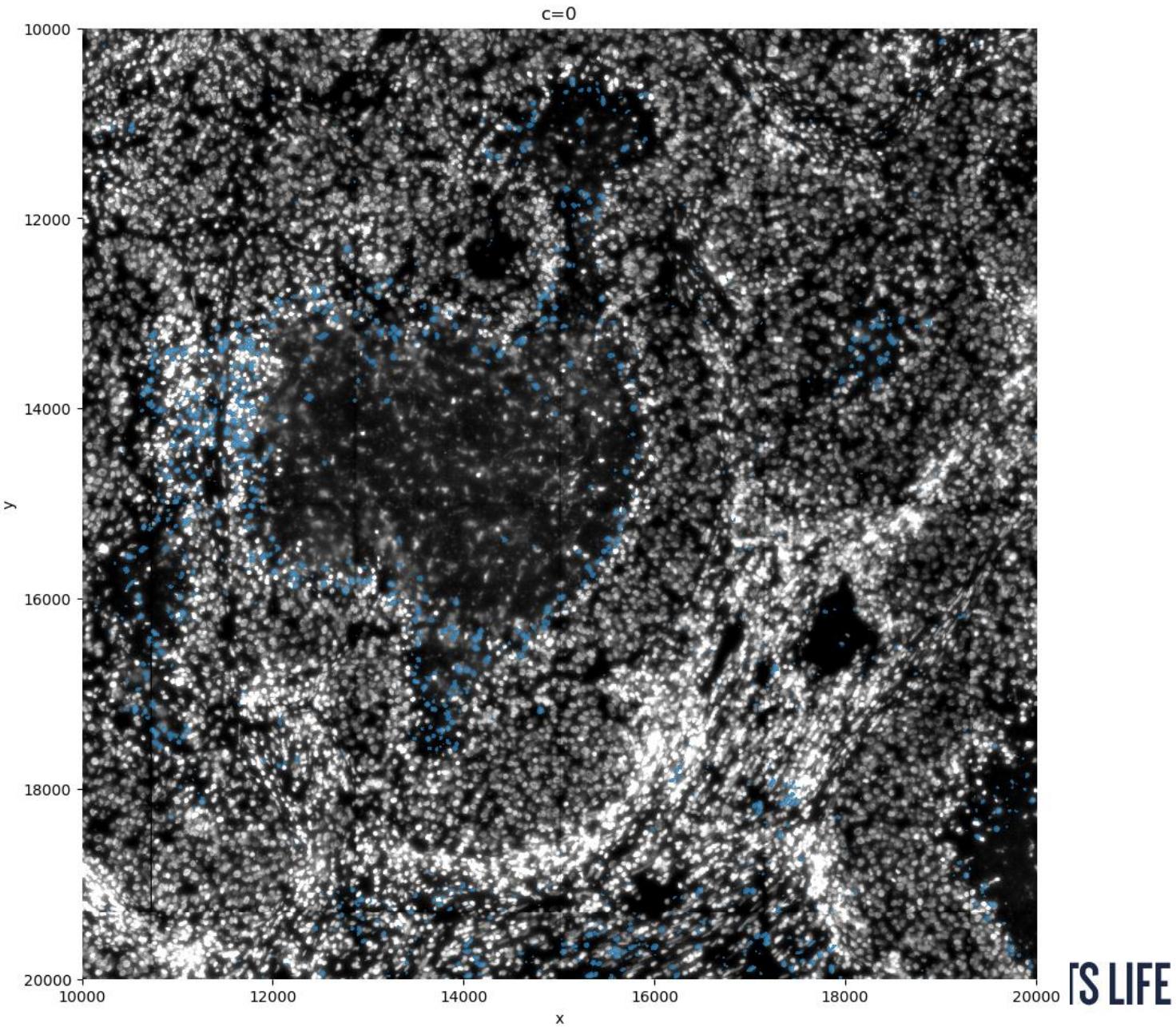
Possible to filter visualize
these cells that are filtered out

Quality control 4



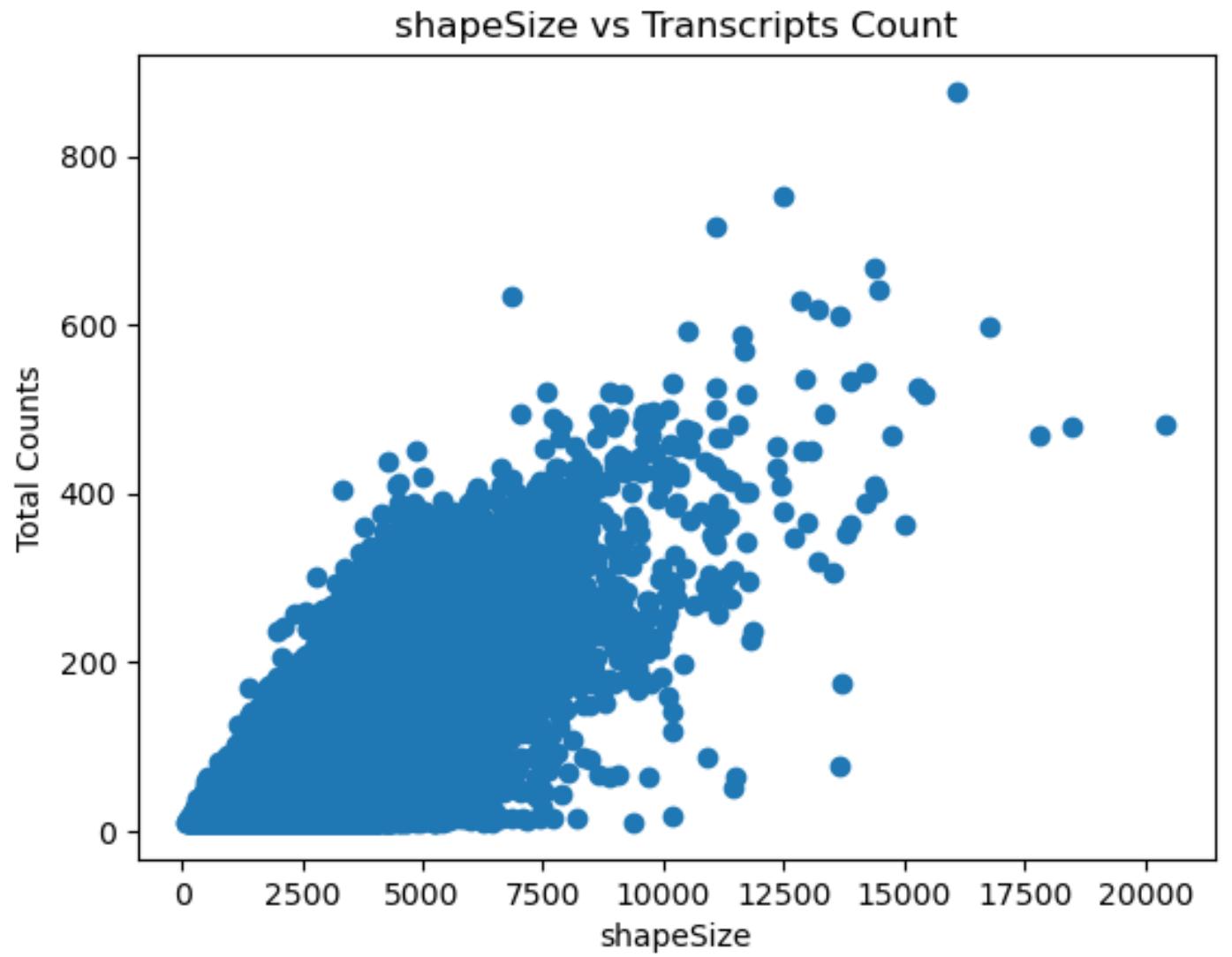
Process the count table:
filter out cells with a gene
count lower than 10.

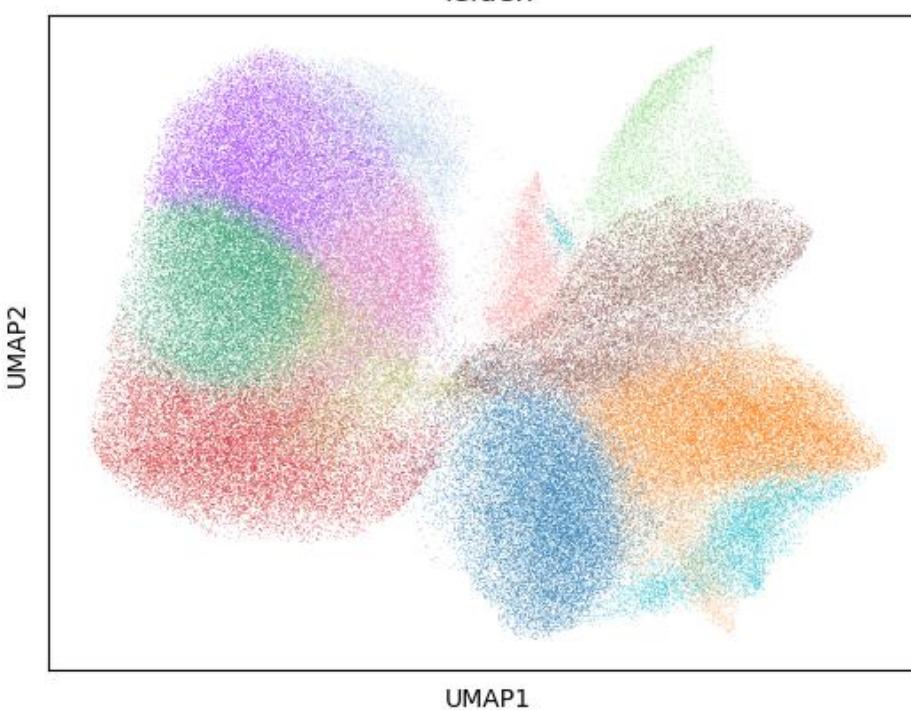
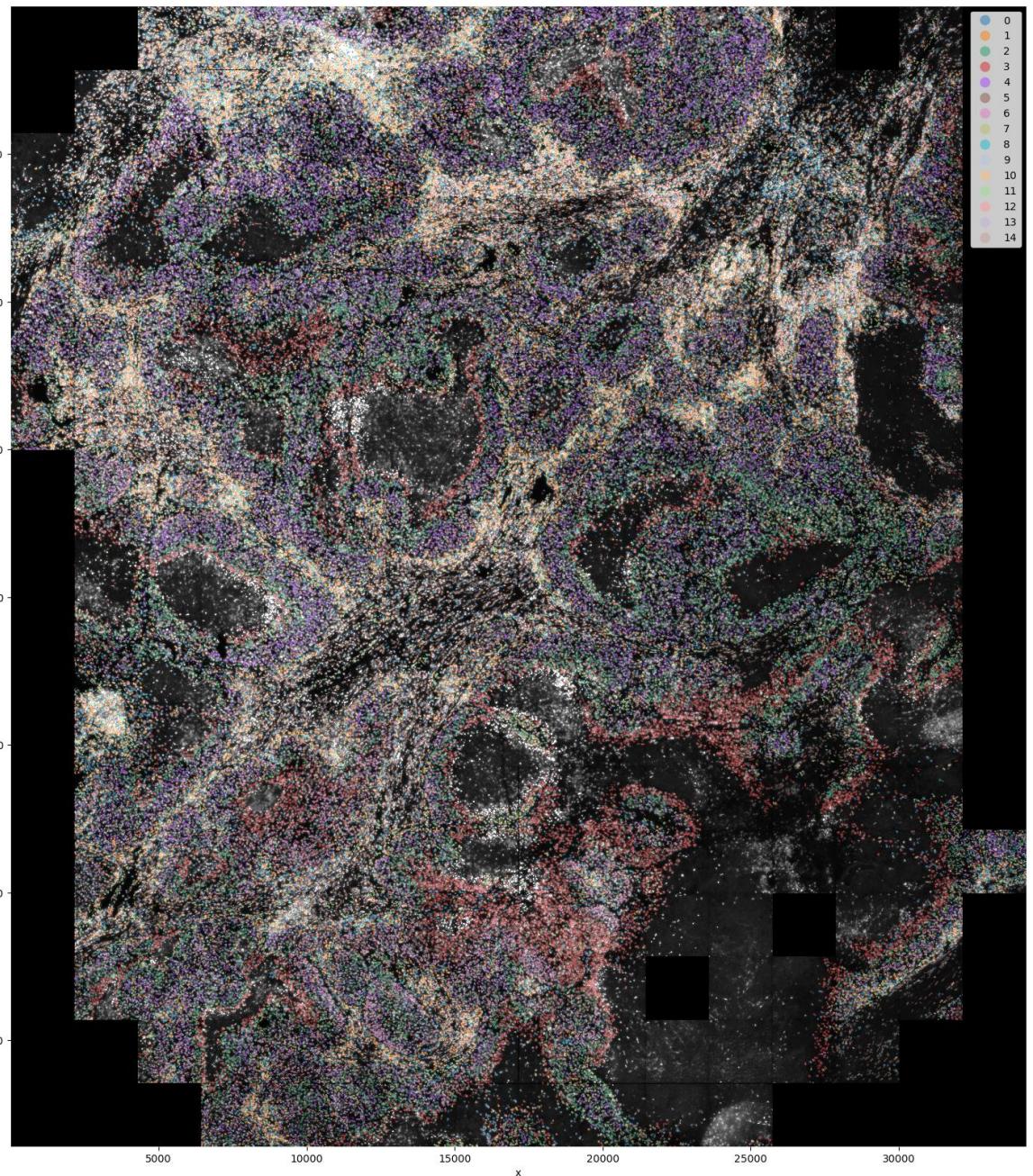
More cells get filtered out,
also possible to visualize

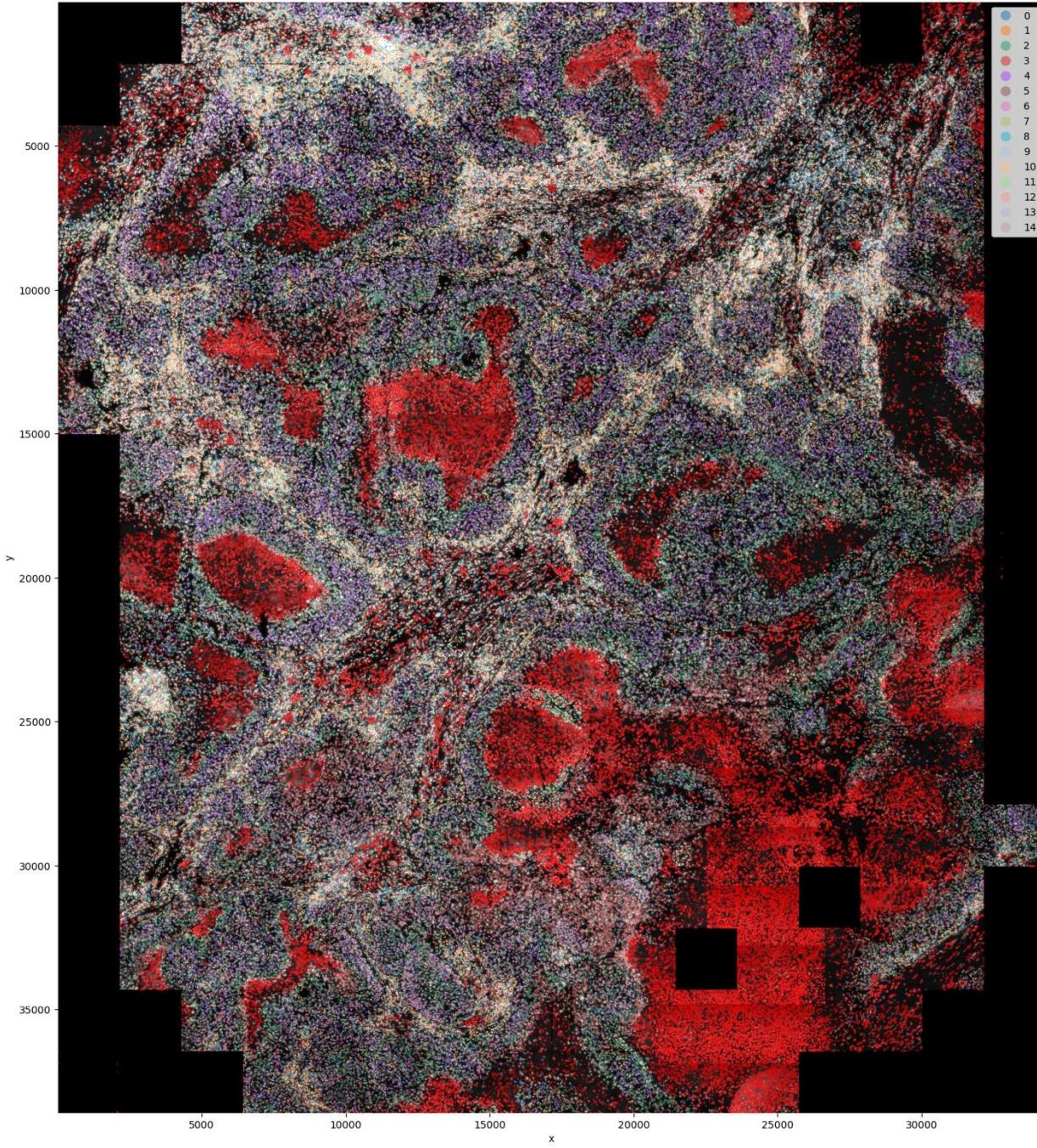


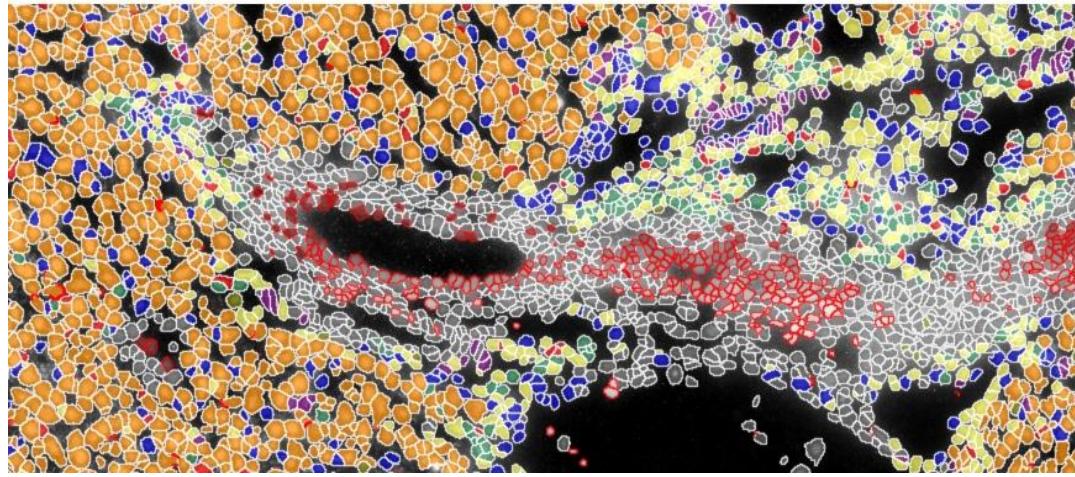
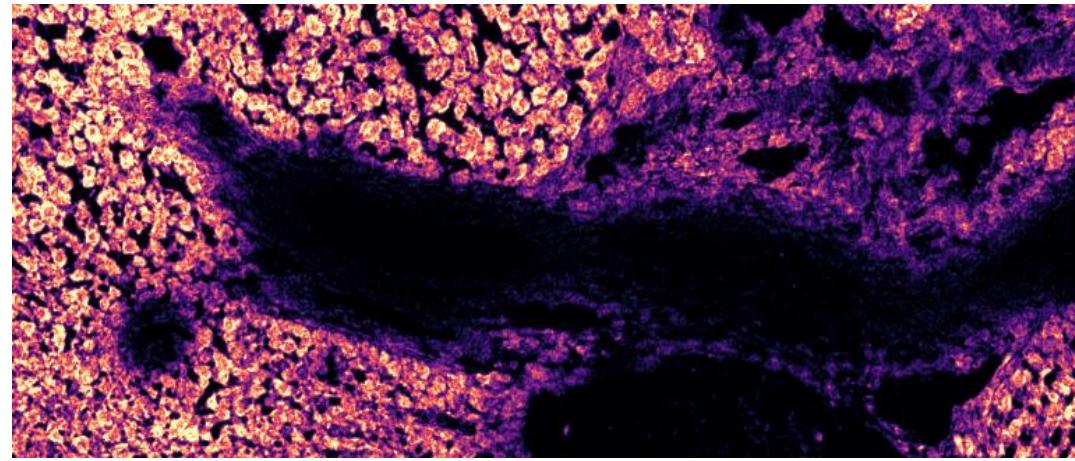
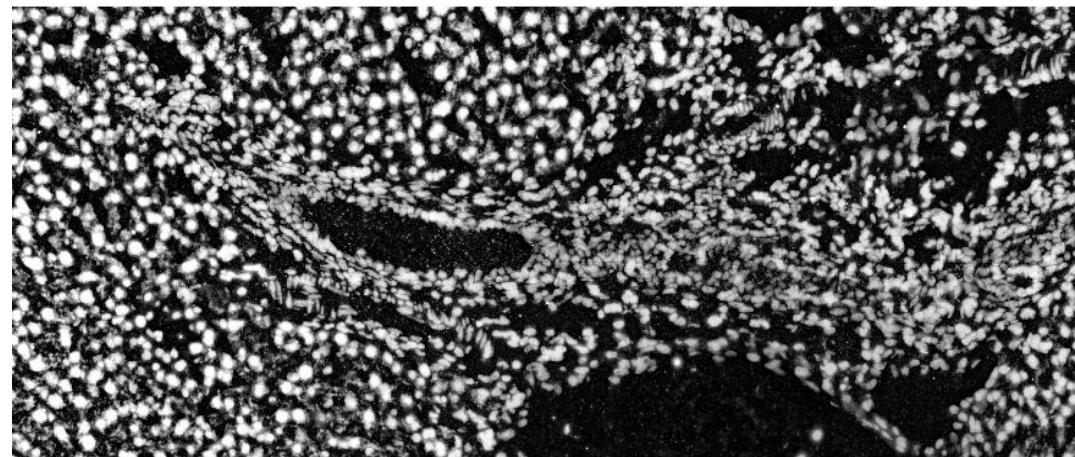
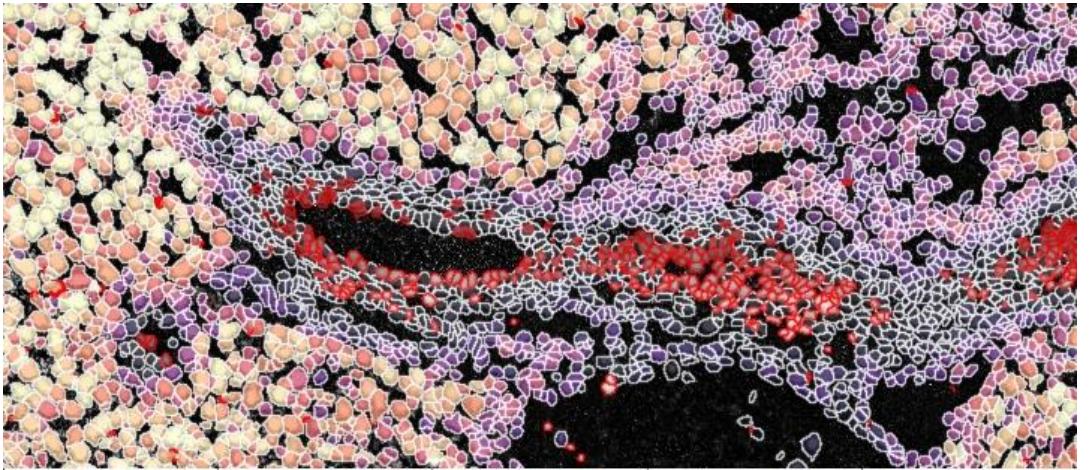
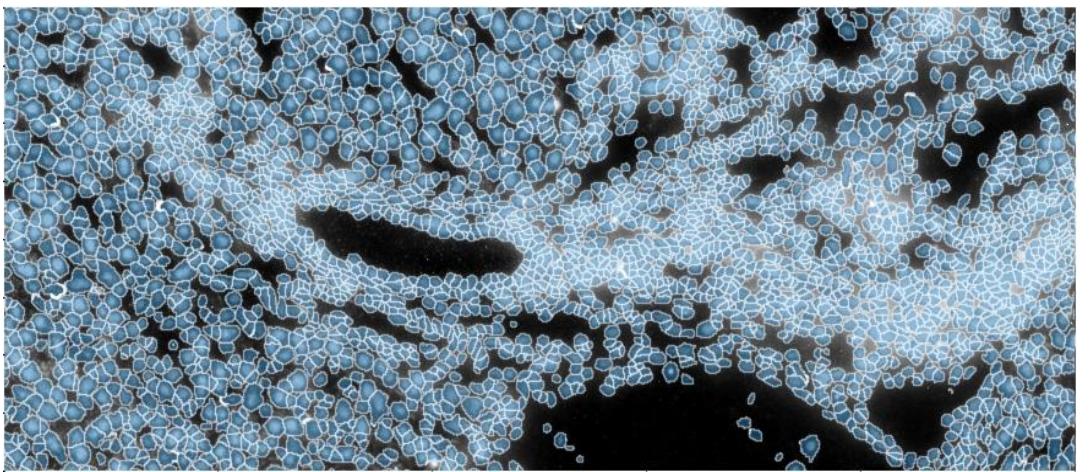
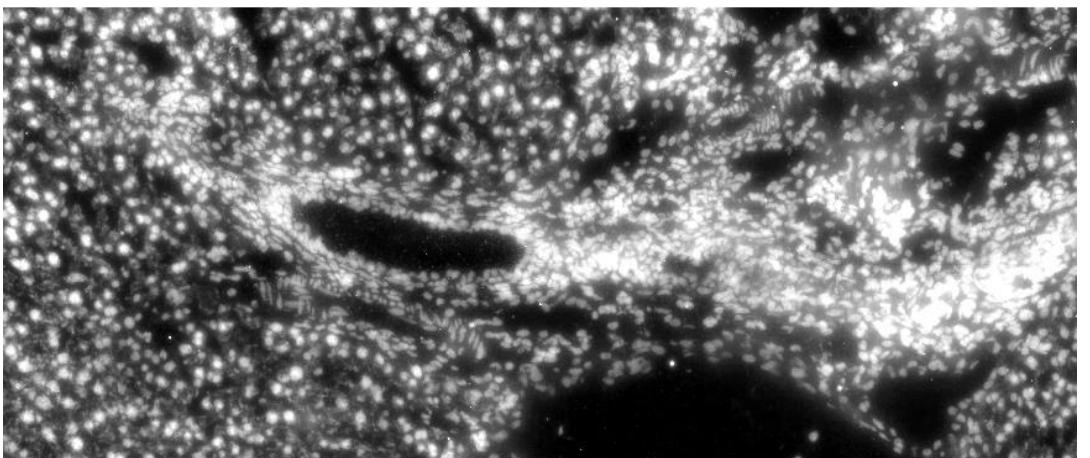
Segmentation isn't perfect:
some cells are too small or
too large and need to be
filtered.

Quality control 5

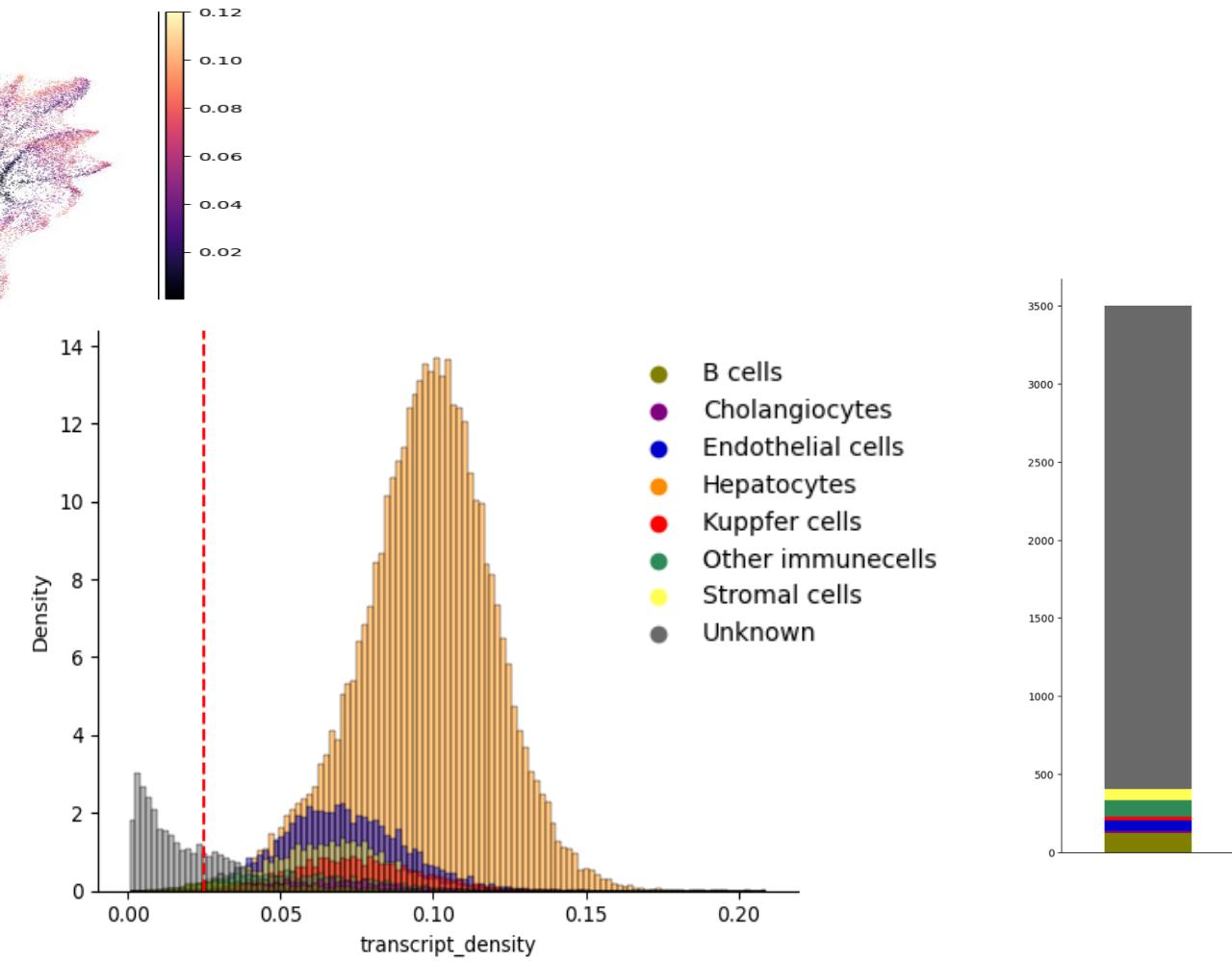
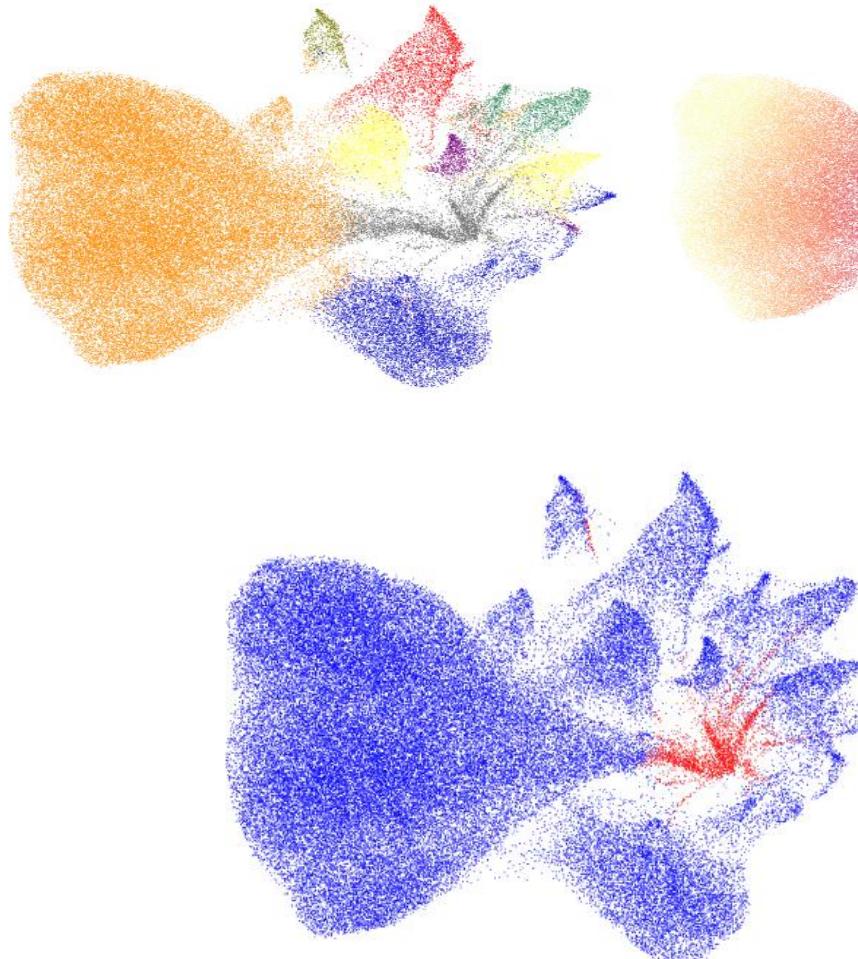




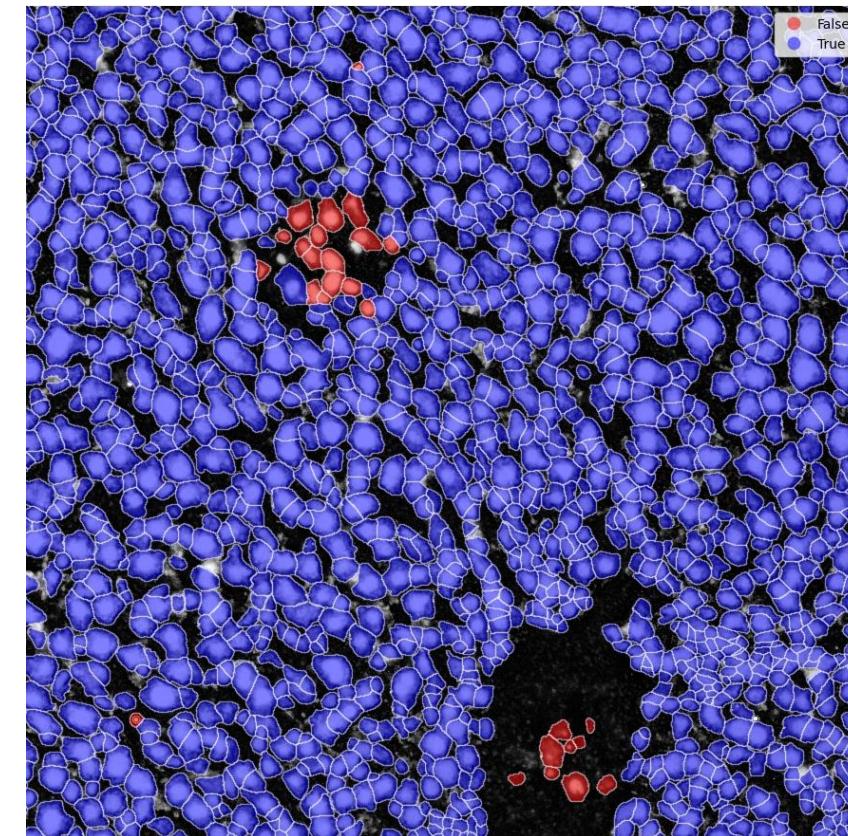
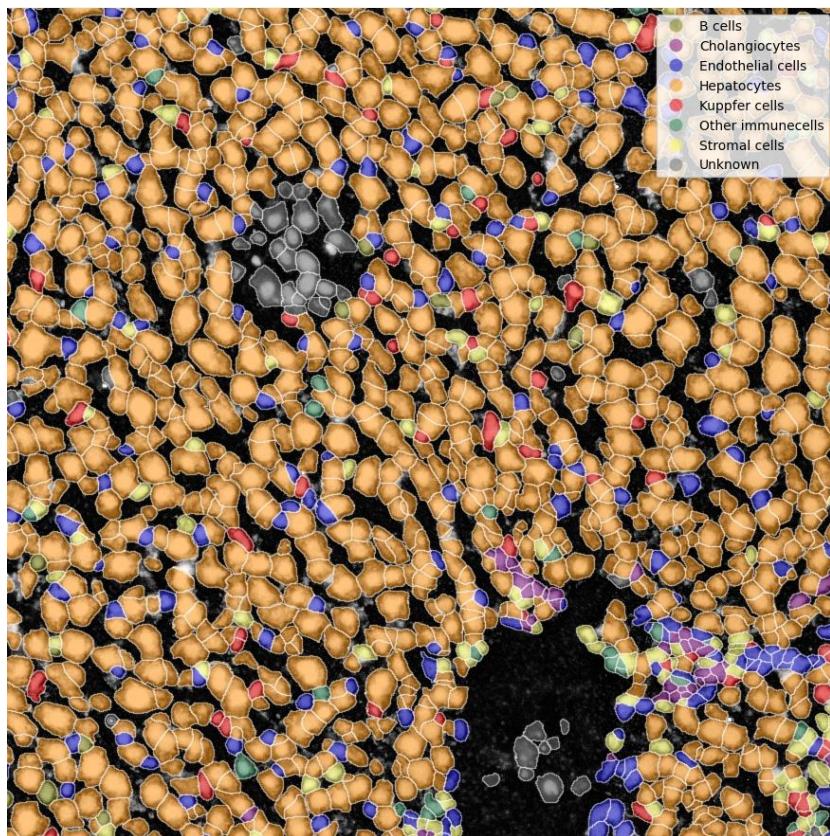




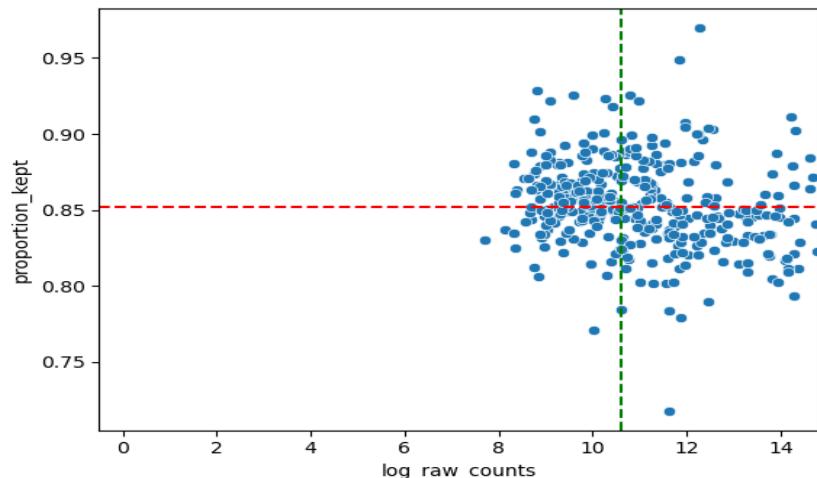
Filter based on transcript density



Filter based on transcript density



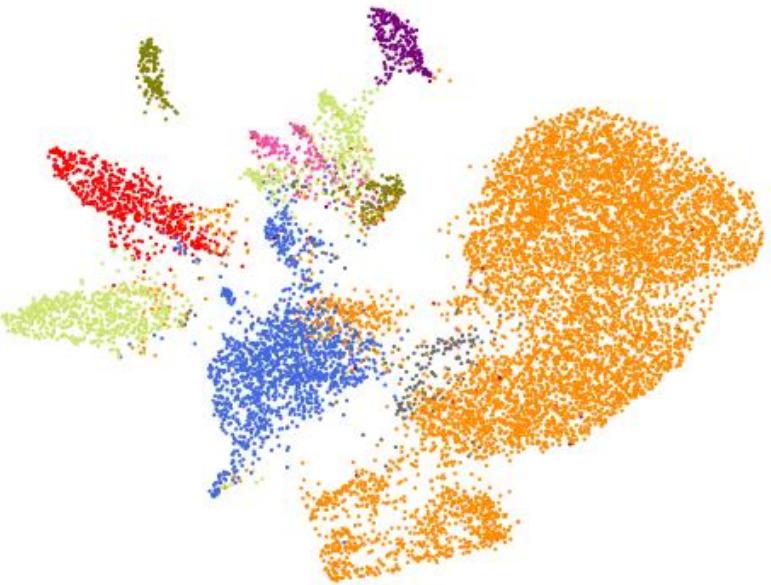
Check filtered genes



gene	proportion_kept	raw_counts
Dpt	0.717878	110555
Nes	0.770773	22144
Alas2	0.778949	144541
Lhfp	0.783647	112201
Sfrp1	0.784296	40092
Sdc3	0.789354	253928
Cald1	0.793574	1616618
Cybb	0.801593	78828
Bpgm	0.801965	104204
Myh10	0.802167	59813

How to annotate Spatial Transcriptomics data?

SPArrOW



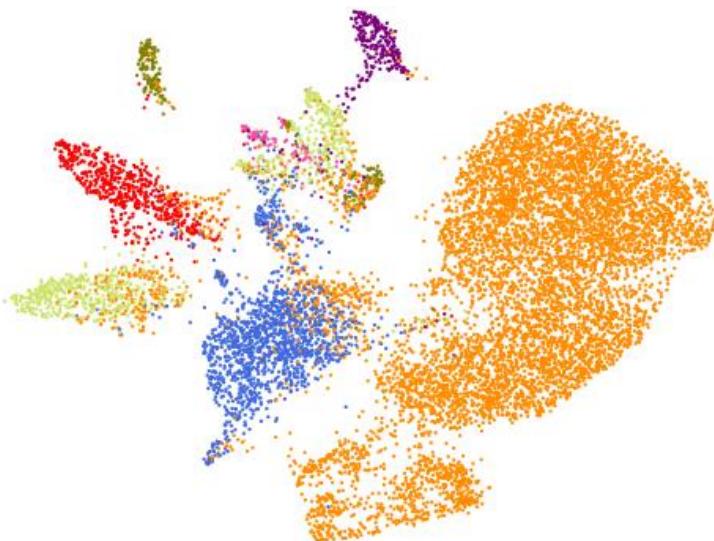
Tacco



NNLS



Score genes



Tangram

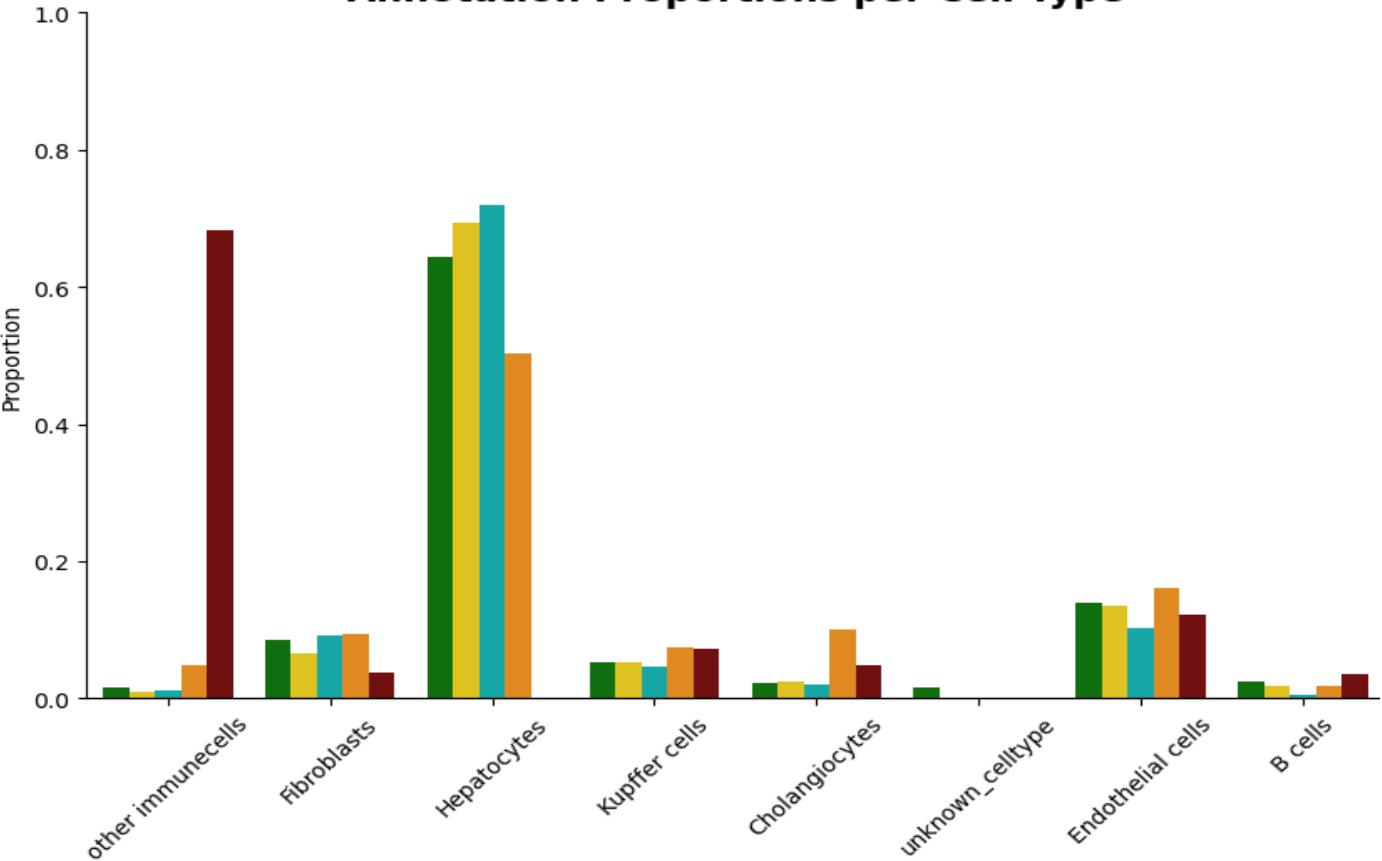


- B cells
- Cholangiocytes
- Hepatocytes
- Kupffer cells
- Endothelial cells
- Other immune cells
- Stromal cells
- Unknown

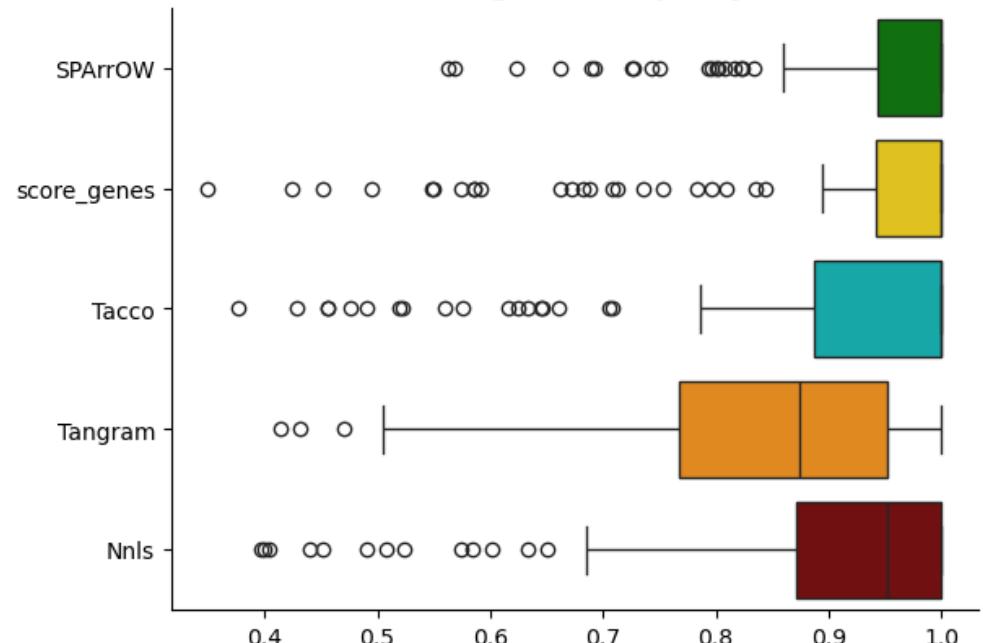


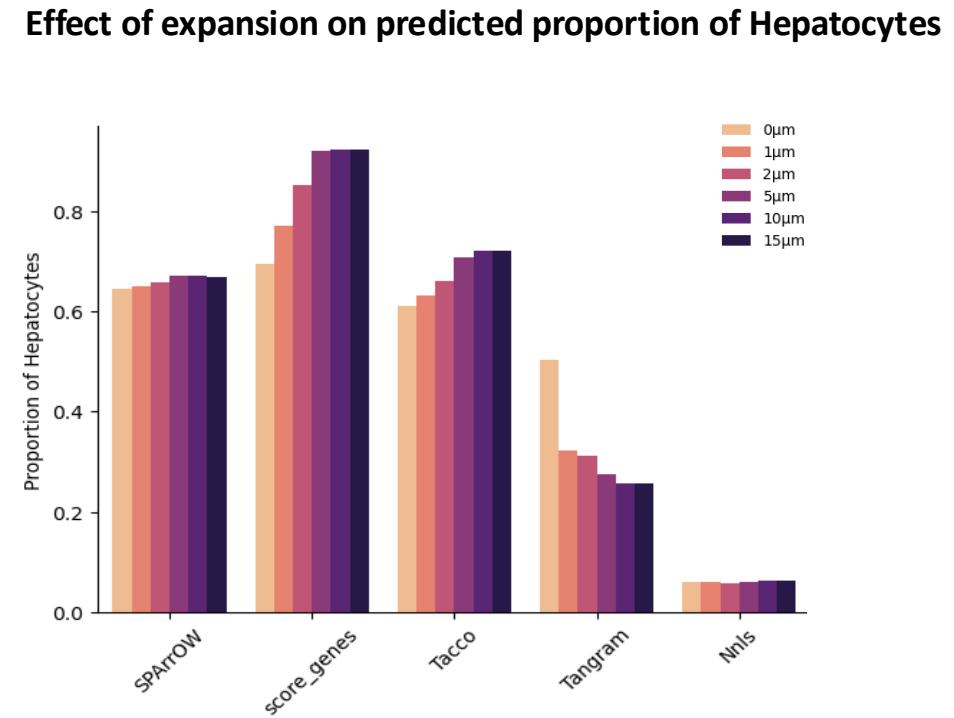
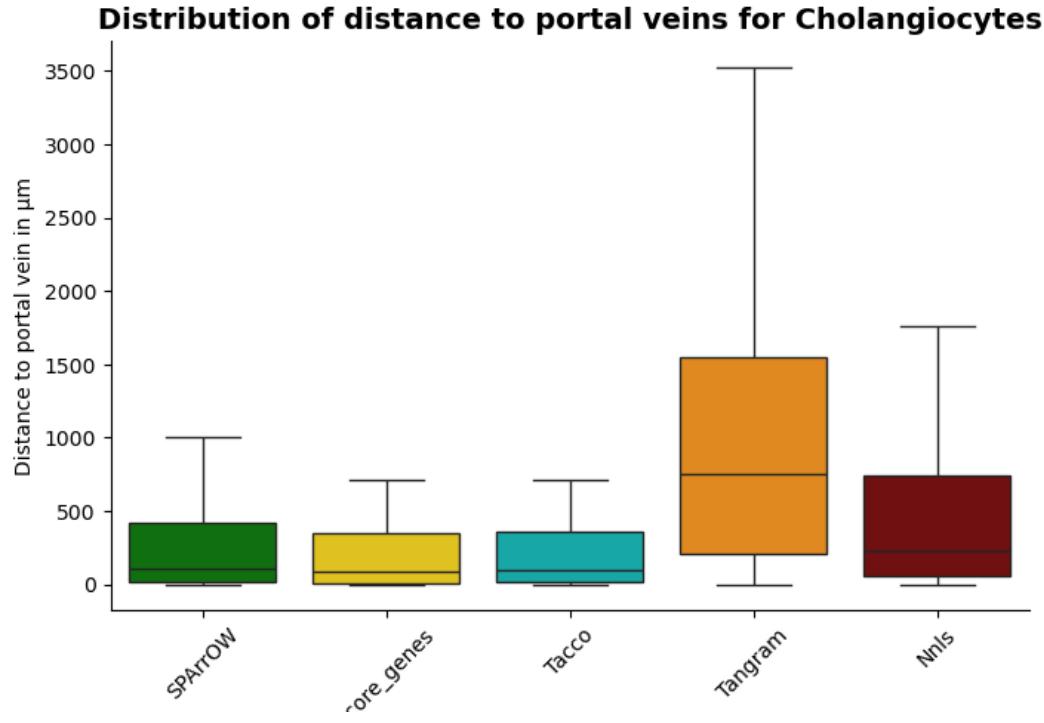
SCIENCE MEETS LIFE

Annotation Proportions per Cell Type

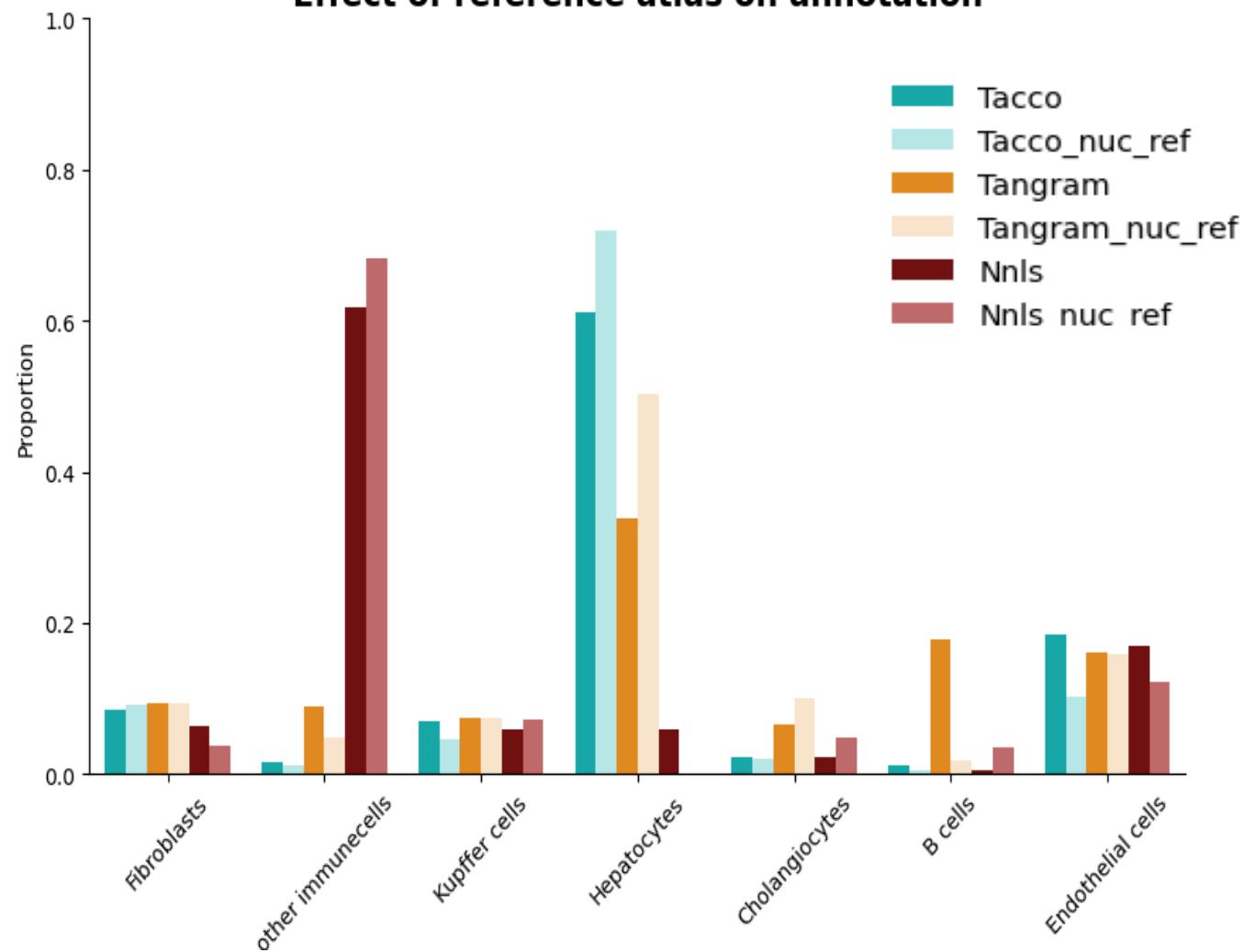


Average Cluster purity





Effect of reference atlas on annotation

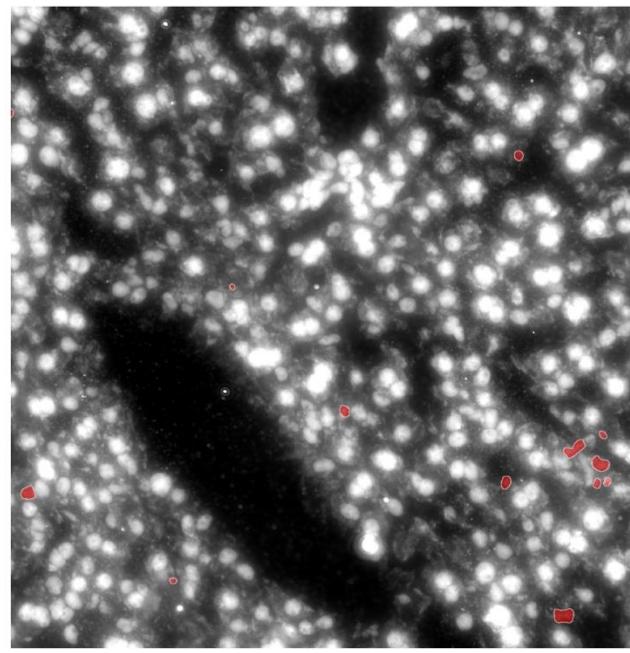
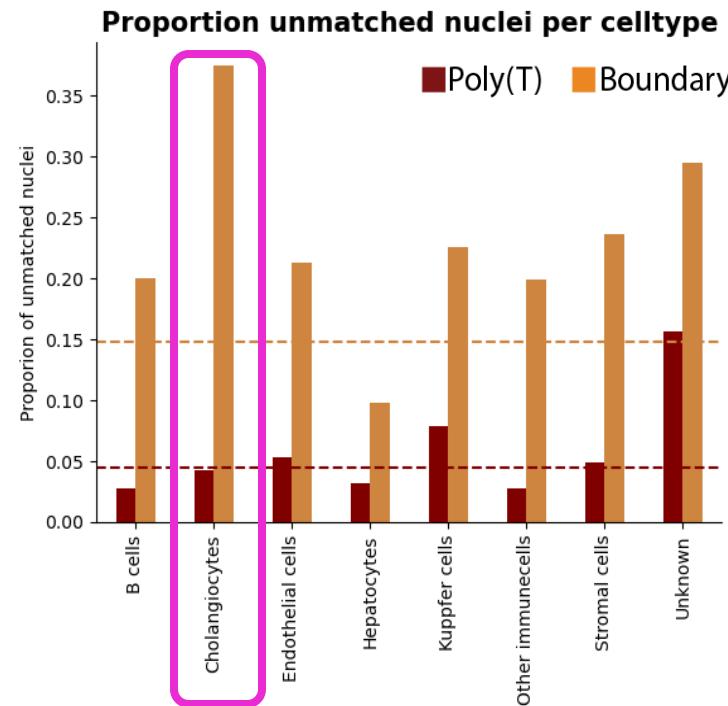


SPArrOW: the effect of different membrane stainings

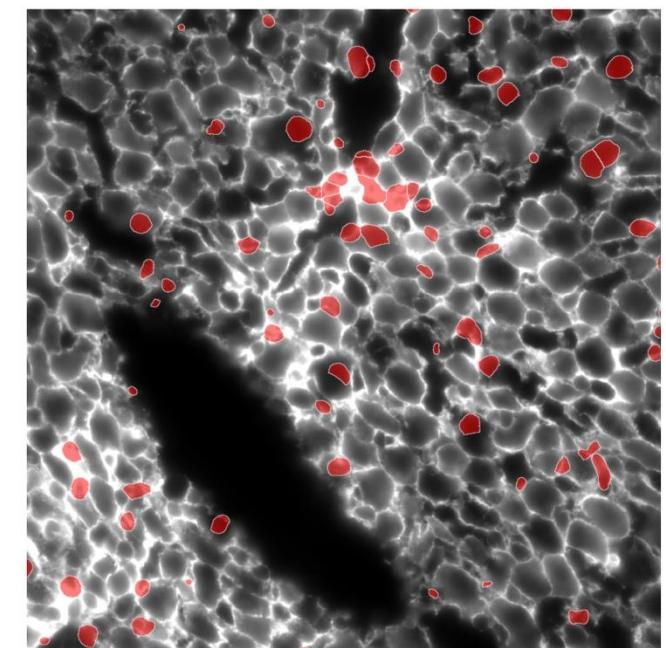
Evaluation of membrane stainings

They might not be as general as you'd want them to be

Celltype specific cell missing per staining

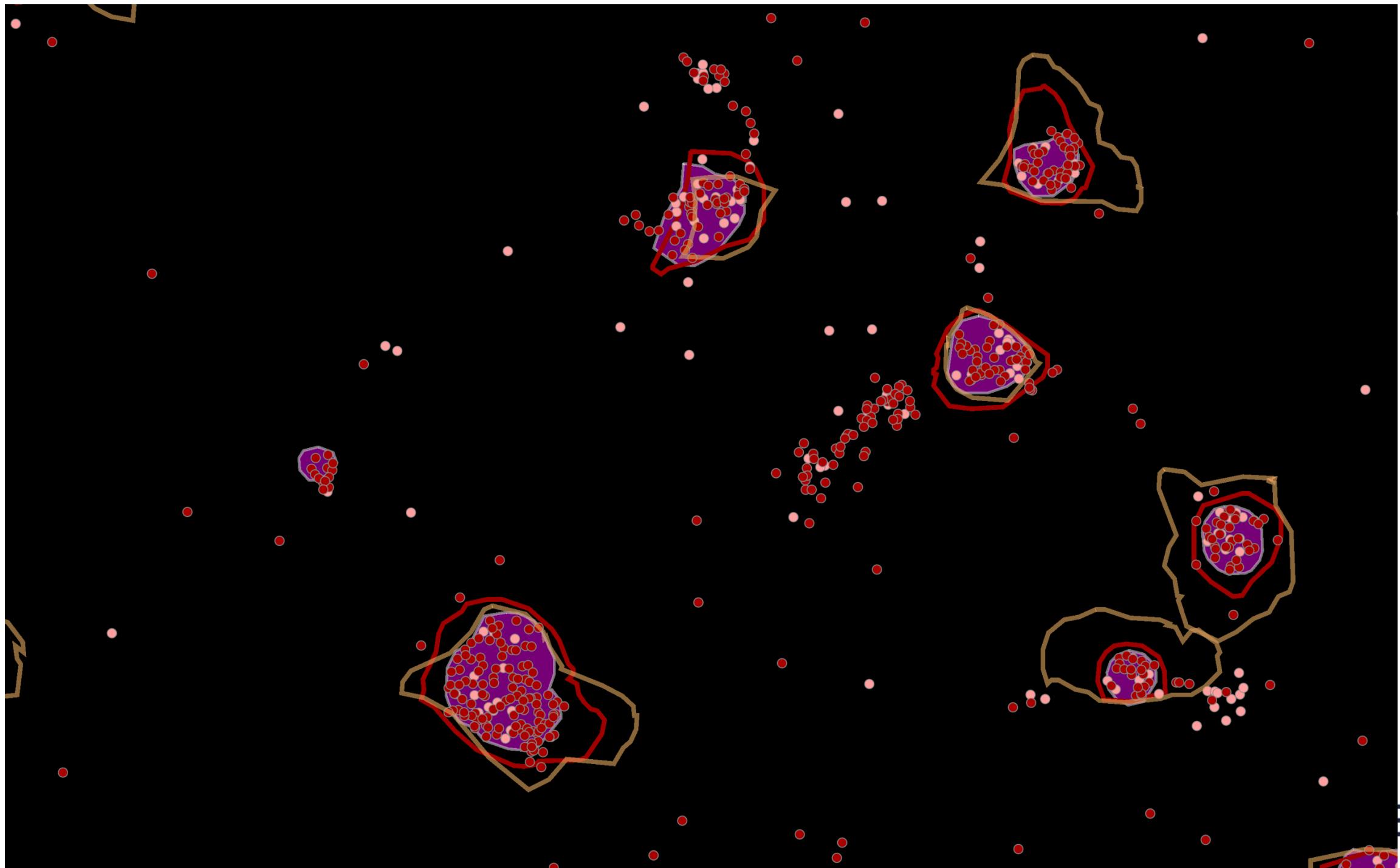


polyT

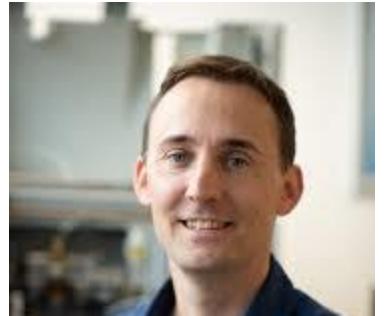


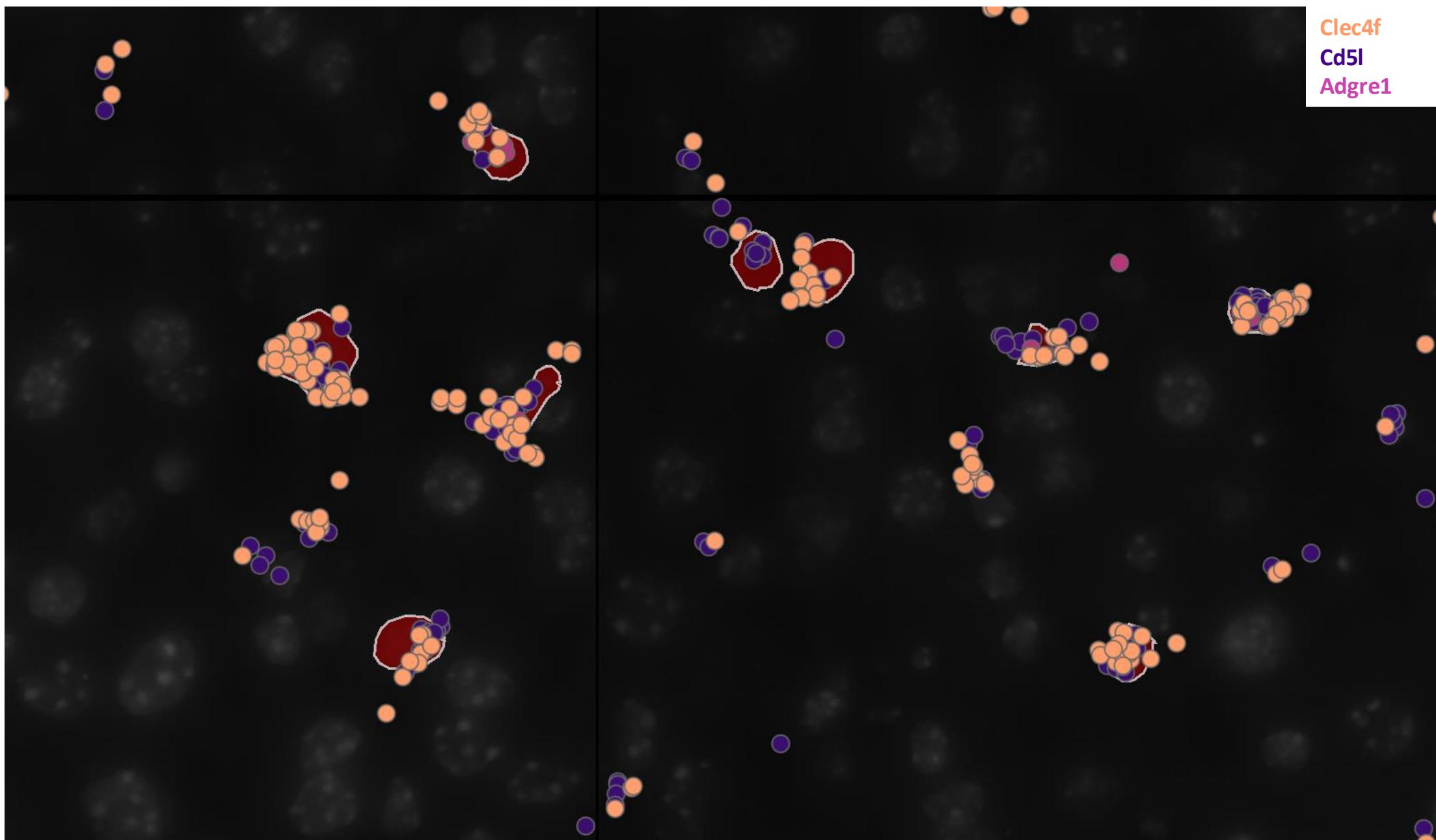
Boundary

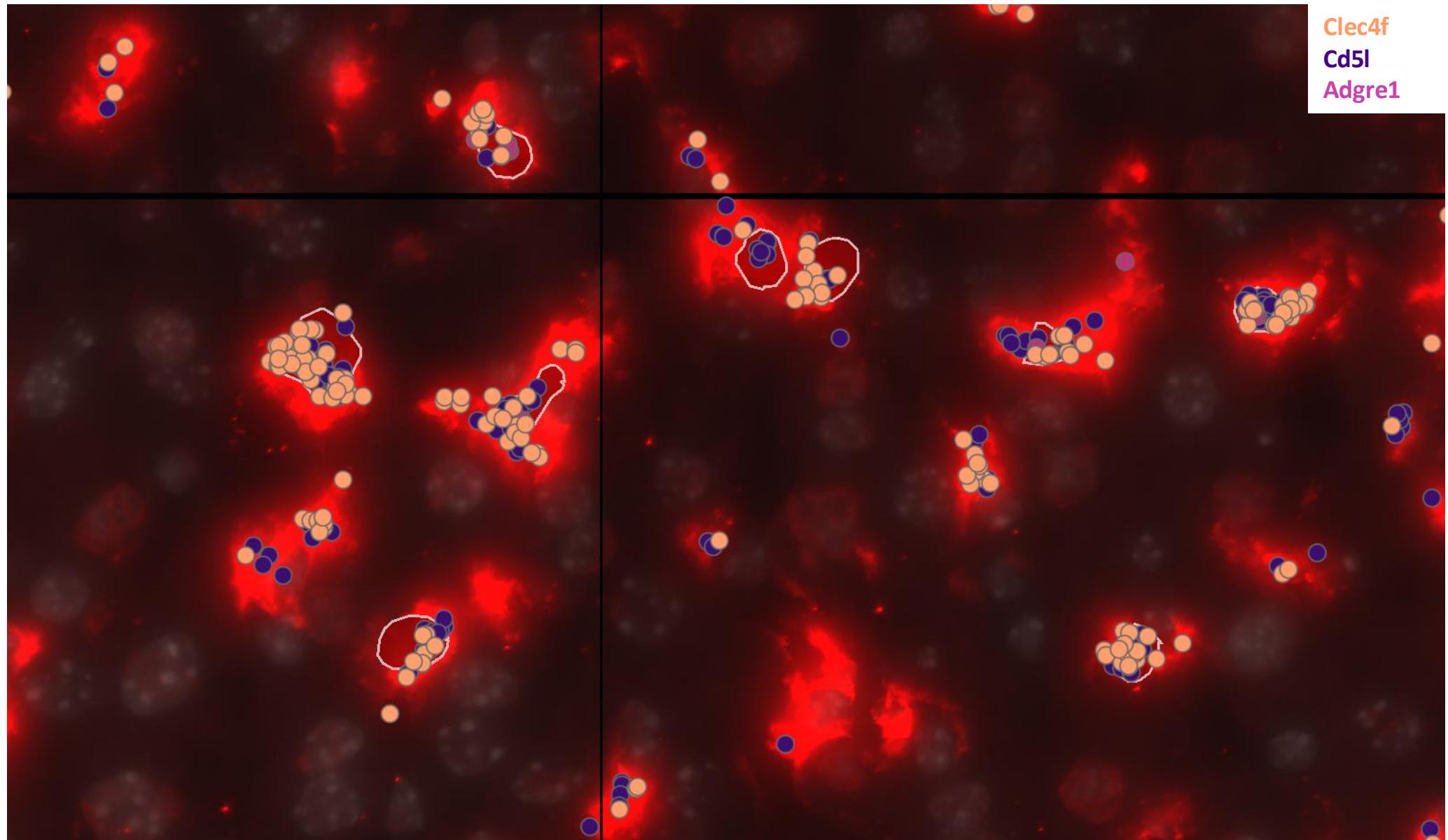
NETS LIFE

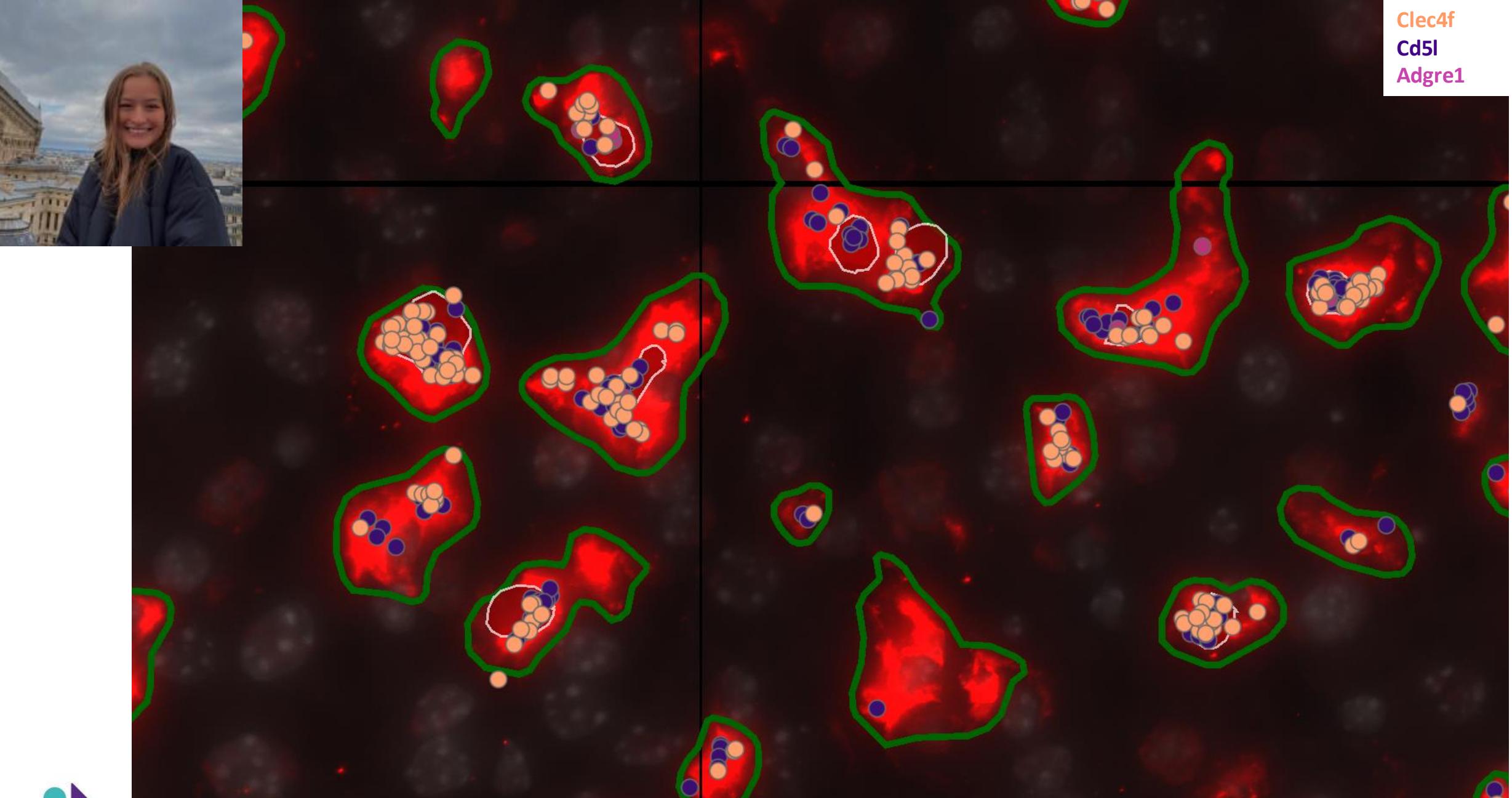


Celltype specific stainings might help!









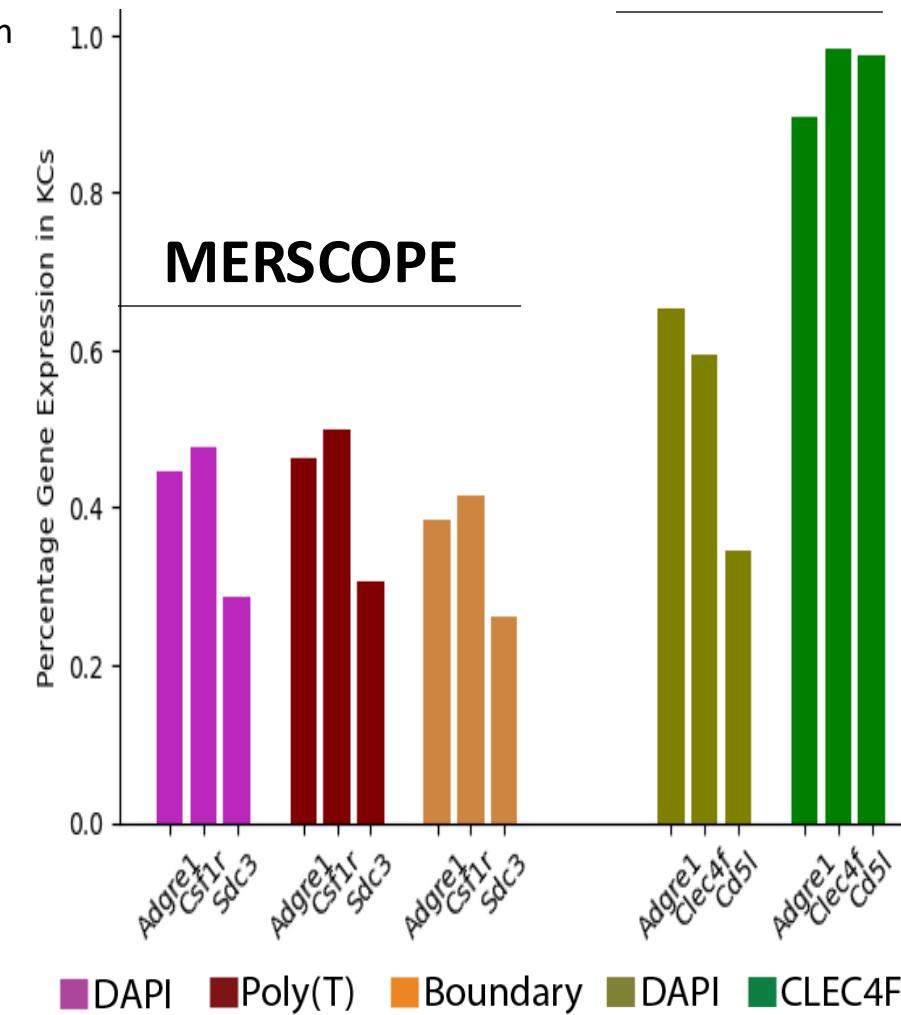
Clec4f

Cd5l

Adgre1

Marker Gene Expression Comparison in KC

Molecular Cartography

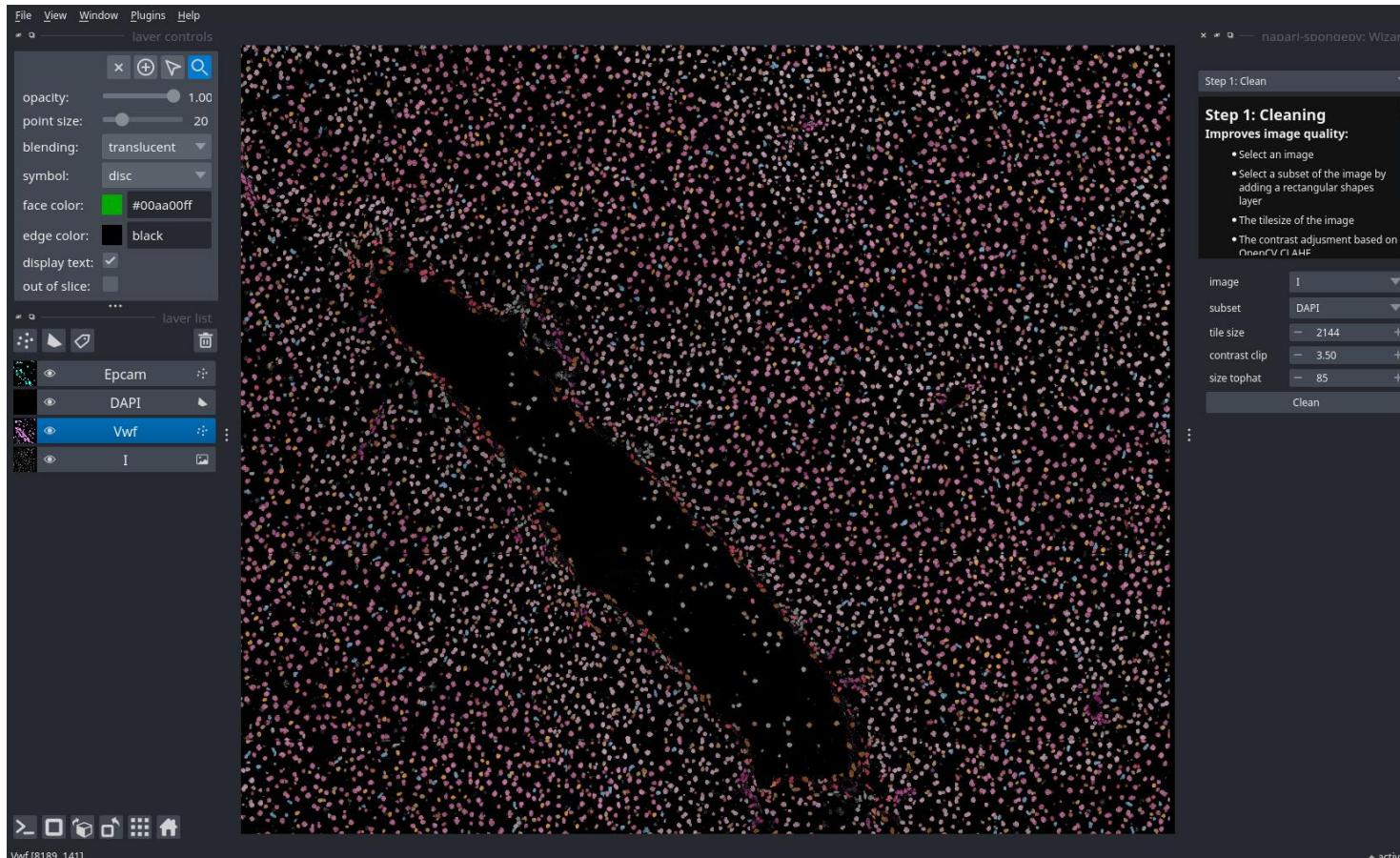


SPArrOW: how to use?

Read information on github page

- 3 ways to interact with SPArrOW
 - ▶ No code in napari
 - ▶ Jupyter notebooks: constant feedback plots
 - ▶ Commandline tool using Hydra

Interactive parameter tuning in napari



Visual, direct feedback, but
very slow on large datasets:
subset selection possible

Less flexible

Note: you can run napari
from notebooks too!

Jupyter notebooks

Similar to R markdowns

Tutorials with explanation and code that you can adapt to your data

Basic usage should be fine for everyone with a bit of coding experience in any language

Direct plotting feedback

Flexible: you can add your own code/ fit in extra steps

Easier to for example run on GPU

How I analyse the datasets

Start with
script_liver_sparrow_generalized_update.ipynb in experiments

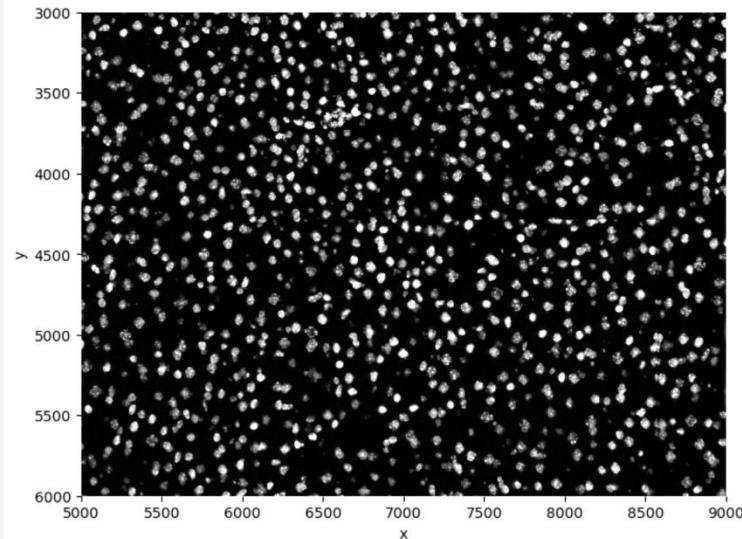
The second step of the preprocessing the data includes a couple of steps:

- RESOLVE has missing data on the borders of two tiles, resulting in black lines on the border between tiles. This can split up cells, and thus influence the analysis. inpainting this black line circumvents this
- Secondly, a min max filter can be added, by defining the size of the min max filter. If the size of the min max filter is not defined, no min max filter is added. The goal of this function is to subtract background noise, and make the borders of the nuclei cleaner, plus it will delete the occasional debris. If you take the size too small, smaller than the size of your nuclei, the function will create donuts, with black spots in the center of your cells. If the size of the min max filter is chosen too big, not enough background is subtracted, so a tradeoff should be made. This might need some finetuning. More information can be found on https://biapol.github.io/blog/ryan_savill/03_background_subtraction/. For nuclei, 45-55 is a great starting point. Bigger for whole cells
- Thirdly, this function enhances the contrast, meaning it makes the white and black more separate. It does this by using histogram equalization (CLAHE function). The amount of correction needed can be decided by adapting the contrast_clip value. If the image is already quite bright, 3.5 might be a good starting value. For dark images, you can go up to 10 or even more.

The result of this function is a cleaned up images, that will give rise to better segmentation results.

```
sdata = sp.im.min_max_filtering( sdata=sdata, output_layer='min_max_filtered' , size_min_max_filter= 45 )
sp.pl.plot_image( sdata, img_layer='min_max_filtered', crd=crd, figsize=( 8,8 ) )
sdata = sp.im.enhance_contrast( sdata=sdata, output_layer='clahe', contrast_clip=3.5, chunks=20000 )
sp.pl.plot_image( sdata, img_layer='clahe', crd=crd, figsize=( 8,8 ) )
```

Python



Command line tool with hydra

- After you optimized parameters to run on whole datasets
- Job based on work stations: GPU clusters or the HPC
- Config files with all the right parameters
- For help, contact the spatial catalyst after reding the github

Acknowledgments

SaeysLab

Lotte Polaris
Ruth Seurinck
Karen Herreman
Benjamin Rombaut
...
Yvan Saeys

VIB Spatial Catalyst

Arne Defauw
Frank Vernaillen
Julien Mortier
Evelien Van Hamme

Guiliams/Scott lab

Bavo Vanneste
Tinne Thoné
Liesbet Martens
Wouter Saelens
Anja Bujko
Jean-François Hastier
Charlotte Scott
Martin Guiliams

Collaborators

Hilde Nelissen lab
Jean-Christophe Marine lab
SpatialData consortium



Questions?

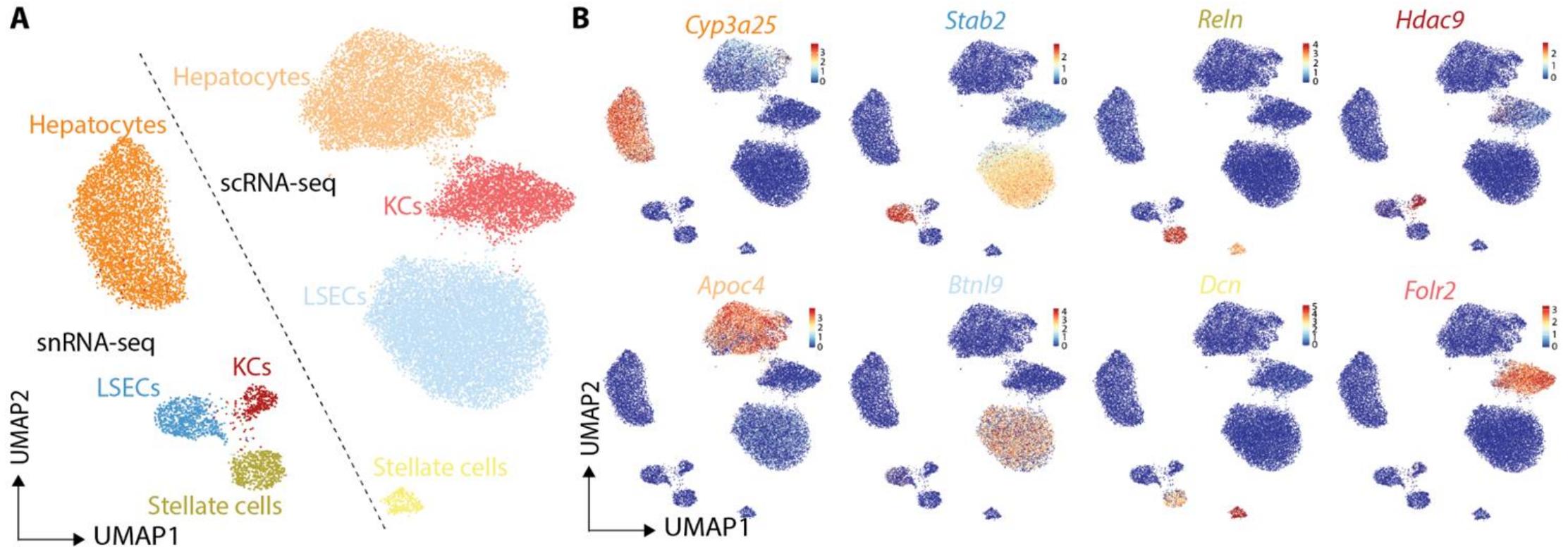
Lotte.polaris@UGent.be

or contact the **VIB Spatial Catalyst**

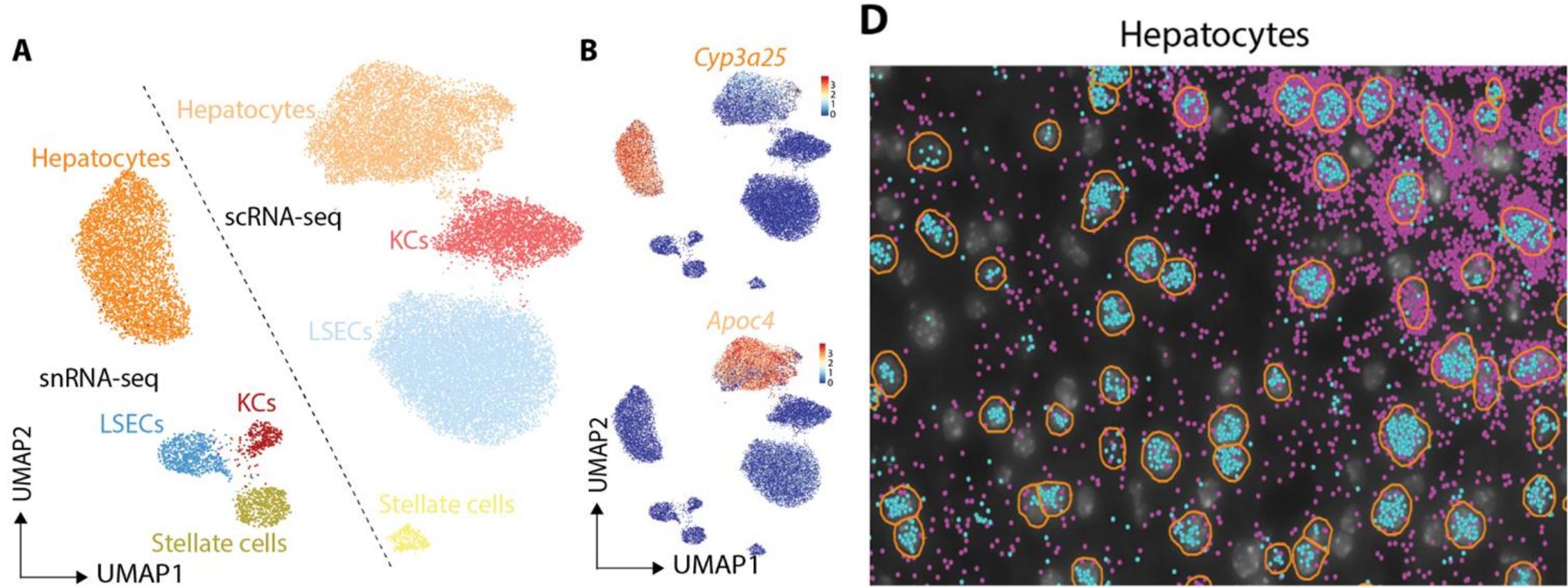
spatial@vib.be

SPArrOW: a Usecase

mRNA molecules are not equally distributed within cells

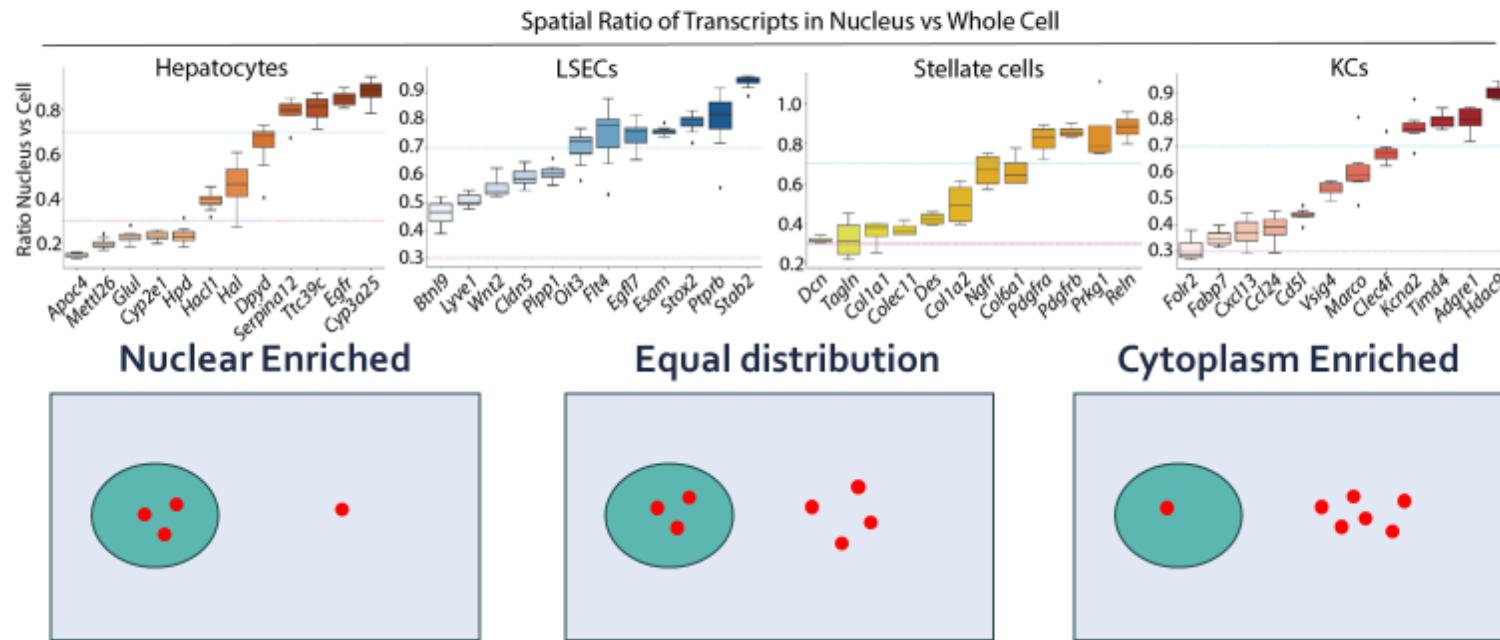


mRNA molecules are not equally distributed within cells



mRNA molecules are not equally distributed

mRNA molecules are not equally distributed





science meets life