Spatial transcriptomics

Thomas D. Otto

Disclaimer

- Slides from many people
- Domenico Somma – GeoMX / Spatial example
- Joy Kabagenyi – Visium / Xenium
- Mahdi Marashi – Comparison

- First time lecture, might be a bit bumpy

Indented learning outcome

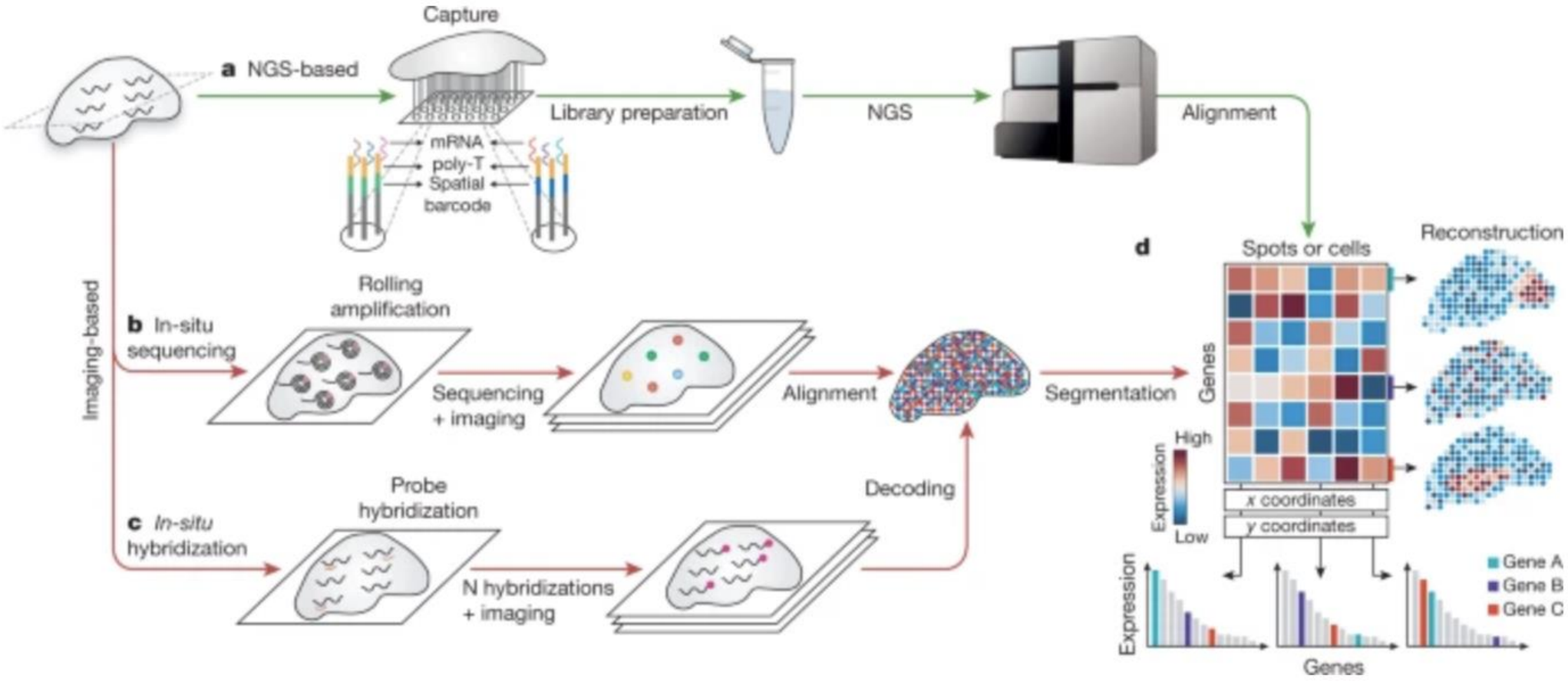
- Critical understanding of spatial biology methods and their need
- Understand limitation

Spatial transcriptomics

- cellular organization in tissues is intimately linked to biological function
- histopathology is often used as a conclusive diagnostic tool, precisely because many diseases are characterized by abnormal spatial organization within tissues
- Infectious and inflammatory processes can drastically change the cellular organization of tissues

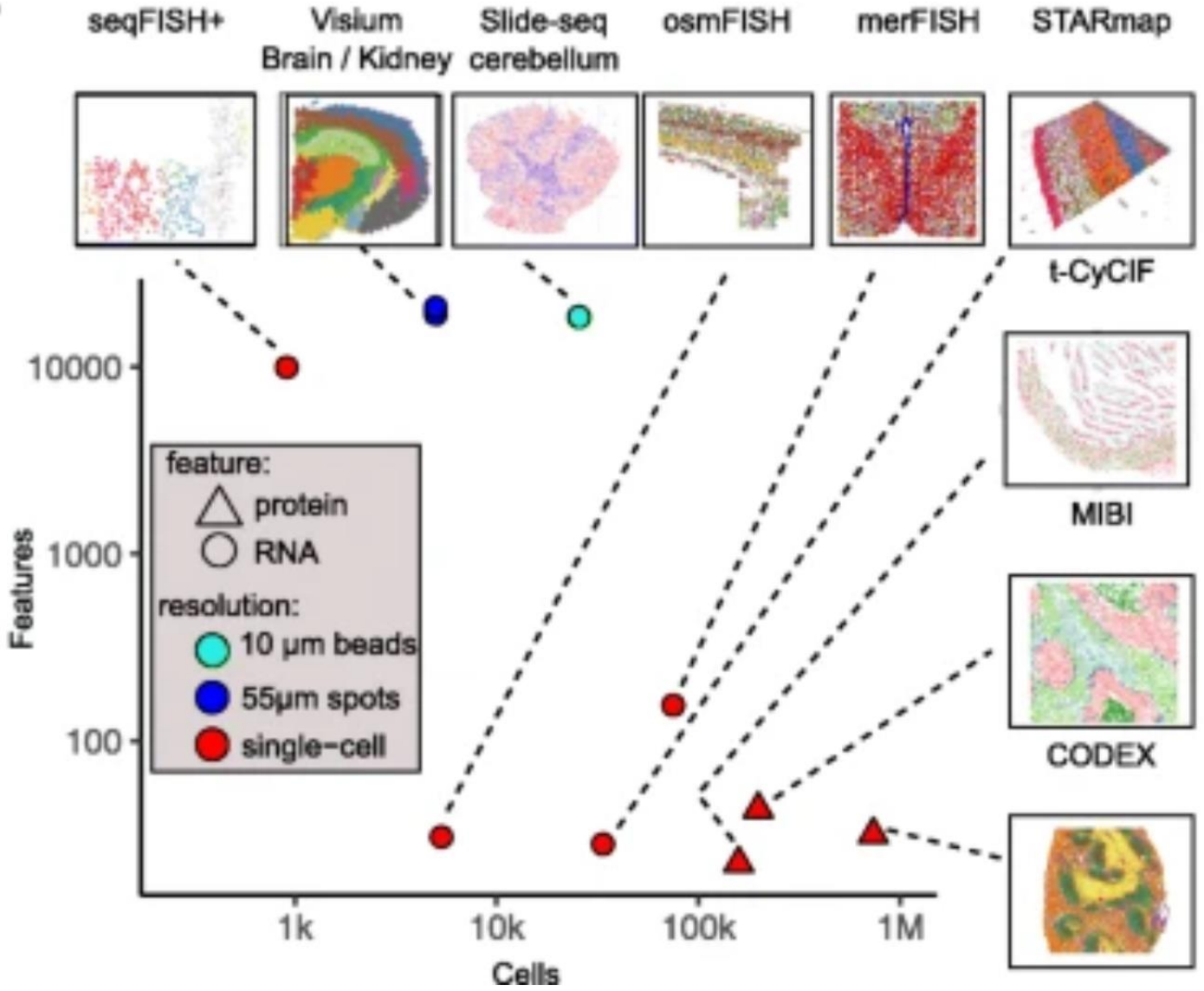
Spatial transcriptomics

Fig. 1: The technologies of spatial transcriptomics provide a gene-expression matrix.



https://www.nature.com/articles/s41586-021-03634-9

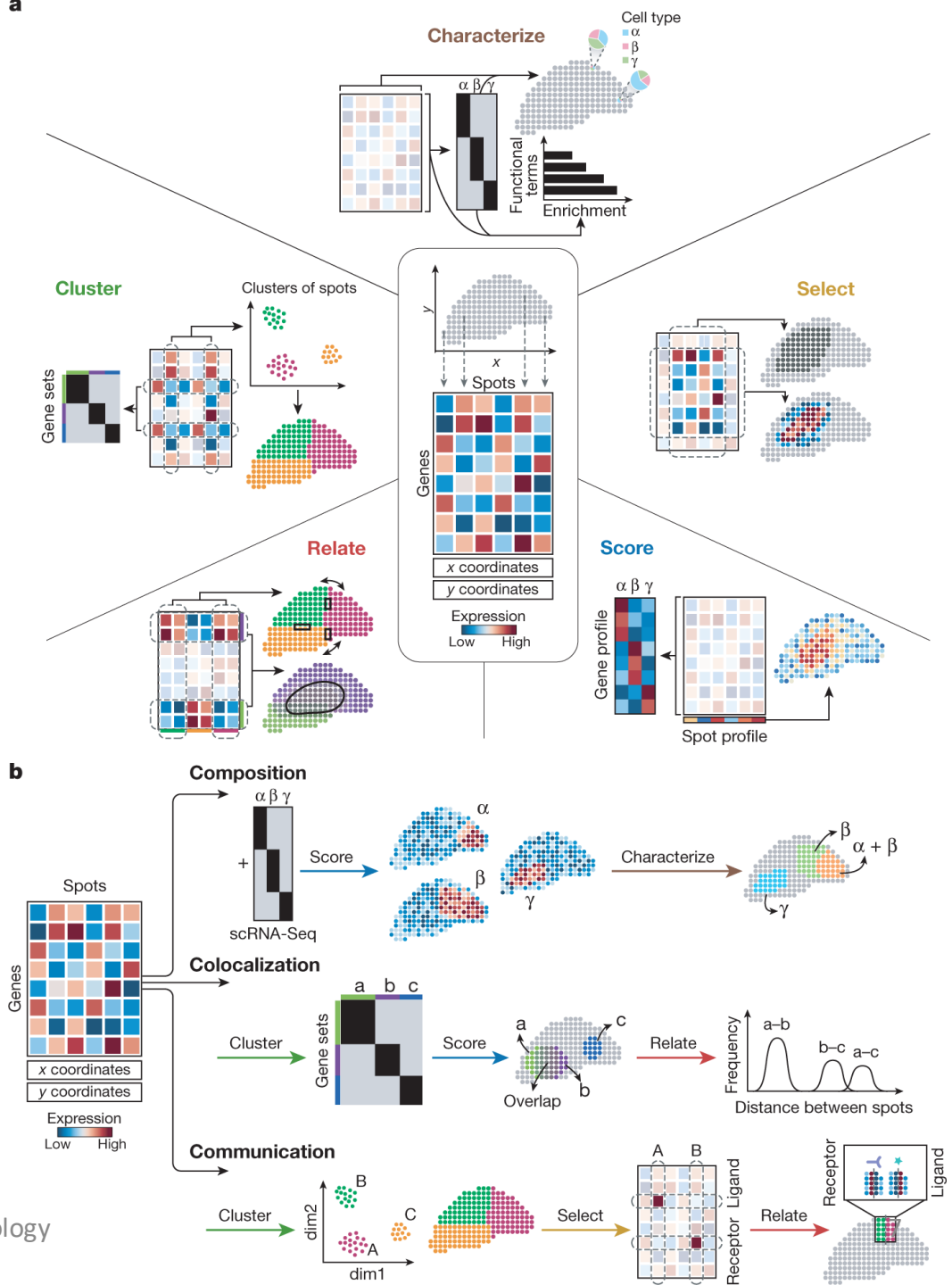
Outlook: Spatial transcriptomics



GiOtto paper

Application

- Spatial data into 2D matrix
- Each row combination of expression of several genes
- Not single cell yet -> next slide
- Visium: ~ 50µm – not good for highly heterogenous tissue: https://www.youtube.com/watch?v=gxvsYeAha50



https://www.nature.com/articles/s41586-021-03634-9/

GeoMx - nanostring

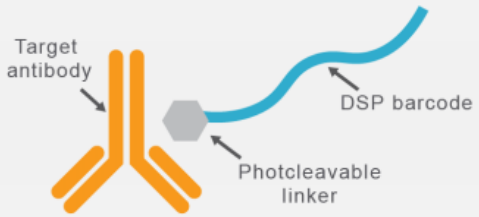
Single cell spatial transcriptomic/proteomic…not really single cell

following slides from Domenico Somma

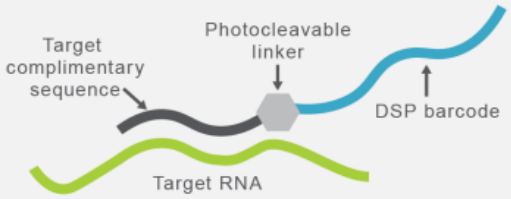
Principle

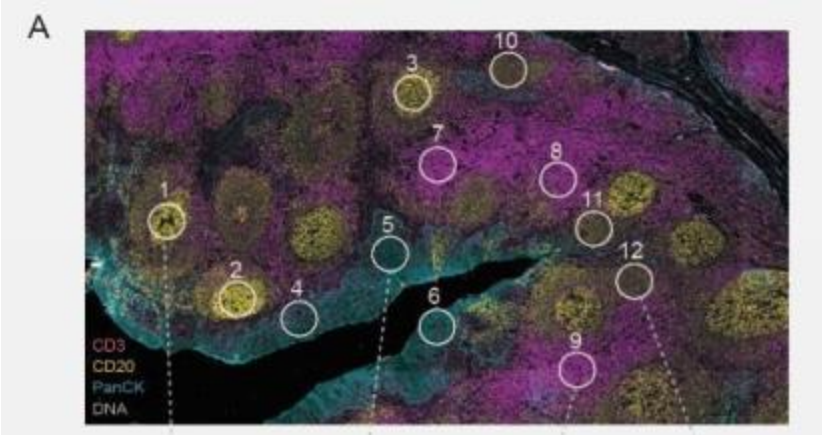
High-Plex Mixtures of Proprietary Reagents

Protein reagents
(UV cleavable linker)



RNA reagents
(UV cleavable linker)

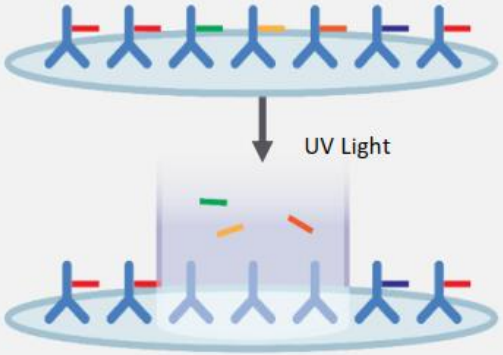




Profile Regions of Interest on FFPE slide

GeoMx™ Digital Spatial Profiler

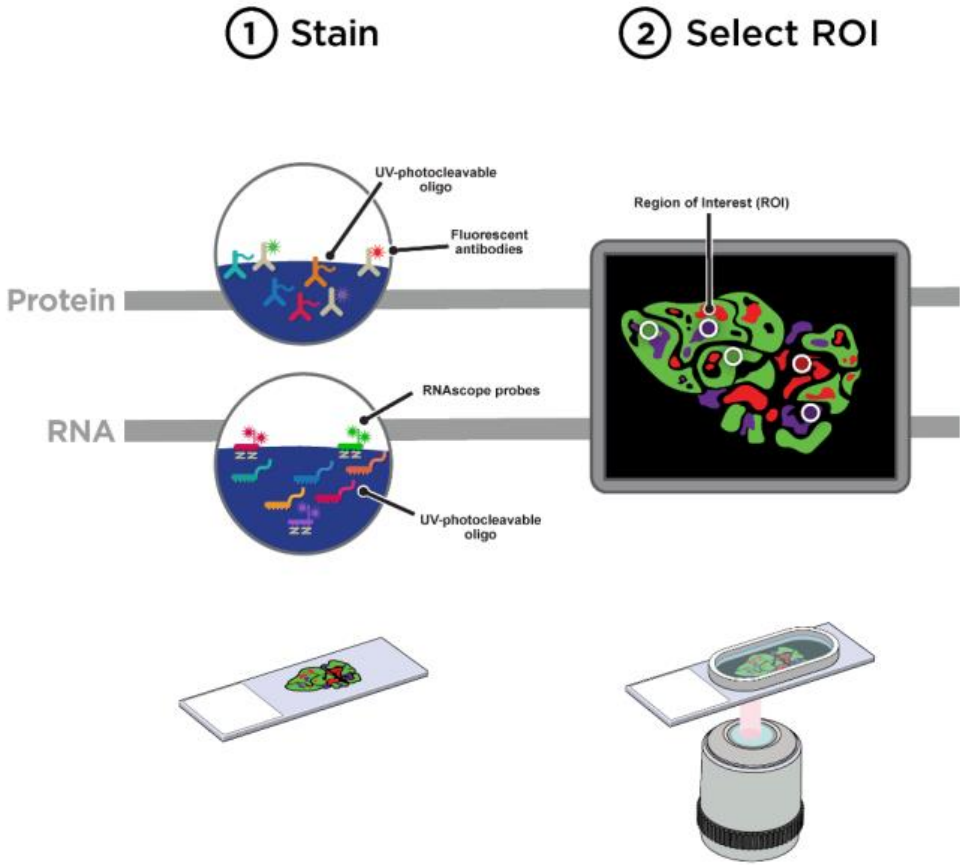
Label FFPE Slide with Probe Mix



Illuminate Region of Interest, as Small as a Single Cell

Workflow

You can use up to 3 markers to select your cells

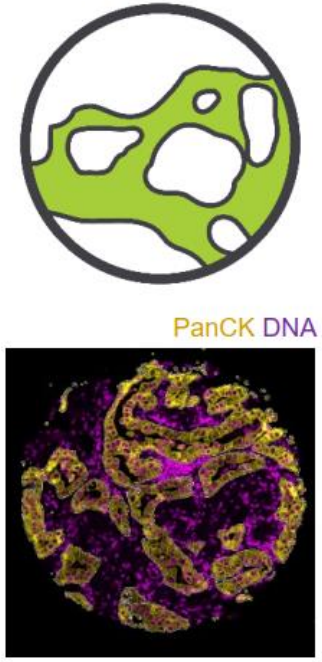


Region of interest

Geometric

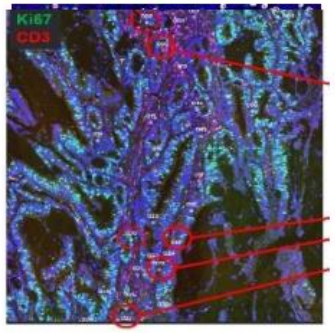
What is the heterogeneity of expression in different regions of my tissue?

Segmentation



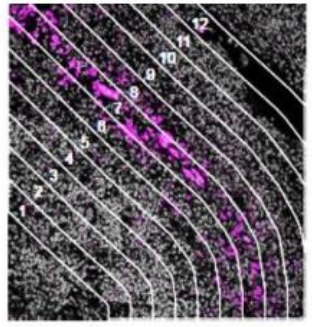
What is the expression profile of distinct biological compartments (e.g., Tumor-TME)?

Cell Type Specific



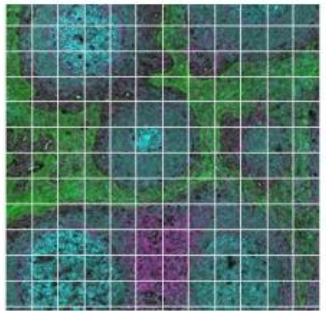
What is the expression profile of a specific cell population in my tissue?

Contour



How does the immune environment change on either side of an infiltrate boundary?

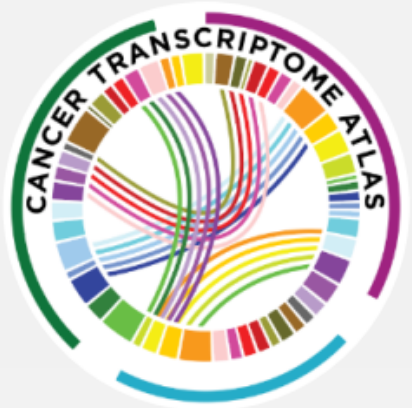
Gridded



What novel targets are uncovered with deep mapping of a specific tissue region?

Genes

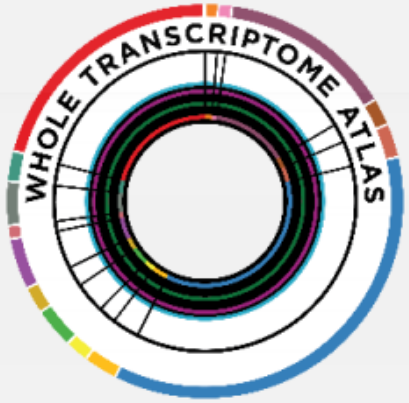
Cancer Discovery Applications



Curated & validated content for cancer research

- **Expansive content:** 1833 genes across 55 pathways

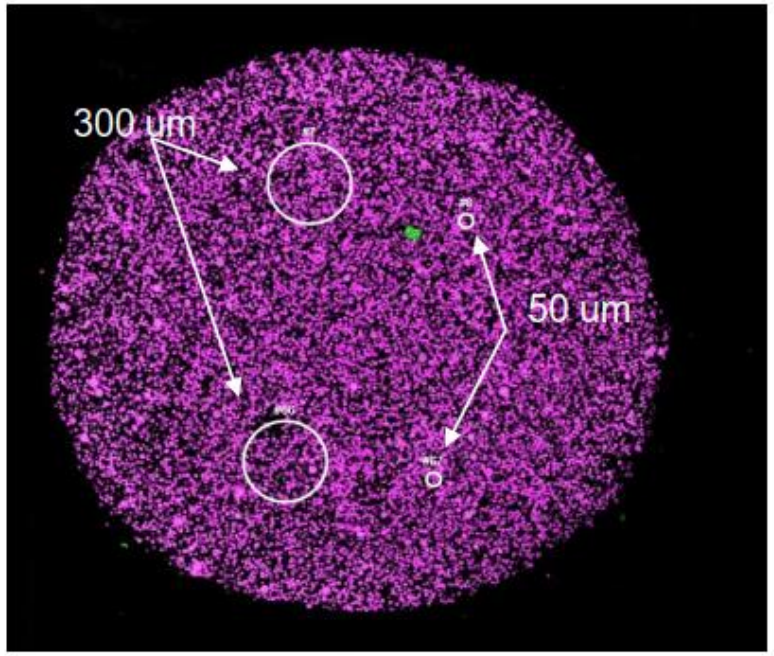
New Biology Discovery Applications



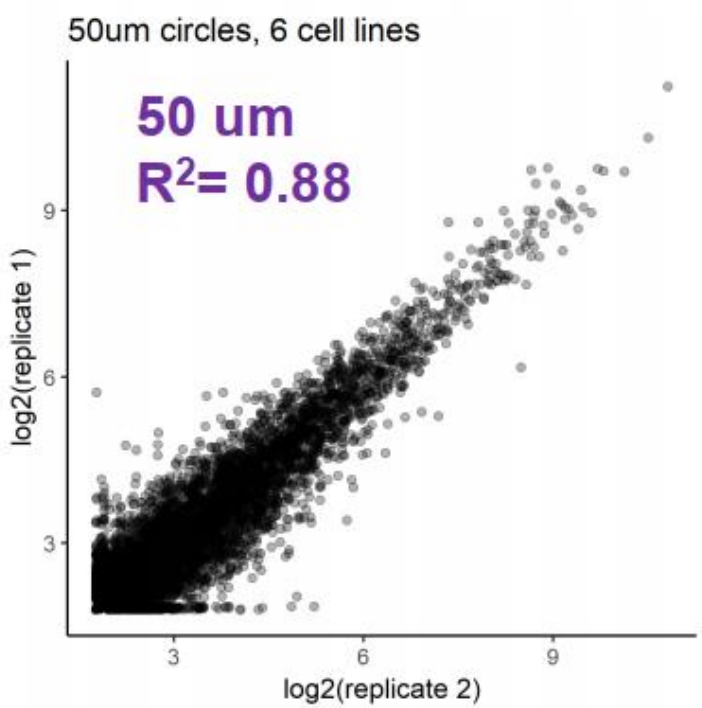
First non- 3'biased whole transcriptome assay for discovery research

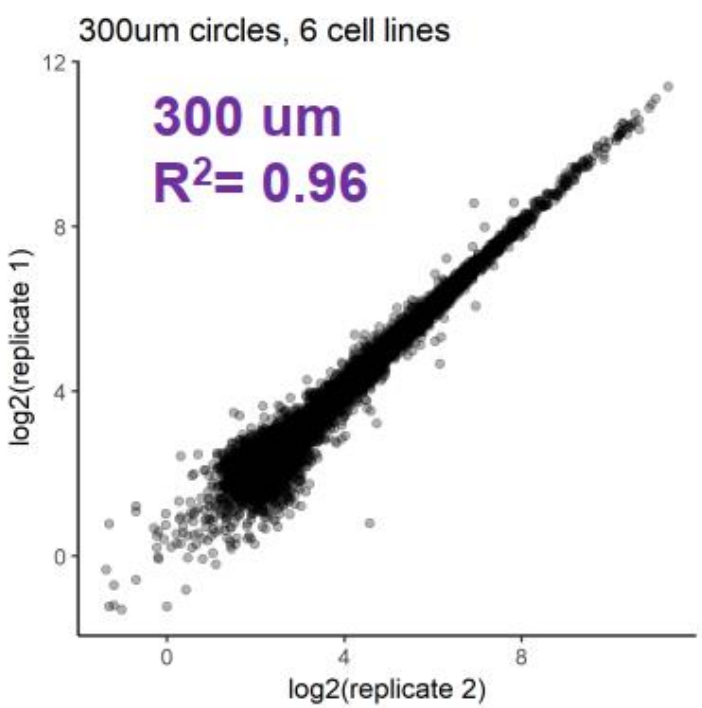
- **Comprehensive content:** >18,000 Gene specific probes designed to every protein coding sequence

Single cells?



FFPE Cell Pellet





Single cells?

	Diameter (µm)	Area (µm²)	
Cell (e.g. T-cell)	***10***	***79***	
	Diameter (µm)	**Area (µm²)**	**# cells**
ROI	700	384845	4900
ROI	600	282743	3600
ROI	500	196350	2500
ROI	400	125664	1600
ROI	300	70686	900
ROI	200	31416	400
ROI	150	17671	225
ROI	141	15615	200
ROI	**100**	**7854**	**100**
ROI	71	3959	50
ROI	55	2376	30
ROI	50	1963	25
ROI	45	1590	20
ROI	32	804	10
ROI	25	491	6
ROI	20	314	4

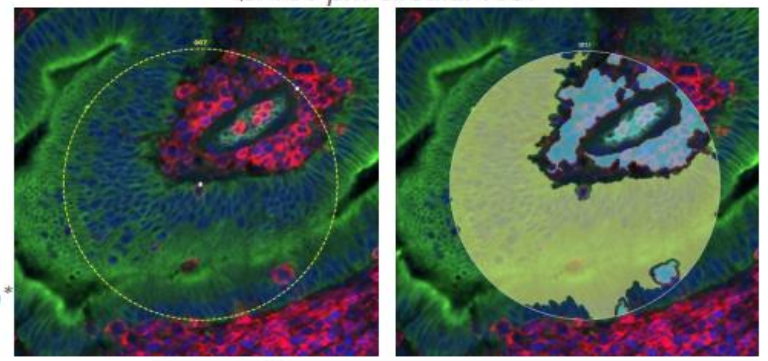
*Advised cut-off for nCounter RNA application**

Advised cut-off for NGS RNA application*

*Advised cut-off for nCounter Protein application**

* Determine empirically whether lower cell numbers can be done.

Ø 150 µm circular ROI



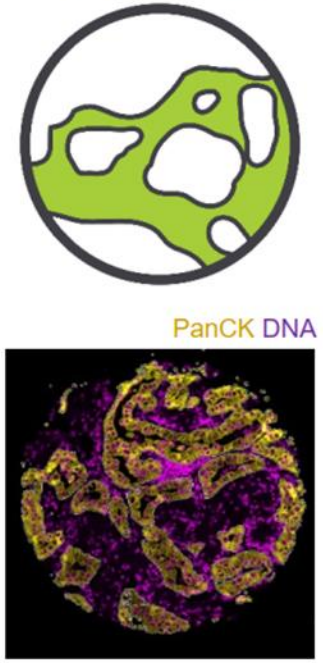
Region of interest - segmentation

You can use up to 3 markers to select your cells

Geometric

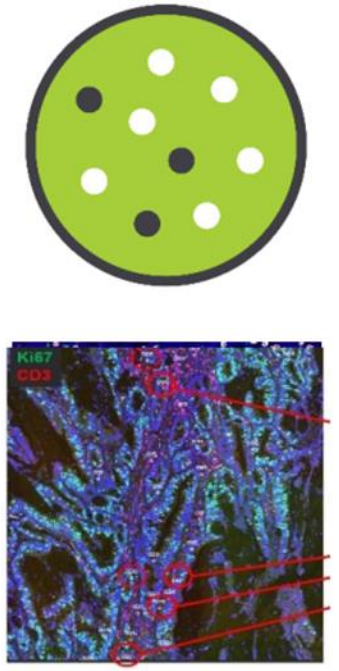
What is the heterogeneity of expression in different regions of my tissue?

Segmentation

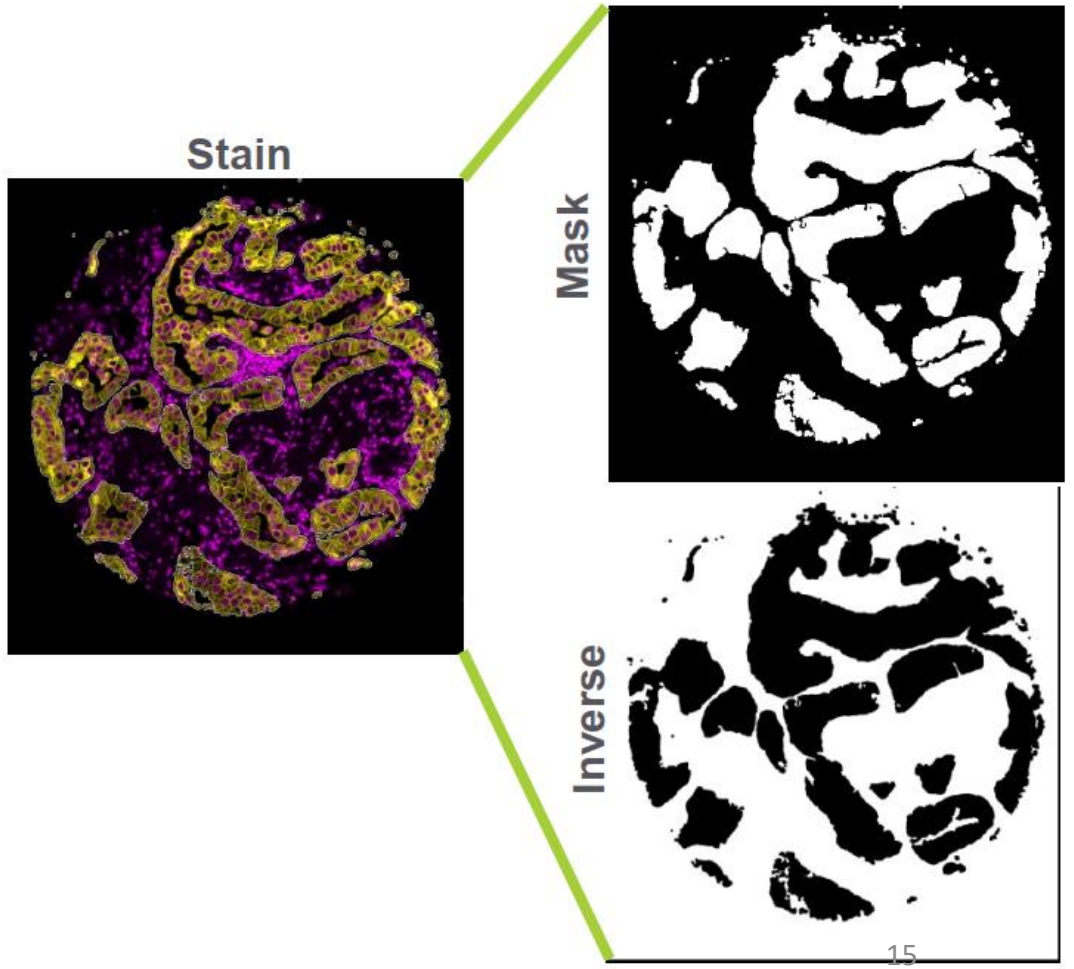


What is the expression profile of distinct biological compartments (e.g., Tumor-TME)?

Cell Type Specific



What is the expression profile of a specific cell population in my tissue?



Summary GeoMx - nanostring

Cons:

- It’s not single cells
- You cannot do transcriptomic and proteomic on the same slide
- You don’t have a sequence to check (only probes, false positive/negative signals?)

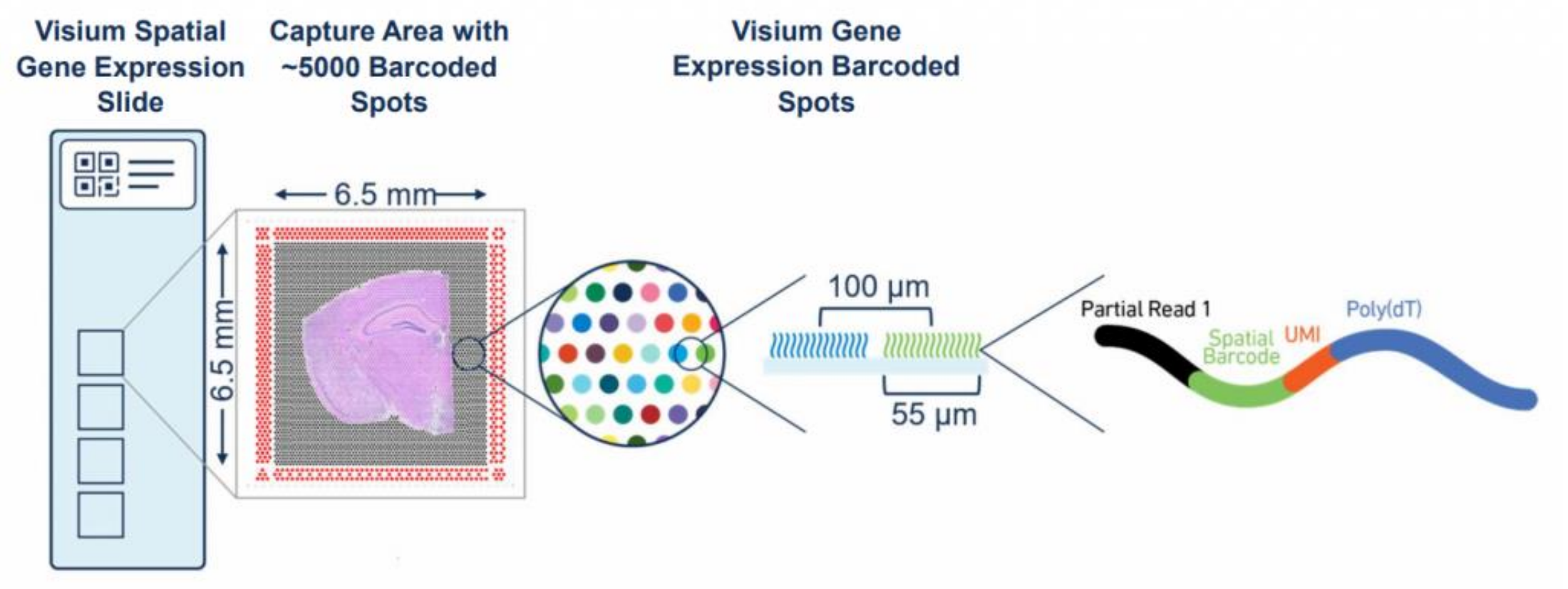
Pros:

- Seems pretty sensitive (20 cells).
- You can select not-consecutive cells to reach n cells
- Use of 3 markers and segment function for all possible combinations
- Proteomic and transcriptomic information (from 2 consecutive slides)
- Deconvolution function works well

10x Genomics - **Xenium**

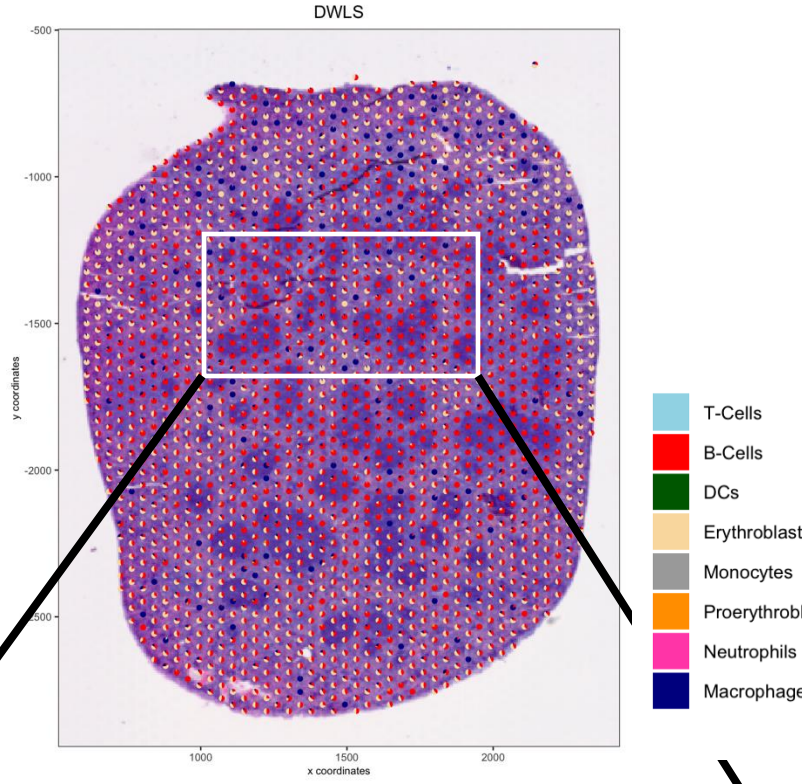
https://pages.10xgenomics.com/tch-2023-03-tech-lit-ra_g-p_xenium-new-dataset-awareness-general_gatedlp.html

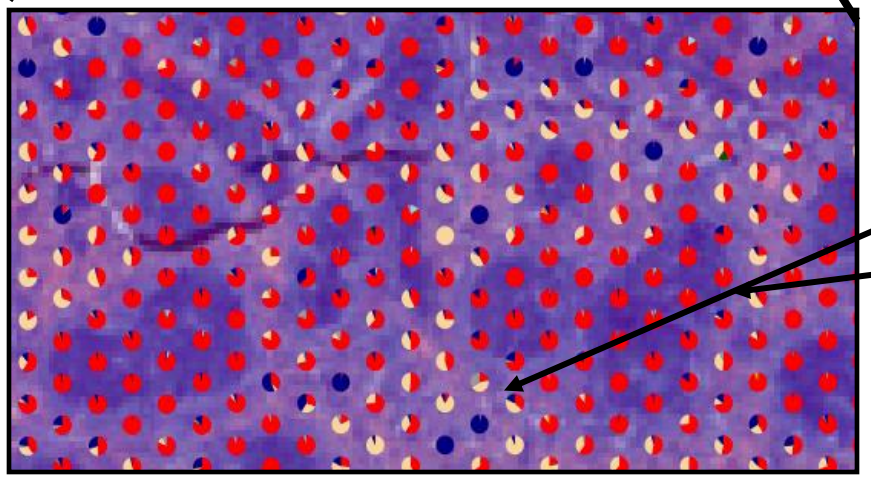
Spatial transcriptomics is needed: 10X Visium



A kind of spatial bulk RNA-Seq

Confirmation? Spatial transcriptomics

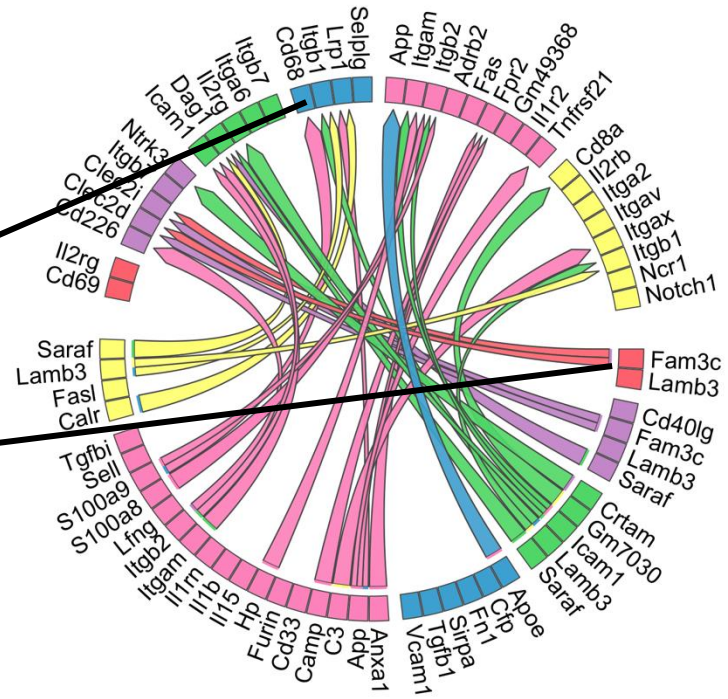




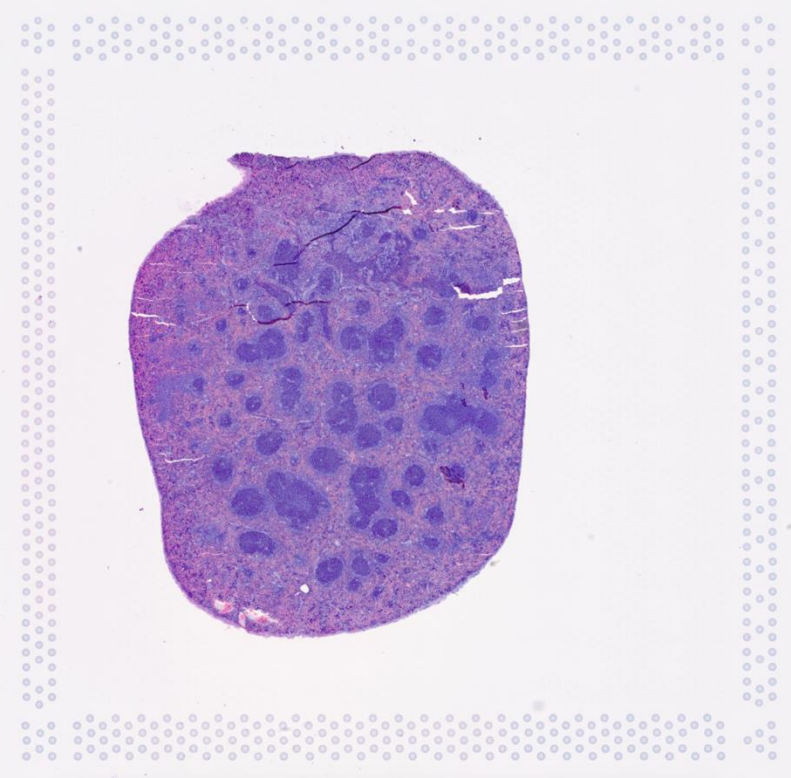
Neutrophils Macrophages B.cells

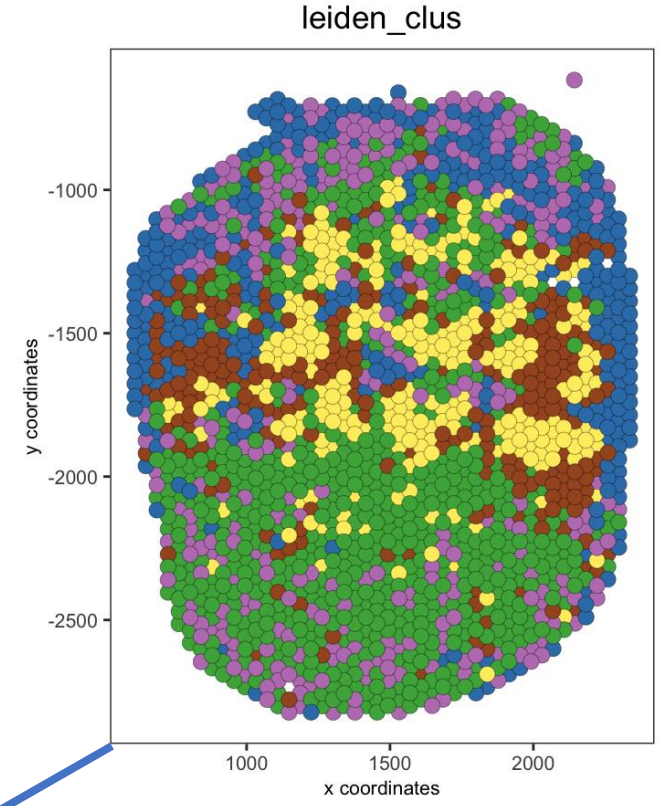
CD8..T.cells CD4..T.cells NK.cells

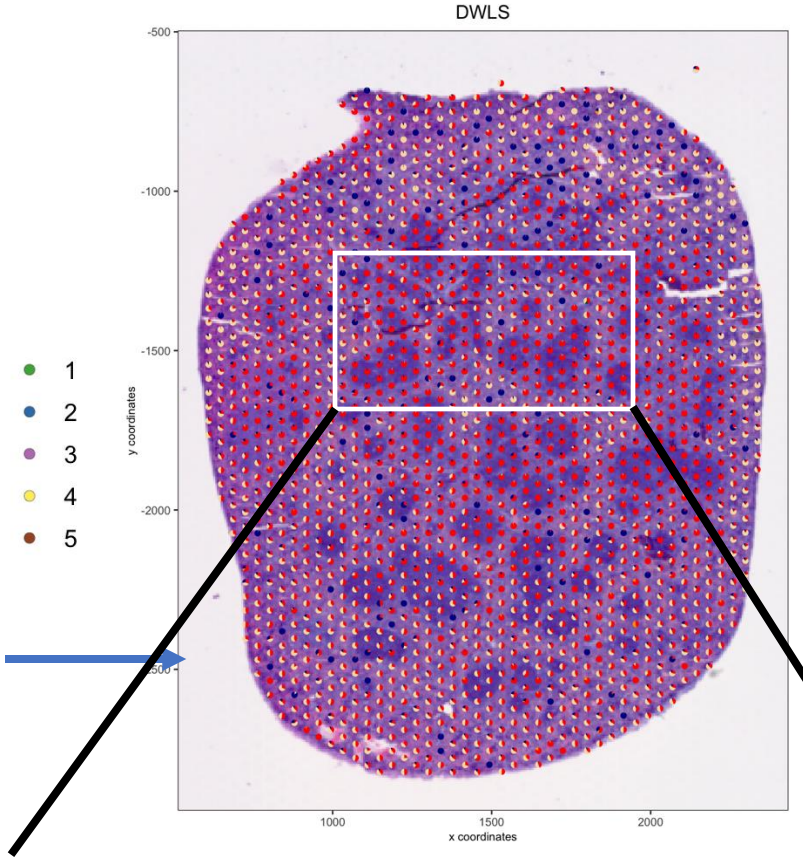
Recrudescence (23 dpi)



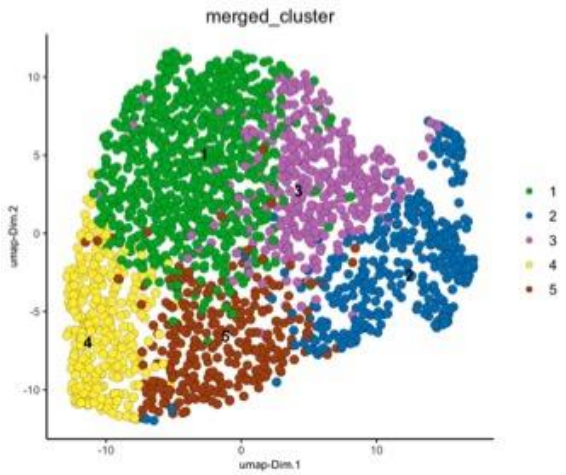
10X Visium on TP1 - Wild type



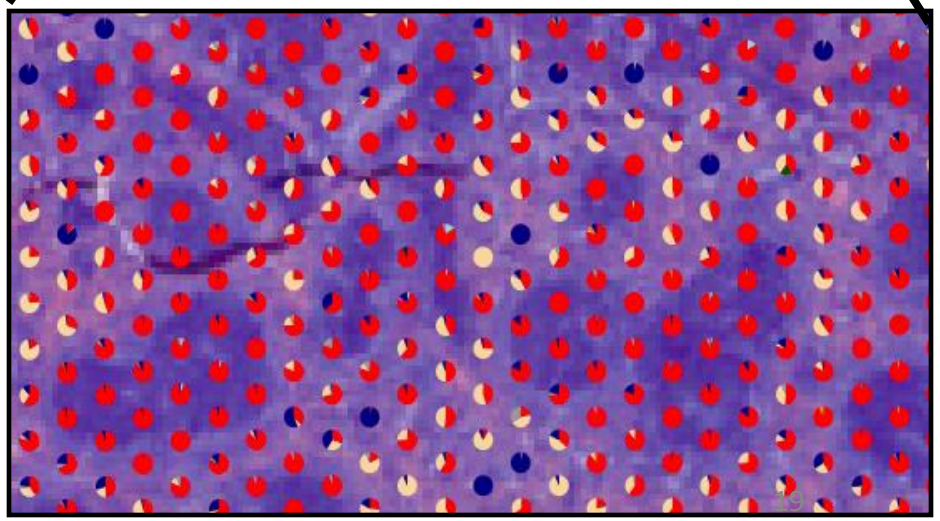




Each dot is a mixture of different cells

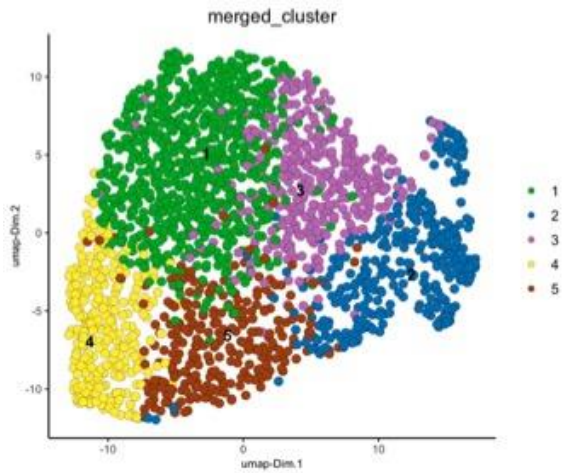


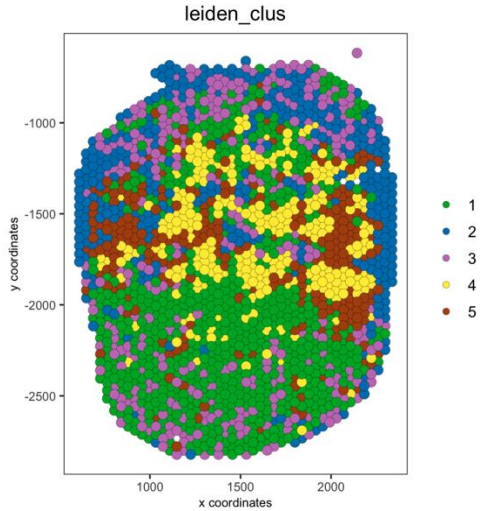
Projection of cell mixtures into 2D space (UMAP)

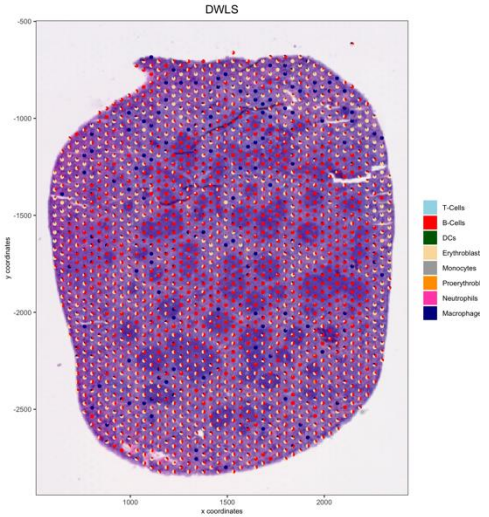


Analysis with Giotto

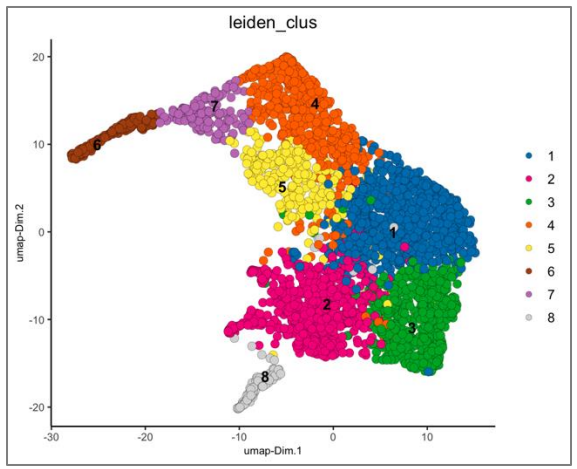
Tp1 – Wild type

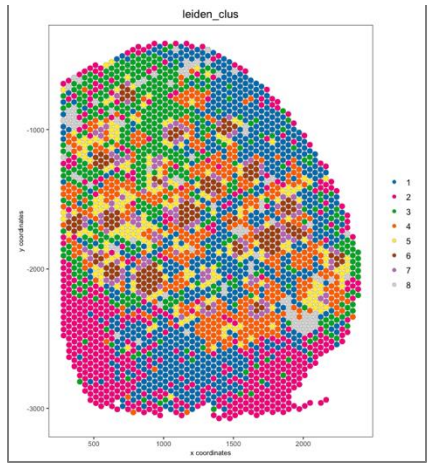


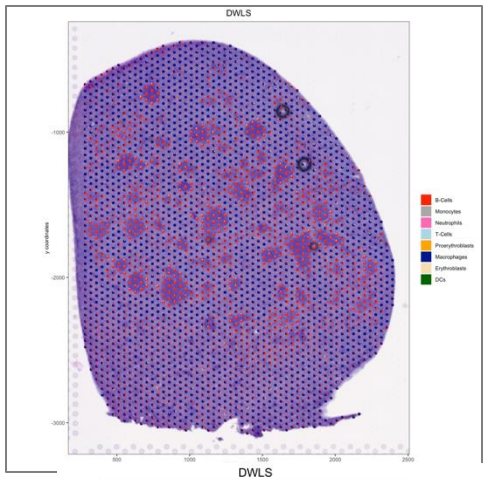




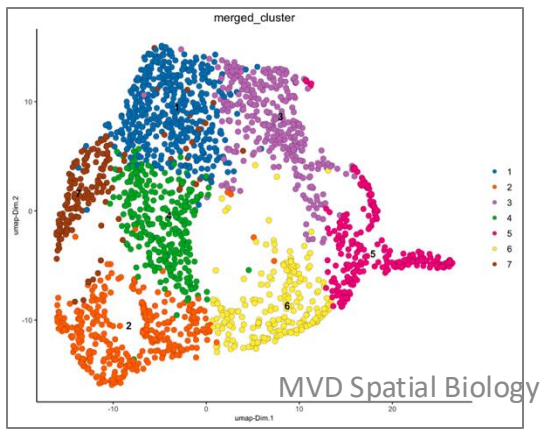
TP-T2 - peak

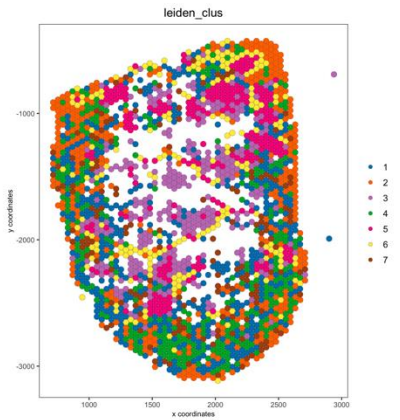


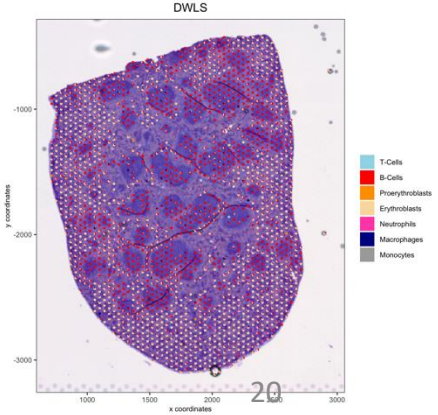




TP4 - resolution







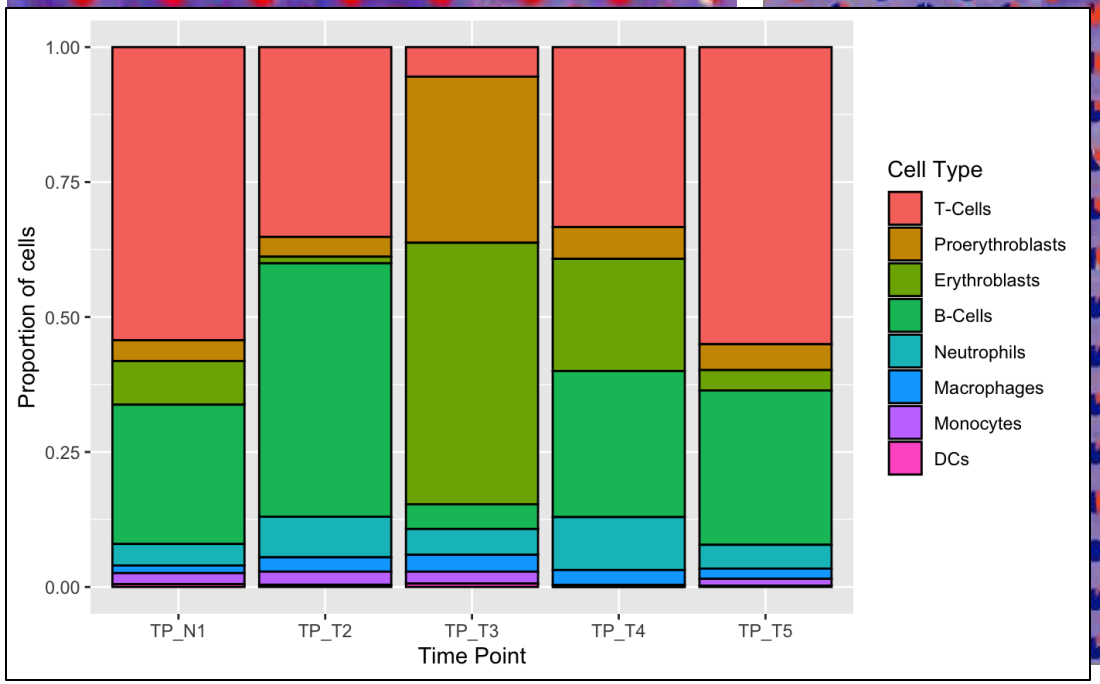
TP1 - wildtype

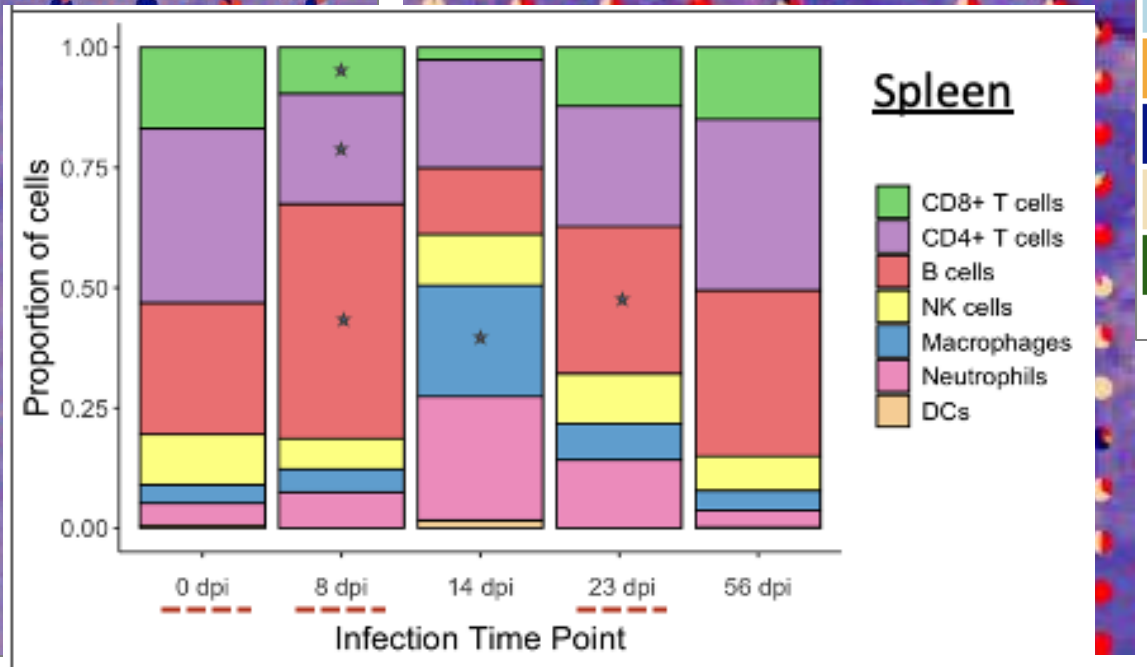
TP2 -8 DPI peak

TP4 – 23 DPI resolution

Immune cells

- B-Cells
- Monocytes
- Neutrophils
- T-Cells
- Proerythroblas
- Macrophages
- Erythroblasts
- DCs

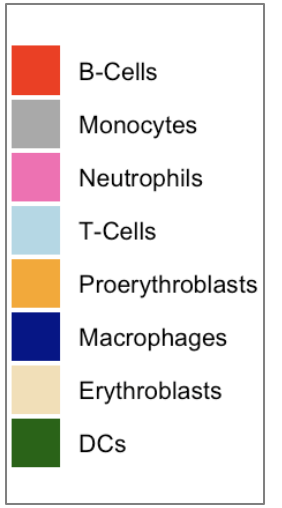


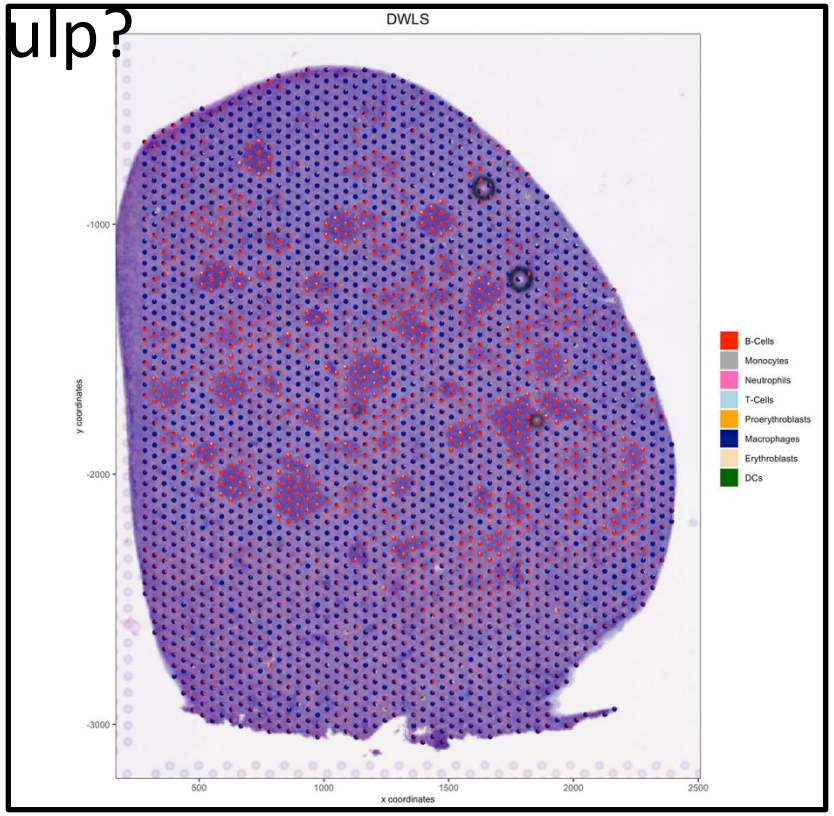


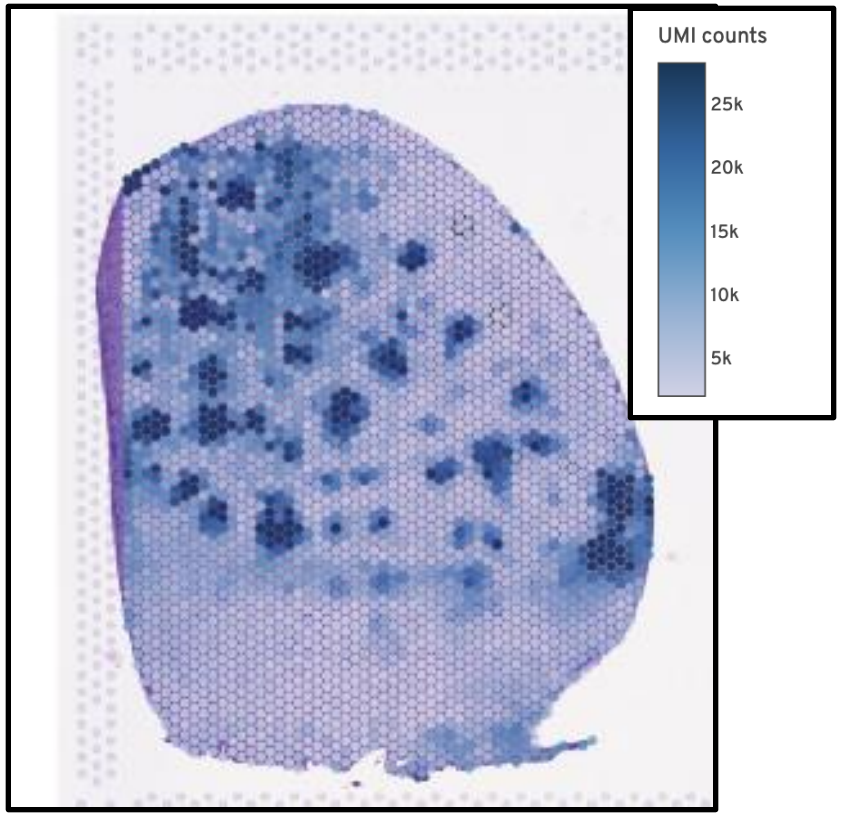
Does not 100% fit with the scRNA-Seq cell frequencies, especially the macrophages…

More cells in the white pulp?

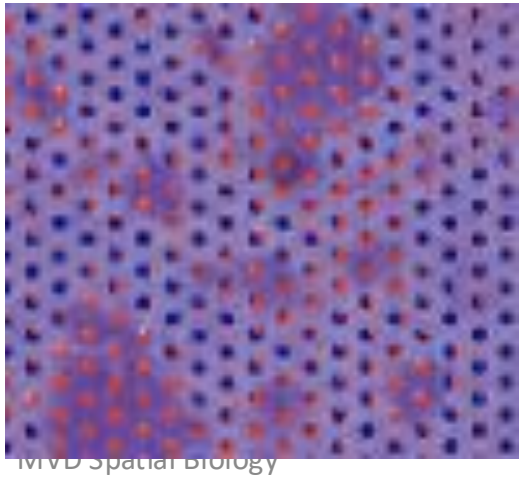
Peak 8 DPI

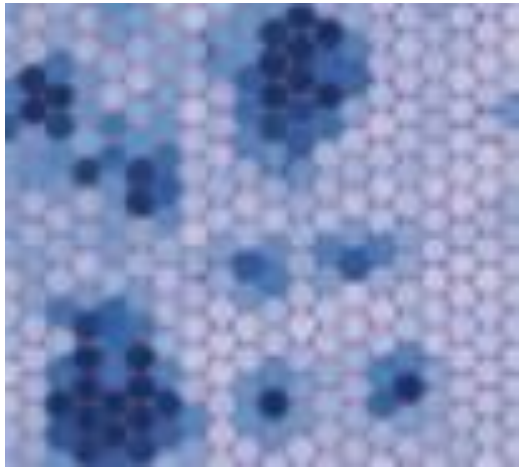






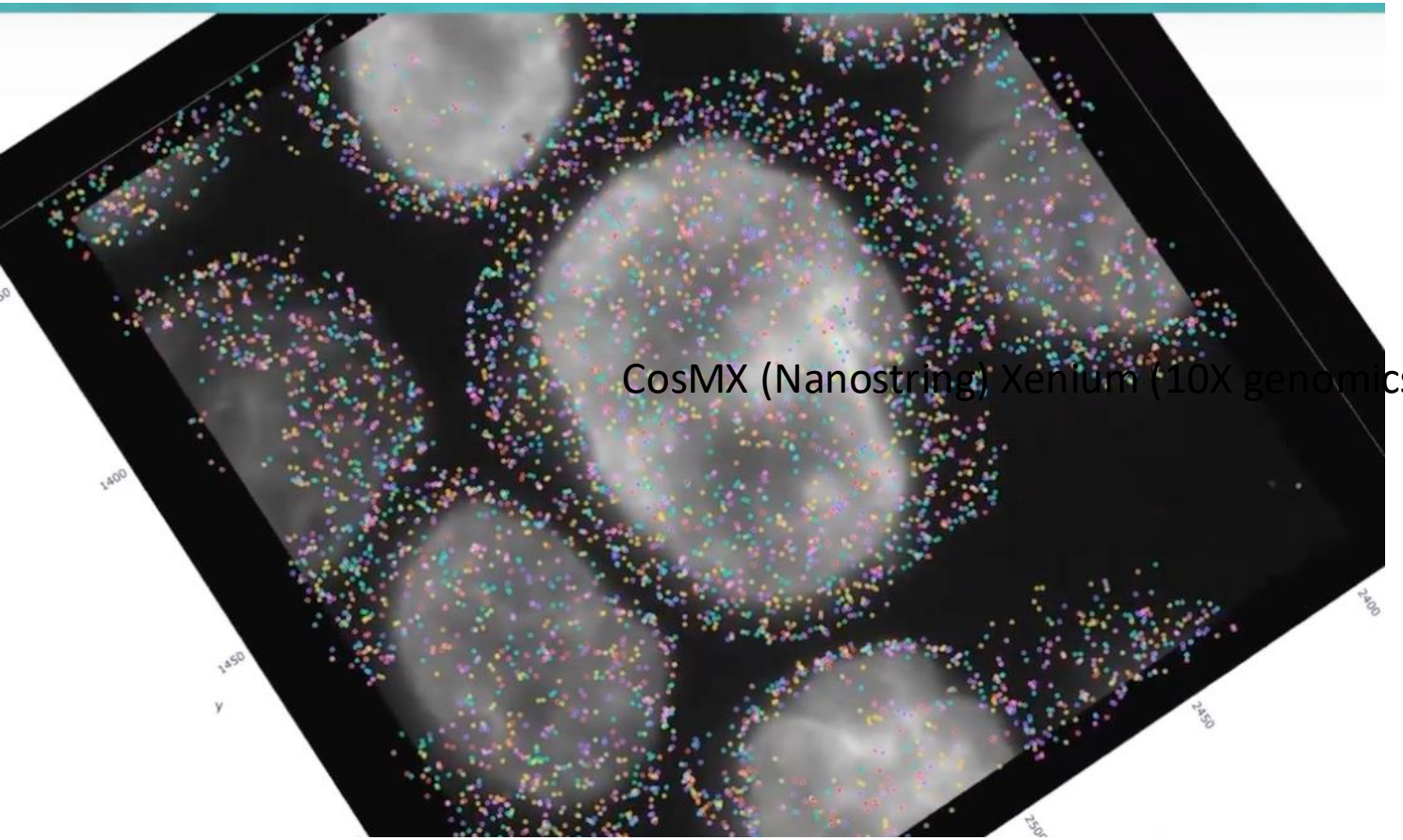
Although “more” macrophages (MK)
more UMI (RNA) in white pulp (B-cells)





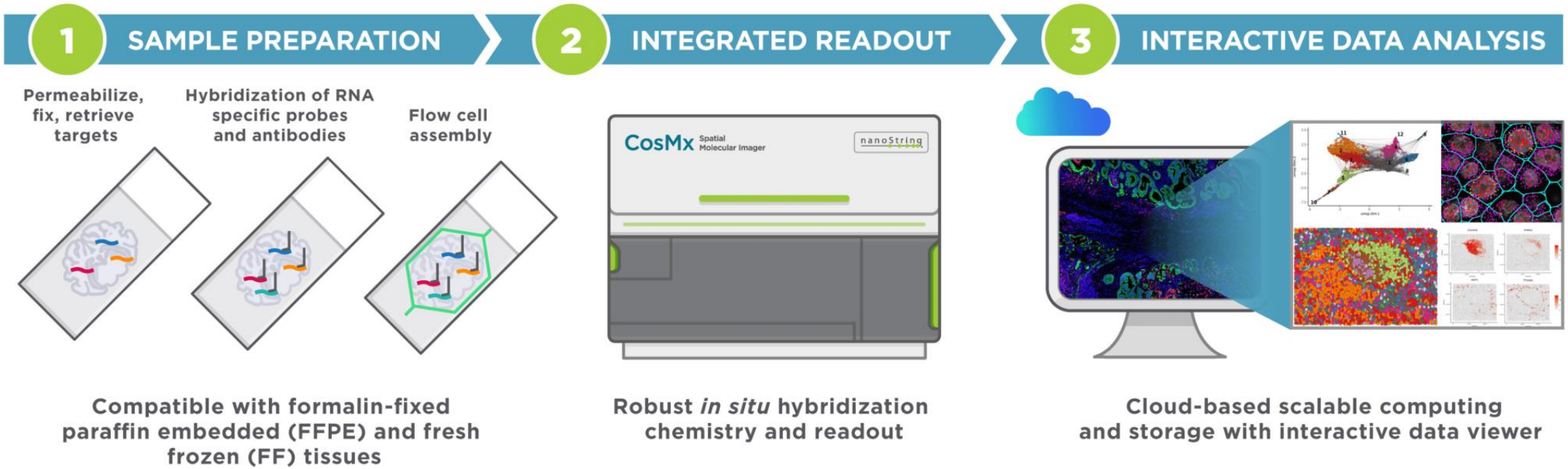
Still no scRNA-Seq….

CosMX (Nanostring) Xenium (10X genomics)



CosMX (Nanostring) Xenium (10X genomics)

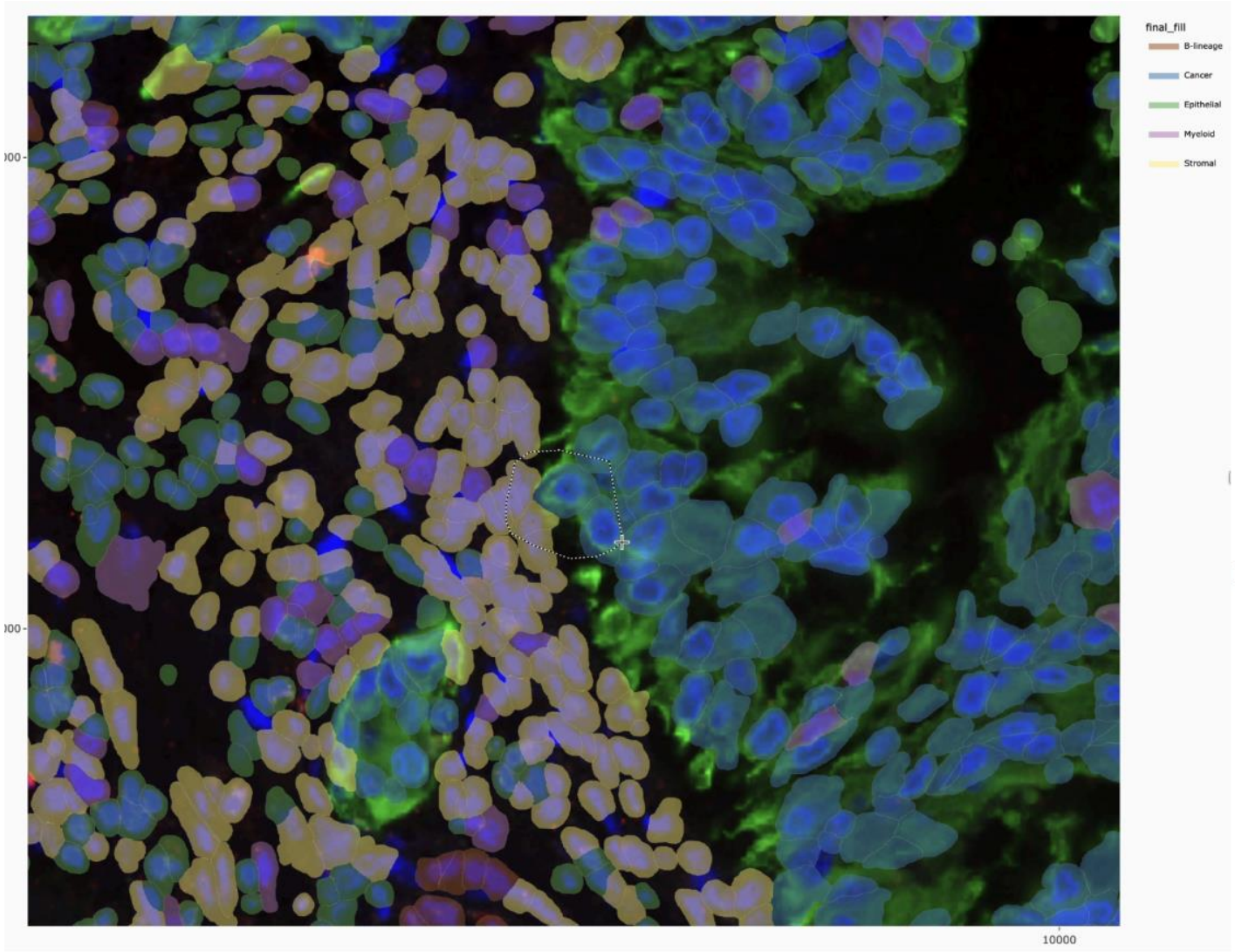
- captures 1000 “host” genes
- CosMx Nanostring – Xenium
- Formalin-Fixed Paraffin-Embedded (FFPE) Tissue
- >1Tb of data per sample

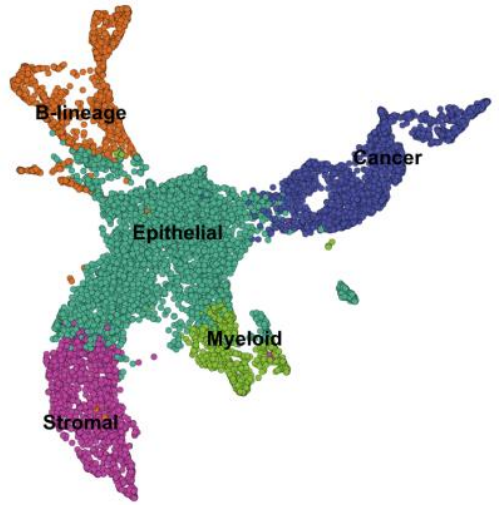


CosMx SMI for single-cell imaging delivers a comprehensive package which includes validated reagents, instrument, and data analysis software for seamless sample-to-result.

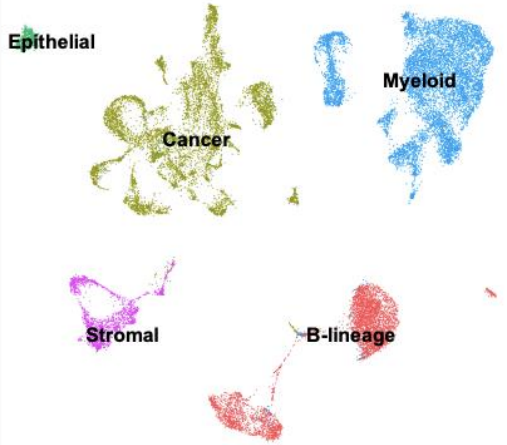
https://nanostring.com/products/cosmx-spatial-molecular-imager/single-cell-imaging-overview/

CosMX data example



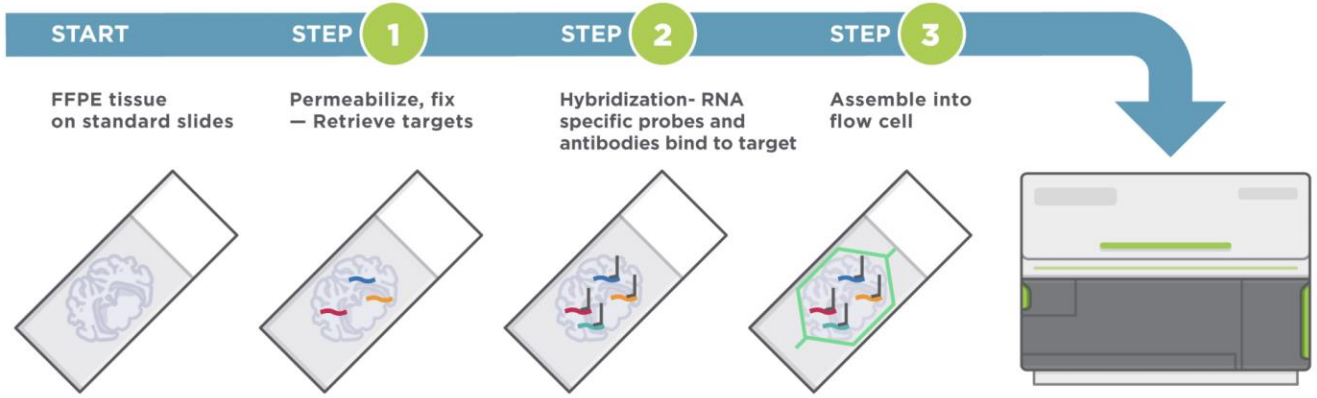


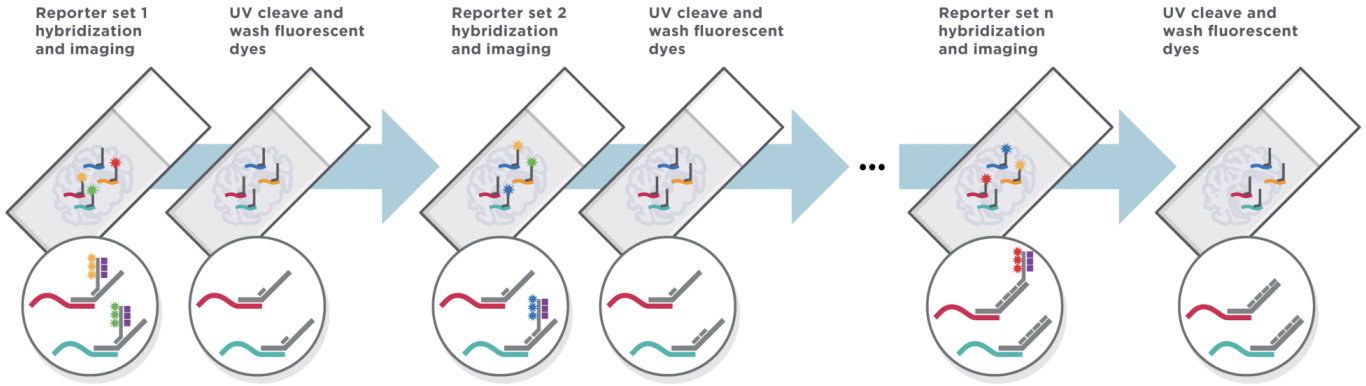
CosMX probe UMAP

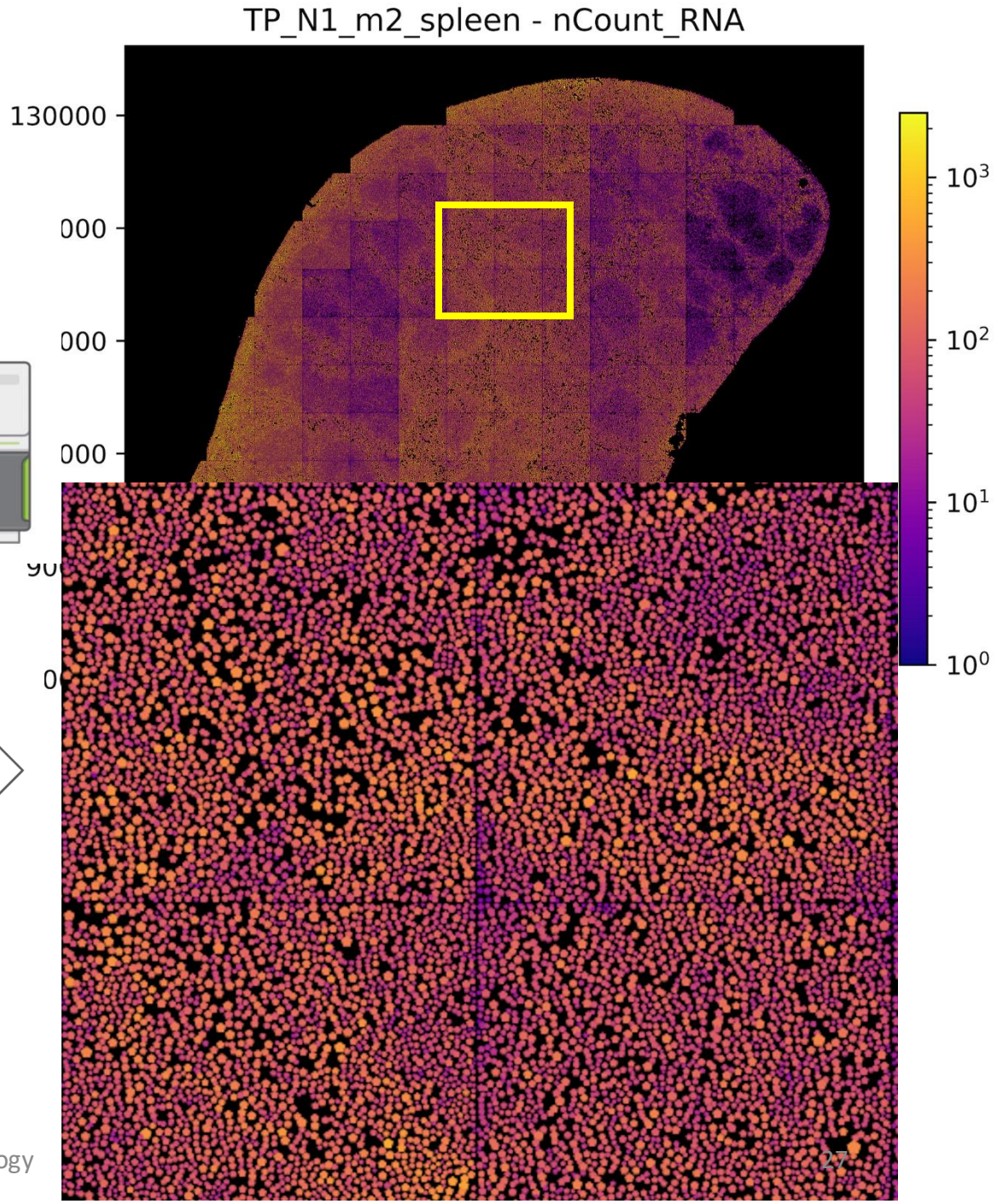


scRNA-Seq UMAP

CosMX – spatial single-cell

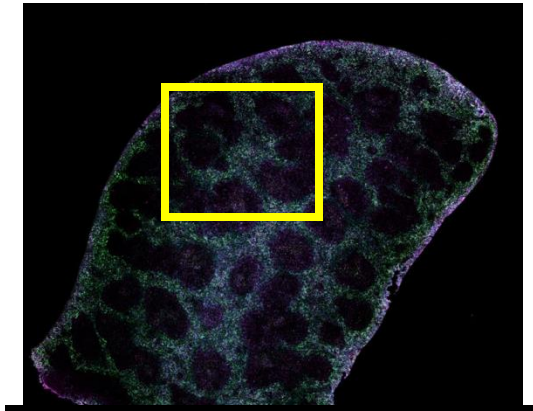


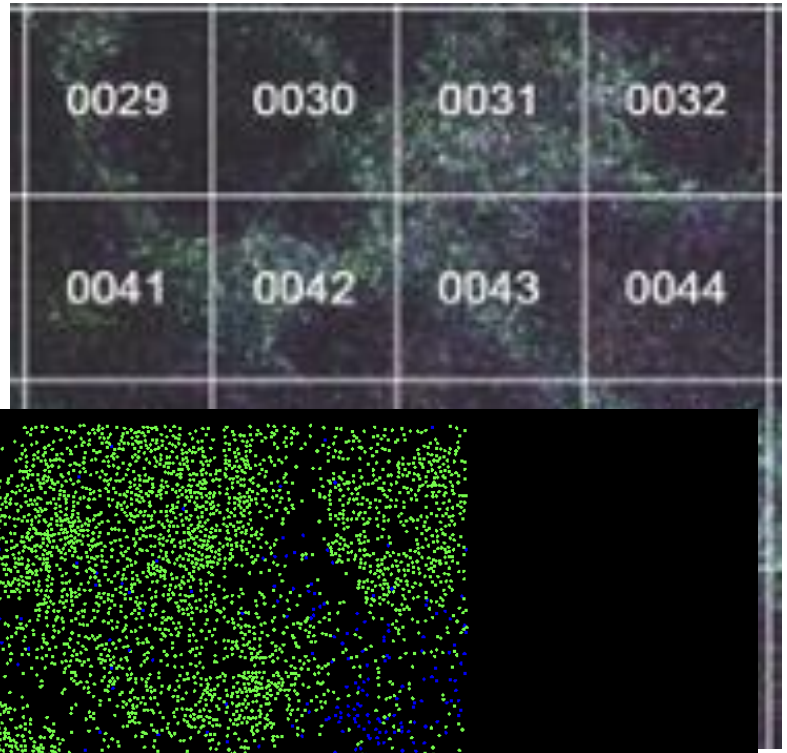


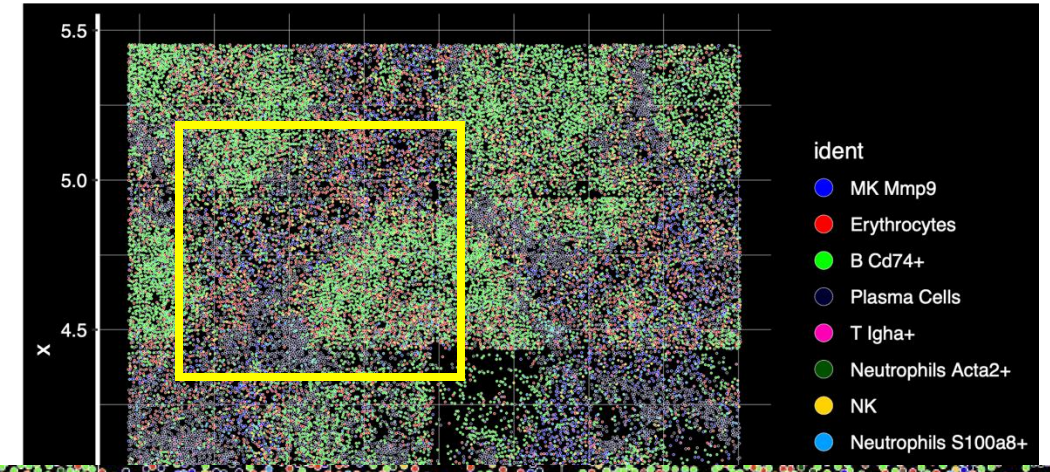


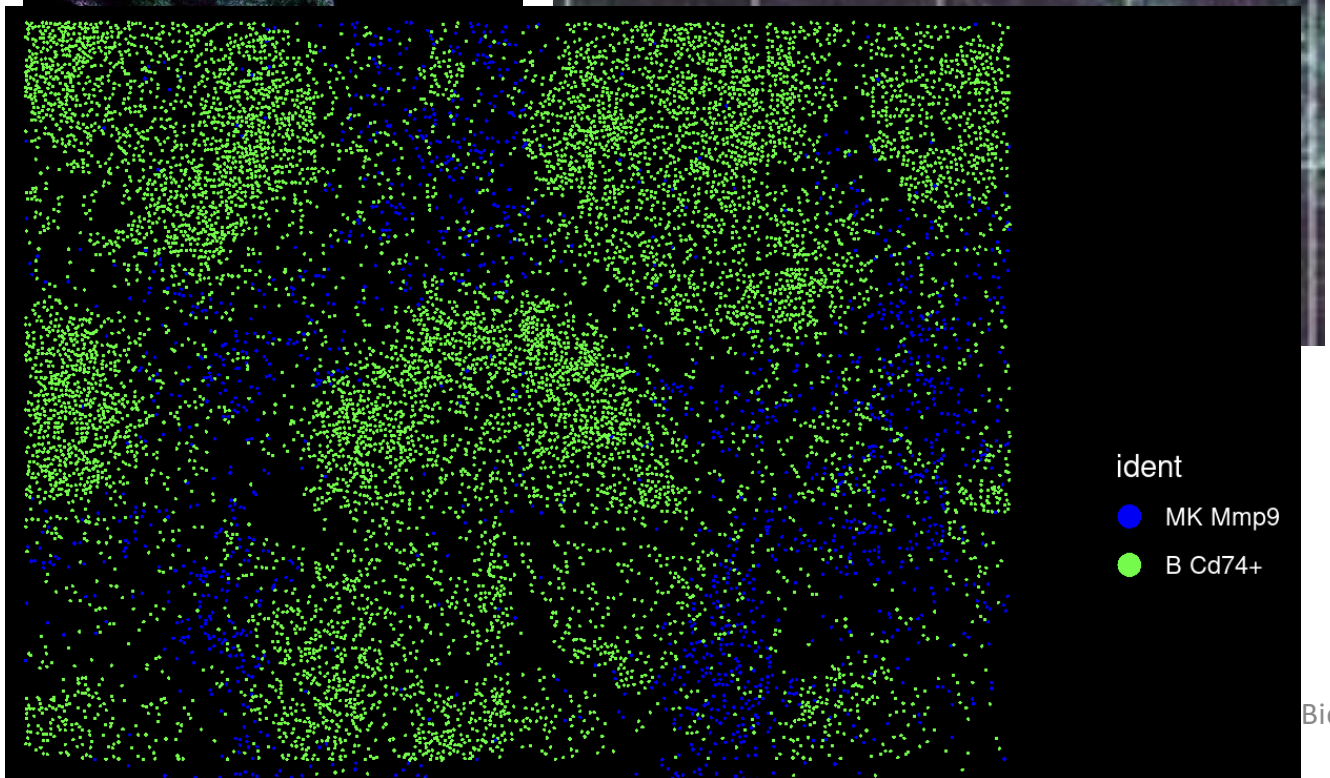
https://nanostring.com/products/cosmx-spatial-molecular-imager

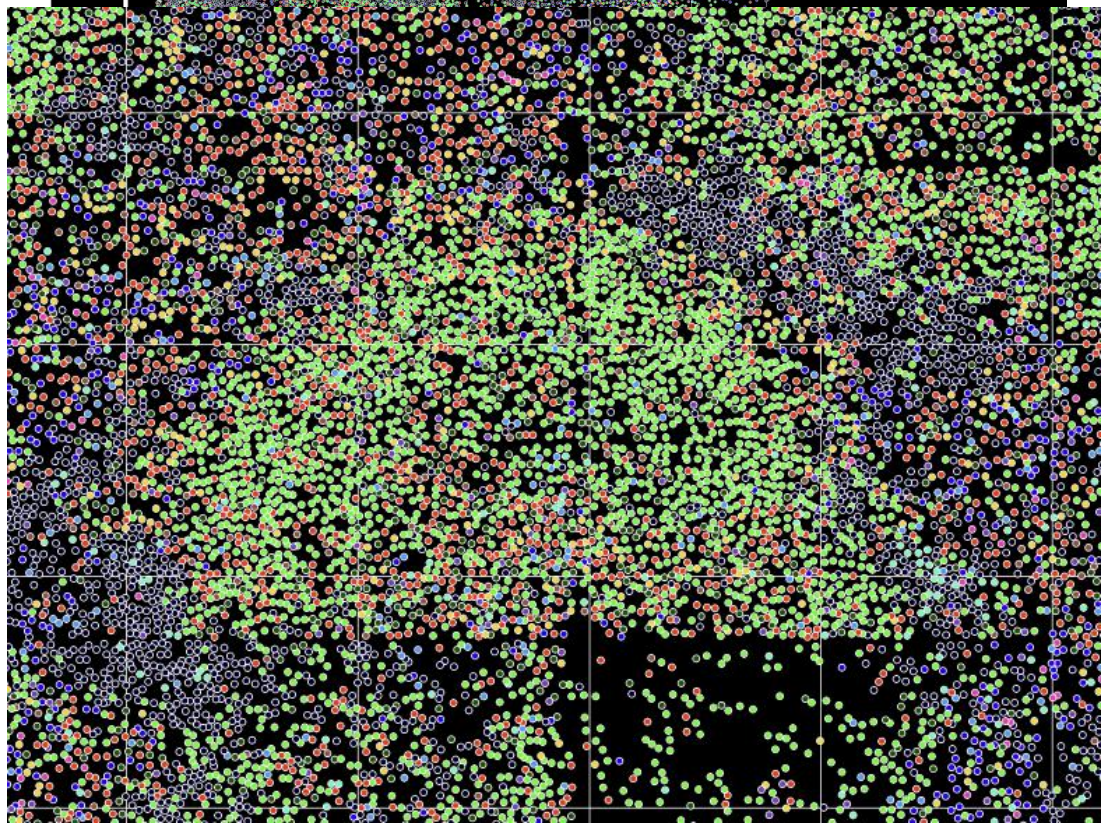
CosMX – it works!





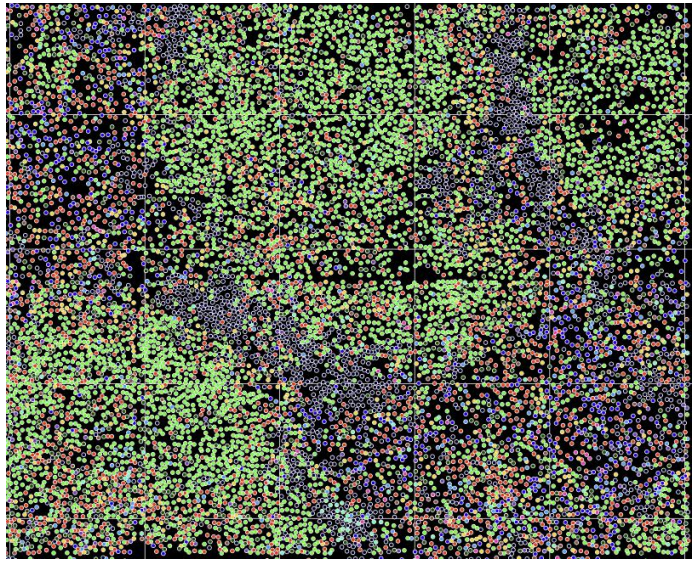




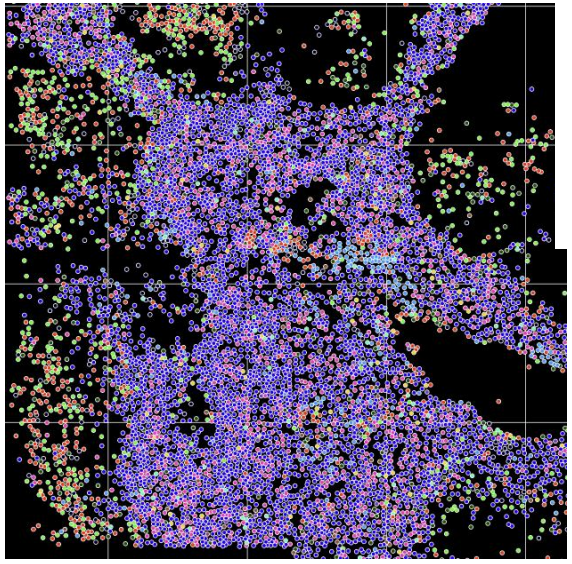


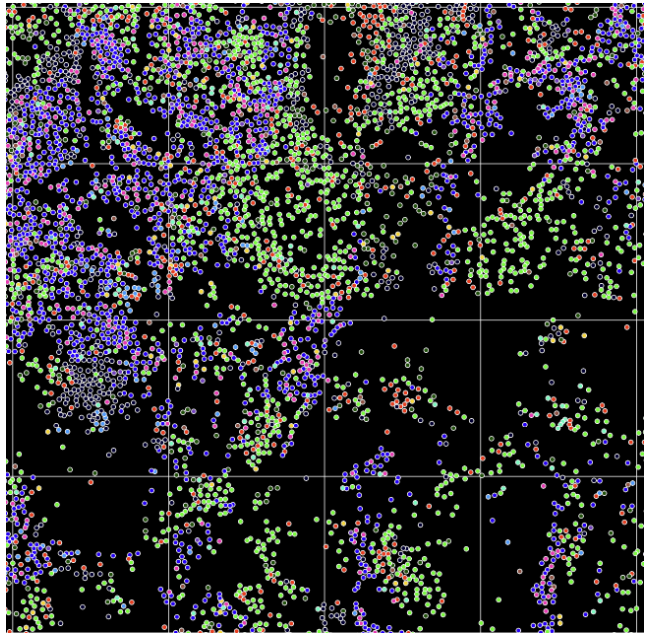
Overall impression

0 dpi - Naive

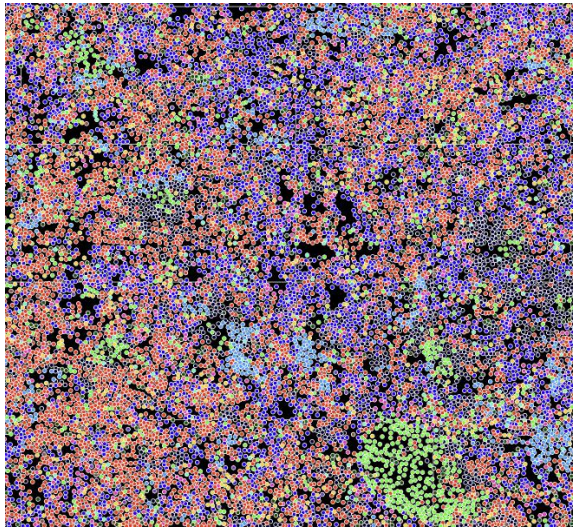


8 dpi - peak

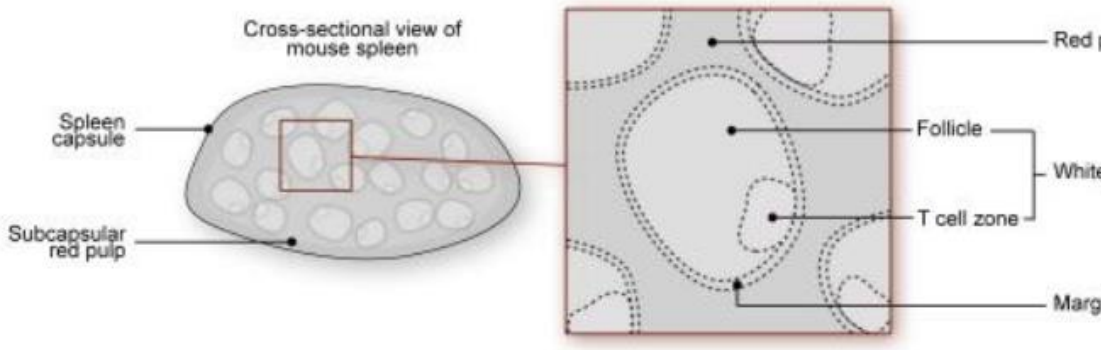


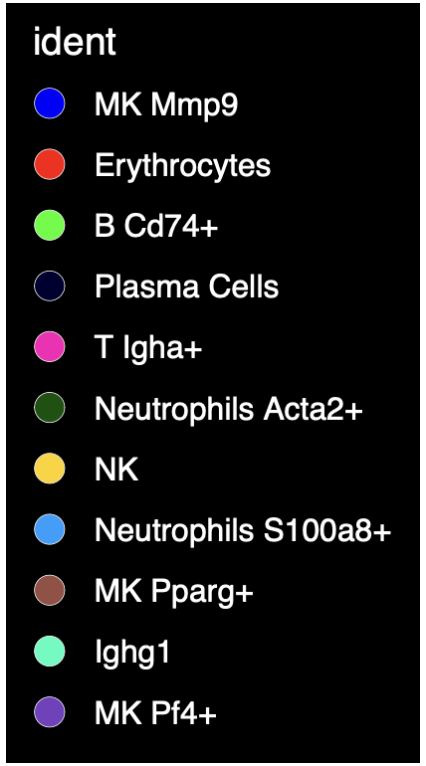


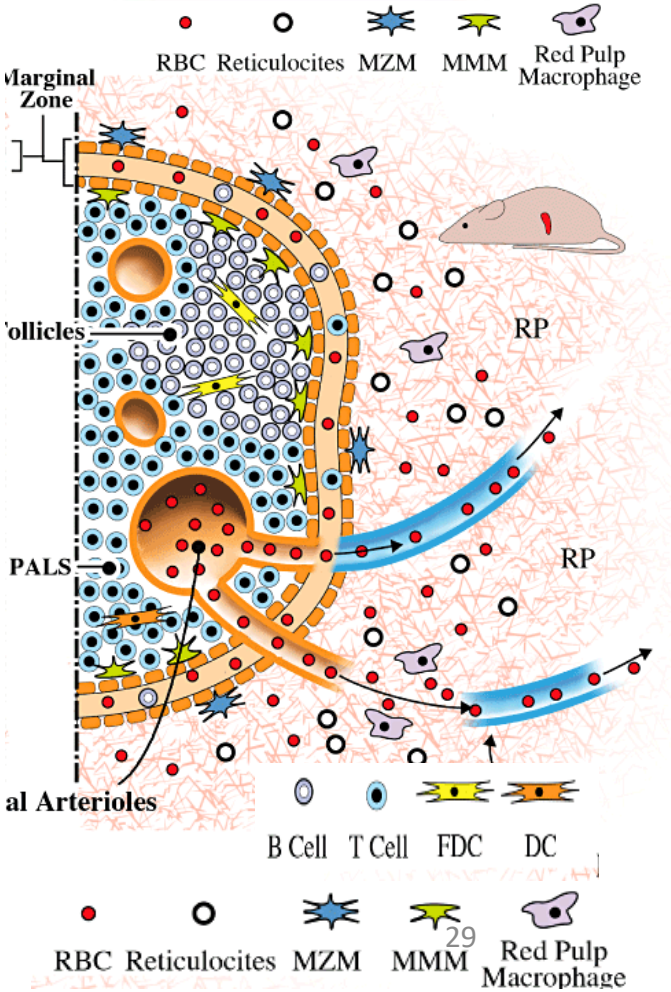
14 dpi



23 dpi

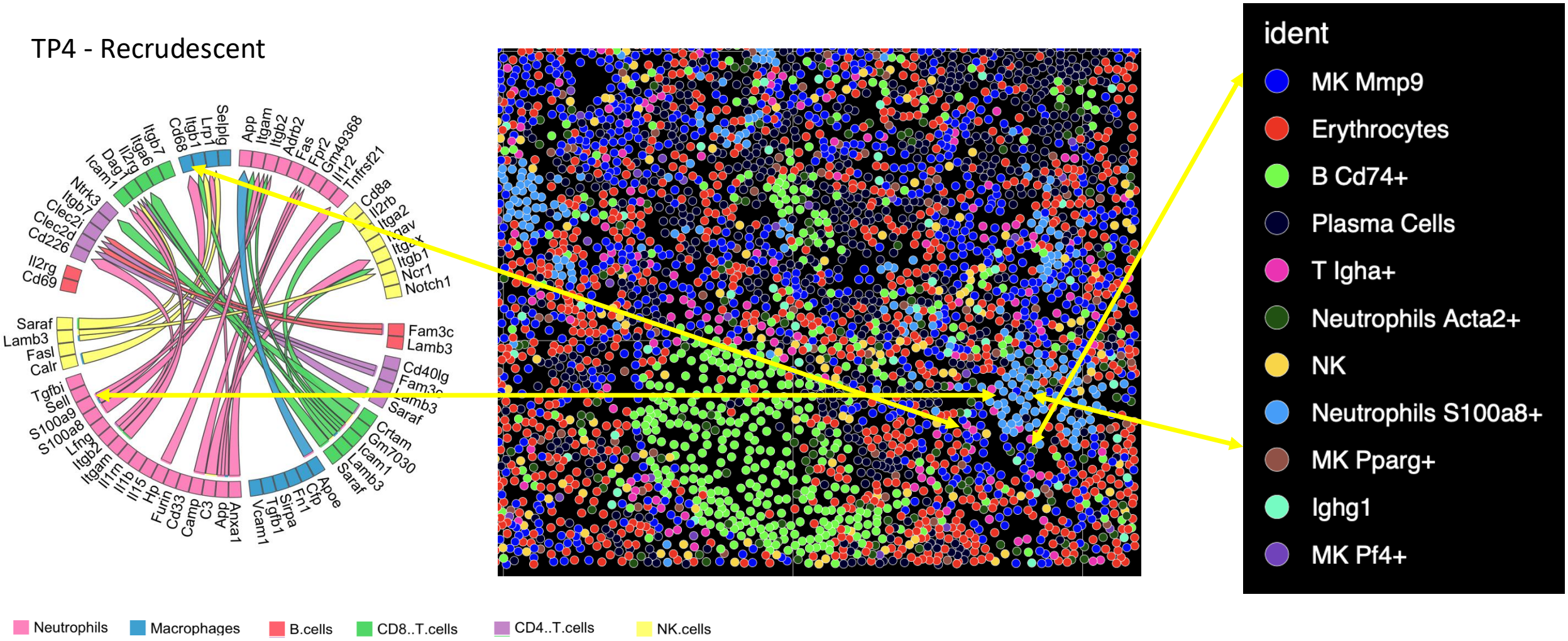






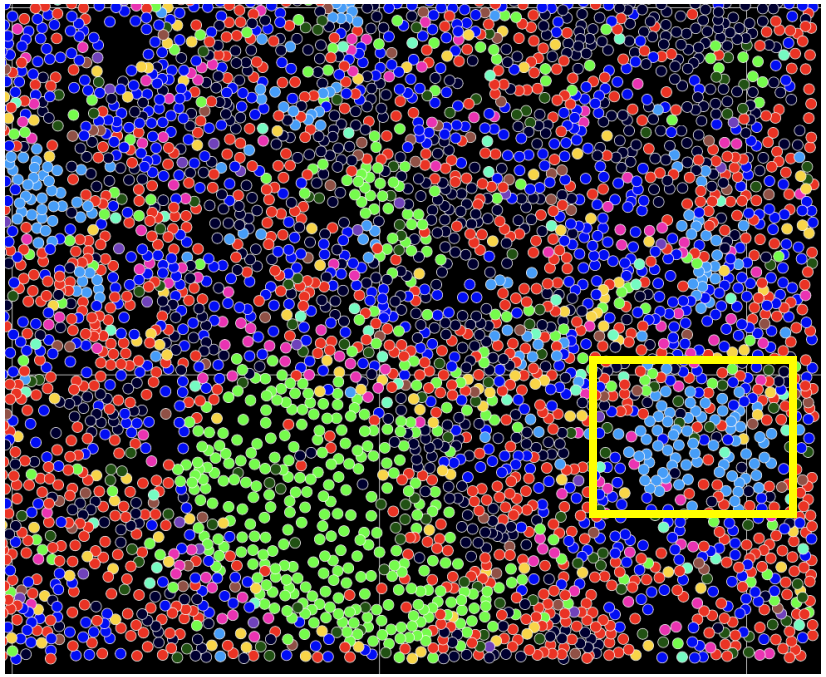
Confirming CCI?

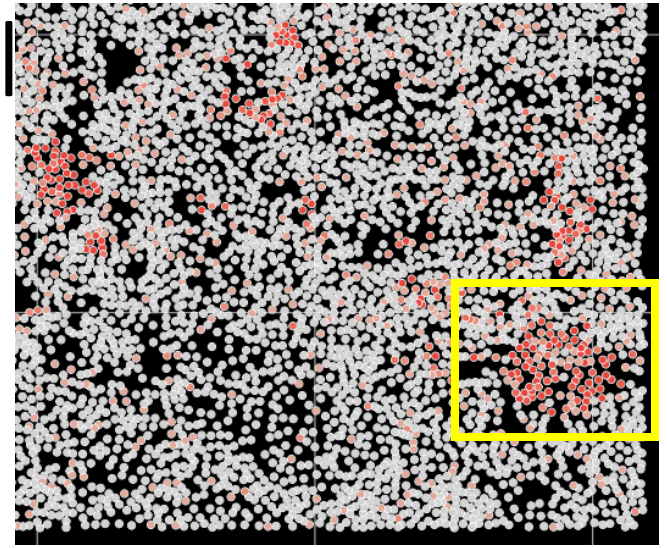
TP4 - Recrudescent

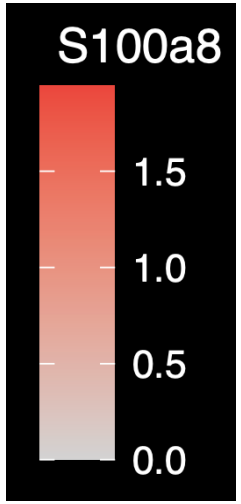


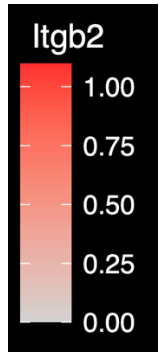
Want to confirm S100a8 (Neutrophils) -> Itbg2 (Neutrophils) / Cd68 (MK)

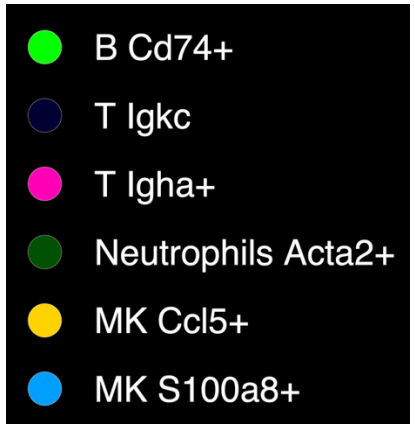
Can we confirm CCI

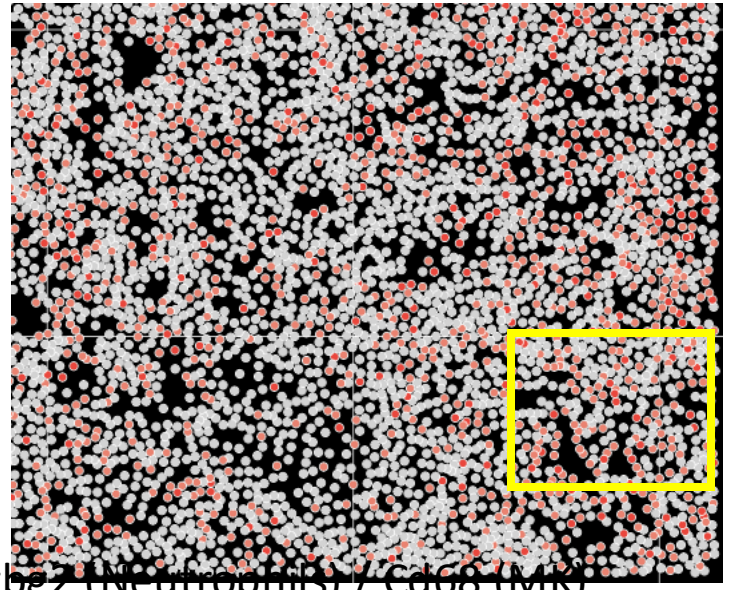


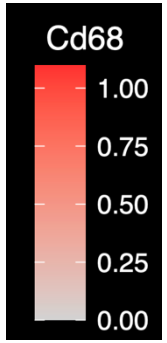


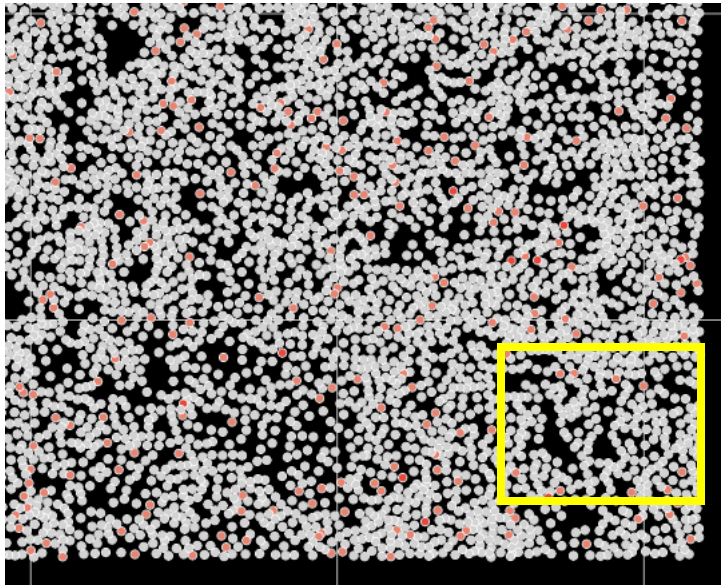












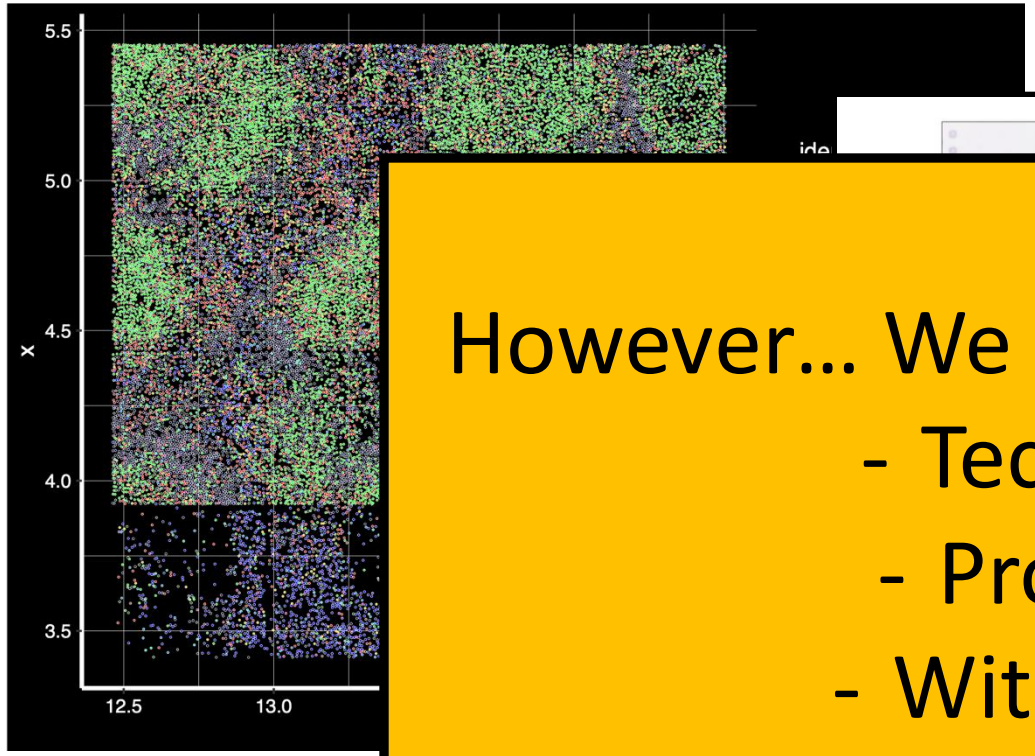
Want to confirm S100a8 (Neutrophils) -> Itbg2 (Neutrophils) / Cd68 (MK)

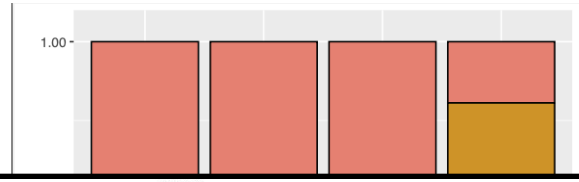
However, aren't they close?

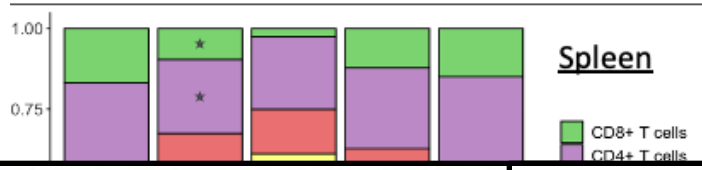
And… spatial biology?

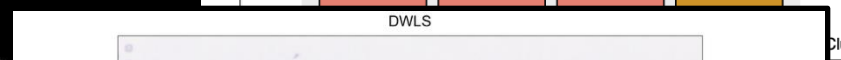
Cell frequency of CosMX data

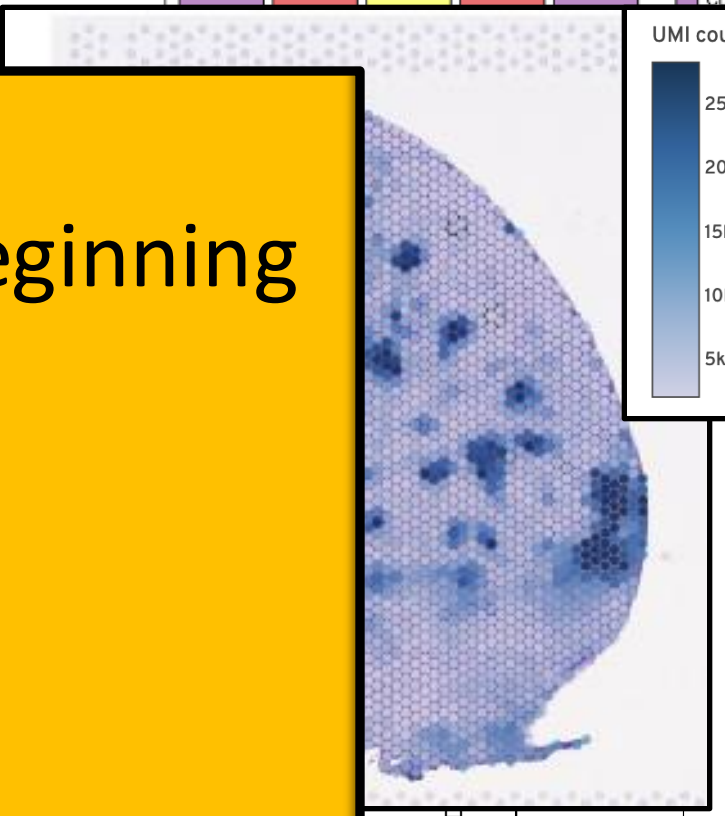
Frequency in scRNA (without Erythrocytes)







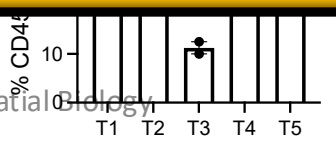


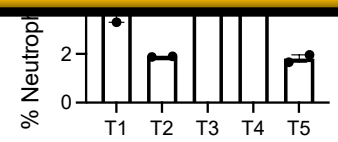


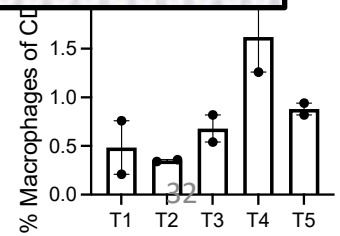
However… We are just at the beginning

- Technology-wise
- Processing wise
- With this datasets

Machine quality inc
Long time
Expensive







Exploring the Xenium Data

Joy Kabagenyi

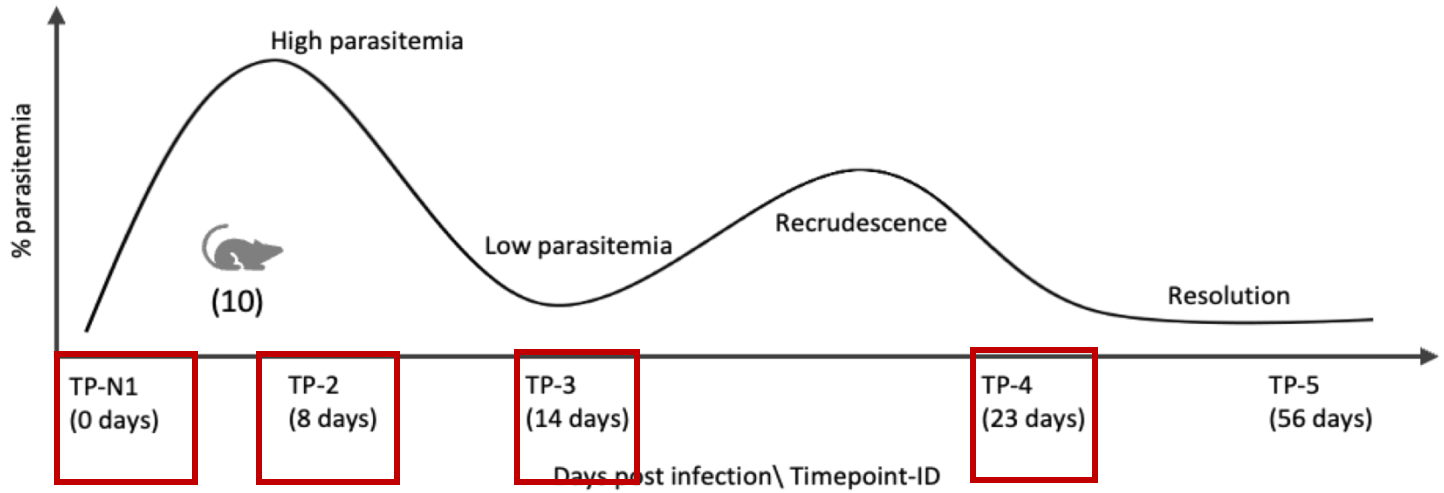
University of Glasgow | Wellcome Trust PhD Student

Supervisors | Prof. Thomas Otto | Prof. James Brewer

Spatial Chat

28-10-2025

Repeat Experiment: Experimental set up



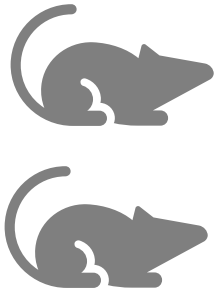
- Spleen, BM – 10x RNA sequencing
- Spleen, BM, LN – Xenium **(in red wraps)**

HTO1 == Mouse1- spleen

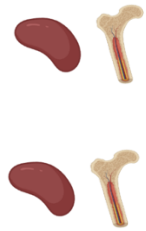
HTO2 == Mouse2- spleen

HTO3 == Mouse1- Bone marrow

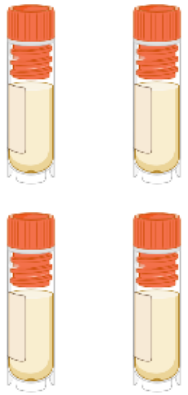
HTO4 == Mouse2- Bone marrow



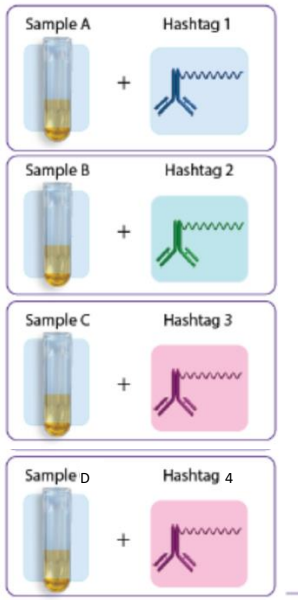
2 mice per timepoint



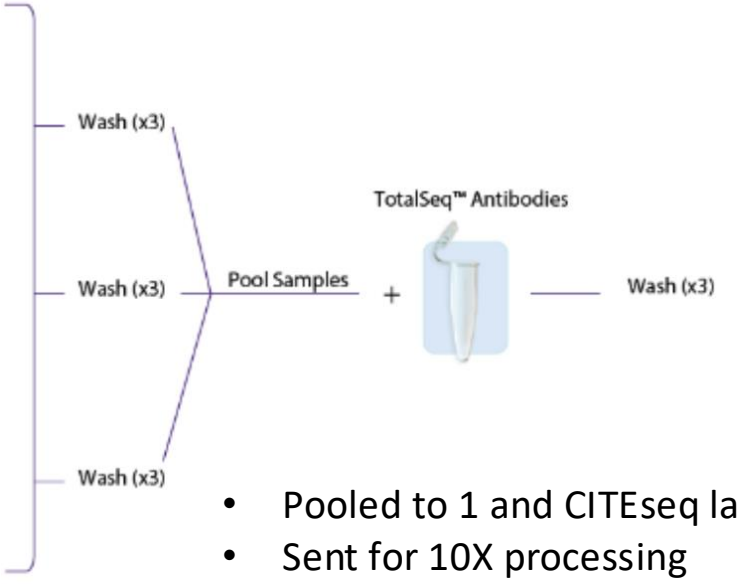
2 mice per timepoint



Into single cell suspensions



Hashtag labelled (1-4)x

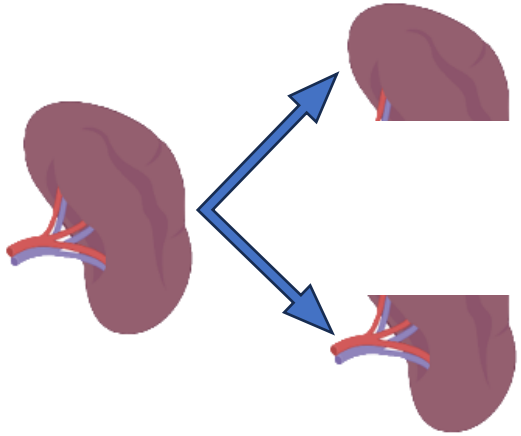


- Pooled to 1 and CITEseq labelled
- Sent for 10X processing

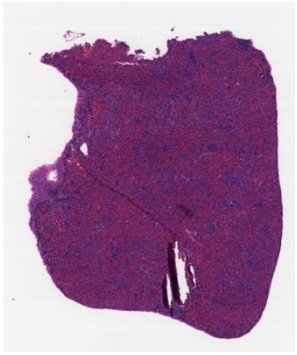
Image adopted from10X genomics

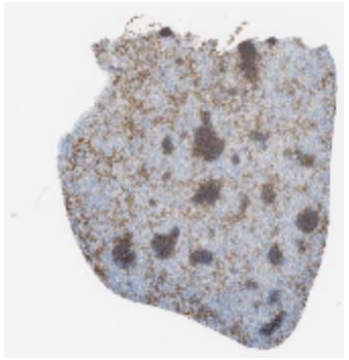
Sample processing

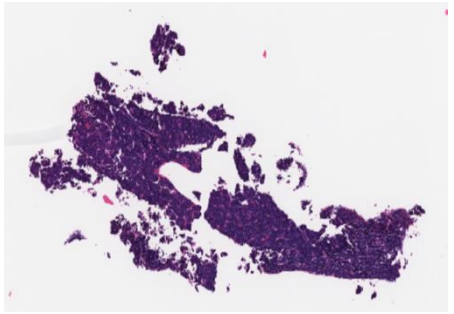
Sectioning

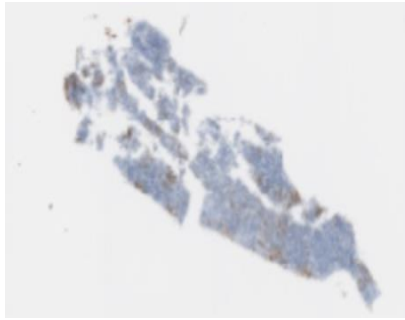


RNAscope

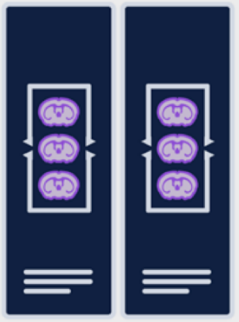




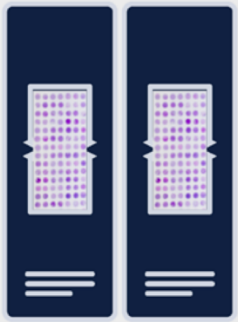




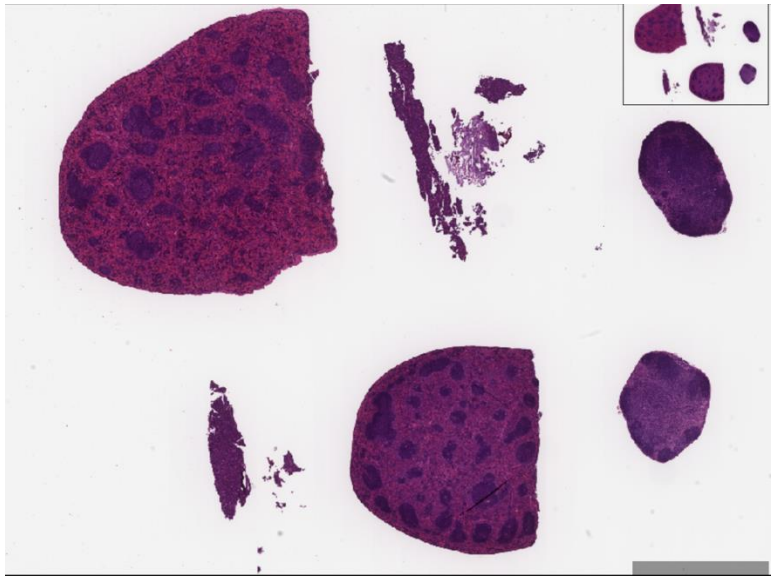
- **Quality control**
  - RNA-scope subset of samples
  - How thick should they cut the tissue
  - Sliced off the top pieces
  - Multiblock of tissues – two slide
  - 5,000 gene panels
  - 50 custom genes – 42-Pchabaudi



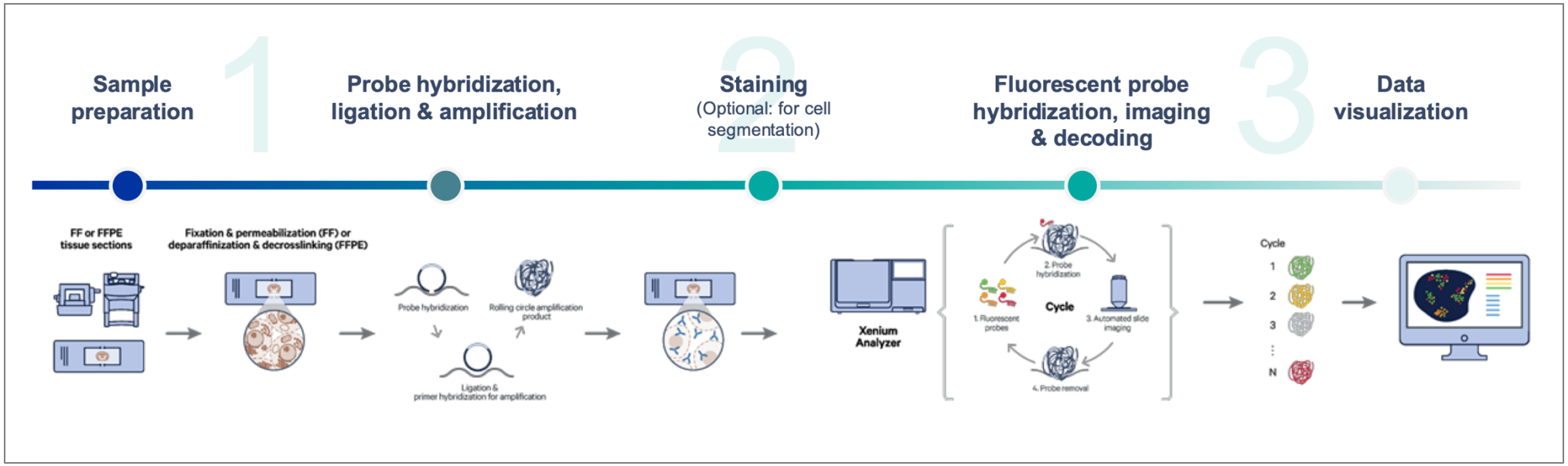
6 mouse brain sections
6 x 10 mm

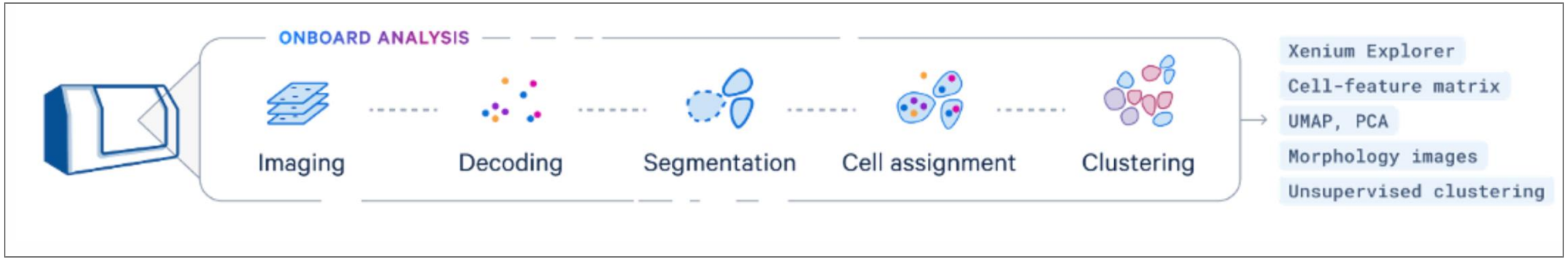


196 biopsies
1 mm tumor microarray



General QC, sample parameters and analysis approach

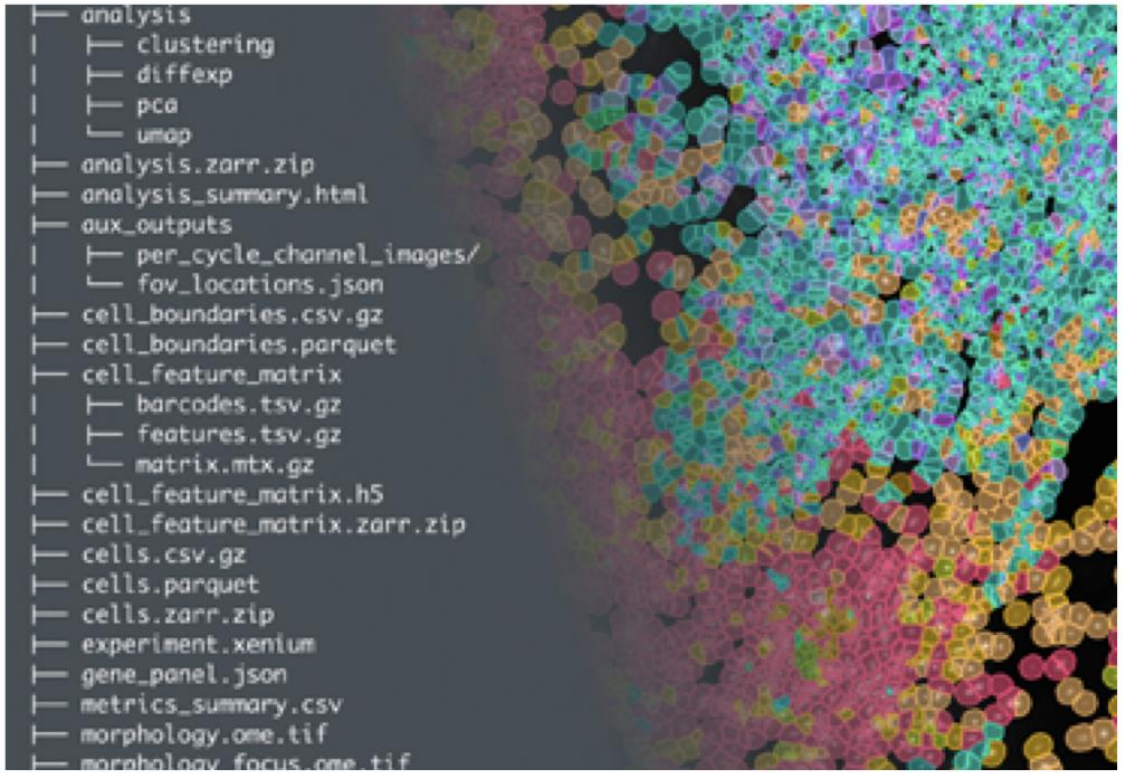




- Decoding: Gene specific probes, in each cycle of imaging, a bit of the gene's barcode, end with code word == transcript
- Transcript call == gene name matched to barcode with corresponding coordinate and qscore

Image adopted from10X genomics

After machine run processing



- Transcripts
  - Decoded transcript calls
  - Run decoding QC
  - Produces decoding plots

- Cell segmentation
  - Multimodal cell segmentation
  - Three approaches
  - Large-cell or multi-scale models: different models tuned for larger cells…?!
  - Import third-party segmentation

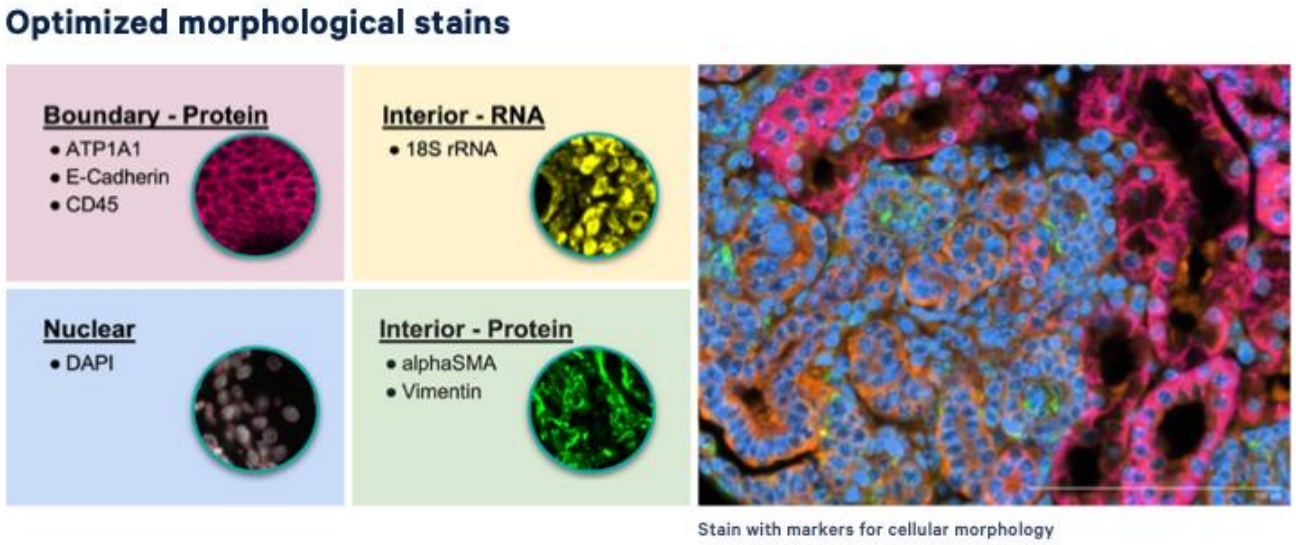
- Assigns transcripts to segmentation (cell / nucleus)
  - Produces per-cell counts (cell × gene)

- Also compute some cell and features QC metrics

- Export outputs for visualization and downstream analysis

Image adopted from10X genomics

Cell segmentation



Stains

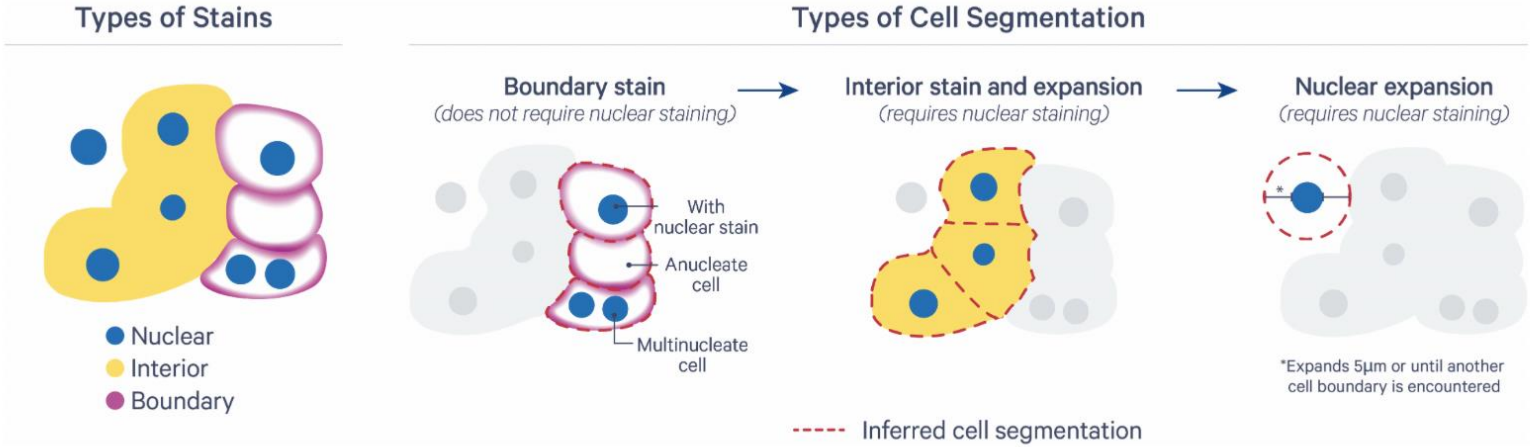
- DAPI - nuclear stain
  - binds DNA, labels the nucleus
- Cell Boundary Stains (ATP1A1, E-Cadherin and CD45)
  - labels cell membranes or cytoplasm
  - helping to define the outline of each cell
- Interior stain (18S rRNA marker)
- Interior protein: alphaSMA/Vimentin
- Nucleus is identified first – 95% pixels. Lower is discarded

Aproach

- Nucleus is identified first – 95% pixels. Lower is discarded
- Based on the DAPI stain

Image adopted from10X genomics

Cell segmentation

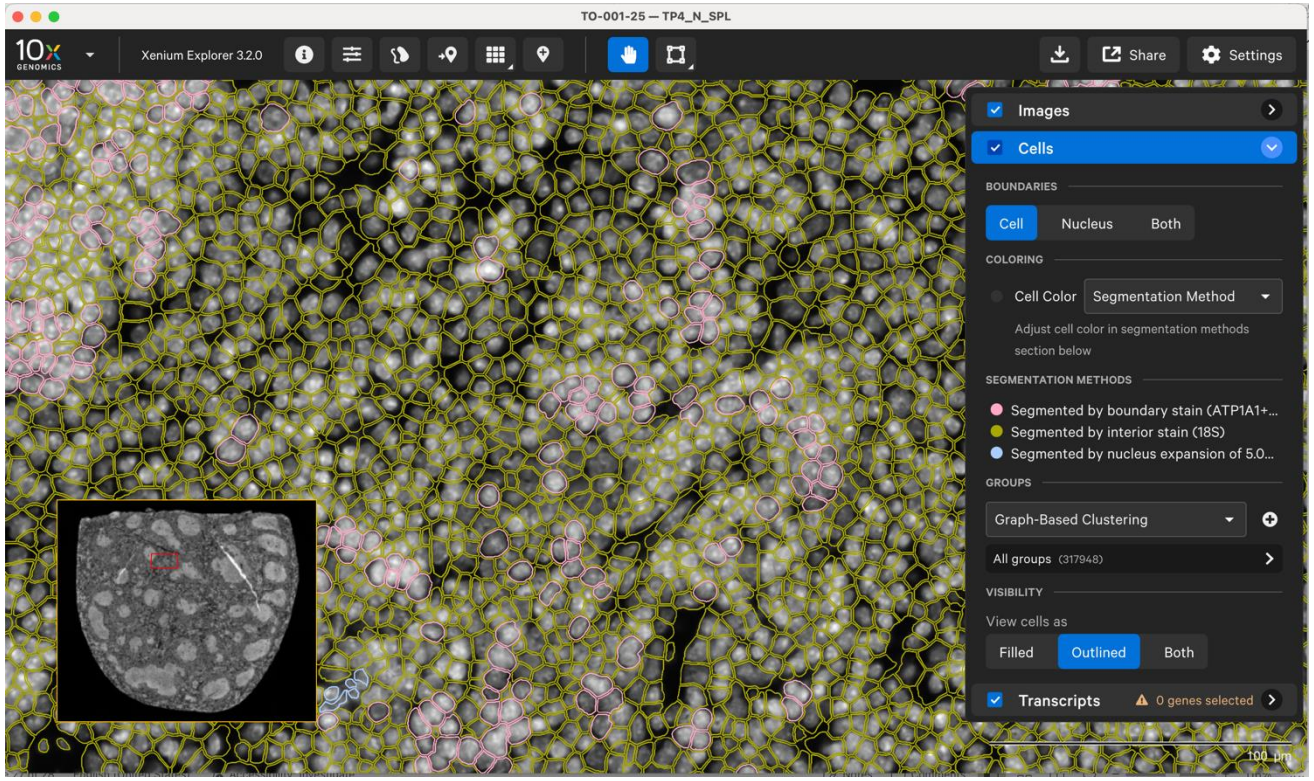


Cell segmentation approaches

- Cell boundaries and cytoplasms
  - epithelial markers (ATP1A1, E-Cadherin)
  - immune markers (pan-lymphocyte: CD45)
- Interior stain (18S rRNA marker) and the DAPI stain for nuclei
- Where cells that do not have boundary or interior stains
  - Cells segmented a nuclear (DAPI) expansion distance of 5 µm or until another cell boundary is encountered

Image adopted from10X genomics

10X Explorer: after Processing in Xenium ranger



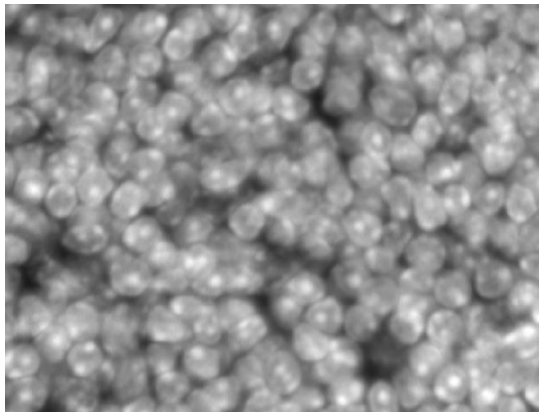
Initial visualisations and data exploration

- Images of cells
- Cell segmentation
- Cell clusters
- Transcript assignment
- General image QC checks

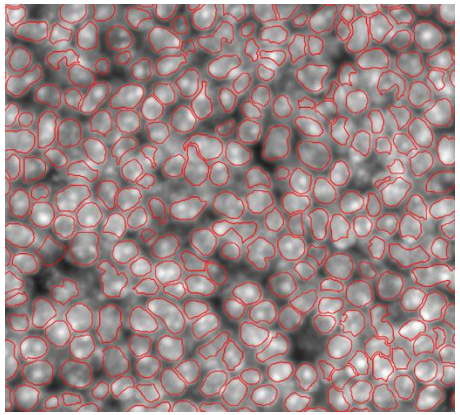
- Later visualization
  - After proper processing – reimport data

Cell segmentation 101

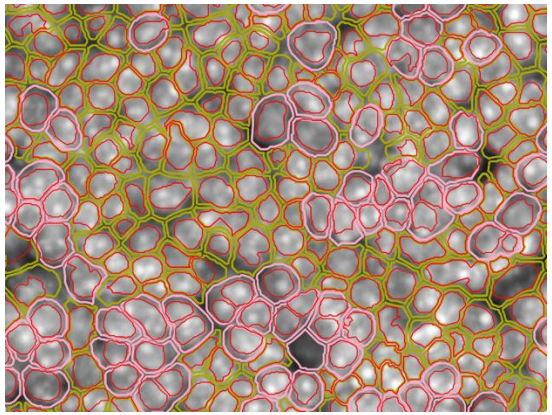
Image- Microscope



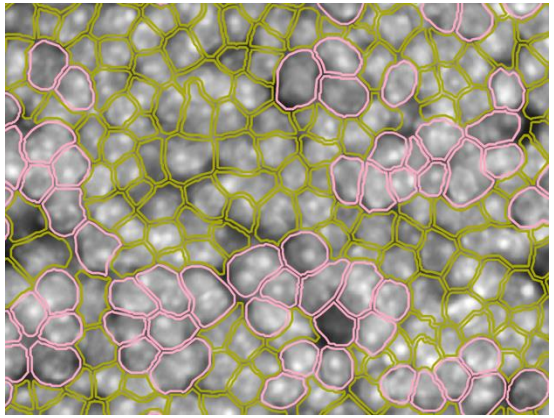
Nucleus



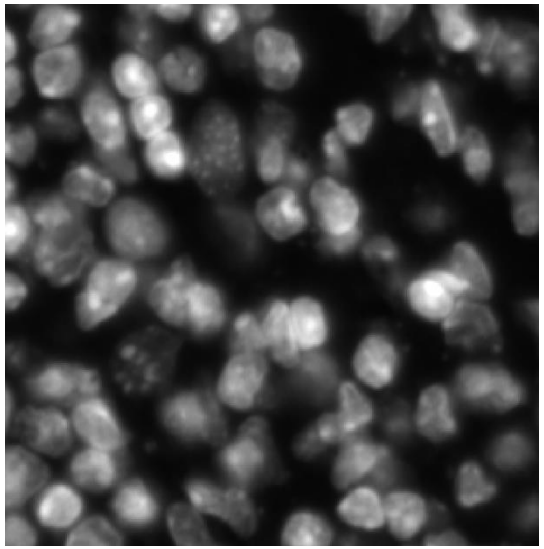
Nucleus and cell



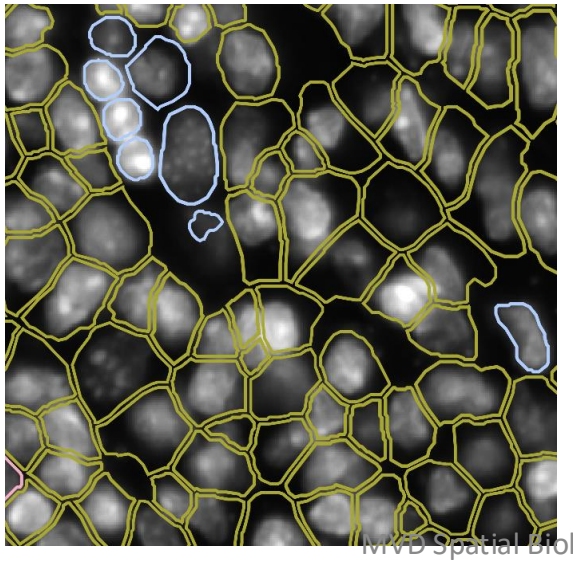
Segmentation method



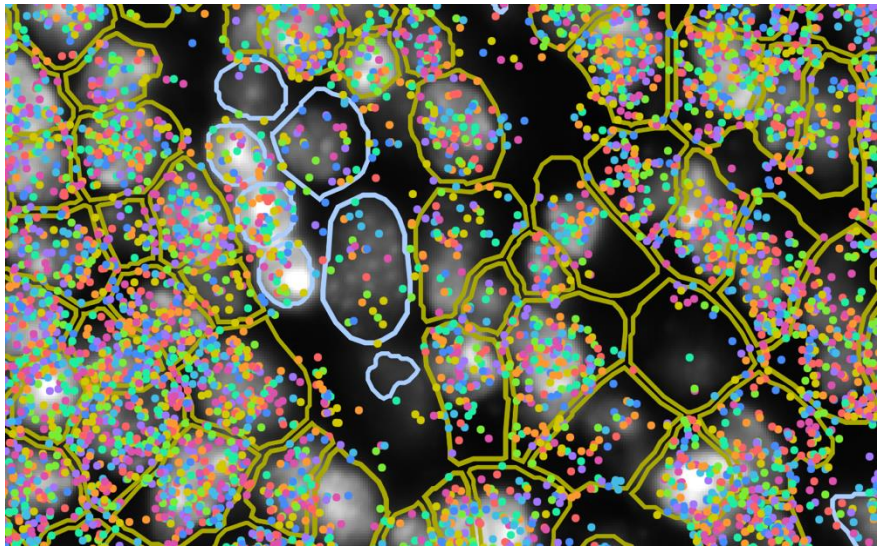
Empty spaces



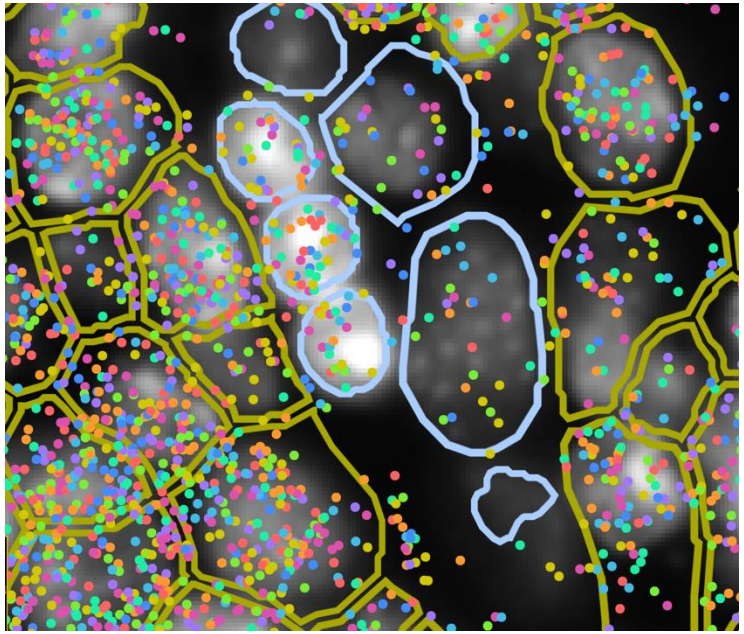
No cells in empty spaces
Nuclear expansion method



Transcripts outside of cells
Cells with no transcripts

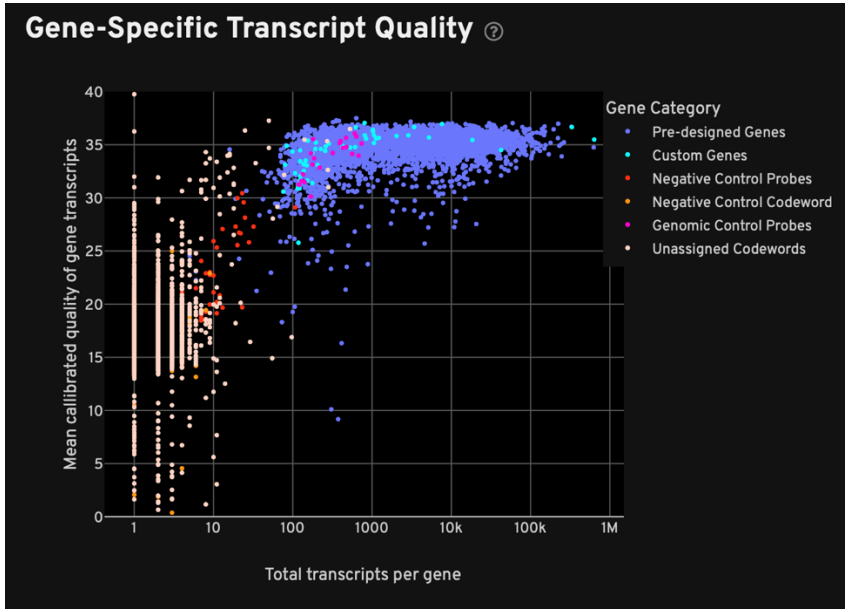


Cell segmentation and transcripts QC summary



- Transcripts outside of cells
  - Filtered at loading files in R
- Cells without transcripts
  - Filtered at processing QC – see later
- > 90% of transcripts assigned to cells – later summary table

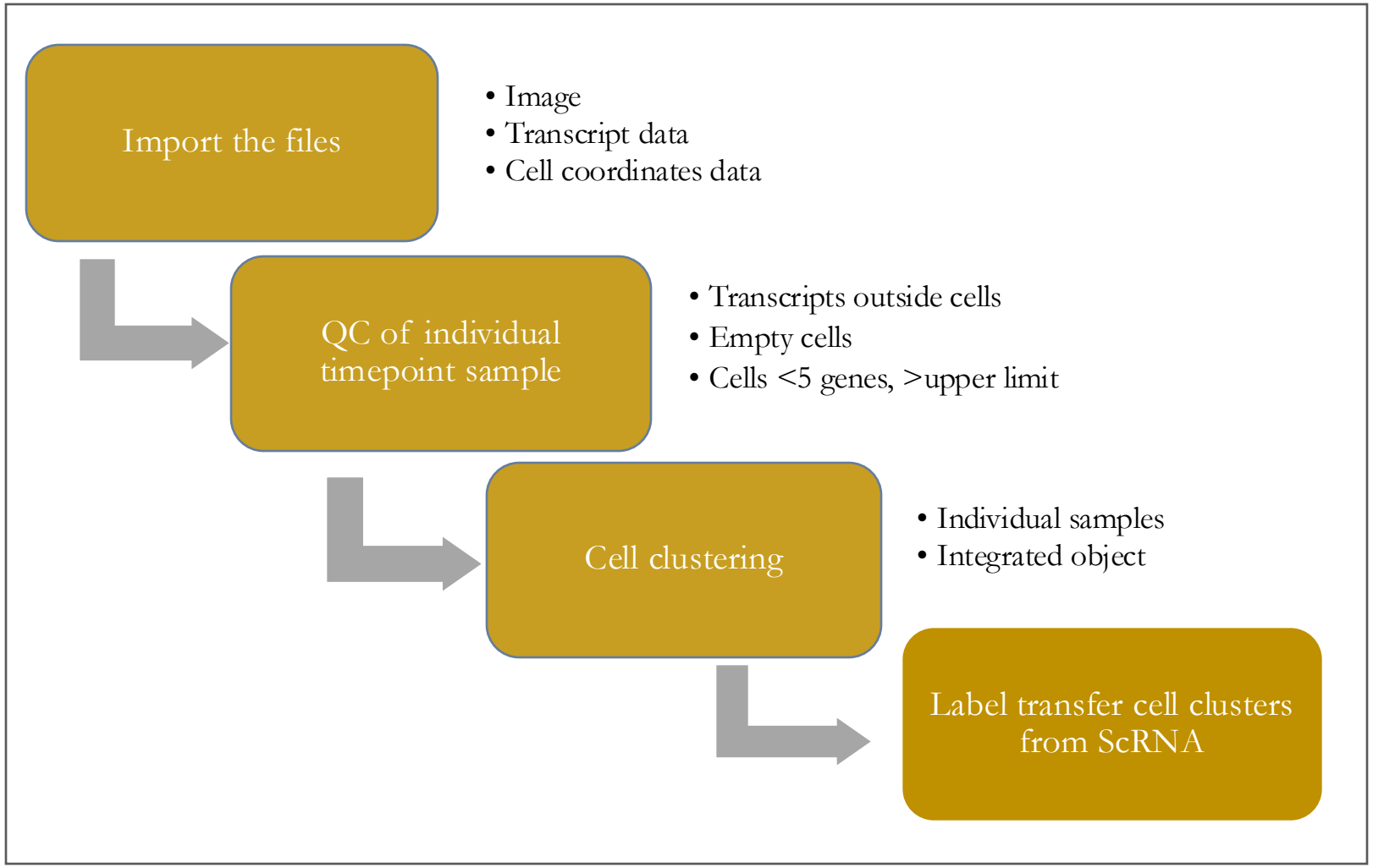
- Other considerations
- Controls in the run
  - Genomic controls
  - Negative controls



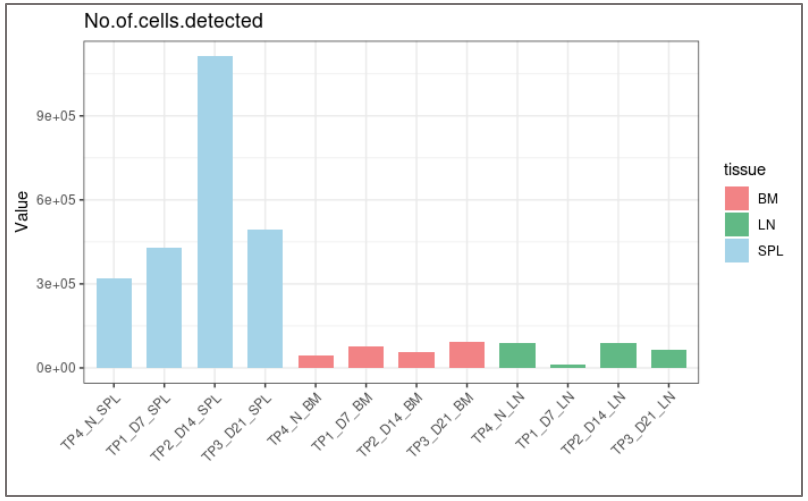
- Other cell characteristics
  - Cell area/size distribution
  - Count of nucleus, how do they relate with count of cells?
  - FOVs
    - Signal consistence
  - ncounts/ nfeatures per cell distribution

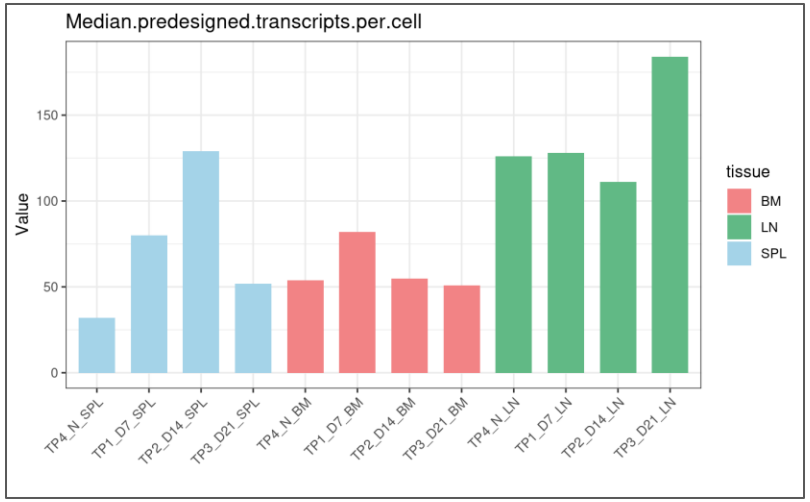
Analysis pipeline

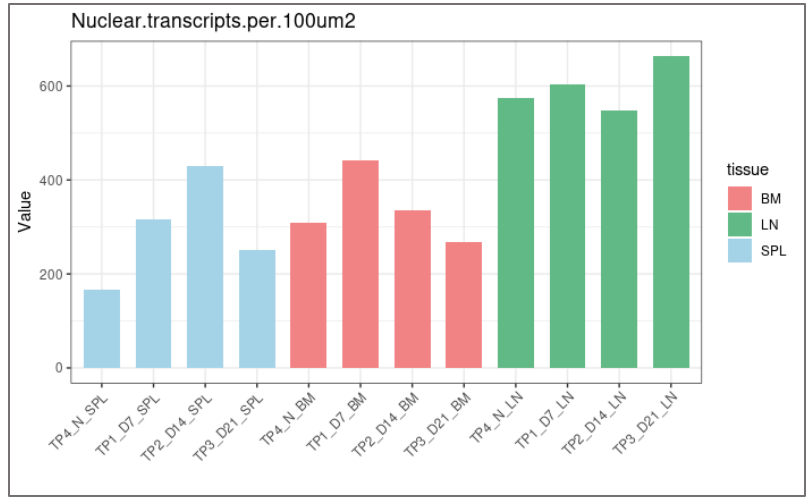
Data analysis approach

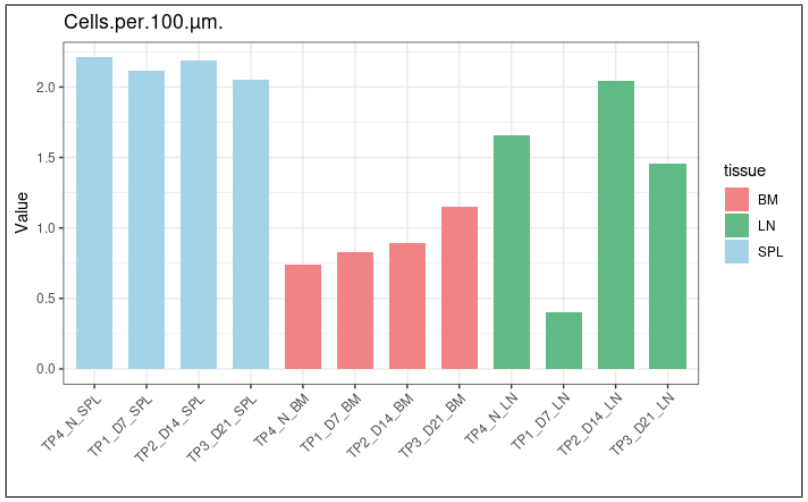


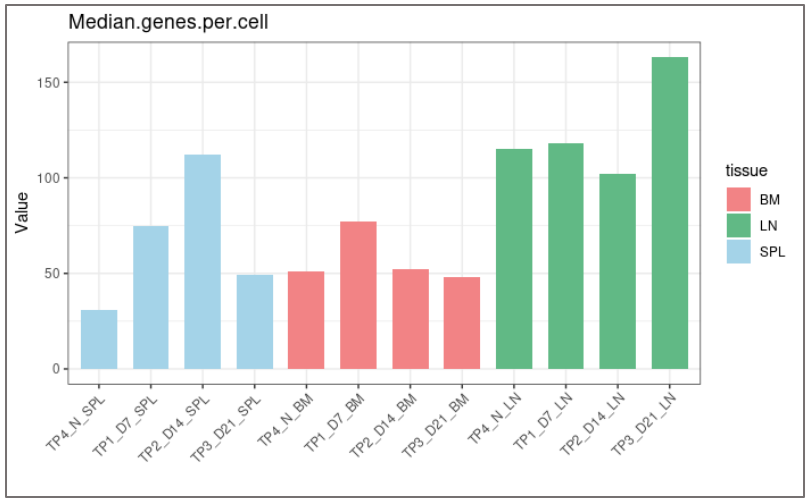
Transcripts QC summary: Expected transcripts per cell

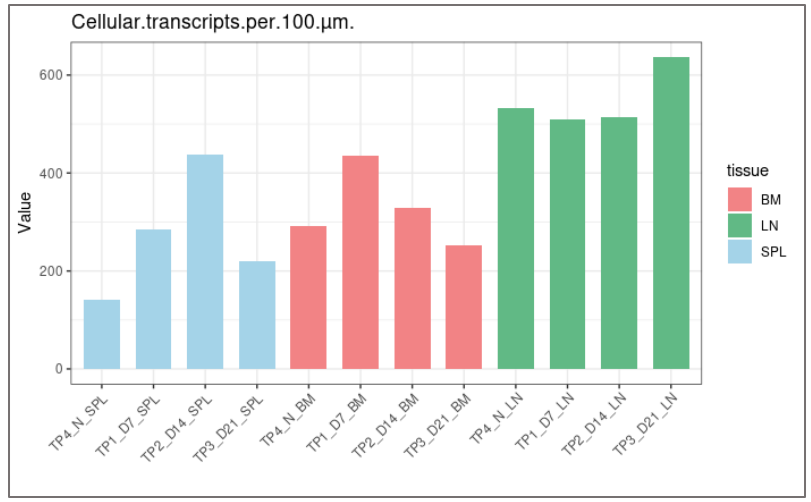




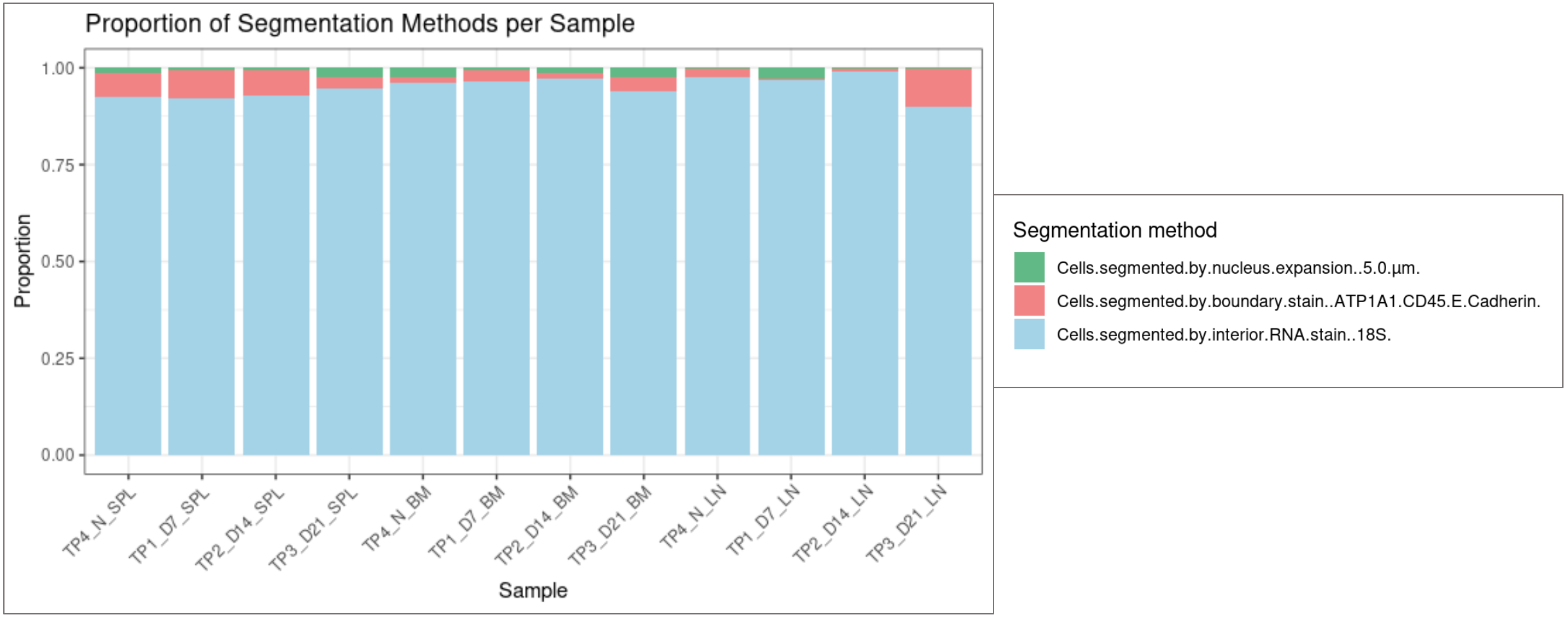








Transcripts QC summary: Expected mean transcripts per cell

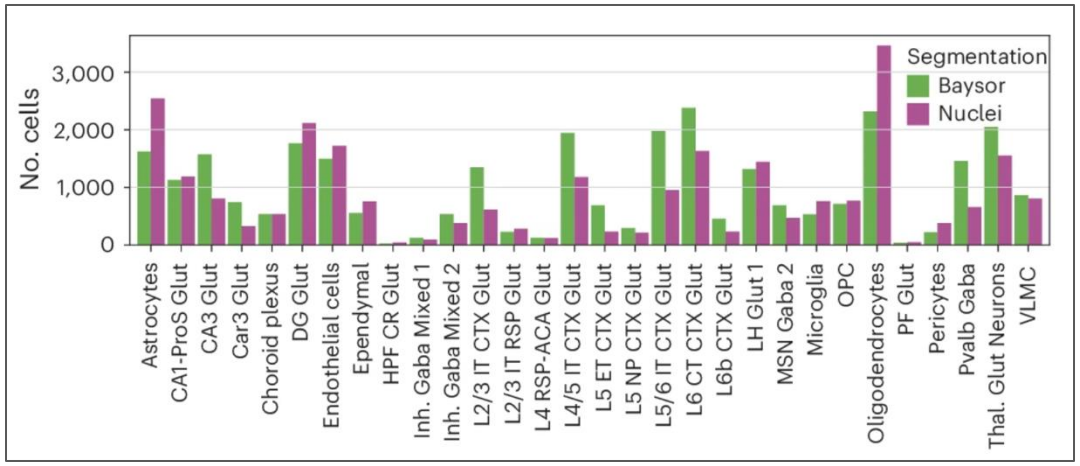


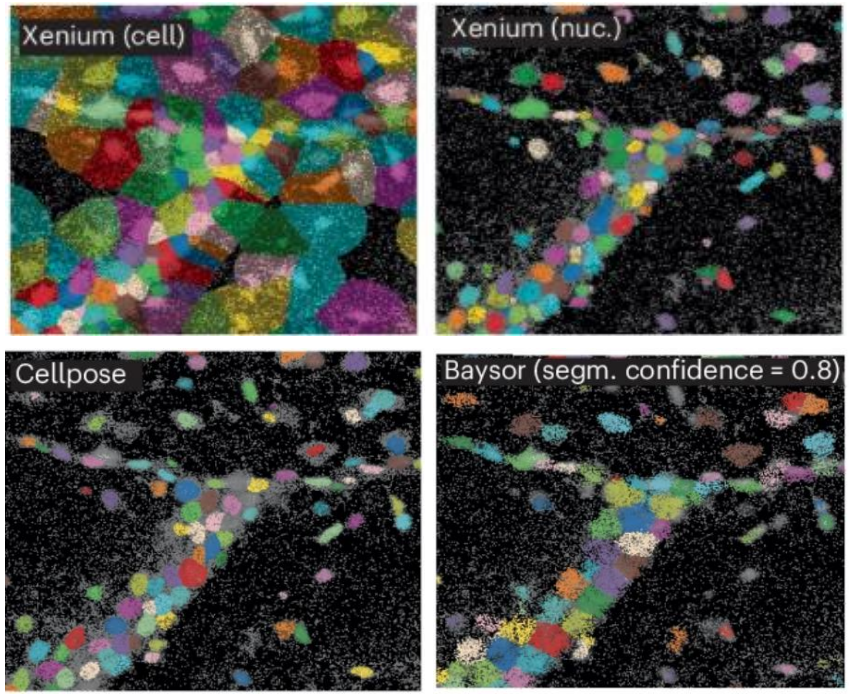
Analysis | Open access | Published: 13 March 2025

Optimizing Xenium In Situ data utility by quality assessment and best-practice analysis workflows

Sergio Marco Salas ✉, Louis B. Kuemmerle, Christoffer Mattsson-Langseth, Sebastian Tismeyer, Christophe Avenel, Taobo Hu, Habib Rehman, Marco Grillo, Paulo Czarnewski, Saga Helgadottir, Katarina Tiklova, Axel Andersson, Nima Rafati, Maria Chatzinikolaou, Fabian J. Theis, Malte D. Luecken, Carolina Wählby, Naveed Ishaque & Mats Nilsson ✉

Nature Methods **22**, 813–823 (2025) | Cite this article

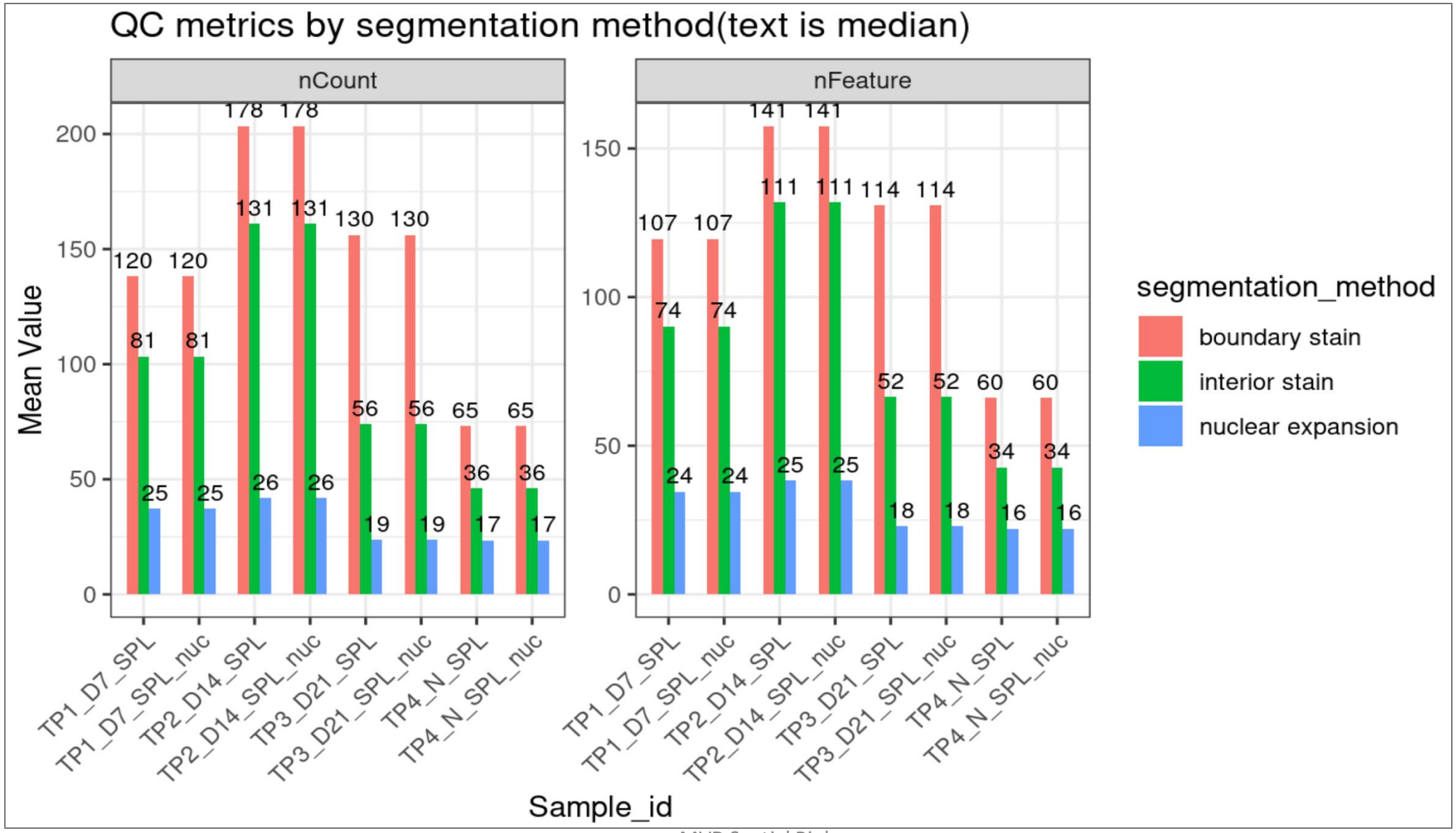




Random curiosity!

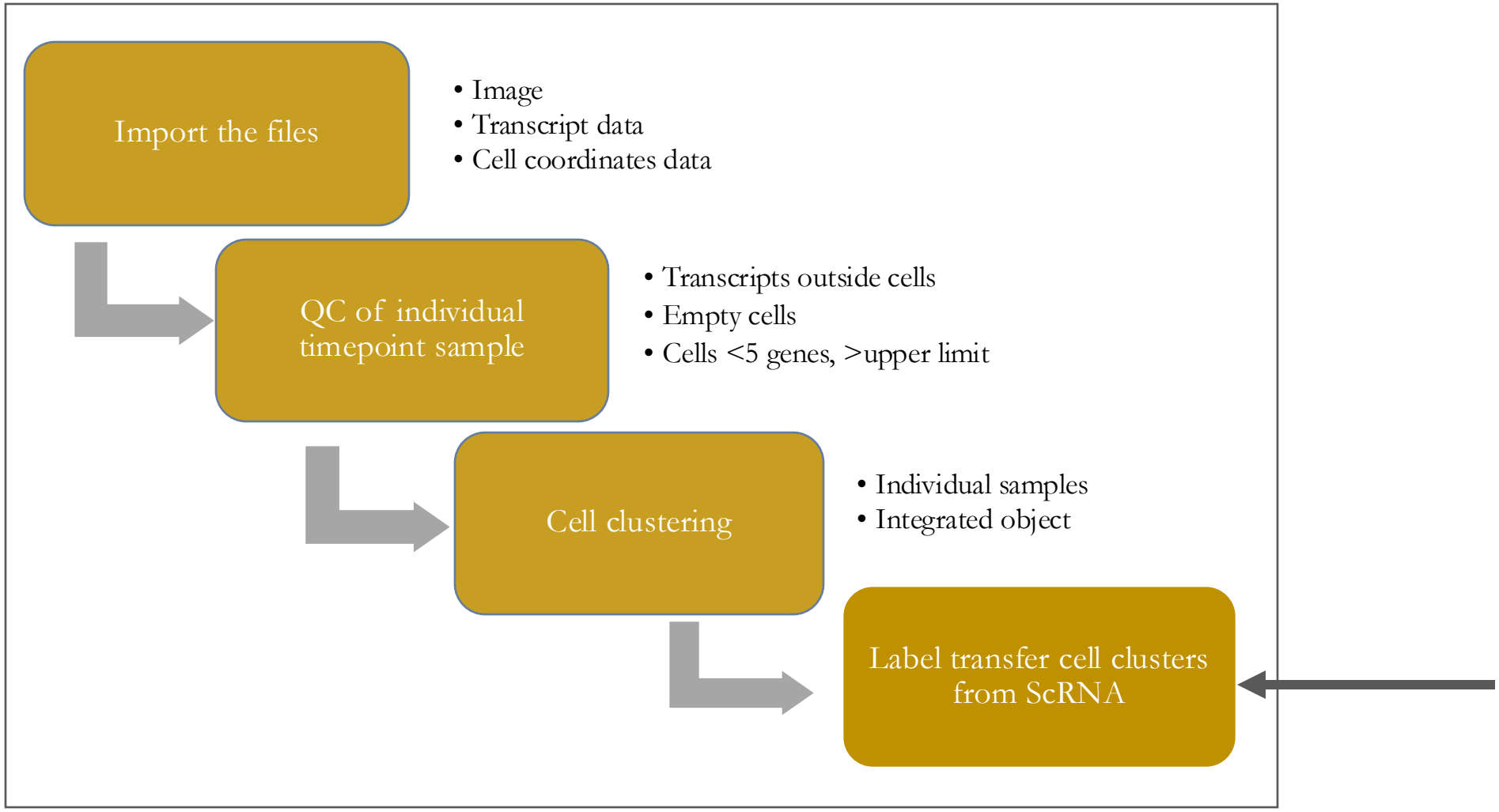
- Nucleus/cells – counts? Transcript assignment?
- New segmentation algorithm
  - All options available
  - How do the differ?
- How would choice affect my downstream CCI analysis
- Imported data with nucleus assignment and the cell assignment

Transcripts QC: Expected mean transcripts per cell by cell call strategy



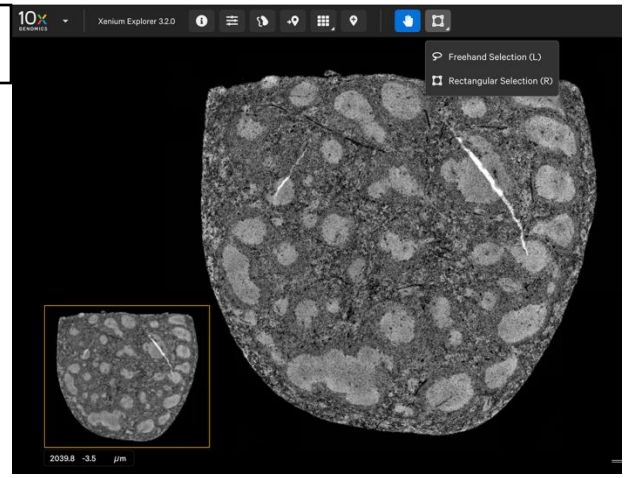
Analysis pipeline

Data analysis approach

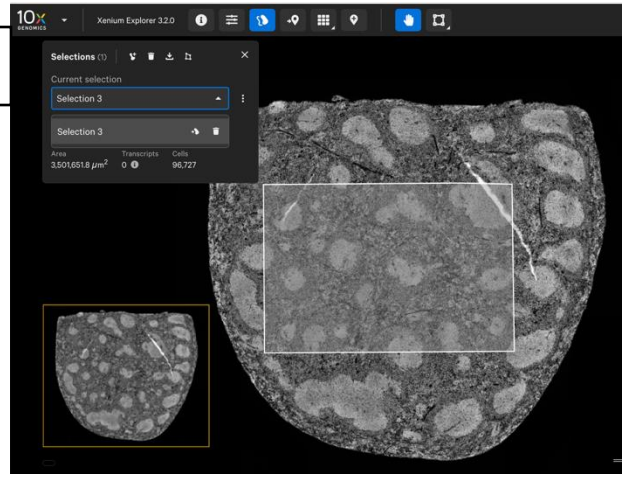


Selection of specific regions

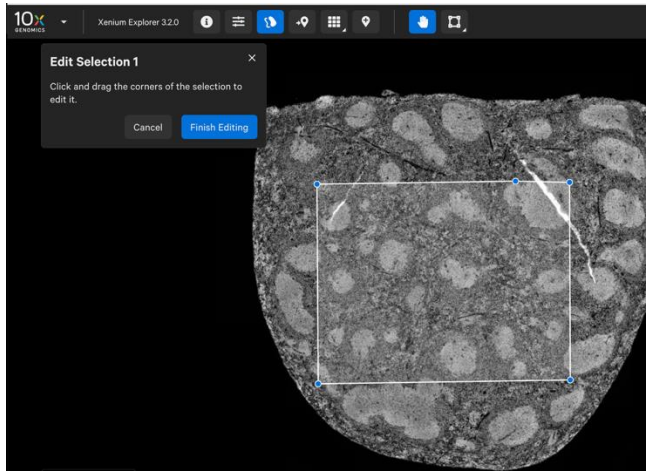
1



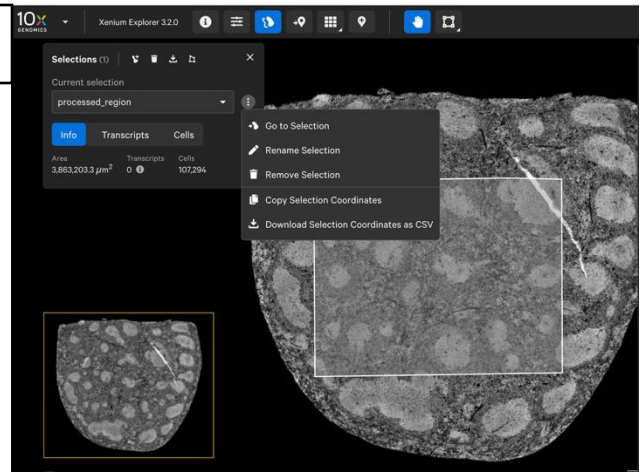
2



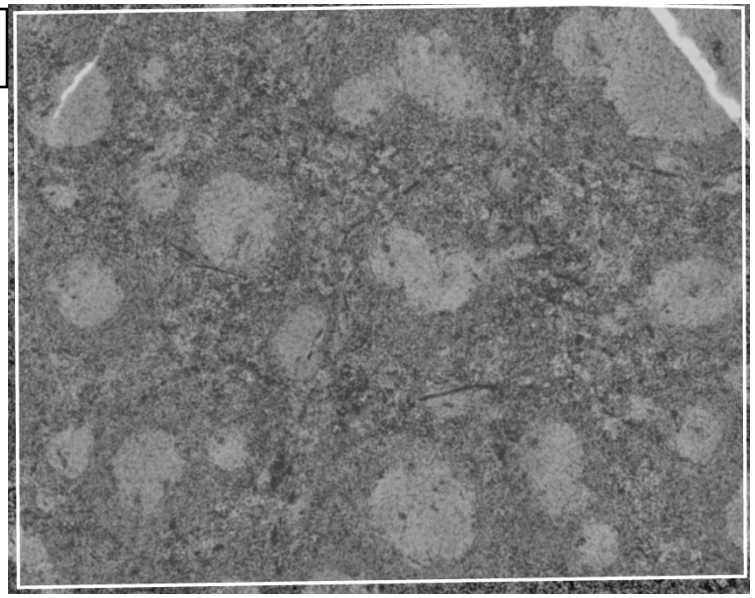
3



4



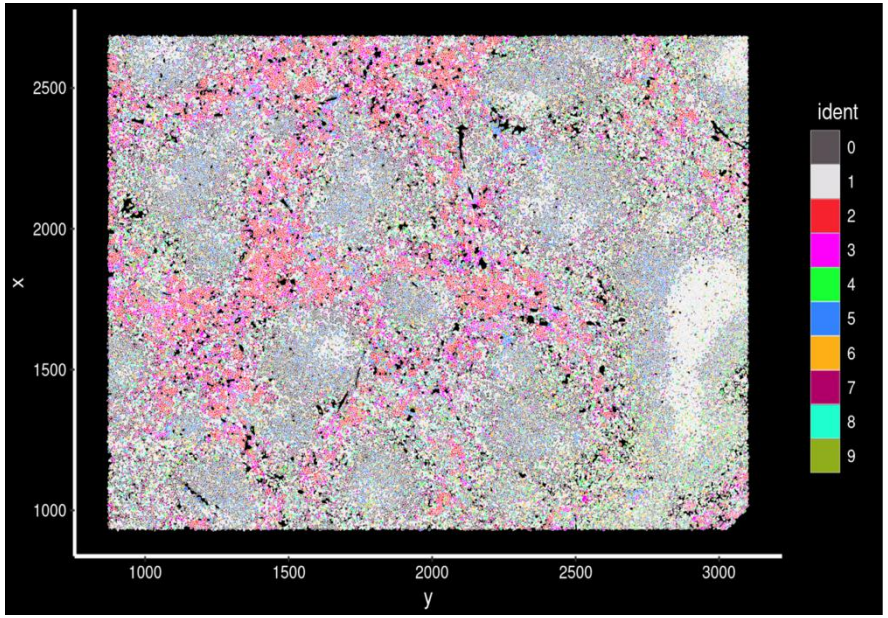
5



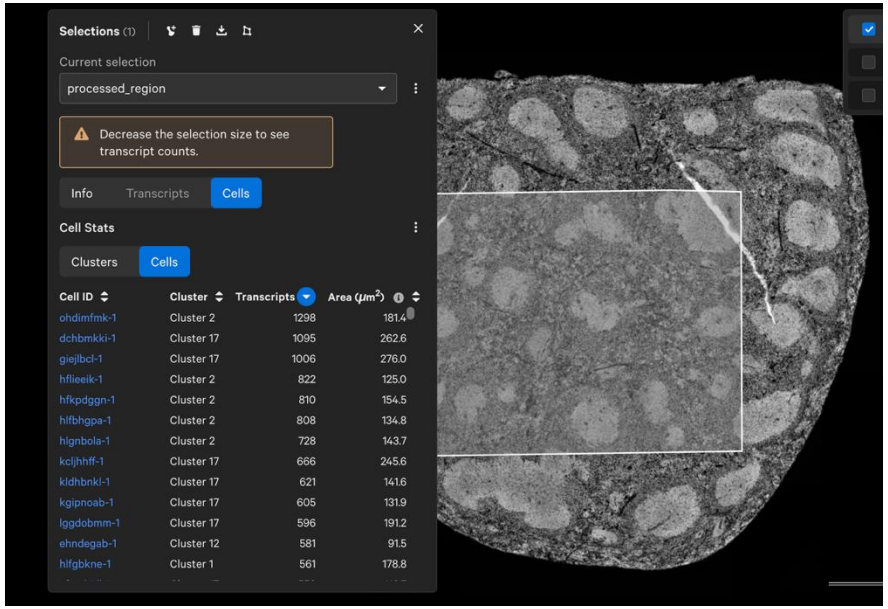
6

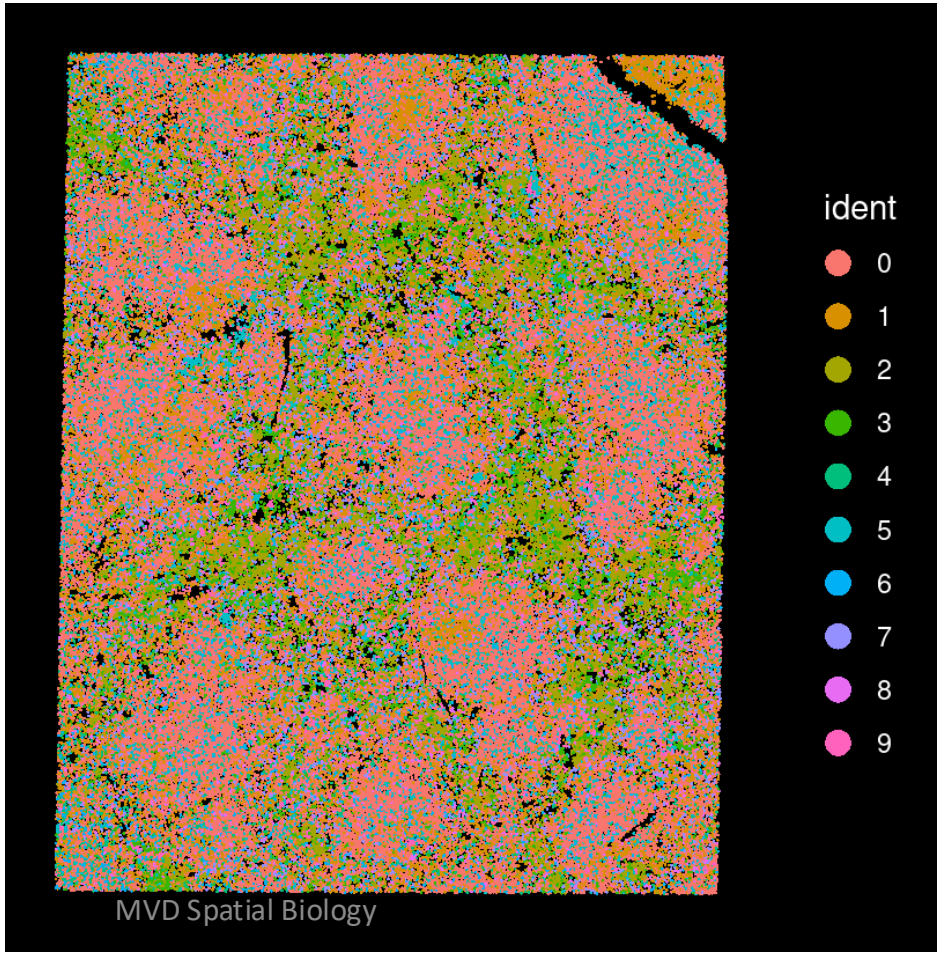
```
# Define coordinates
x_coords <- c(3097.35, 888.38, 877.81, 2617.36, 3090.39)
y_coords <- c(2679.55, 2713.33, 963.86, 937.70, 948.16)

# Define crop boundaries
x_min <- min(x_coords)
x_max <- max(x_coords)
y_min <- min(y_coords)
y_max <- max(y_coords)

x_min; x_max; y_min; y_max

# Crop the image
cropped.coords <- Crop(dpi_0[["fov"]],
                       x = c(x_min, x_max),
                       y = c(y_min, y_max),
                       coords = "plot")
dpi_0[["zoom"]] <- cropped.coords

# visualize cropped area with cell segmentations & selected molecules
DefaultBoundary(dpi_0[["zoom"]]) <- "segmentation"
ImageDimPlot(dpi_0, fov = "zoom", axes = TRUE, border.color = "white", border.size = 0.1, cols = "polychrome",
    coord.fixed = FALSE,
    #molecules = c("Gad1", "Sst", "Npy2r", "Pvalb", "Nrn1"), nmols = 10000
    )
```

Extract the cell IDs in your region of interest



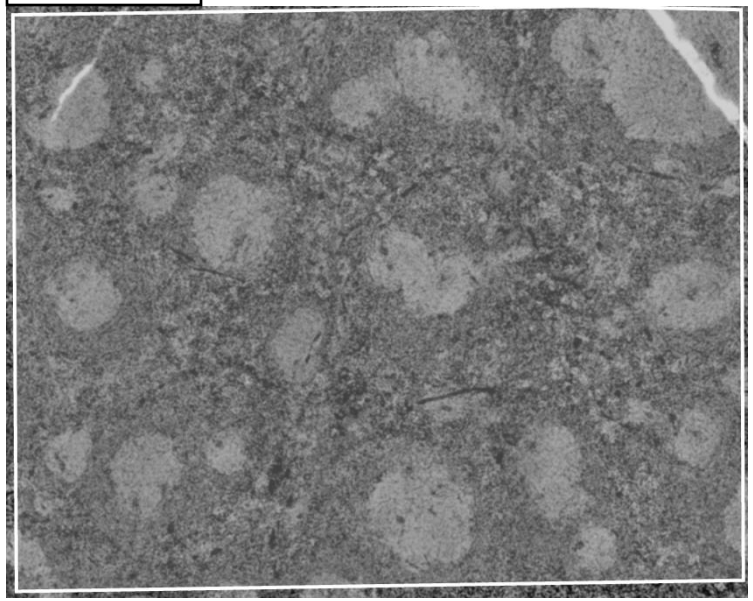

```
head(dpi_0@meta.data)
library(readr)
dpi0_region <- read_csv("/datastore/2670578k/Xenium_data/dpi0_processed_region_cells_stats.csv")
View(dpi0_region)

# create meta data to subset on
dpi_0$cell_id <- colnames(dpi_0)

# subset the cells
dpi_0_subset <- subset(dpi_0, subset = cell_id %in% dpi0_region$`Cell ID`)
```



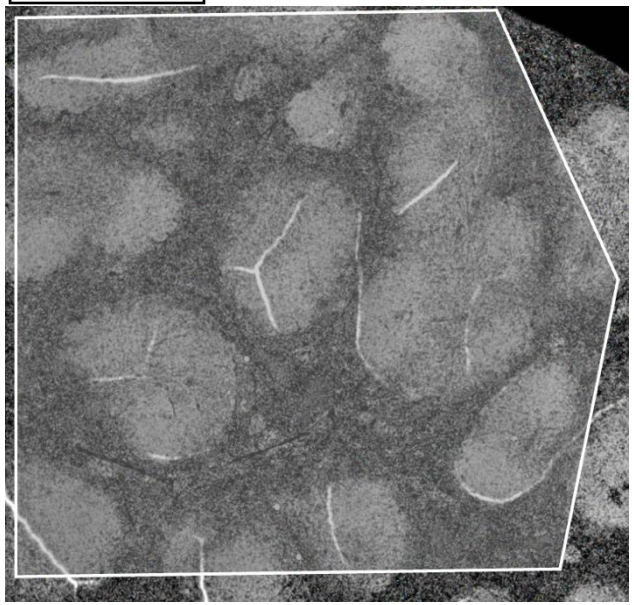


Process the selected data

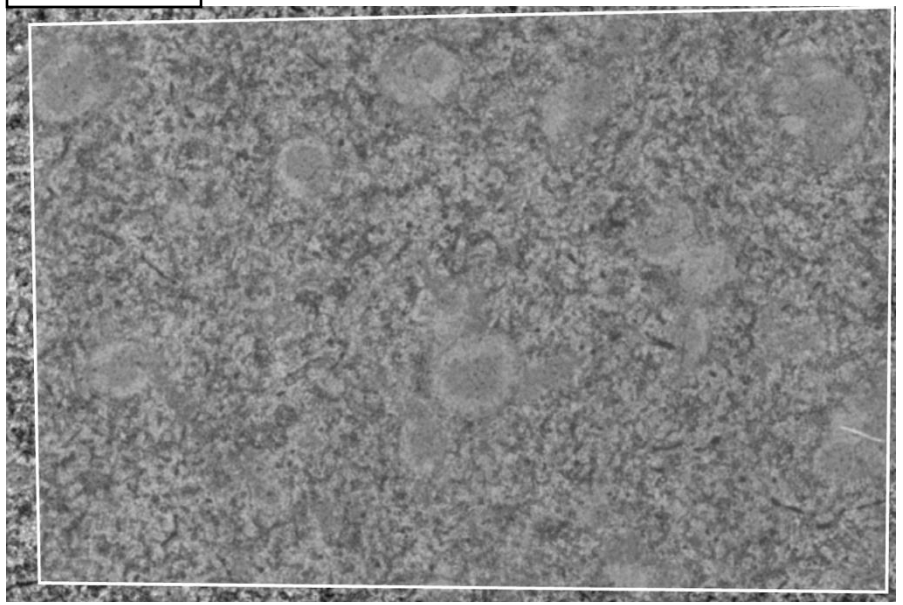
Dpi_0



Dpi_7



Dpi_14

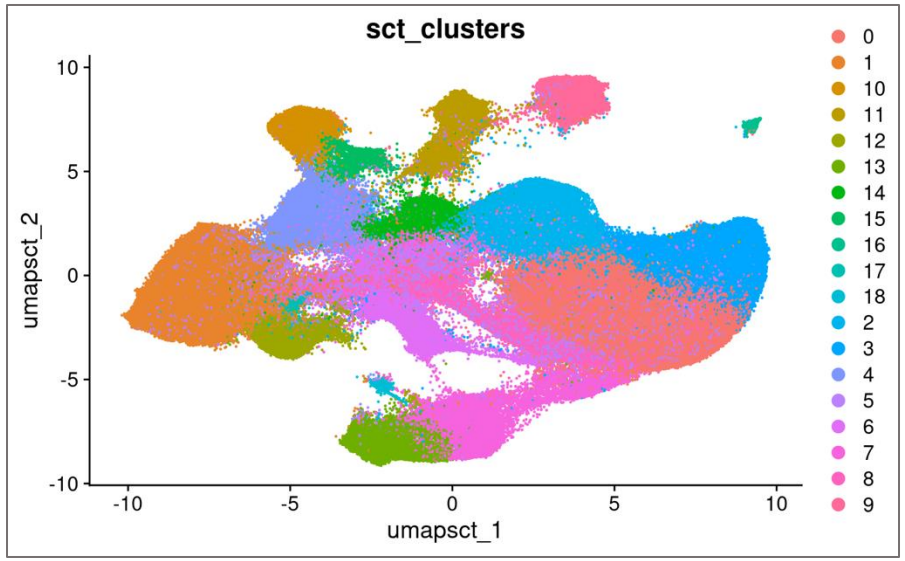


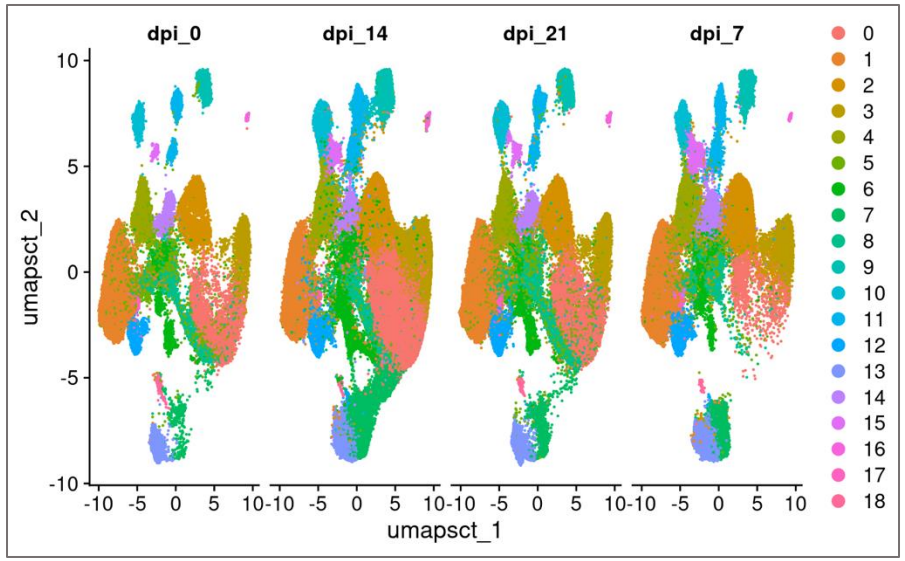
Normal Seurat pipeline

- Merge
- Normalise - Normal
- SCTransform
- Integrate – Harmony
- Cluster labeling

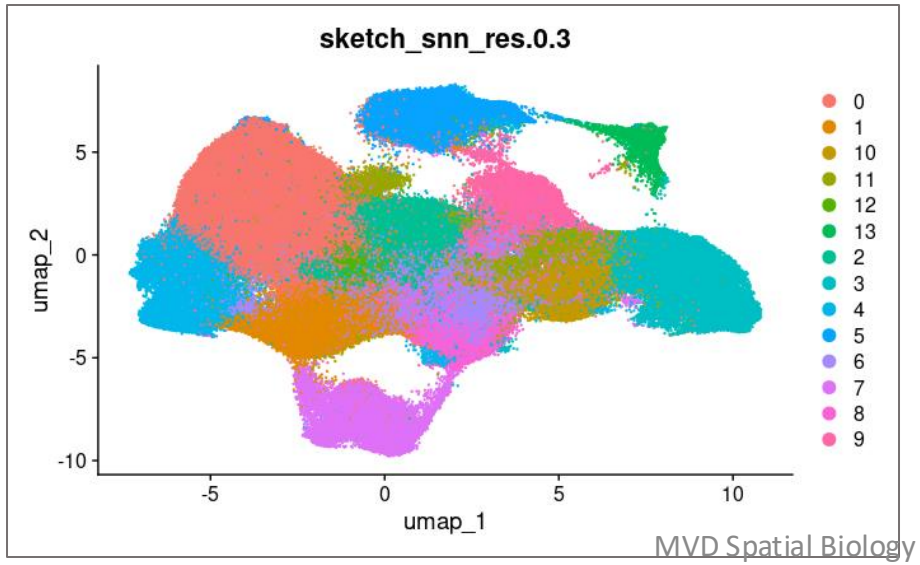
Integration: script vs rstudio, Label transfer: 10x pipeline may not work

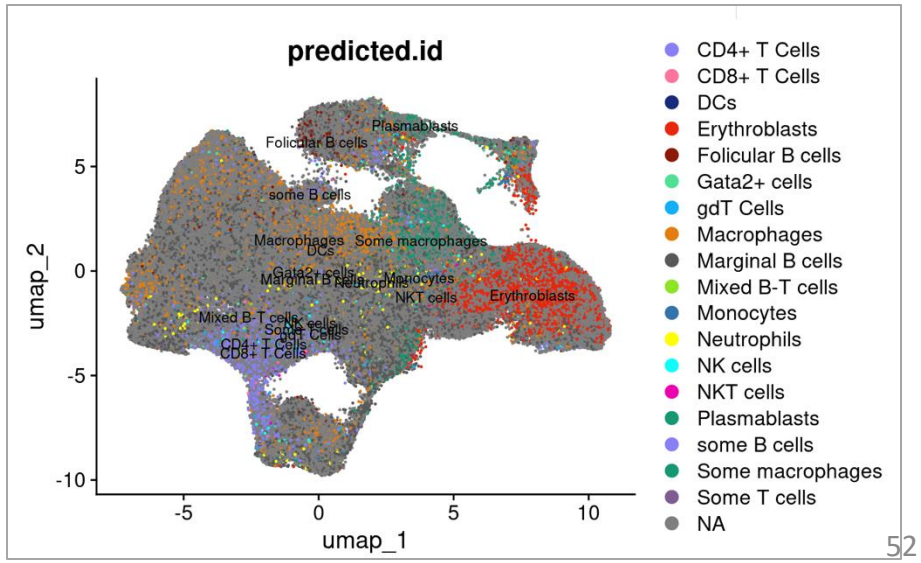
INTEGRATION





LABEL TRANSFER

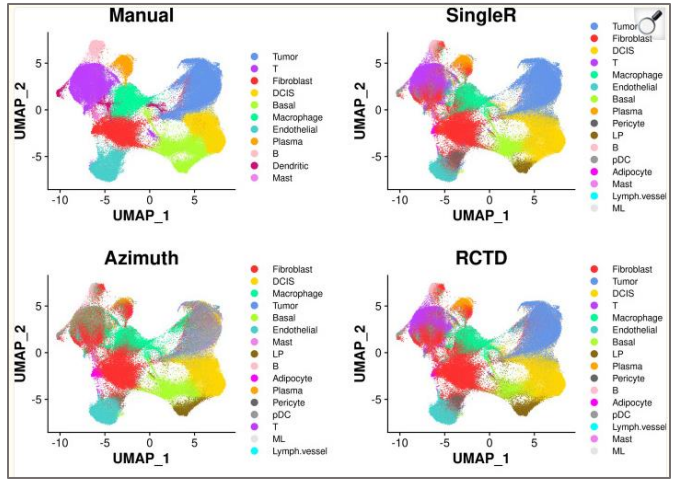


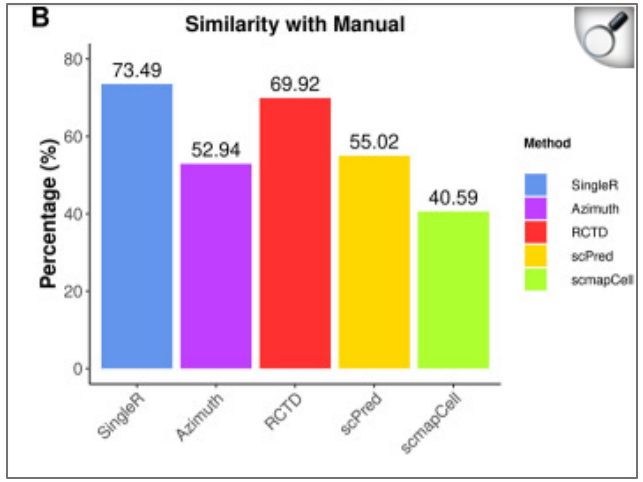


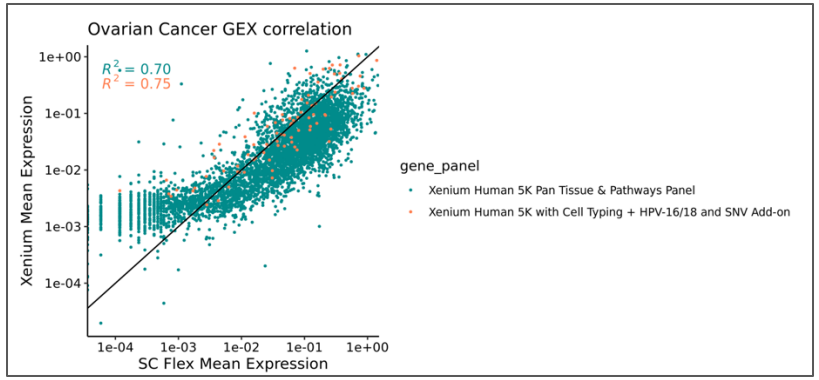
BMC Bioinformatics. 2025 Jan 20;26:22. doi: 10.1186/s12859-025-06044-0

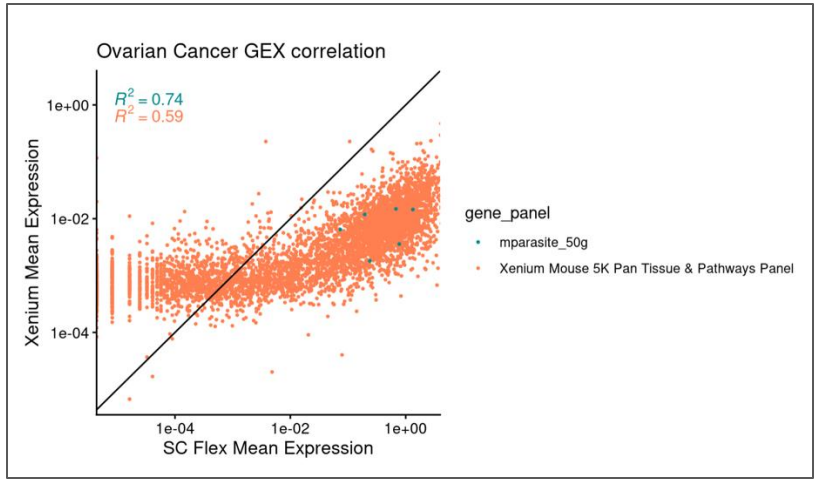
Benchmarking cell type annotation methods for 10x Xenium spatial transcriptomics data

Jinming Cheng [1,2], Xinyi Jin [1,2], Gordon K Smyth [1,3,✉], Yunshun Chen [1,2,4,✉]









- Poor correlation for my data
- Integrate with scRNA-seq data and manually annotate
- Marker genes shared between dataset – use for determining cell Ids
- Hashtags out- Try again

Take home

1. May consider RNAscope, multiblock of samples
2. Attention to QC paramenters – Decide what to do with nuclear expansion cells
3. To choose cell/nucleus and how does it affect transcript assignment – any works the same
4. Subsetting helps - server work around. Also, no parallelization (esp. SCTransform)
5. Manual annotation vs label transfer

Analysis pipeline

- Similar to scRNA
- QC, integration, PCA etc
- Annotation of cluster
- Projections to images

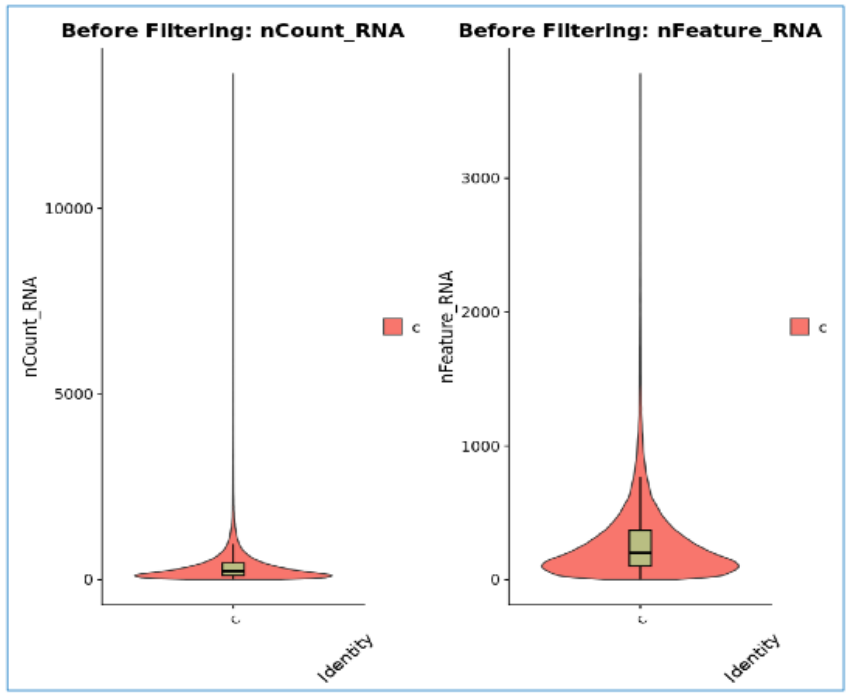
- Vignette on Seurat page

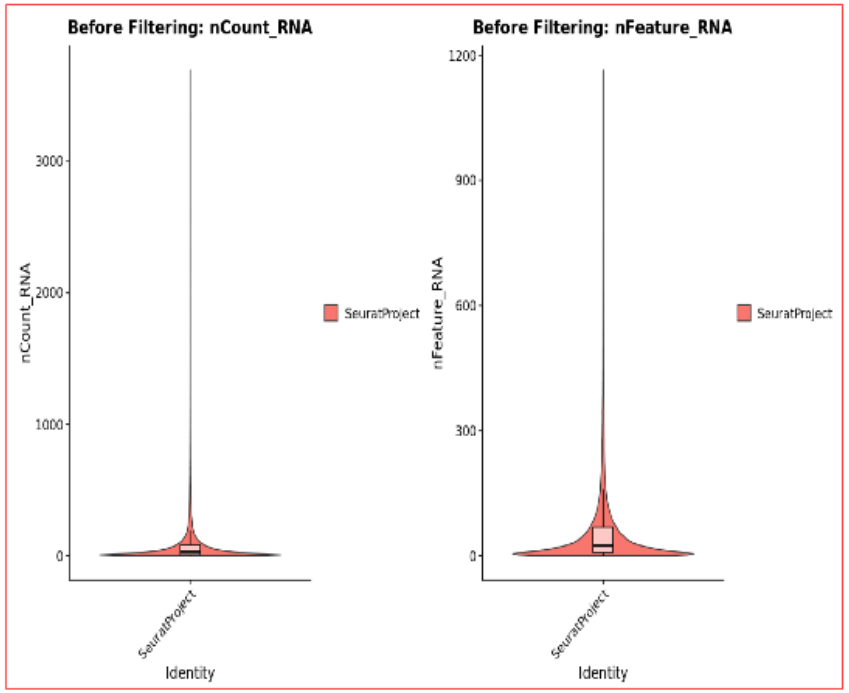
Show samples with parasites in Explorer!

Comparison: what is better?

- Depends who you asked, for us

QC summary (CosMX vs. Xenium)

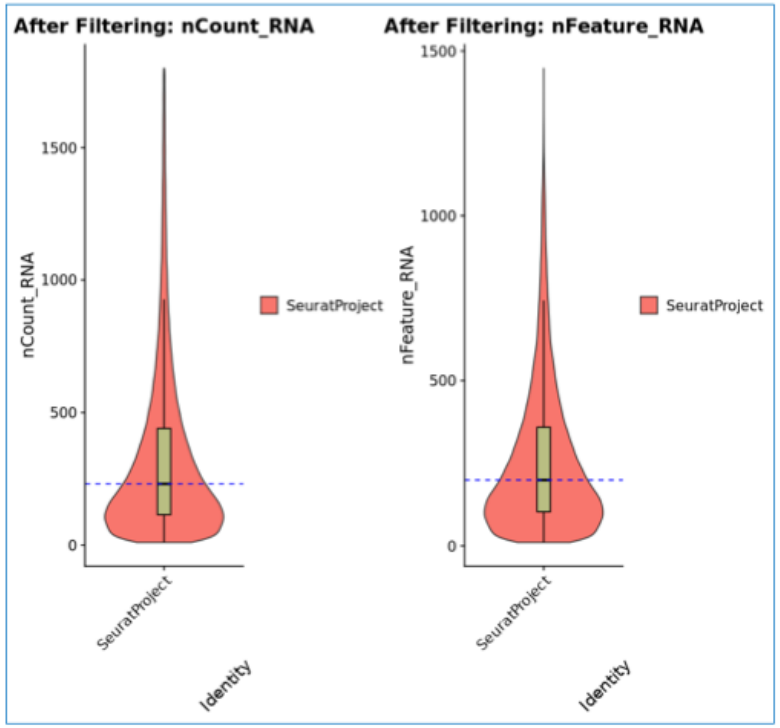


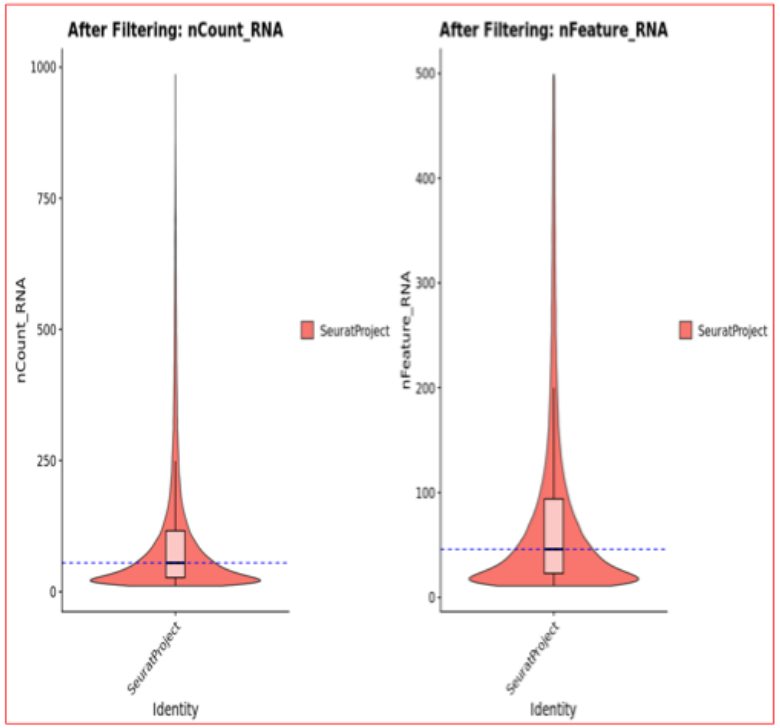


Platform	Parameter	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Total cells
CosMX	nCount_RNA	5	114	232	366.7	452.5	13626	221011
	nFeature_RNA	2	102	201	281.2	369	3781	
Xenium	nCount_RNA	0	9	30	79.94	83	3694	141682
	nFeature_RNA	0	8	25	60.07	69	1165	

1

QC summary (CosMX vs. Xenium)





Platform	Parameter	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Total cells
CosMX	nCount_RNA	10	115	231	330.3	440	1800	214711
	nFeature_RNA	10	103	199	263	359	1447	
Xenium	nCount_RNA	11	27	55	101.5	116	986	97339
	nFeature_RNA	11	23	46	78.37	94	499	

2

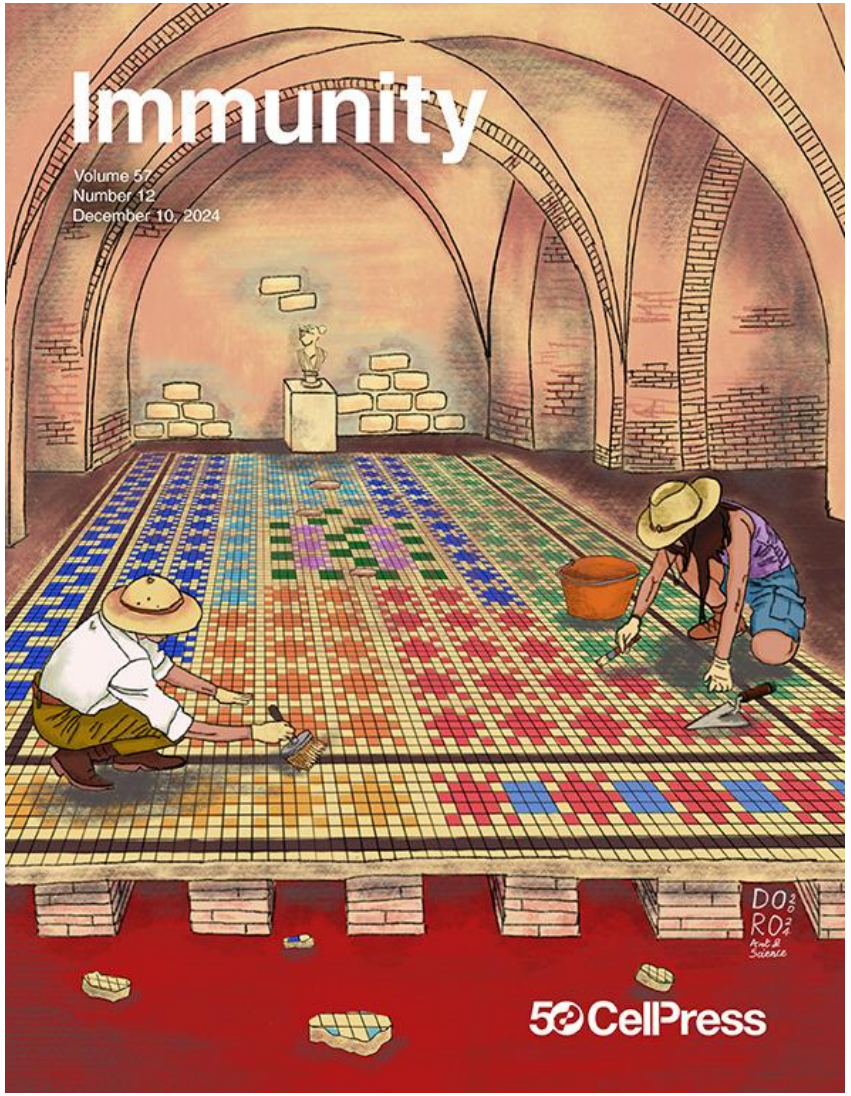
RA examples *again*

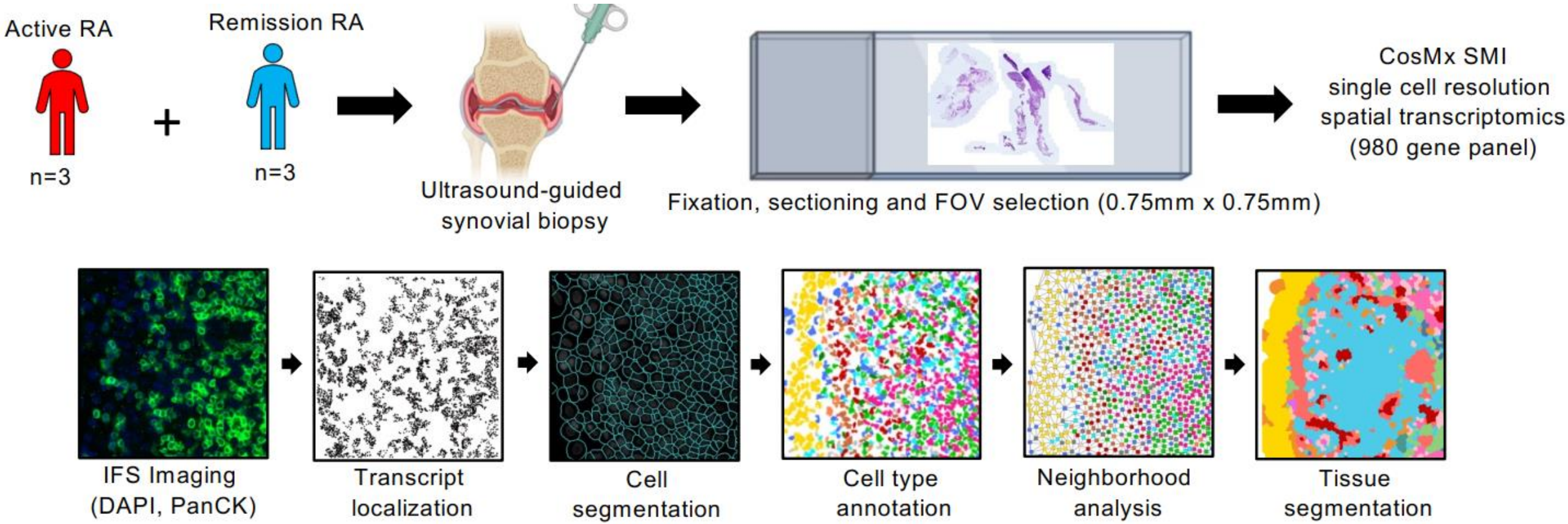
Immunity

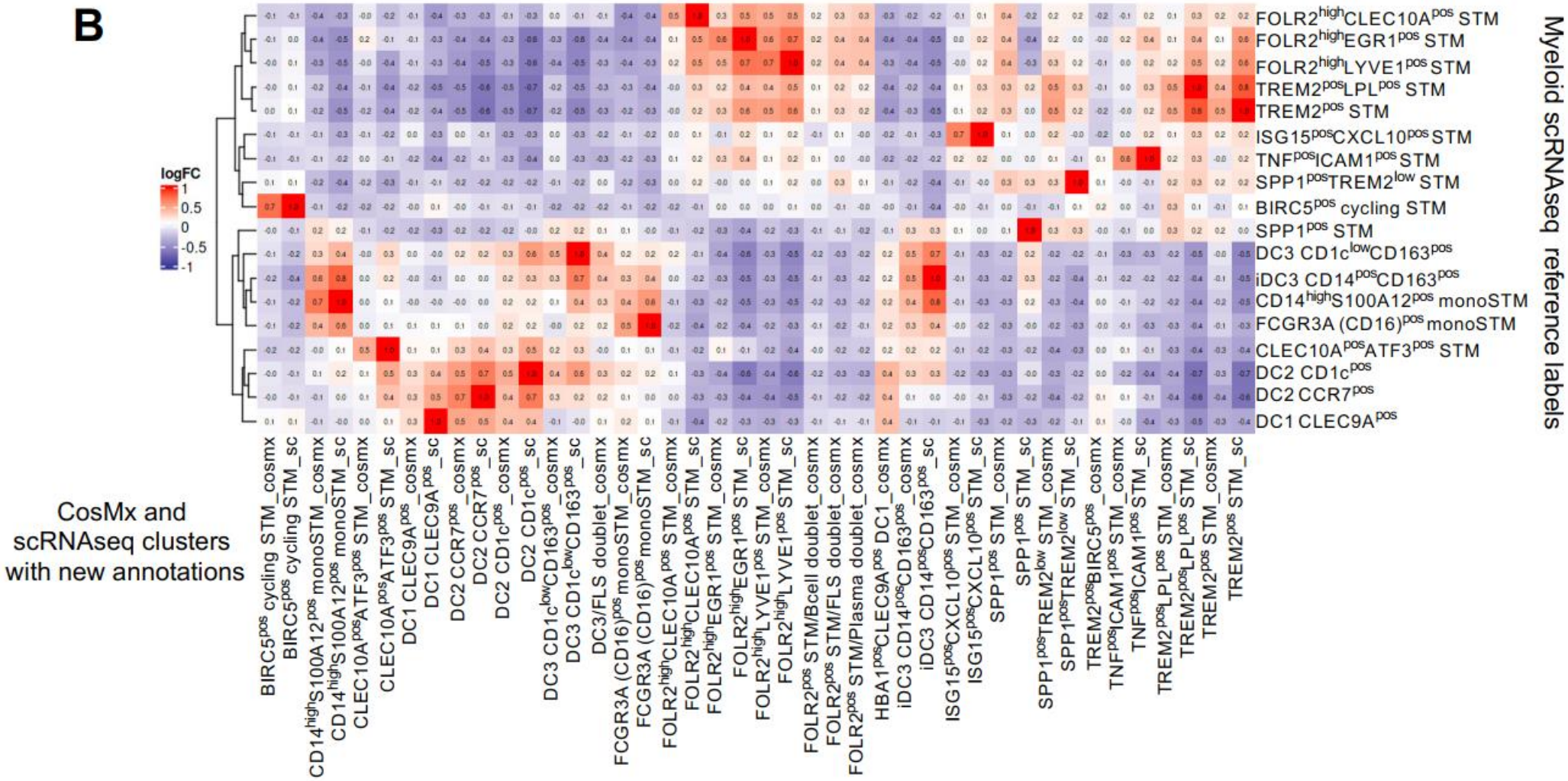
CellPress
OPEN ACCESS

Article

Synovial tissue myeloid dendritic cell subsets exhibit distinct tissue-niche localization and function in health and rheumatoid arthritis



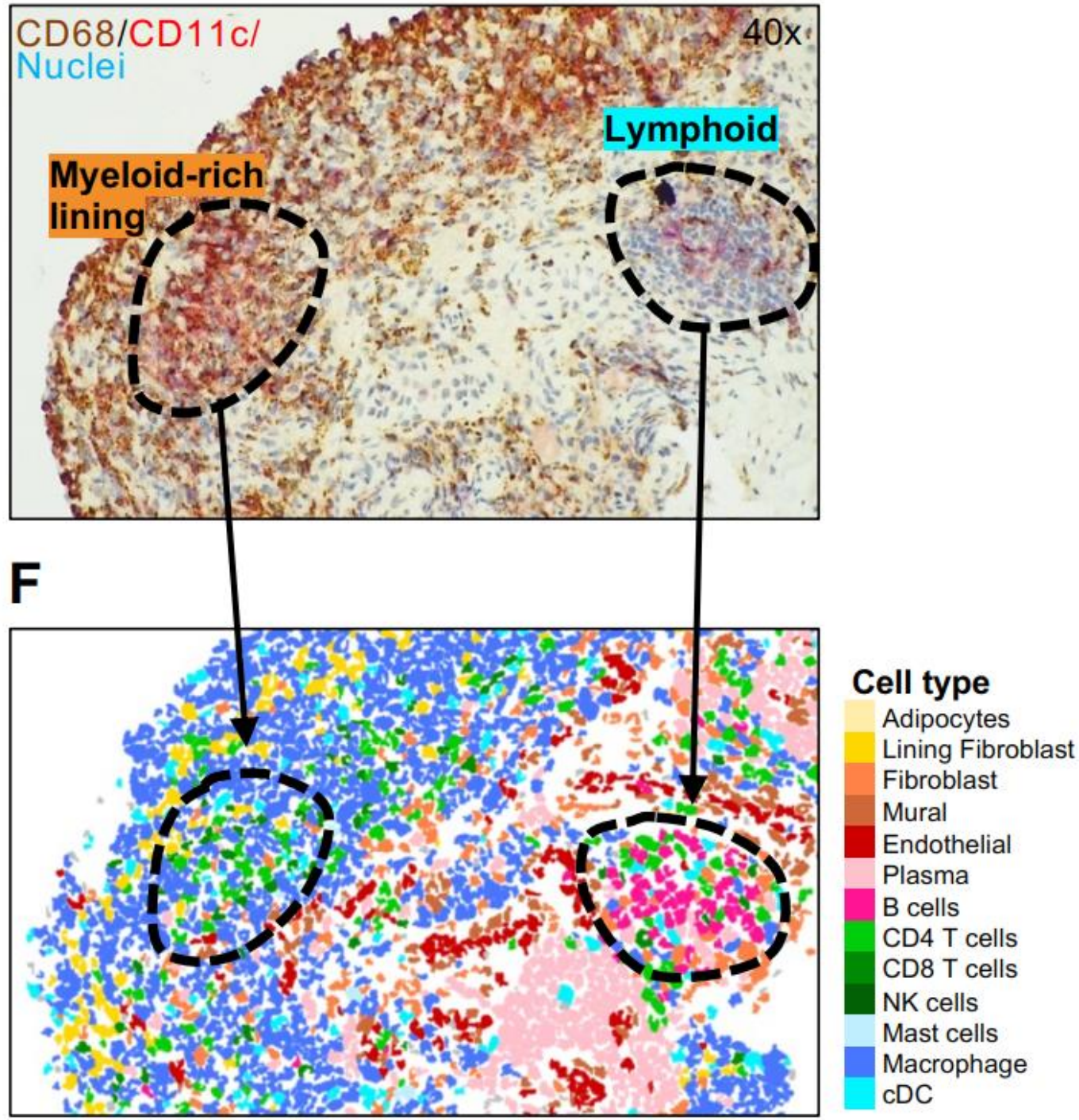

Active RA
n=3
+
Remission RA
n=3
Ultrasound-guided
synovial biopsy
Fixation, sectioning and FOV selection (0.75mm x 0.75mm)
CosMx SMI
single cell resolution
spatial transcriptomics
(980 gene panel)
IFS Imaging
(DAPI, PanCK)
Transcript
localization
Cell
segmentation
Cell type
annotation
Neighborhood
analysis
Tissue
segmentation



CosMx and scRNAseq clusters with new annotations

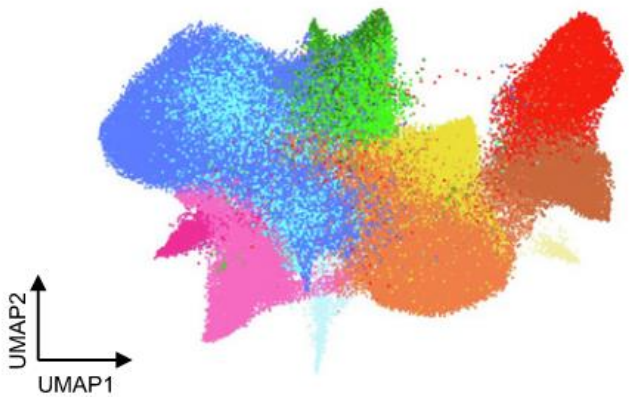
Celltype annotation was done by Harmony integration

Heatmap visualization of myeloid cells marker gene correlation between reference synovial tissue scRNAseq clusters (rows) and clusters identified from harmony integration of CosMx spatial transcriptomic and scRNAseq data (columns).


CD68/CD11c/
Nuclei
40x
Myeloid-rich
lining
Lymphoid
F
Cell type
Adipocytes
Lining Fibroblast
Fibroblast
Mural
Endothelial
Plasma
B cells
CD4 T cells
CD8 T cells
NK cells
Mast cells
Macrophage
cDC

A

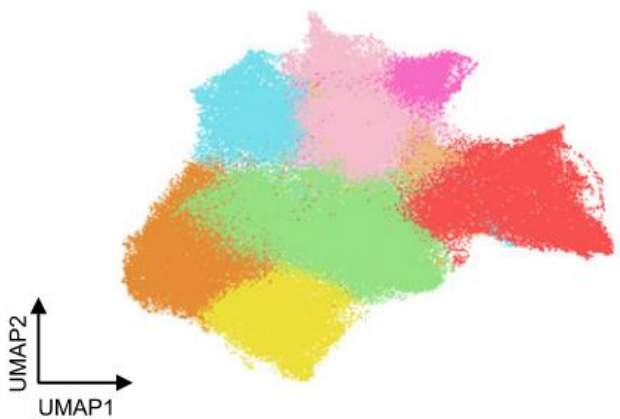
***127,199** cells expressing **960** transcripts across **60** fields of view (FOV)*



Cell type

- Adipocytes
- Lining Fibroblast
- Fibroblast
- Mural
- Endothelial
- Plasma
- B cells
- CD4 T cells
- CD8 T cells
- NK cells
- Mast cells
- Macrophage
- cDC

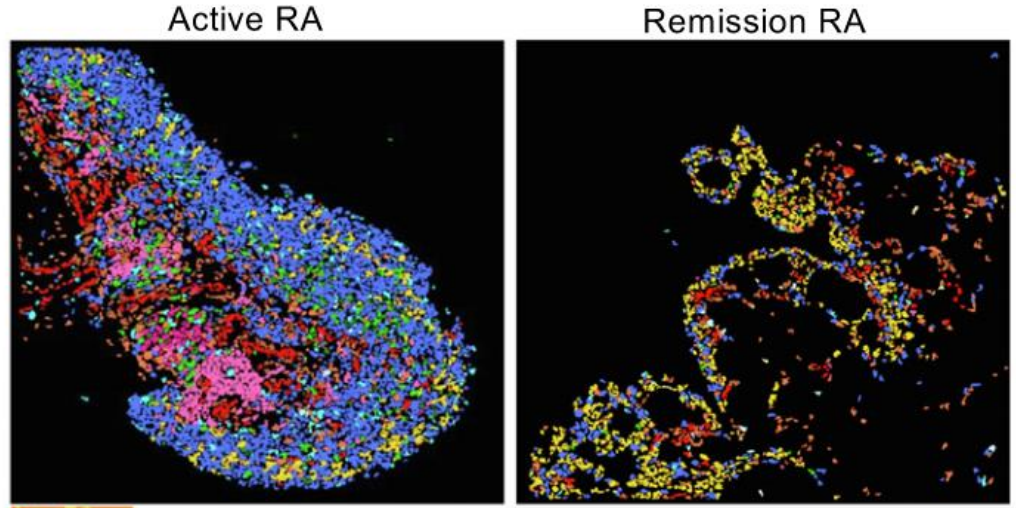
C



Niches:

1. PRG4^{pos} Lining
2. Myeloid-rich Lining
3. FLS/Myeloid Sublining
4. Adipocyte-rich
5. Perivascular
6. TGFβ^{pos} Myeloid-rich
7. Plasma-rich
8. Lymphoid

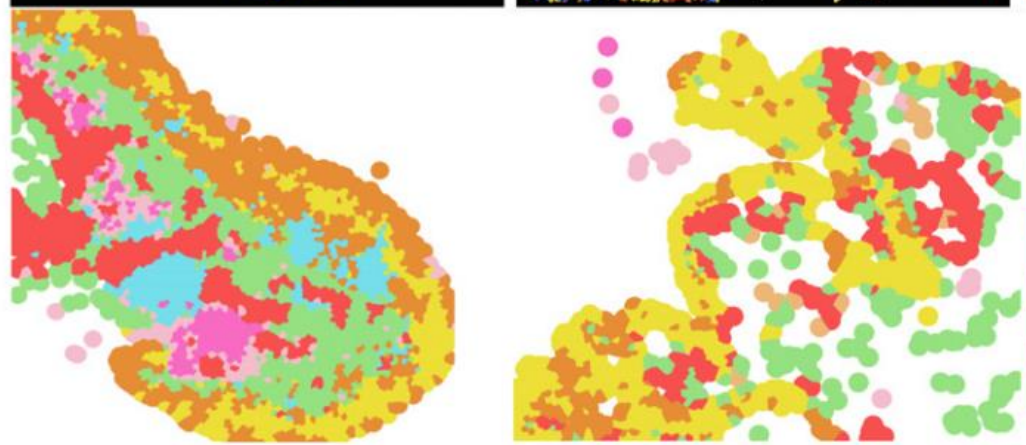
B



Cell type

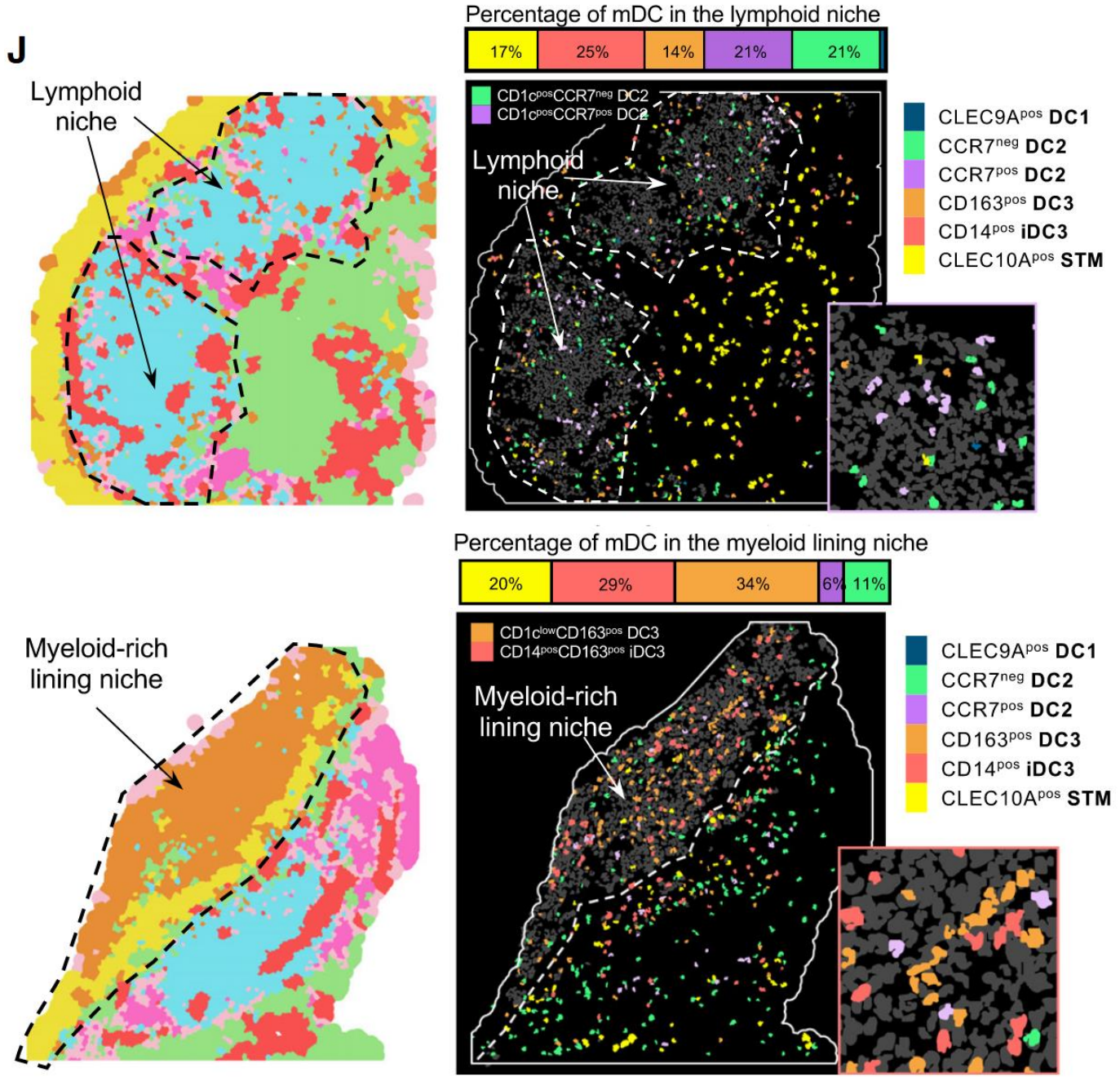
- Adipocytes
- Lining Fibroblast
- Fibroblast
- Mural
- Endothelial
- Plasma
- B cells
- CD4 T cells
- CD8 T cells
- NK cells
- Mast cells
- Macrophage
- cDC

D



Niches:

1. PRG4^{pos} Lining
2. Myeloid-rich Lining
3. FLS/Myeloid Sublining
4. Adipocyte-rich
5. Perivascular
6. TGFβ^{pos} Myeloid-rich
7. Plasma-rich
8. Lymphoid


J
Lymphoid niche
Percentage of mDC in the lymphoid niche
17%
25%
14%
21%
21%
CD1c^pos CCR7^neg DC2
CD1c^pos CCR7^pos DC2
Lymphoid niche
CLEC9A^pos DC1
CCR7^neg DC2
CCR7^pos DC2
CD163^pos DC3
CD14^pos iDC3
CLEC10A^pos STM
Myeloid-rich lining niche
Percentage of mDC in the myeloid lining niche
20%
29%
34%
6%
11%
CD1c^low CD163^pos DC3
CD14^pos CD163^pos iDC3
Myeloid-rich lining niche
CLEC9A^pos DC1
CCR7^neg DC2
CCR7^pos DC2
CD163^pos DC3
CD14^pos iDC3
CLEC10A^pos STM

Conclusion

- Huge opportunity with spatial
- Nice open computational task for us
- Still needs a lot of technical improvement – which will come

- VERY EXPENSIVE…