THE UNIVERSITY OF CHICAGO

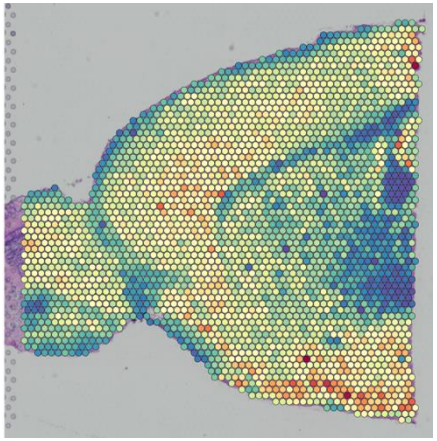
Biological Sciences
Bioinformatics Core

CENTER FOR RESEARCH INFORMATICS

BIOINFORMATICS CORE

Bioinformatics Workshop Learning Series

Spatial Transcriptomics 2026 Session 2: Visium and Xenium

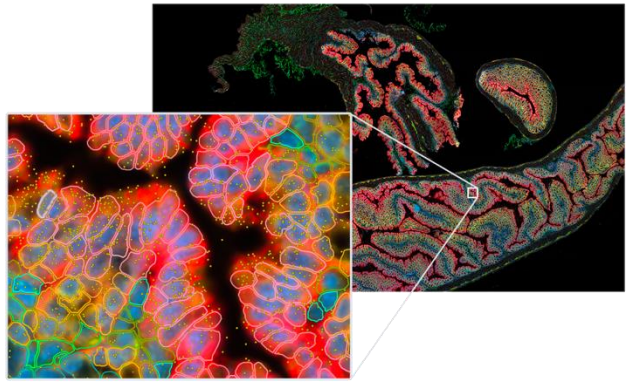


Jason Shapiro, Ph.D.

Bioinformatics Core

CRI, University of Chicago

May 22nd, 2026



OUR AWESOME TEAM

Mengjie Chen, PhD

Faculty Director

Wenjun Kang, MS

Technical Director

Jason Shapiro, PhD

Bioinformaticians

Diana Vera Cruz, PhD

Evan Wu

Katie Aracena, PhD

David Tieri, PhD

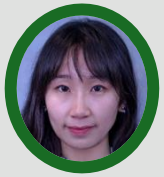
Yan Li, PhD

Associate Director

Houxiang Zhu, PhD

Qiaoshan Lin, PhD

Yildiz Koca, PhD



Zhongyu Li, MS

Geetha Priyanka, MS

Genomics and proteomics data analysis

BiCF applies appropriate and state-of-the-arts statistical and bioinformatic methodologies to analyze genomics data generated from standard and emerging assays.

Consulting, grant writing and training

BiCF provides consulting services for experimental design or data analysis; grant writing assistance, including bioinformatics development, cost analysis, and documentation of tools to complete the research.

Data management system development

BiCF offers enterprise solutions for project and study management, for data production , sharing and integration .

Contact us: **bioinformatics@bsd.uchicago.edu**

Submit a project request: **https://biocore.cri.uchicago.edu/**

Bioinformatics Core Services

- NGS sequencing data analysis
  - bulk sequencing data analysis
    - bulk RNA-seq , bulk ATAC-seq, Cut&Run, Library screen CRISPR etc.
    - standard 10 hours workflow analysis
    - 2 weeks results turn around time
  - single cell data analysis
    - scRNA-seq, scATAC-seq, OMICS etc.
    - standard 50-60 hours workflow analysis
    - 4-6 weeks results turn around time
  - spatial data analysis
    - Visium, Xenium, GeoMx, CosMx.
- Grant writing assistance: Aims, power computation etc.
- Consultation: first hour free consultation, in person office hours
- Application development

Contact us: **bioinformatics@bsd.uchicago.edu**
Submit a project request: **https://biocore.cri.uchicago.edu/**

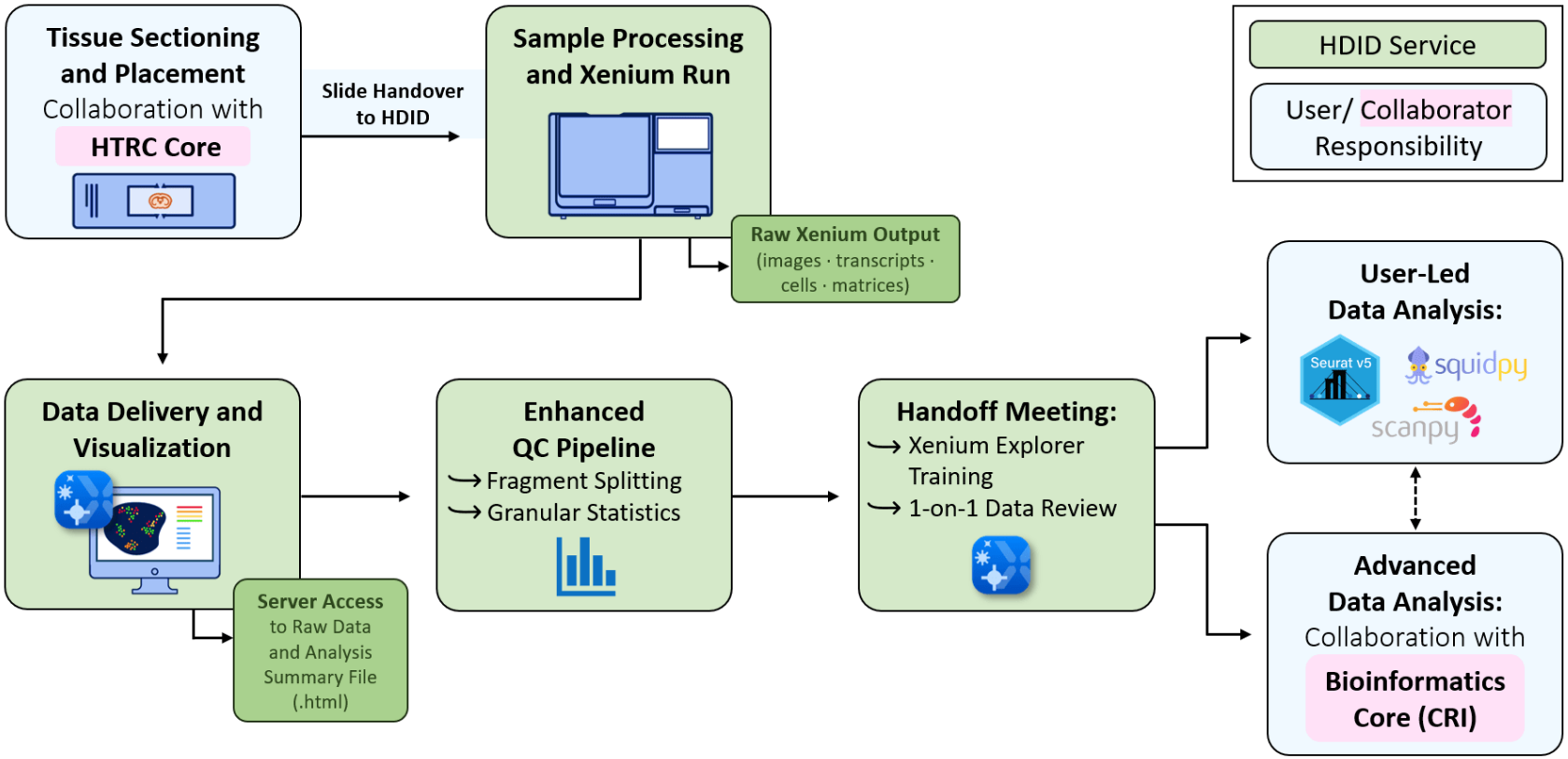
Weekly In-Person Free Consultations

- HPC system team: Randi Support
  - 🕒 **Every Tuesday afternoon** 12:00 PM – 2:30 PM
  - Location: Peck Pavilion N161
  - 📎 Slack Group: https://join.slack.com/t/criscientific-dzi9891/shared_invite/zt-2kghy4392-1ELPfgn8pL5BcXk4oF9D4g

- In-person office hours bioinformatics
  - 🕒 **Every Tuesday afternoon** 12:30 PM – 3:30 PM
  - Location: Peck Pavilion N161
  - Bioinformatics related questions
  - Experimental design questions

- Biostatistical Questions
  - 🕒 **First Tuesday of Each Month** 12:30 PM – 3:30 PM
  - Location: Peck Pavilion N161

Xenium Core at UChicago



Additional information: https://voices.uchicago.edu/hdid/xenium-analyzer/
Contact: Cezary Ciszewski (Technical Director)

Today's Agenda

- **Visium**
  - Data Preparation
  - Quality Control
  - Spatially variable gene detection
  - Spatial Clustering
  - Deconvolution
  - Working with multiple samples
  - Cell-cell communication
  - Differential Expression

- **Xenium**
  - Data Preparation
  - Xenium Explorer
  - Basic sample processing
  - Clustering and deconvolution

Resources

- All code examples for today are available on GitHub:
- https://github.com/CRI-Biocore/Spatial_Transcriptomics_Visium_Xenium_2026
- Datasets can be downloaded from separate sources (see instructions in the page above)
- From 10X:
  - Space Ranger
  - Loupe Browser
  - Xenium Explorer
  - Data: https://www.10xgenomics.com/datasets

Comparing Technologies

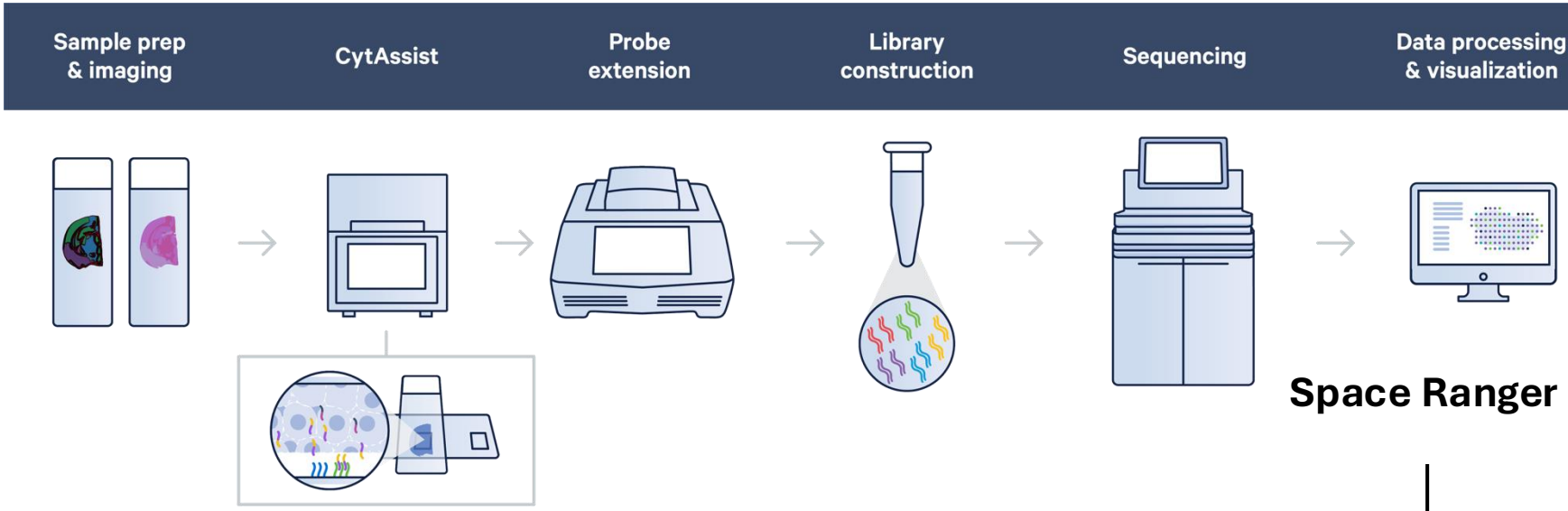
Name	Company	Strategy	Resolution	Depth
Visium v2 (CytAssist)	10X	Sequencing	1-10 cells per 55 µm spot	Whole transcriptome
Visium HD	10X	Sequencing	Single-cell (2 µm)	Whole transcriptome
Xenium	10X	Imaging	Sub-cellular (50 nm)	Gene panels (300 - 5000)
Slide-seq (“Seeker”)	Curio Bioscience	Sequencing	Single-cell (10 µm)	Whole transcriptome
GeoMX	NanoString	Imaging and sequencing	10 to 100 of cells, regions defined by user	Whole transcriptome
CosMX	NanoString	Imaging (FISH)	Sub-cellular (50-120 nm), cell segmentation method.	1000-6000, genes, Whole transcriptome

Analysis Example: Visium

- **Visium**
  - Data Preparation
  - Quality Control
  - Spatially variable gene detection
  - Spatial Clustering
  - Deconvolution
  - Working with multiple samples
  - Cell-cell communication
  - Differential Expression

- **Xenium**
  - Data Preparation
  - Xenium Explorer
  - Basic sample processing
  - Clustering and deconvolution

Visium



Space Ranger

Downstream Analyses

Images from 10X Genomics

File Checklist for Space Ranger

- Directory with Raw Reads
- Reference: typically human or mouse, but you can create a custom reference. Also requires probe set information.
- Images (TIFF or JPEG)
- Image Alignment files (JSON)
  - Note: optional for Visium HD
  - Most users will want to use **Loupe Browser** to create alignments, either using manual or automatic features.
  - **Always check automated alignments.**

Image Alignment

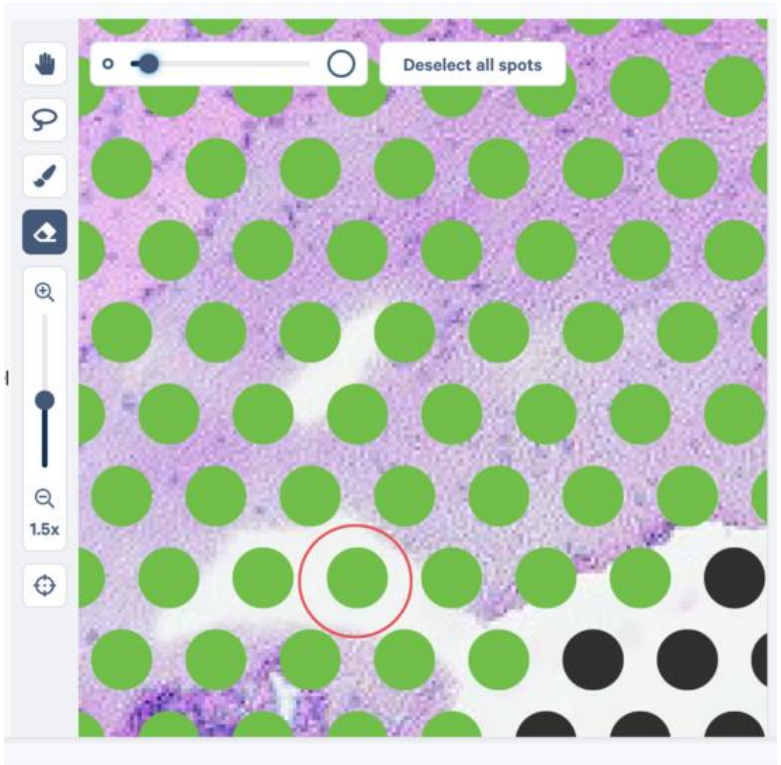
- Different options are available depending on the Visium version you are working with:
  - Visium v1/v2: Manual alignment requires choosing individual spots to include or exclude. Automated alignment may be more error-prone, but is still an option.
  - CytAssist Microscope Images: A separate step to align the CytAssist image with microscopy. (Note: many labs use CytAssist chemistry without all of the microscopy features, and this step is not required)
  - Visium HD: Automated alignment is recommended, and the manual alignment is still largely automated.

Loupe Browser Example

- Next, we'll do a quick tour of how to use Loupe Browser to create an alignment.
- We will use the cloupe file for the mouse anterior brain section from 10X.
- Note: this is the same dataset that is used in Seurat's Spatial Vignette and is often used for benchmarking.

Image Alignment

- Tips for common issues:
  - If using automated alignment, review the result and touch up the selected spots manually if needed.



Images from 10X Genomics

Space Ranger

- 10X-provided command line software for processing raw reads
- Similar to running Cell Ranger for Chromium scRNA data
- Please see our GitHub for an example code block for running Space Ranger for a single sample on the HPC
- Key Outputs:
  - /outs/spatial folder
    - contains images and coordinate information
  - filtered_feature_bc_matrix.h5 (same data also saved in a folder)
    - contains counts, features, and barcodes
    - “filter” restricts results to just the spots in the alignment
  - web summary report

Quality Control

- Always check the “web_summary.html” file produced by Space Ranger (in the “outs” folder)
  - Includes key information about reads per spot and alignment to the reference
- Often, it is not necessary to filter out any spots
- Visualize the read and feature depth per spot in R
  - Keep an eye out for any visible irregularities in the count distributions. For instance, a “banding” effect with low or high counts in a solid rectangle could indicate a problem with sample preparation.

Space Ranger Report

Example from a good quality sample

-Well over 10,000 reads per spot.
-Over 1,000 genes per spot.
-High fraction of valid barcodes/UMIs
-High fraction of reads mapped to reference.
-High fraction of reads mapped to spots.

2,695
Number of Spots Under Tissue

124,414
Mean Reads per Spot

6,228
Median Genes per Spot

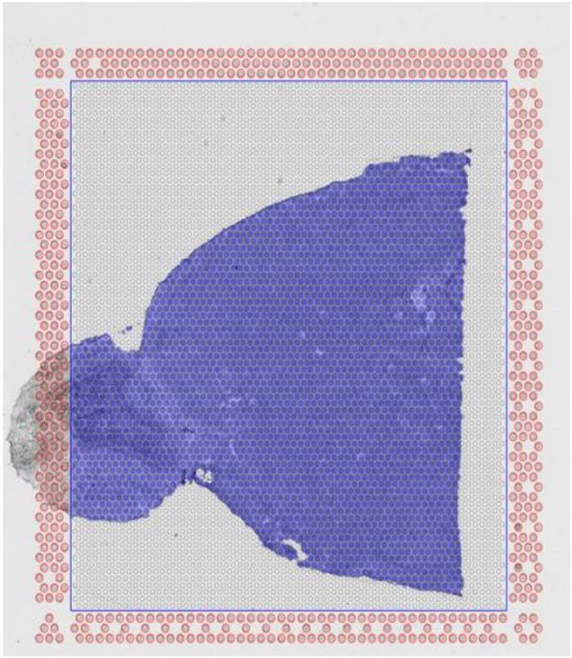
Sequencing

Number of Reads	335,294,592
Valid Barcodes	97.5%
Valid UMIs	100.0%
Sequencing Saturation	73.6%
Q30 Bases in Barcode	97.2%
Q30 Bases in RNA Read	94.6%
Q30 Bases in UMI	97.2%

Mapping

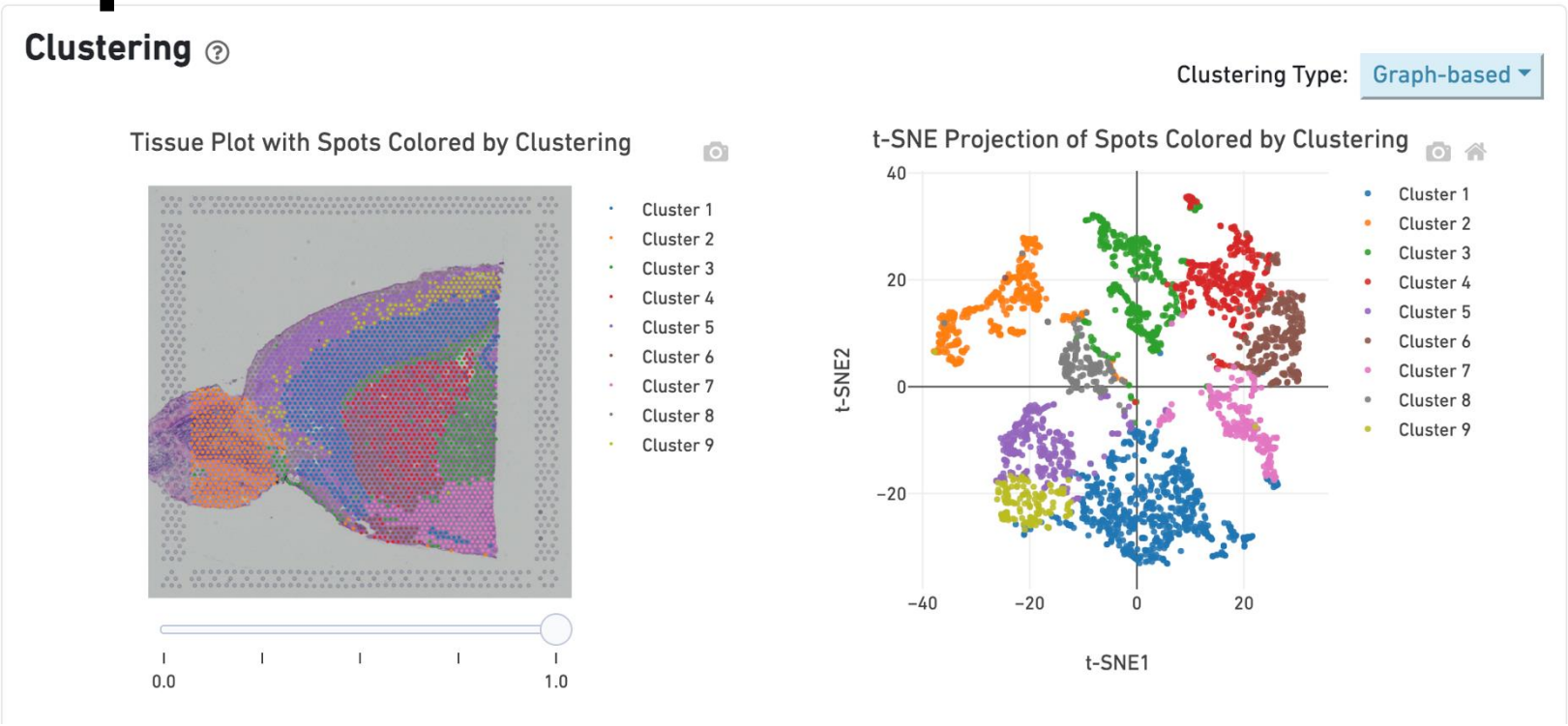
Reads Mapped to Genome	96.0%
Reads Mapped Confidently to Genome	93.4%
Reads Mapped Confidently to Intergenic Regions	3.5%

Spots



Fraction Reads in Spots Under Tissue	92.9%
Mean Reads per Spot	124,414
Median Genes per Spot	6,228
Total Genes Detected	21,334
Median UMI Counts per Spot	25,888

Space Ranger Report



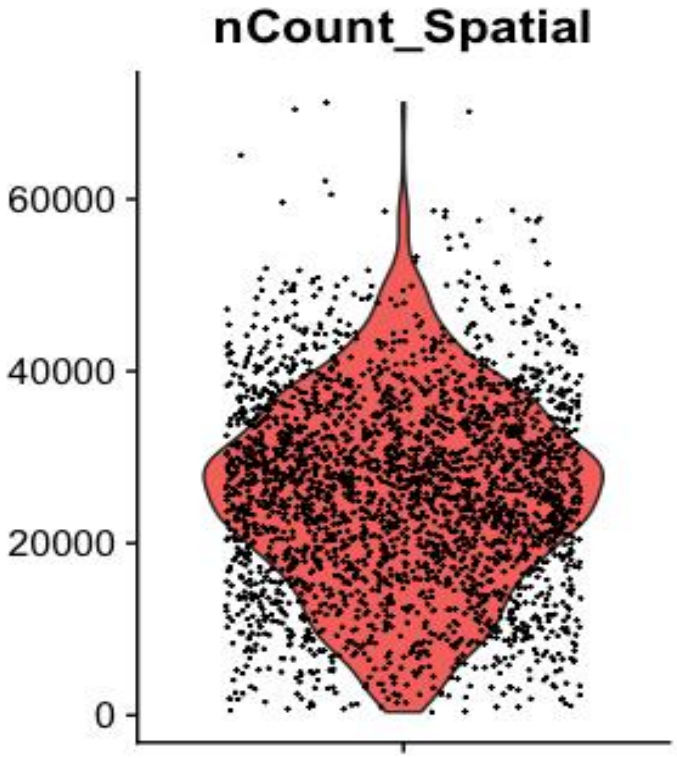
Space Ranger also runs dimensionality reduction and clustering, and results are viewable with their interactive tool, Loupe Browser.

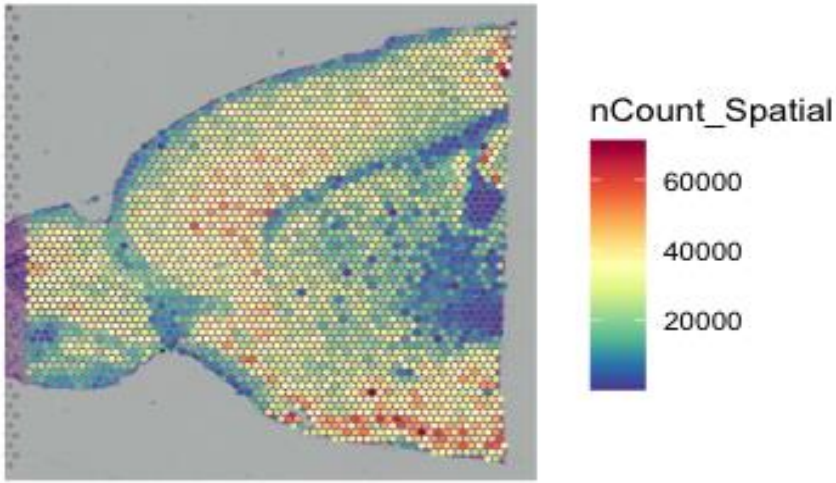
These clusters might not be identical to clusters we find with our own downstream analysis.

Loupe Browser

- We will typically do all of our analysis in R or Python, but the Loupe Browser can be useful for exploring data if you want a more interactive option.
- Key functionality is the ability to manually select spots and export the barcodes, import feature lists to explore gene signatures with interactive visualizations. You can also import selected barcodes.
- If you have run an analysis in R, you can also export annotations and clusters and read them back into Loupe. The package LoupeR will also let you save your own cloupe file from R.
- **Break for a quick Loupe Browser demo**

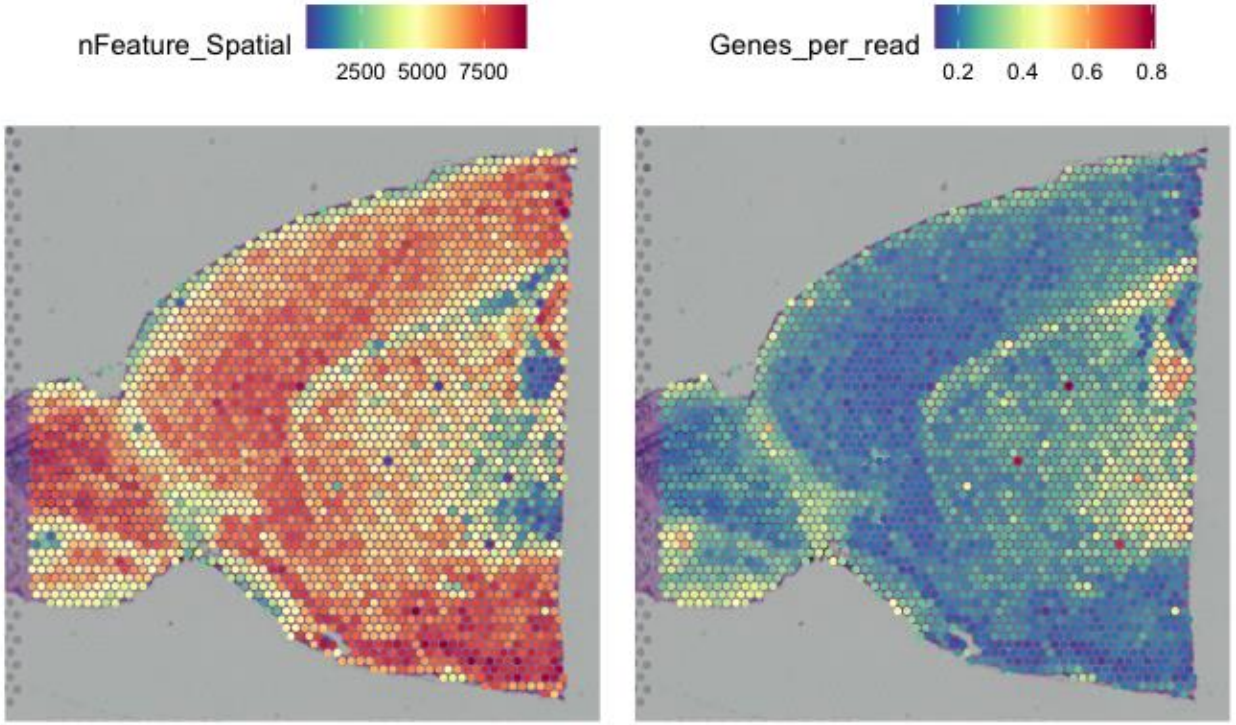
Quality Control in R





We typically will not filter out individual spots in the same way that we would remove individual cells for scRNA. Instead, we check for global issues with the data.

Quality Control in R

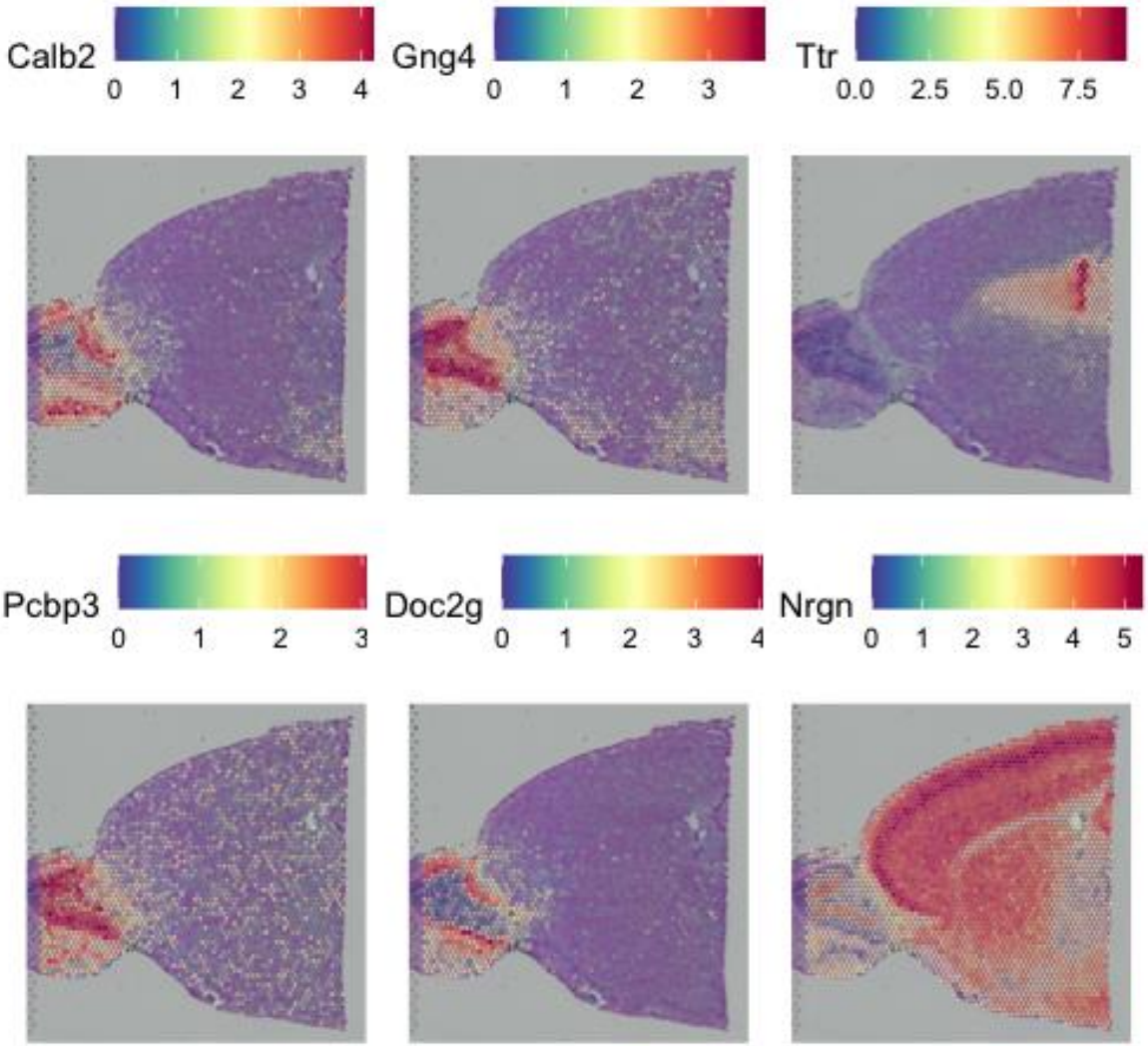


Having low total expression is generally not a good indicator that a spot lacks information.

Spatially Variable Genes

- Spatially variable genes are genes that show a spatial pattern in their expression.
- These genes can be useful to identify on their own for biological interpretation
- They can also form the basis for identifying spatial clusters/domains
- Seurat has its own method for identifying these genes (FindSpatiallyVariableFeatures), as do other packages. Many rely on SPARK-X.

Spatially Variable Genes: Seurat



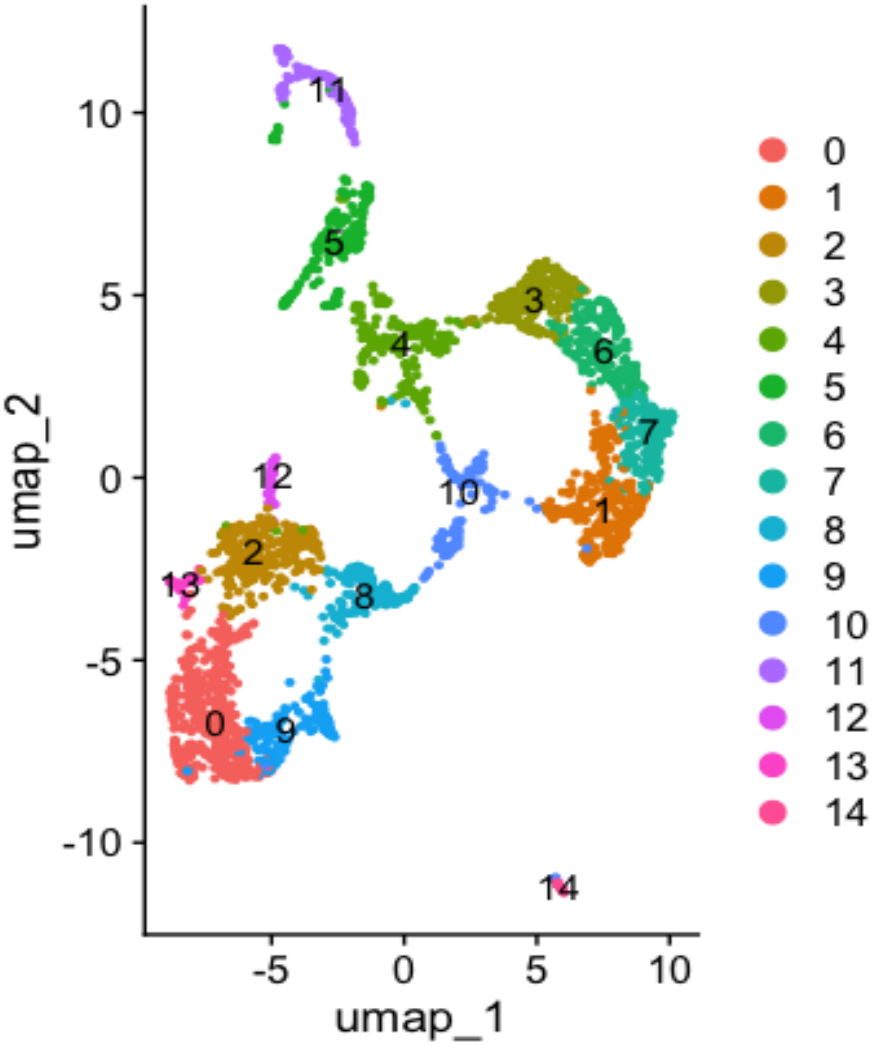
Clustering

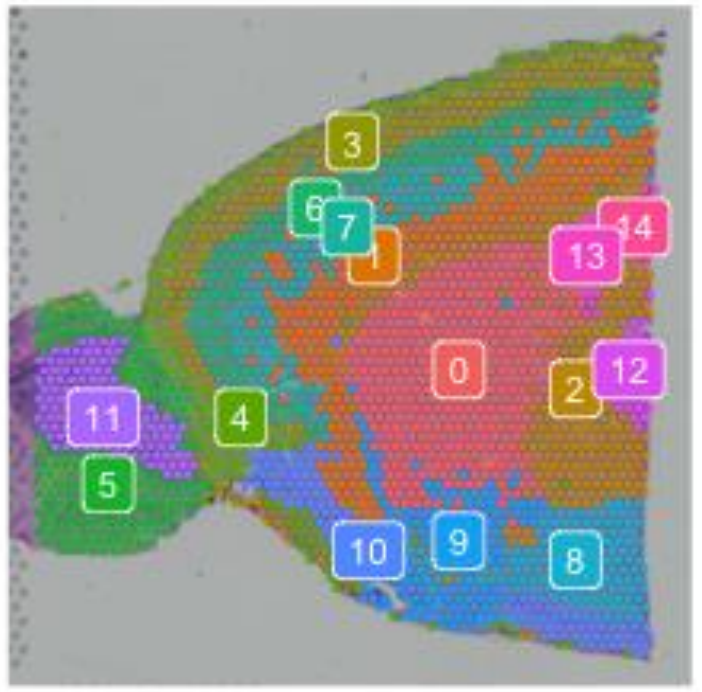
- Clusters define groups of spots that share information.
- Seurat's FindClusters function relies solely on gene expression data, but this tends to still include some spatial context for Visium data.
- Many alternatives explicitly include spatial information during clustering, which often provides a more accurate representation of the biology.

Clustering

- There are dozens of available tools:
  - Seurat (R): identical workflow if you have done scRNA
  - BayesSpace (R)
  - MERINGUE (R)
  - BASS (R)
  - MEcell (R)
  - GraphST (Python)
  - Banksy (Python)
  - PRECAST (R)
  - And at least 20 more…
- Note: not all questions/data depend on spatial clustering. The simplest approach might suffice, and sometimes we will prefer to define regions with marker genes.

Clustering: Seurat

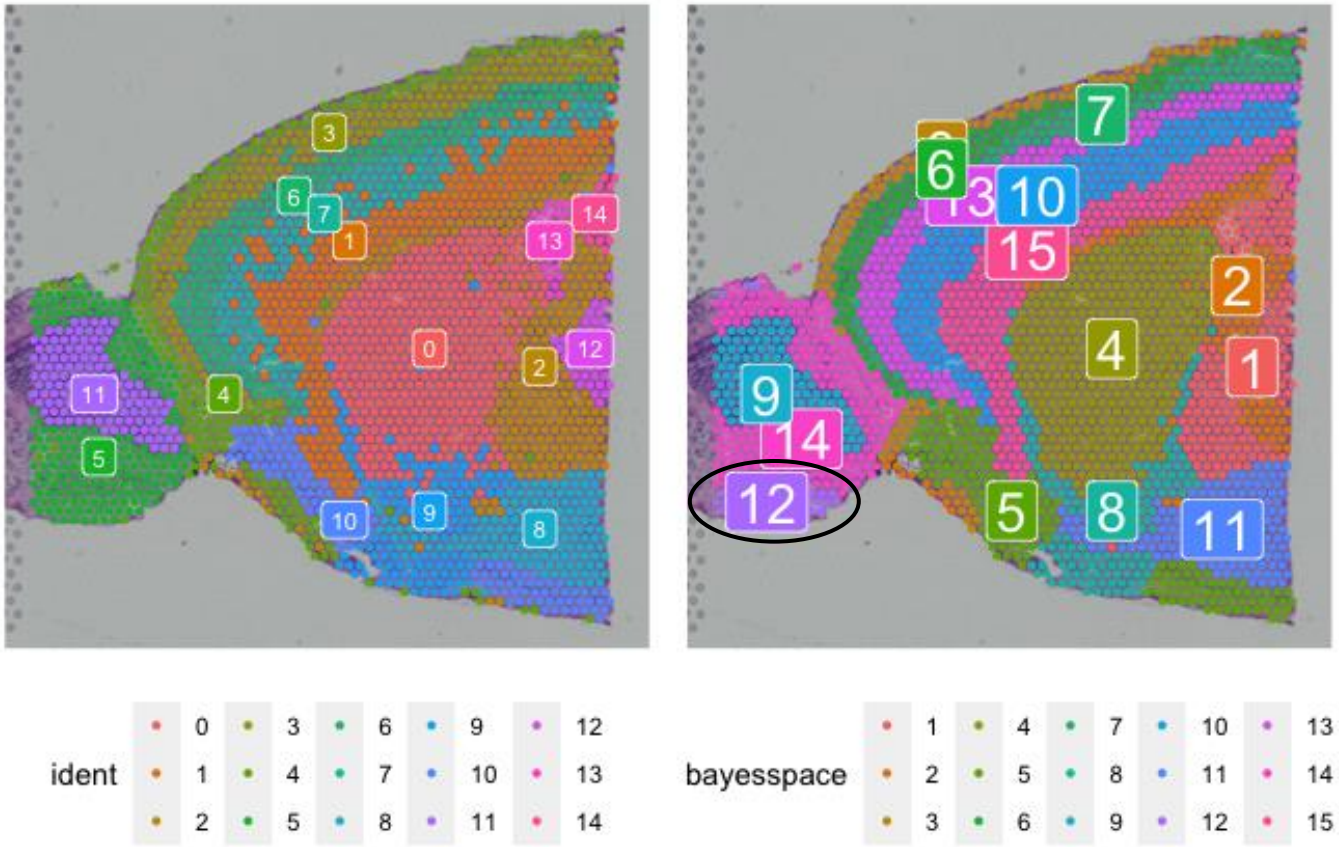




Seurat clustering works fairly well, but some regions are noisy

* How well this works depends on your tissue type. A highly structured brain section will have cleaner clusters than a tumor with immune infiltration.

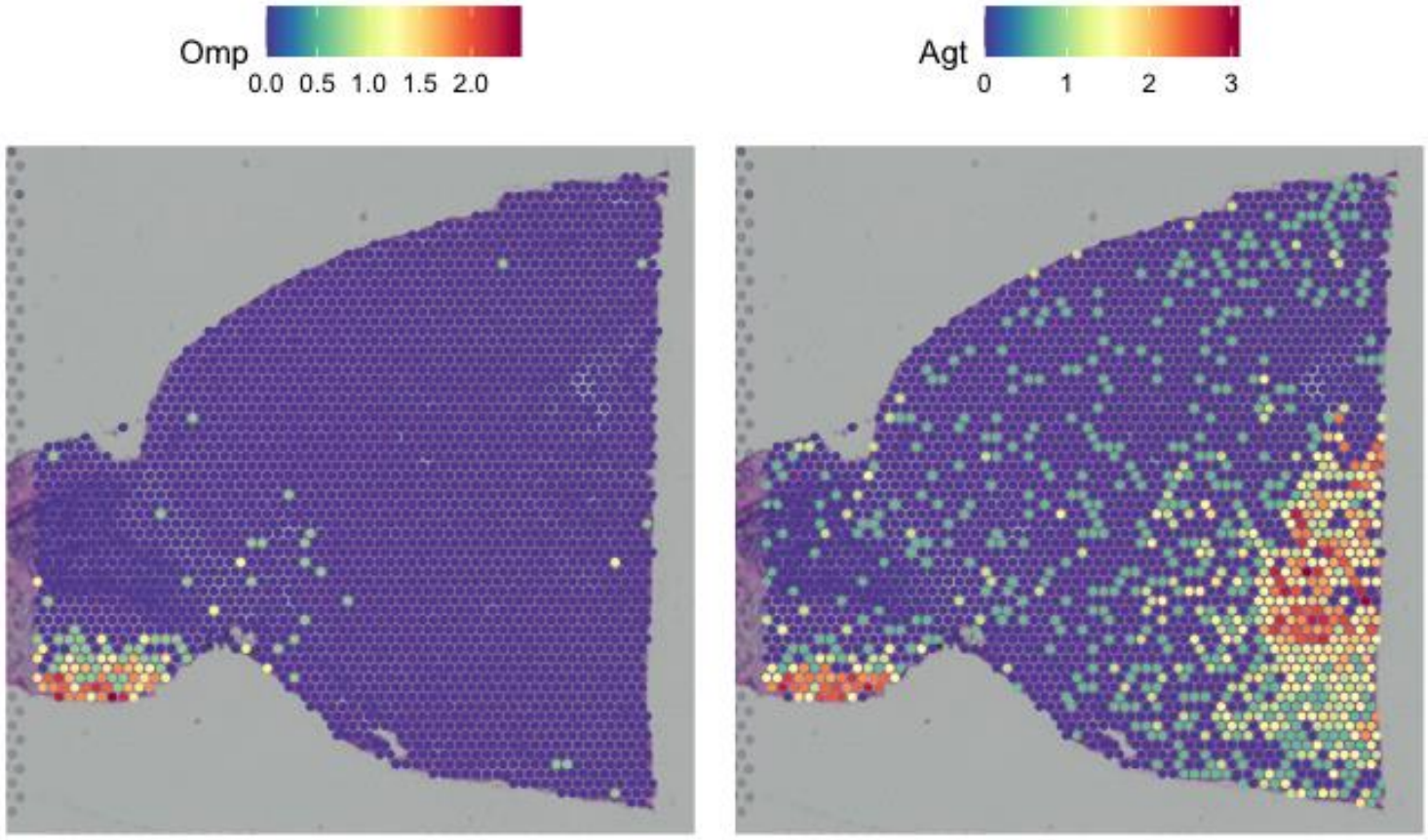
Clustering: BayesSpace



BayesSpace is a common option for spatial clustering, and it improves the definition of separate regions in this case.

We can look for region markers to validate the differences

Clustering



We can check for markers specific to the new BayesSpace clusters and confirm that there are genes that correspond to these distinct regions.

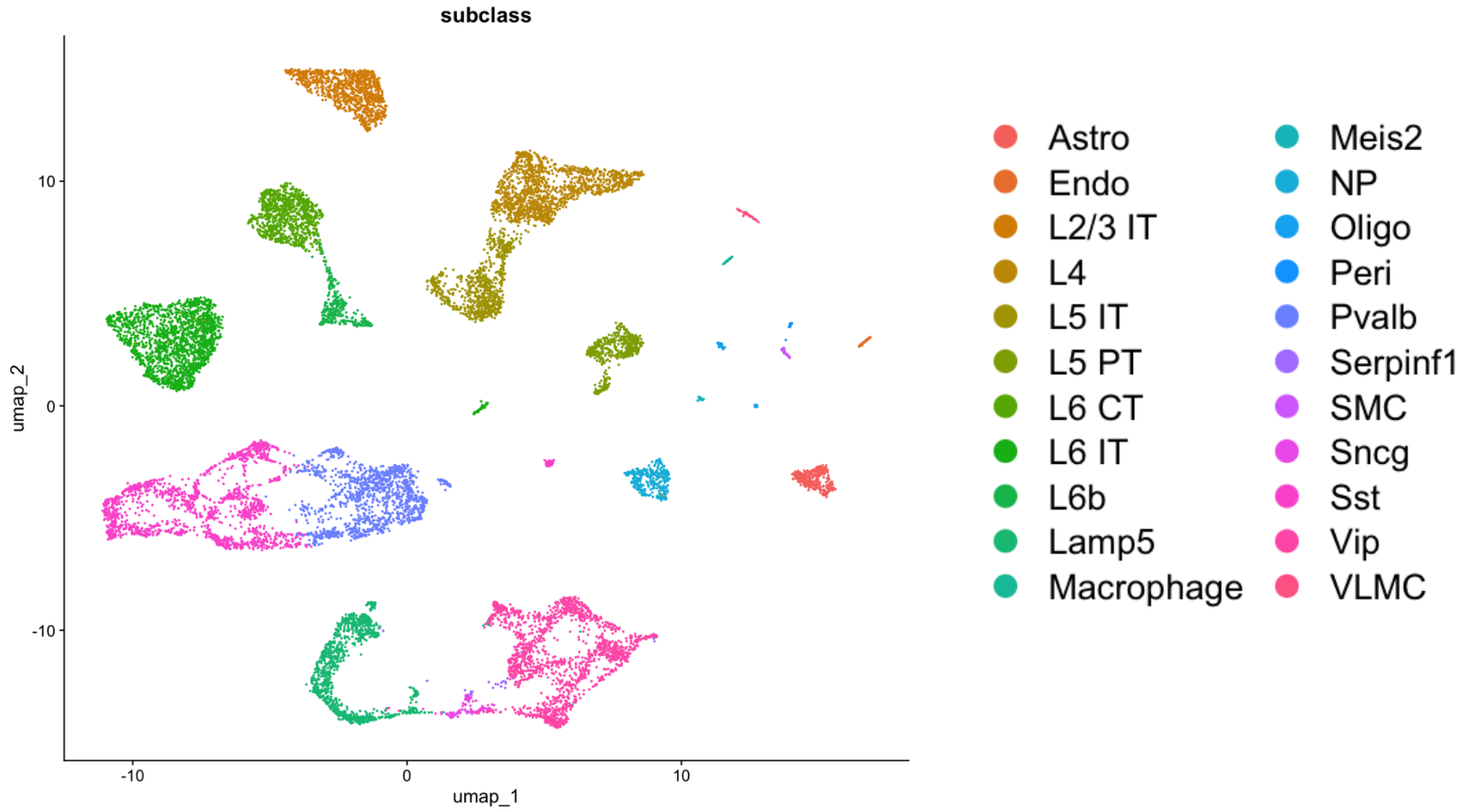
Deconvolution

- Each spot in Visium data contains ~1 - 10 cells, depending on the tissue (> 20 cells in some cases).
- Deconvolution is the process of inferring cell type proportions for each spot.
- **In general, a cluster ≠ a cell type.**
- Most methods require a scRNA reference that is comparable to the source of the spatial data
  - The reference should include **all** cell types found in the tissue
  - The quality of the reference annotation will also affect results
  - Even with an excellent reference, it can be difficult to resolve subpopulations in the spatial data due to overlapping markers.

Deconvolution Options

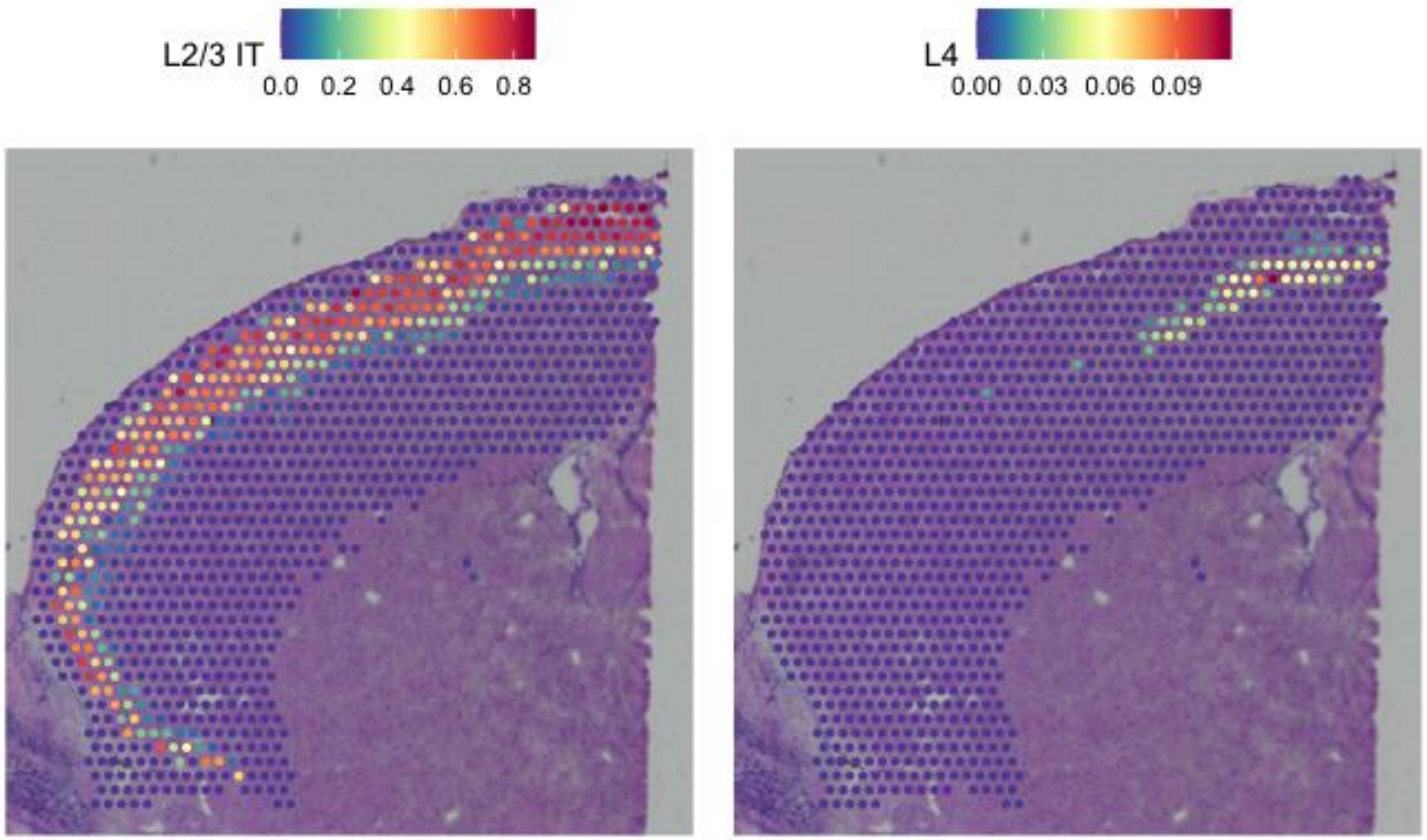
- As with clustering, there are many options available.
- **Today: Seurat and spacexr (RCTD)**
- Other R options:
  - CARD, SPOTlight, SpaCET, STdeconvolve, Giotto
- Python options:
  - cell2location, Tangram

Deconvolution



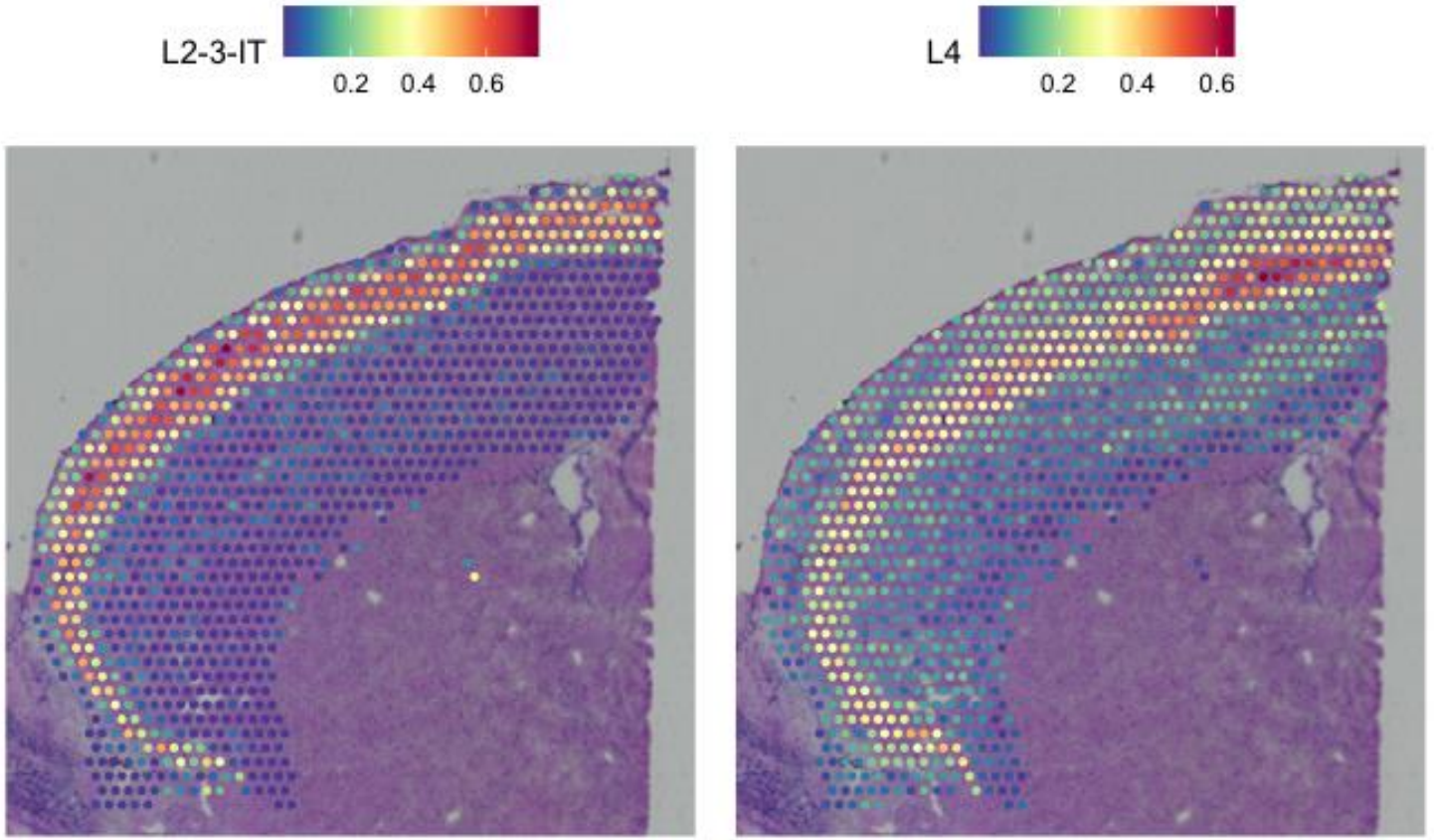
We will use the same cortex reference dataset (Tasic et al (2016)) as in the Seurat tutorial

Deconvolution: Seurat



We only apply the reference to the cortex region of the Visium data.
Seurat provides prediction scores, rather than true proportion estimates.

Deconvolution: RCTD

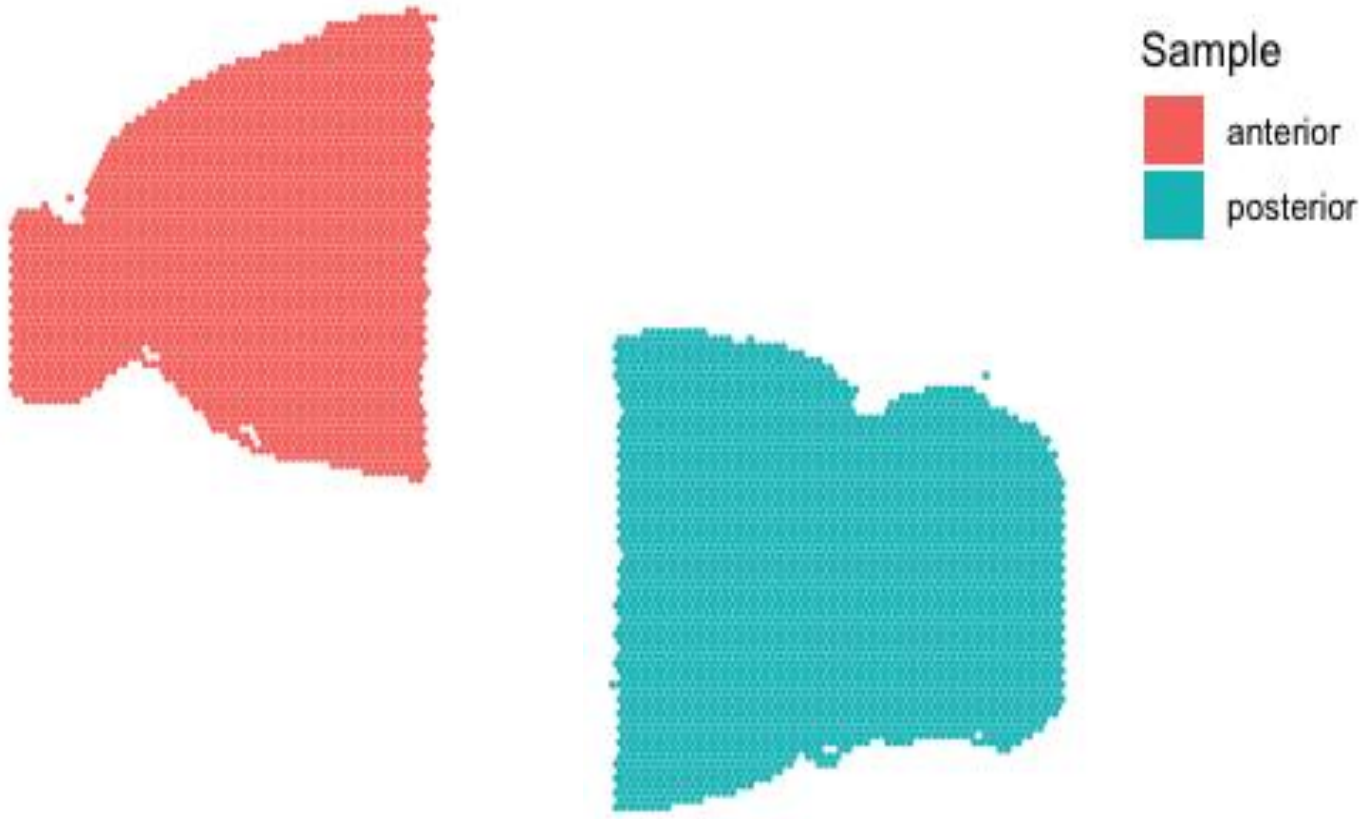


RCTD returns weights for each cell type, which can be converted to proportion estimates. These values may be preferred when working with other tools downstream, such as C-SIDE for differential expression or SpatialCellChat for cell-cell communication.

Working with multiple samples

- Often, you will have more than one sample (I hope)
- For some methods, we can artificially stitch the data into one, mega-sample and use single-sample tools
- For other analyses, we will integrate the samples together, similar to working with scRNA data.

Stitching samples

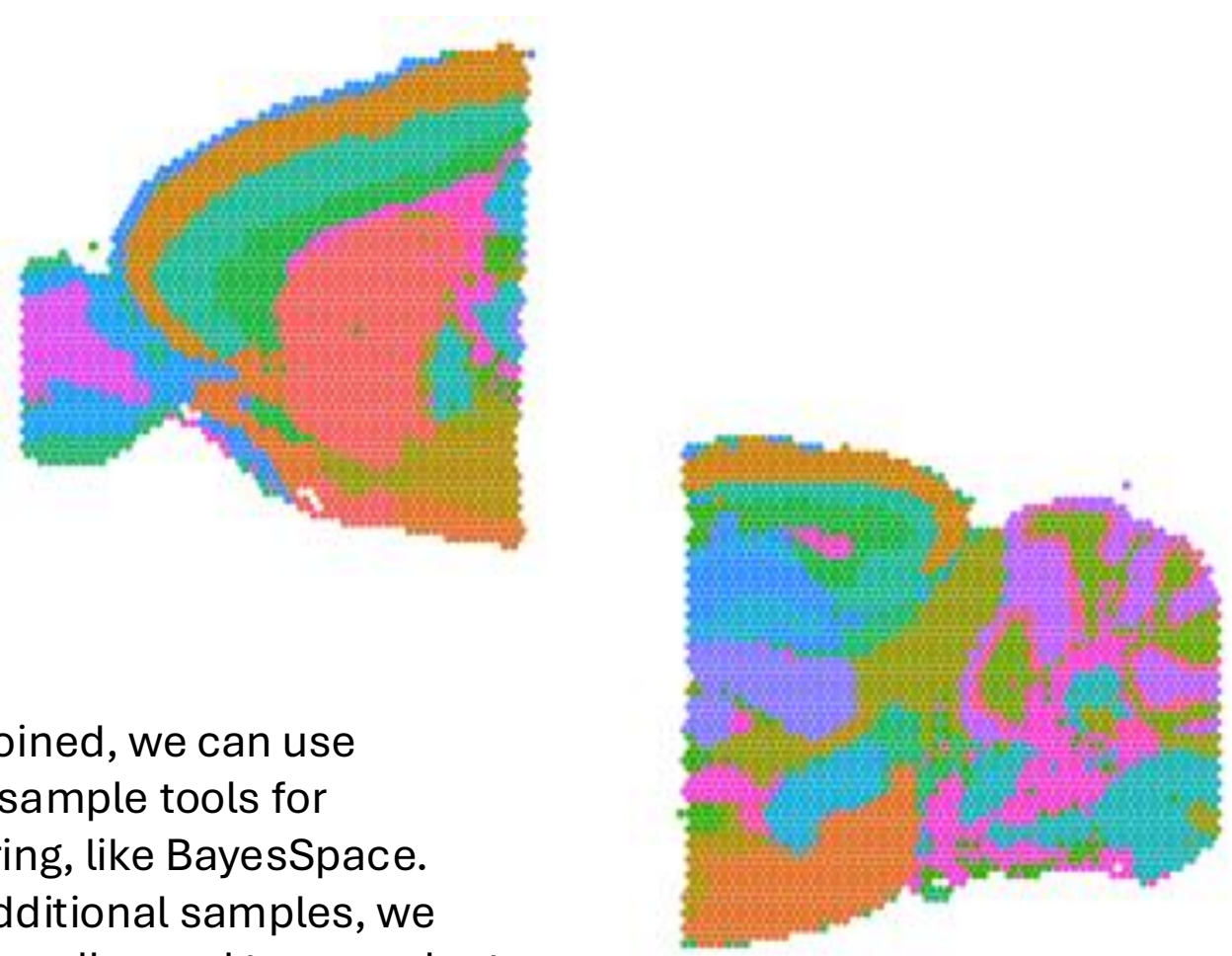


Here, we create a merged object starting from separate samples. This can be done with Seurat or with SingleCellExperiment. The key is to nudge the coordinates for each additional sample to prevent any accidental overlaps.

Integration

- Standard integration methods from scRNA can be applied to gene expression from multiple Visium samples
- These include CCA, RPCA, and Harmony. Often, Harmony provides the best balance for Visium data of finding similarities between samples while preserving unique clusters.
- Even when working with stitched coordinates, we will still want to integrate the gene expression data prior to clustering.

Stitching samples



Once joined, we can use single-sample tools for clustering, like BayesSpace. With additional samples, we will generally need to re-evaluate parameters, such as expected clusters, based on additional cell types that might be present.

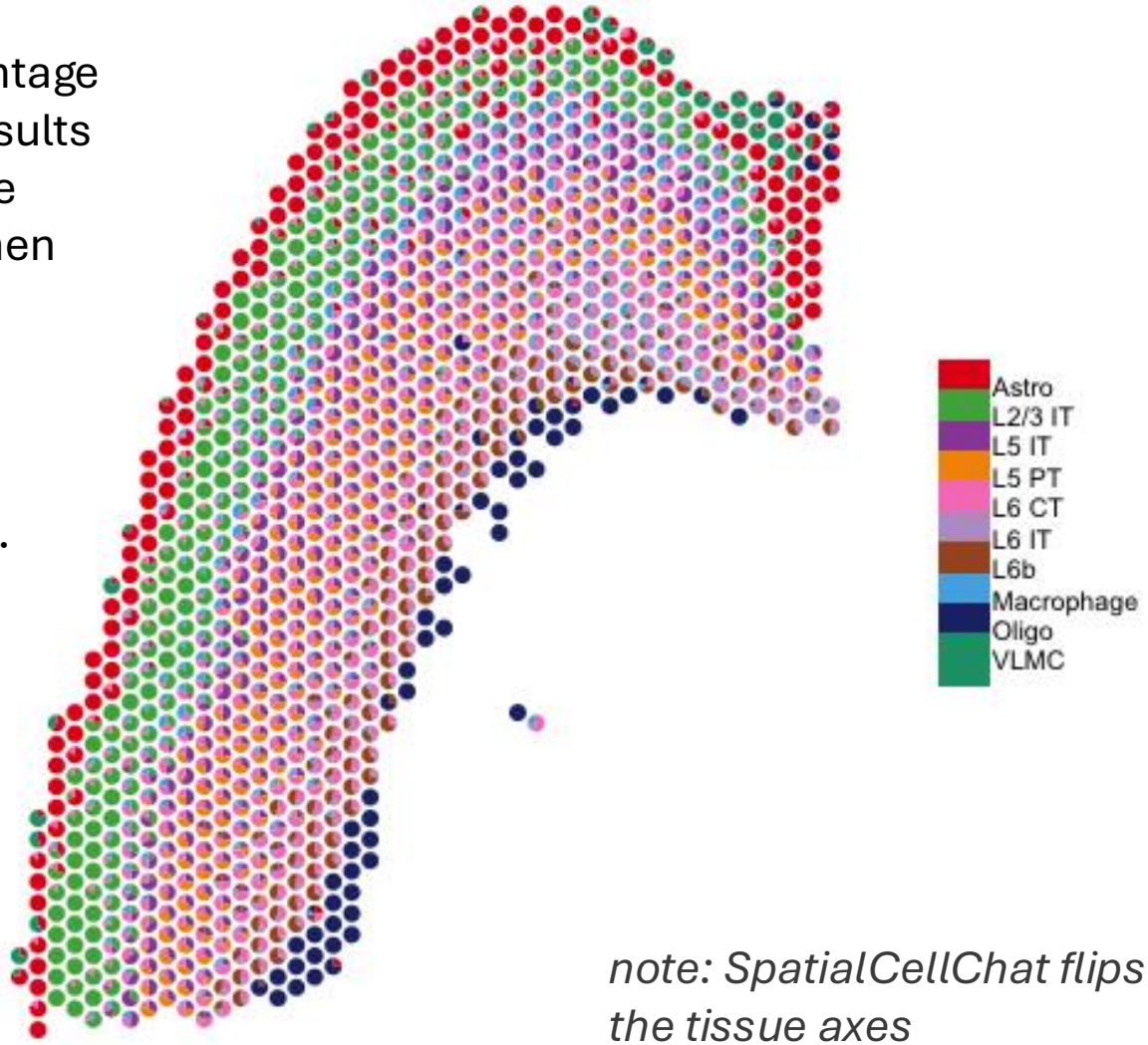
Cell-cell communication with SpatialCellChat

- A key use of spatial data is identifying spatial relationships between different cell types.
- scRNA tools for cell-cell communication, like CellChat and CellPhoneDB can both be applied.
- SpatialCellChat is a recent update to CellChat specifically designed to incorporate spatial context into the analysis.

SpatialCellChat

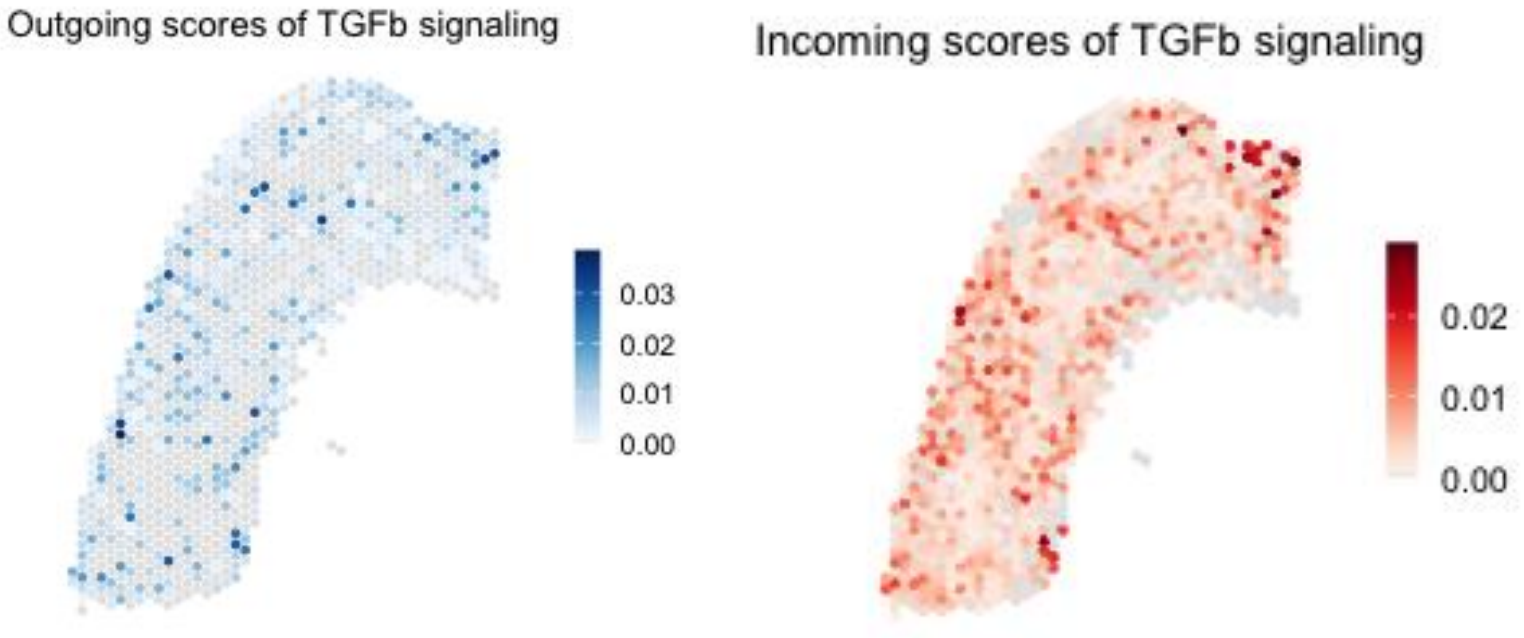
SpatialCellChat takes advantage of existing deconvolution results and incorporates the relative weights of each cell type when averaging gene interactions between cell types.

It also has some helpful plotting functions of its own.



note: SpatialCellChat flips the tissue axes

SpatialCellChat



It will build a network of proposed interactions for various pathways, and can provide estimates of where signals originate and where they are received.

Differential Expression

- If working with a single sample, FindMarkers from Seurat is often sufficient to identify key differences between regions/clusters
- When working with multiple samples, we will often use pseudobulk approaches, similar to scRNA analysis (e.g. DESeq2, limma).
- A few tools are available that incorporate spatial information or deconvolution estimates into DE analysis, such as C-SIDE, DESpace2, and NicheDE.

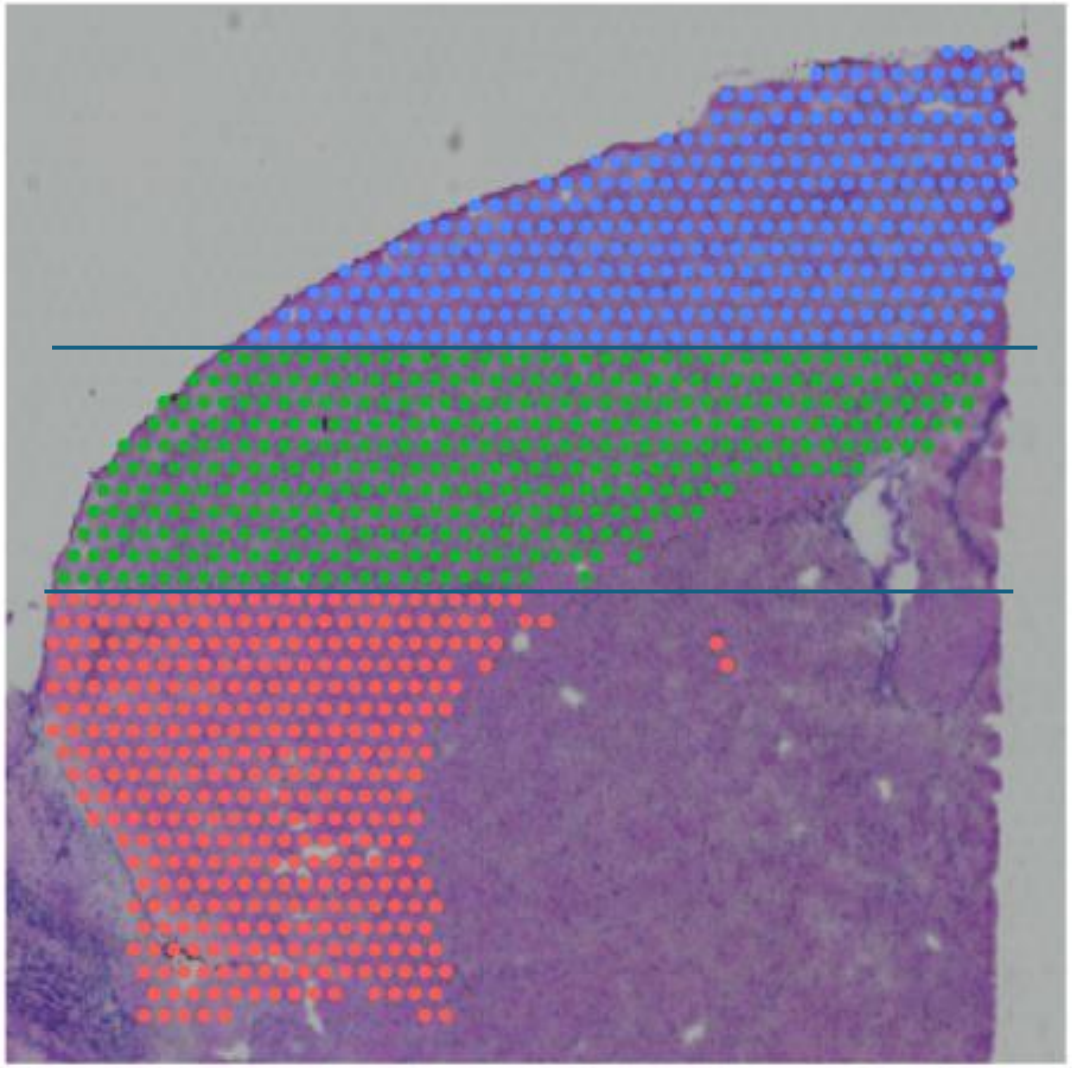
Differential Expression: C-SIDE

- C-SIDE is part of the spacexr package that provides RCTD.
- Given the deconvolution weights and a set of regions to compare, C-SIDE looks for cell-type specific DEGs.

Differential Expression: C-SIDE

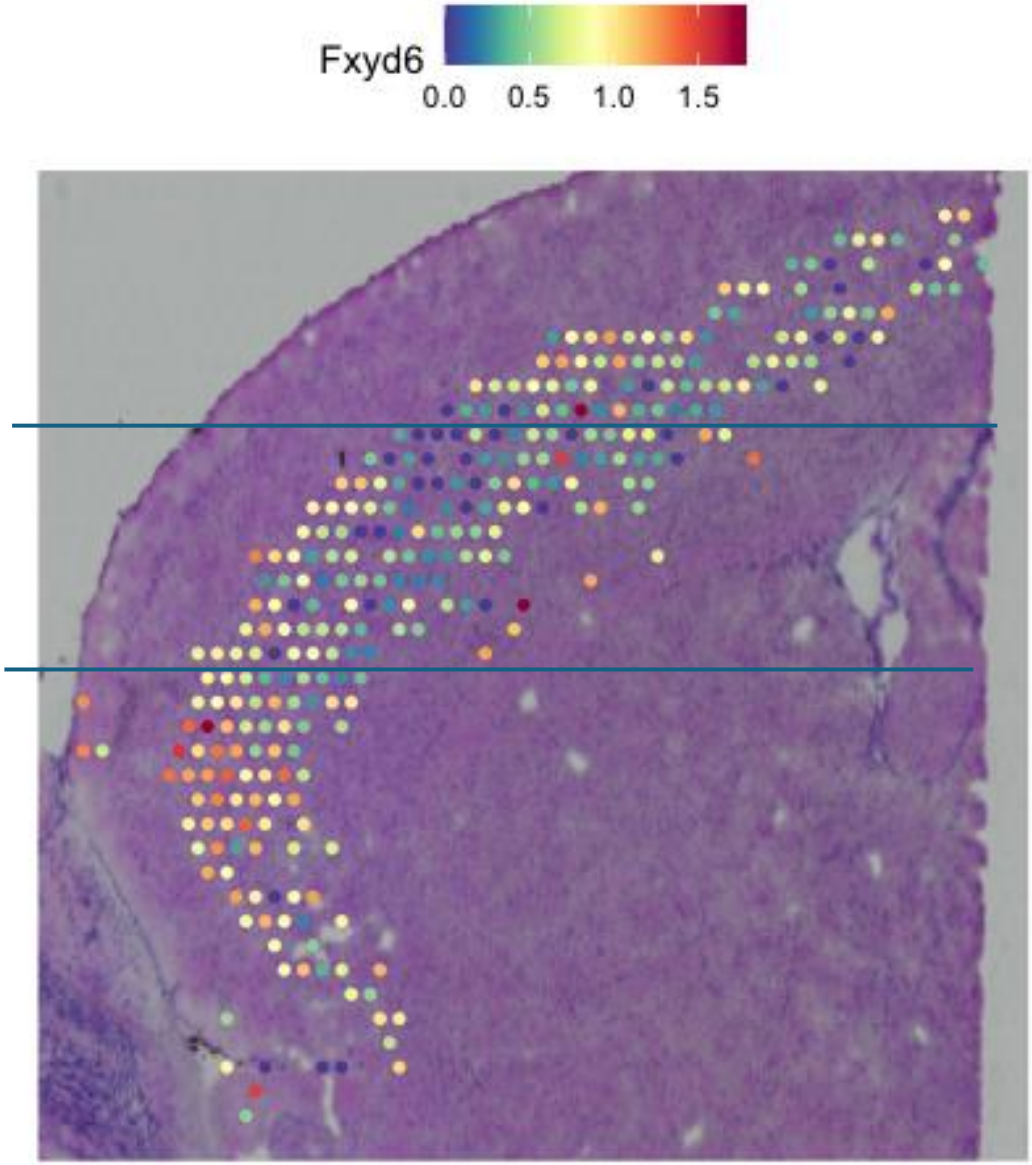
In a toy example, we divide the cortex into three regions and then run C-SIDE.

C-SIDE provides separate DE results for each cell type.



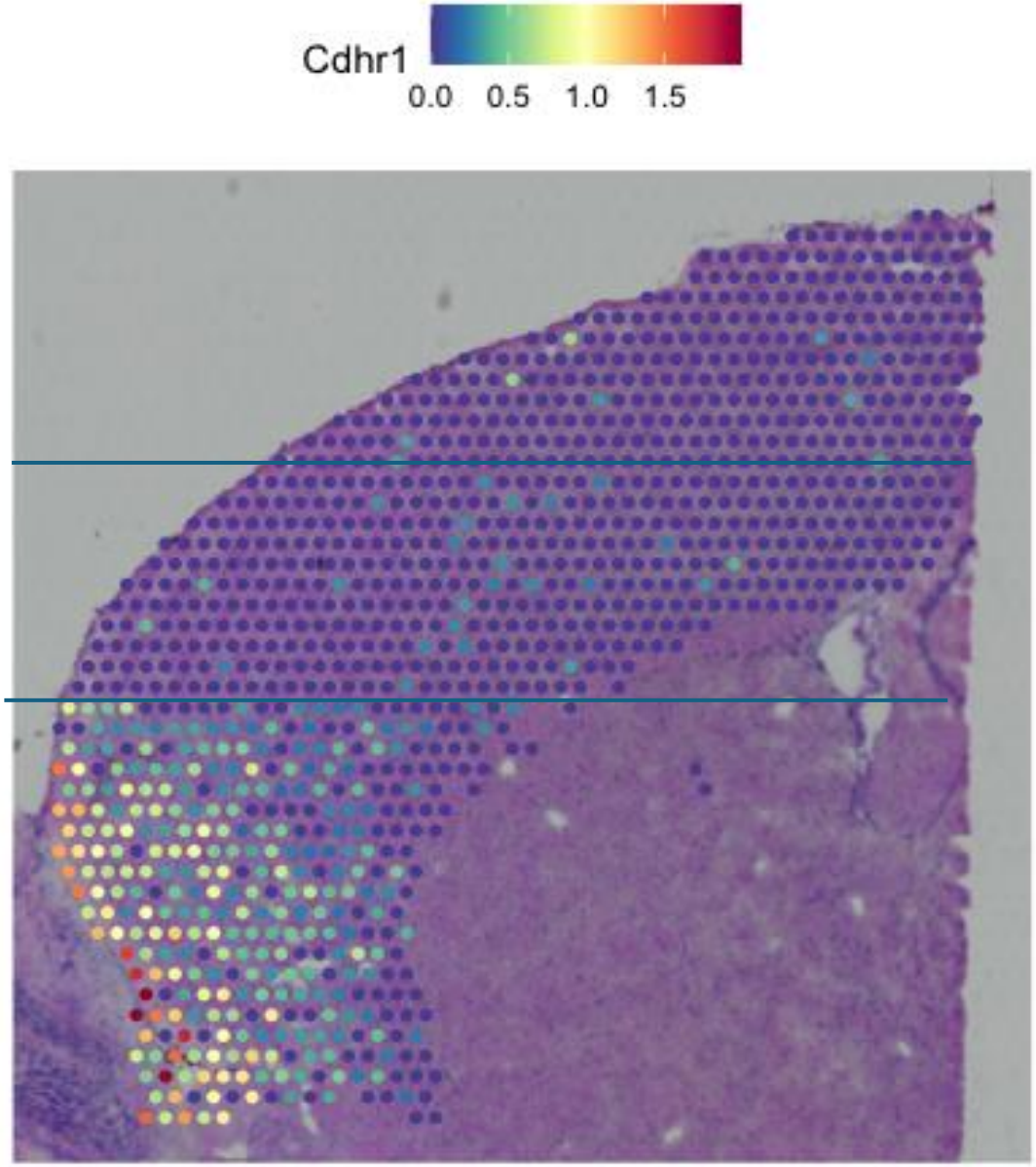
Differential Expression: C-SIDE

Within “L5 IT”, it identifies *Fxyd6* as a gene that varied between regions, and we can see that expression is elevated in the lower 1/3 of the tissue.



Differential Expression: FindMarkers

If we had used FindMarkers to compare the regions, ignoring cell types, the top difference would have been Cdhr1, which is much more specific to the lower 1/3 but varies across multiple cell types.



Visium Tutorials

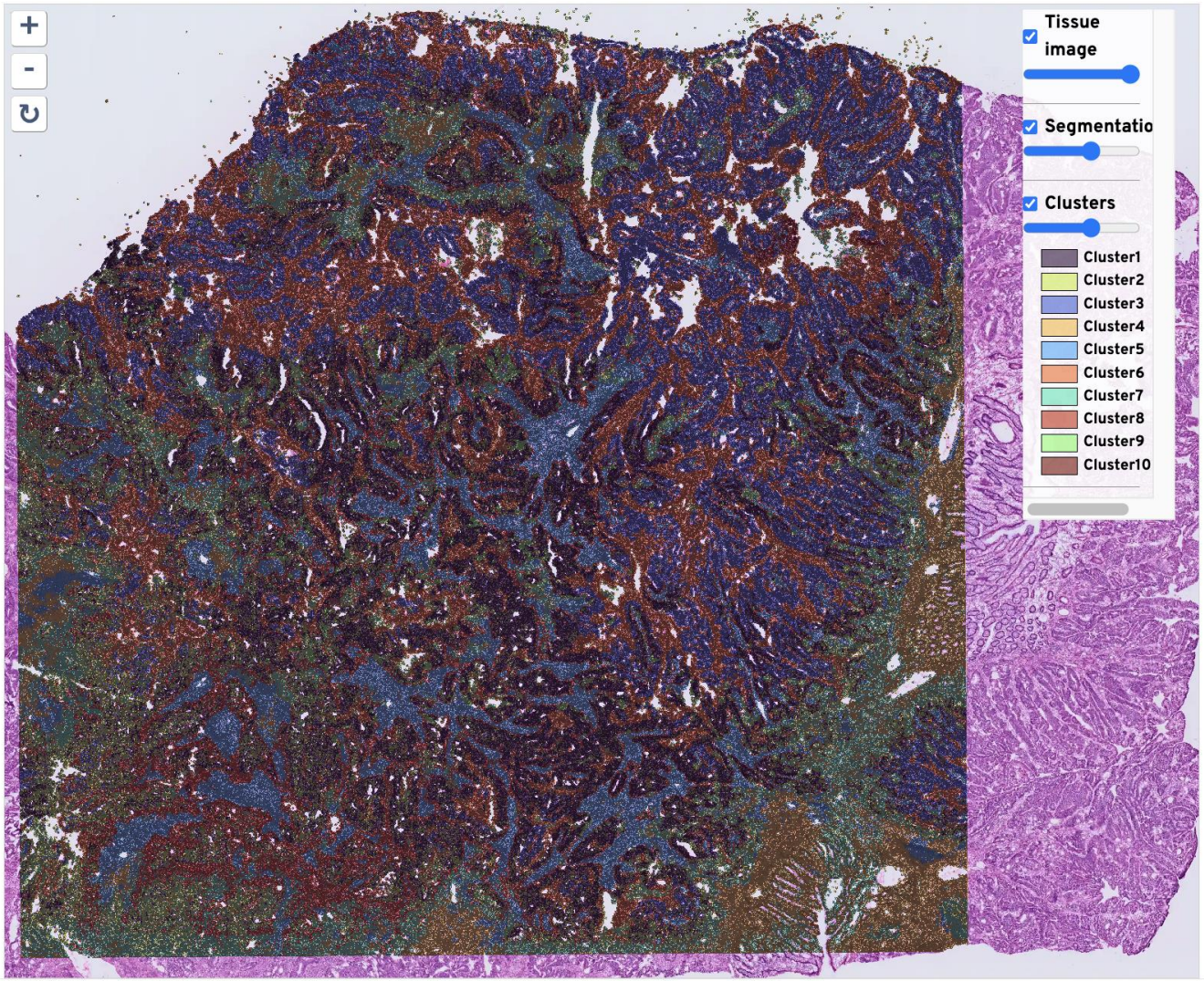
Additional details for each of these approaches (plus some alternatives) are available on GitHub:

[Preprocessing]
[Clustering]
[Deconvolution]
[SpatialCellChat]
[Differential Expression]

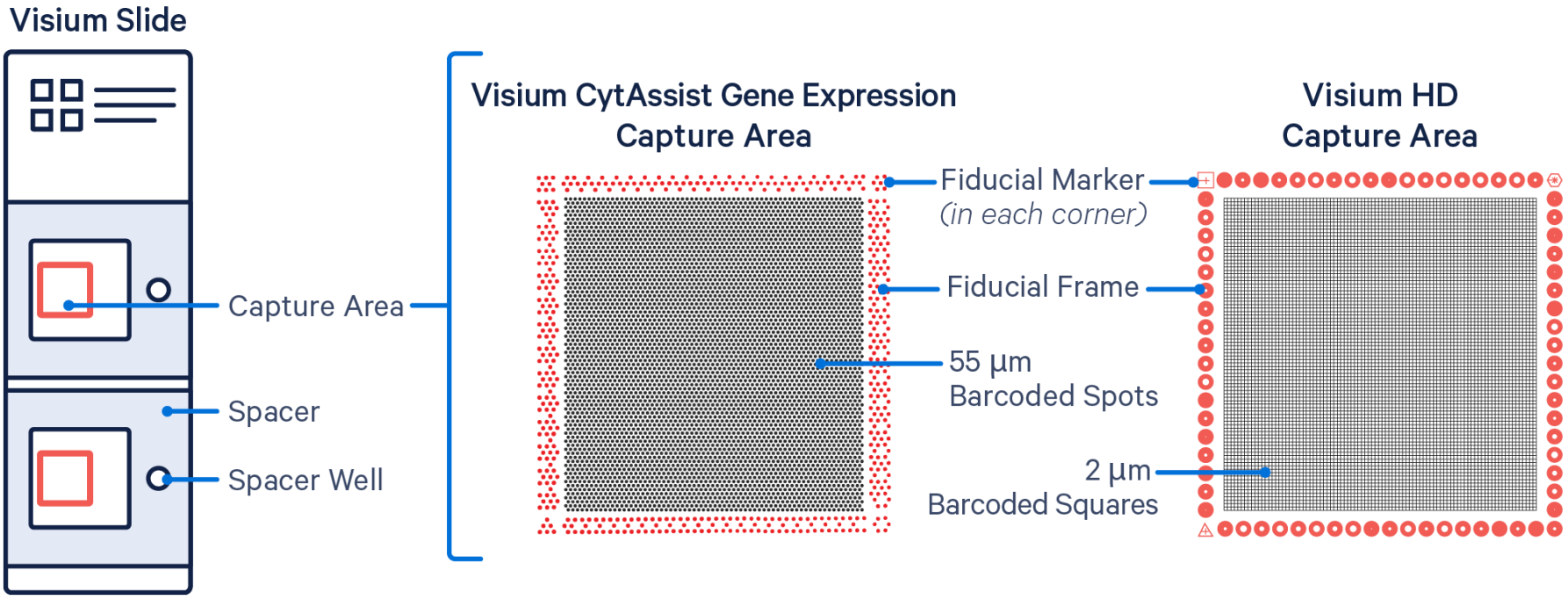
Common Issues

- SpatialFeaturePlot makes an image, but it doesn't show the spots:
  - Check the pt.size.factor. In Seurat v5, it sometimes interprets tissues as having larger spots, and the default pt.size.factor of 1 will make the spots too small to see. Try increasing values by factors of 10 to find the right range.
- You did everything right, but a package fails
  - With new versions coming out, it is very possible to have unexpected incompatibilities. There may be solutions on GitHub in some cases
  - Strategy 1: Check the GitHub for a package to see if there are known issues/solutions. Sometimes, you just need to downgrade a version
  - Strategy 2: create a fresh conda environment just for the specific tool that caused the problem
  - Strategy 3: Check for an alternative tool

Visium HD



Visium HD



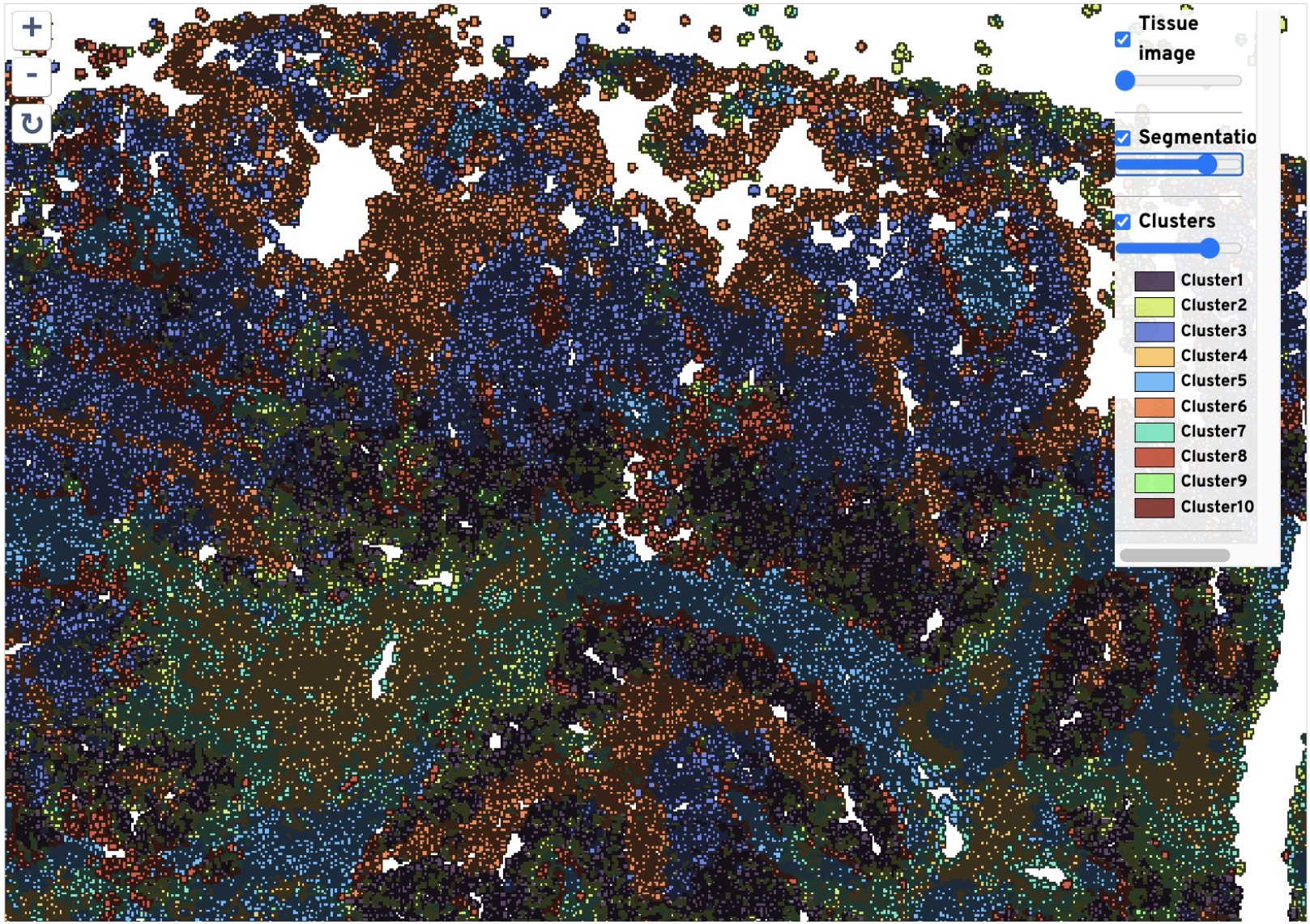
Visium HD binds tissues to a higher resolution barcode grid with 2 µm squares rather than 55 µm spots.

It then bins data in 8 µm windows and extrapolates to approximately single-cell resolution.

Visium HD

- (potential) single-cell resolution
- Data include segmentation from 10X, but there are 3rd party tools available that might offer improvements.
- Doublets are still possible. RCTD can help resolve overlapping cell types.
- Because of the higher resolution, sequence saturation for each cell tends to be lower than for each Visium spot.
- Aside from segmentation, most methods from Visium or scRNA analysis can be applied.
- Visium HD web summary example

Segmentation



Segmentation

- Segmentation is the process of determining cell boundaries in spatial data
- In Visium HD, this is done by binning transcripts in the high-resolution grid and extrapolating to separate cells.
- Xenium relies on staining to infer the nuclei, interior, and borders of each cell.
- Errors in segmentation may lead to doublets, errors in cell type annotation, and additional downstream mistakes.

Assessing Segmentation

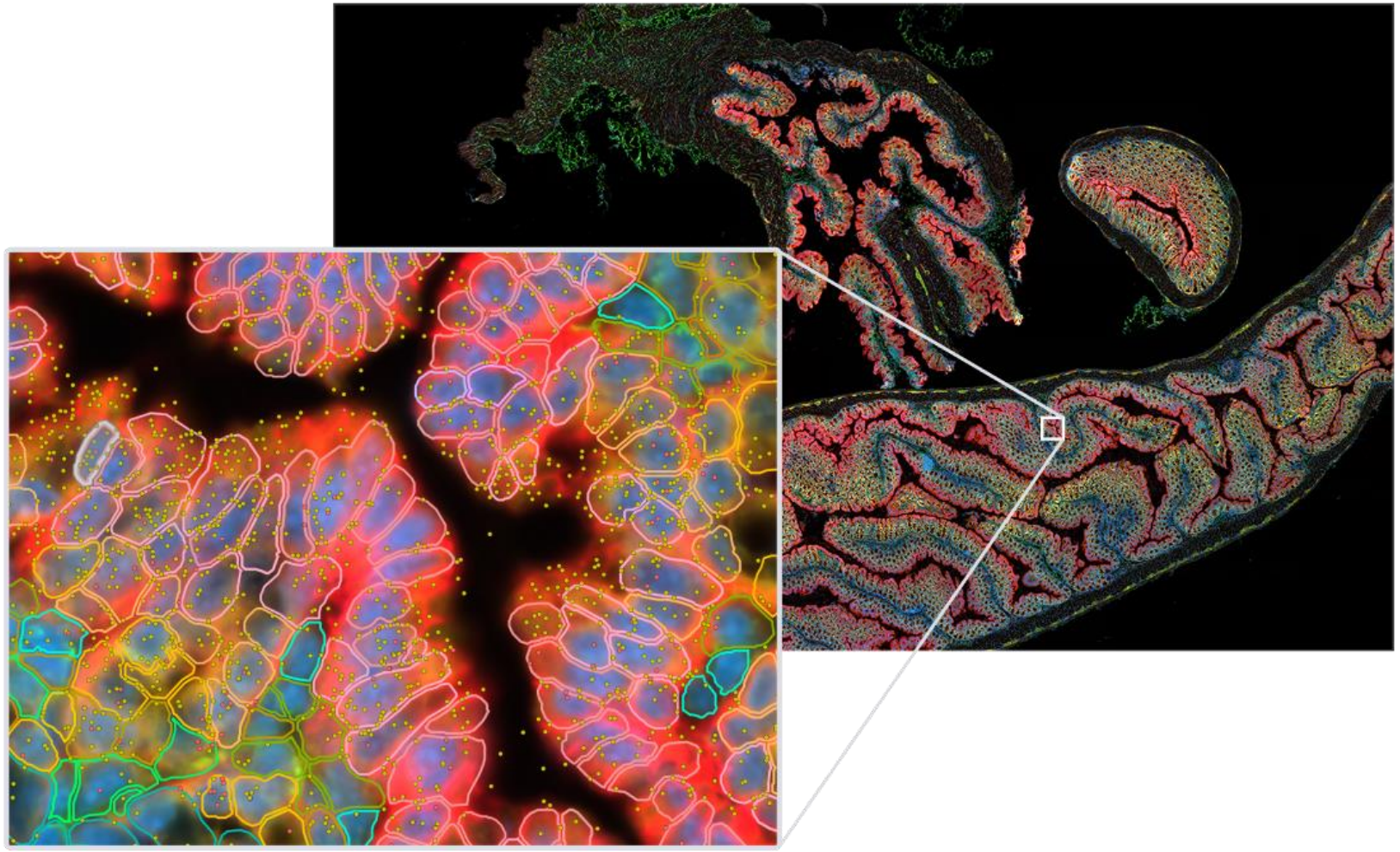
- Mutually Exclusive Co-expression Rate (MECR):
  - For a pair of genes that should be mutually exclusive, the MECR is the fraction of cells that co-express them. It is essentially a proxy for doublet rate.

- Deconvolution specificity:
  - We can also assess segmentation by examining deconvolution weights for each cell type (e.g. using RCTD). When segmentation is accurate, each cell should be assigned to a single type.
- Packages for Visium HD: bin2cell, ENACT

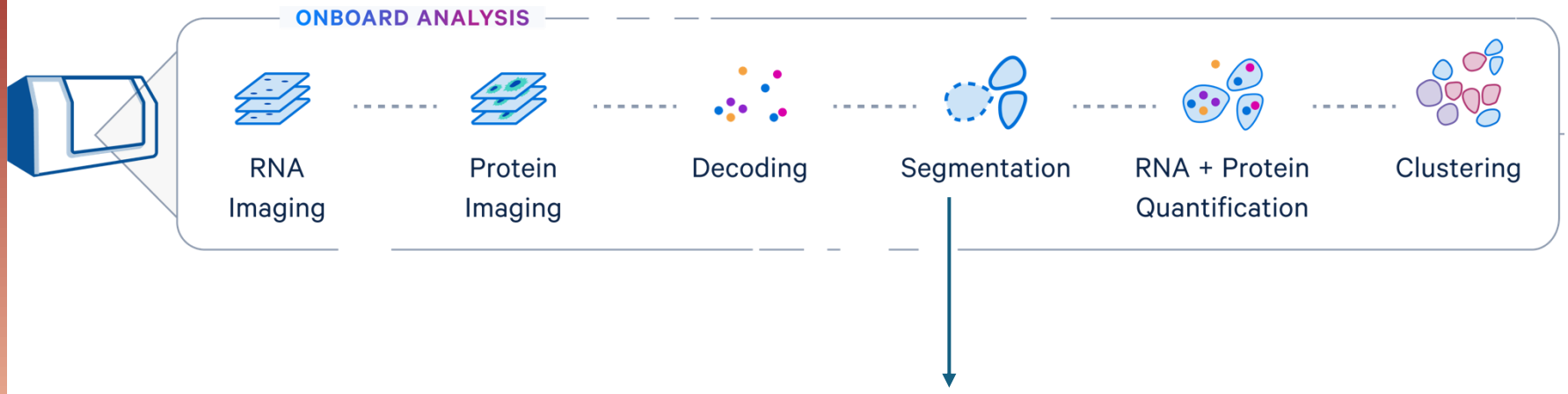
Questions and Break

- We will take some time now for any questions on Visium, and then we will take a 5 minute break before discussing Xenium.

Xenium



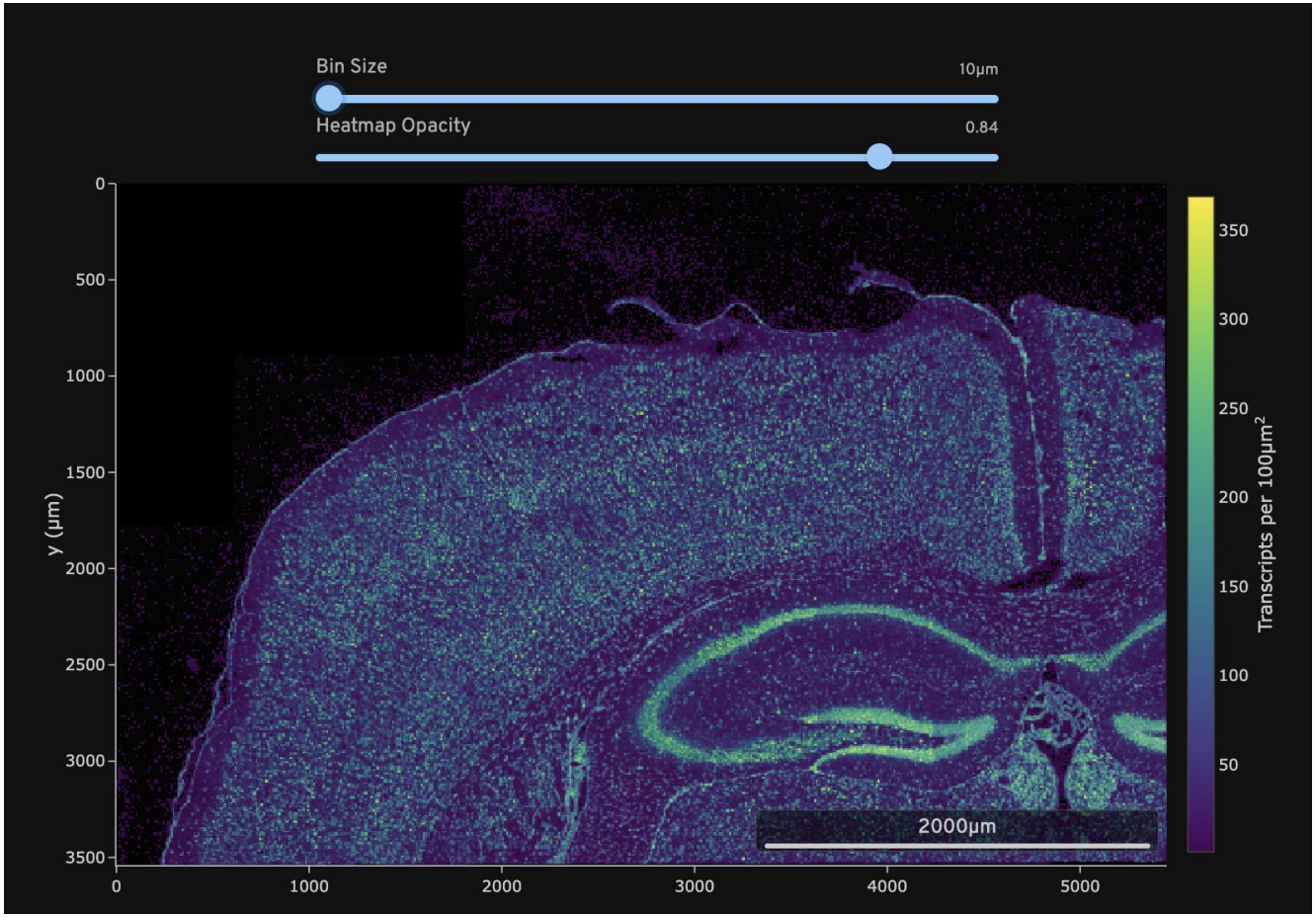
Xenium



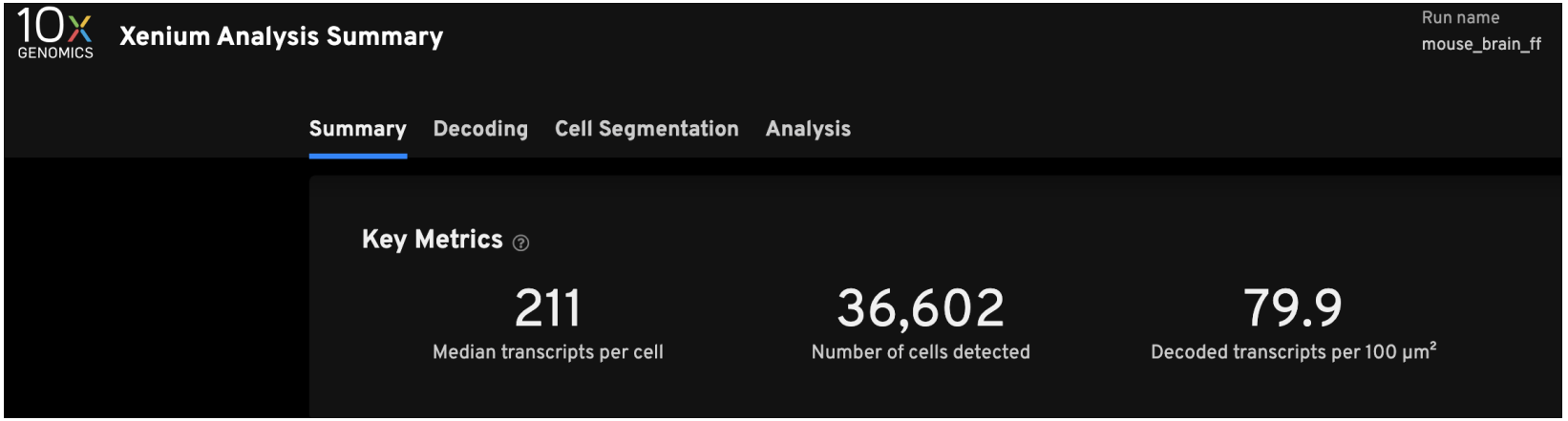
Based on fluorescent stains.
Often more accurate than Visium HD

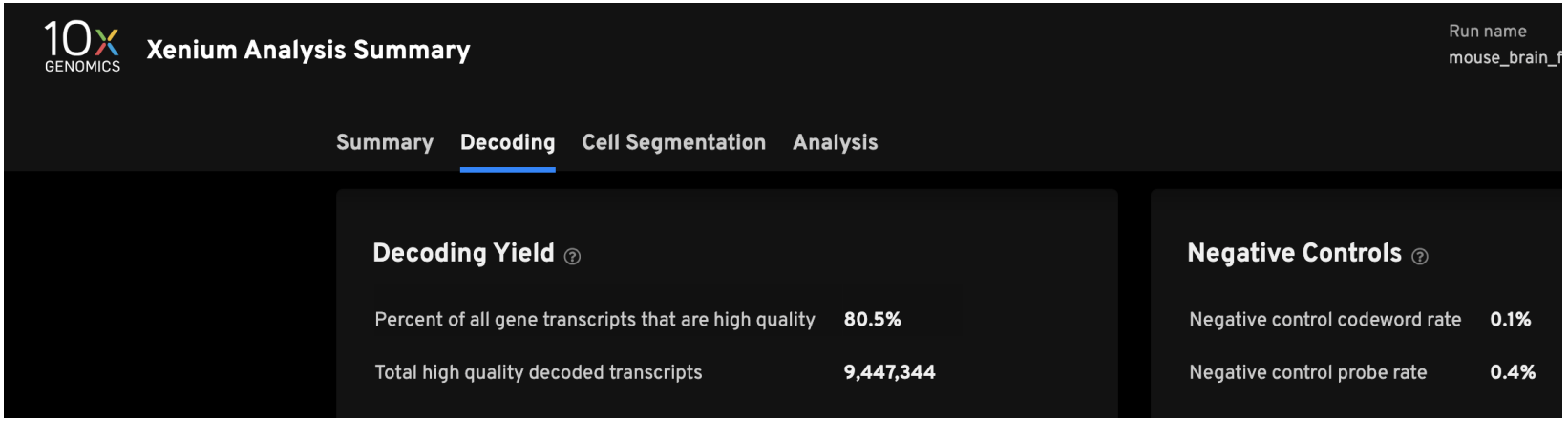
Images from 10X Genomics

Xenium



Xenium QC



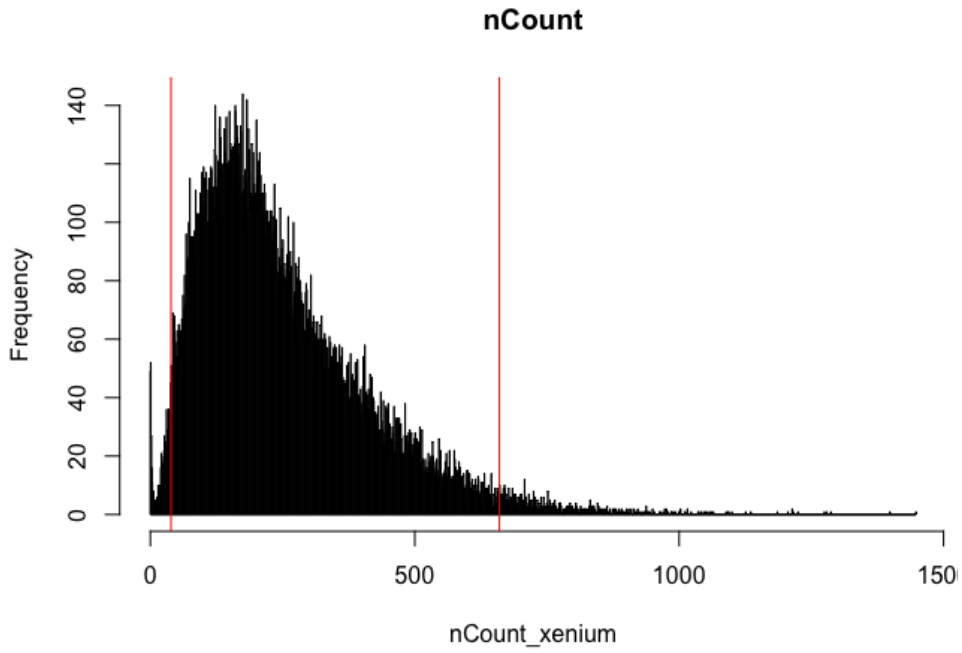


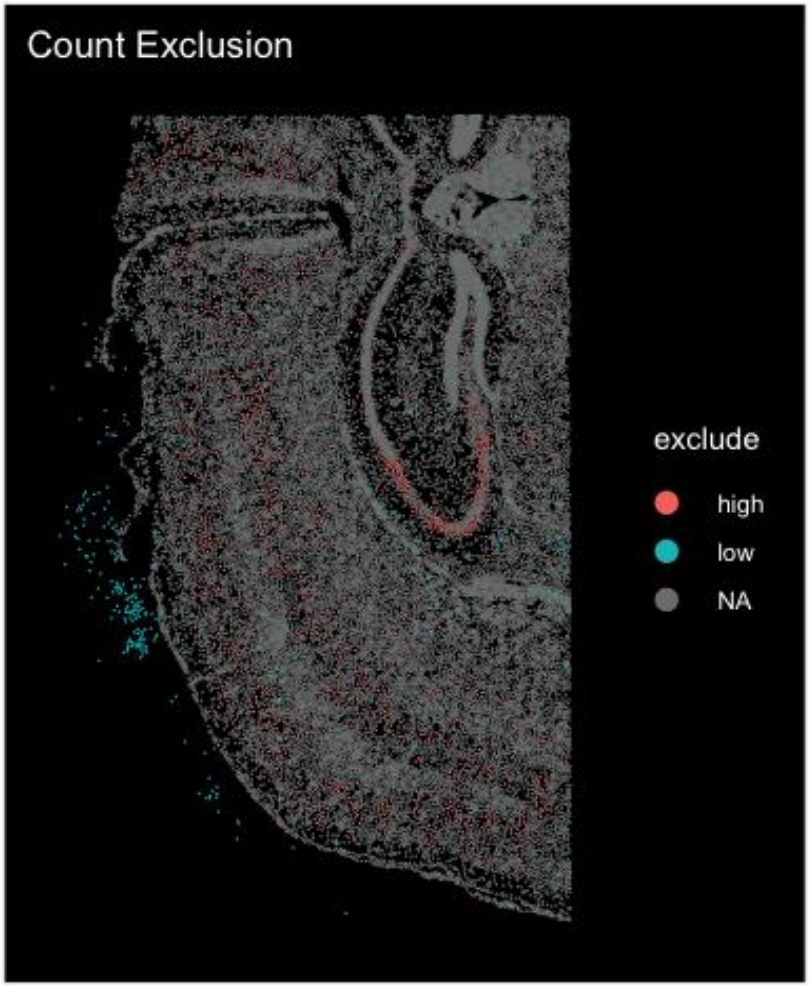
**These screenshots are from a report from an earlier version of Xenium. The newer version has similar information with a different layout*

Xenium Explorer

- Replaces the Loupe Browser when working with Xenium data
- Requires .xenium files (provided with Xenium output)
- Enables visualizing each transcript for each cell, along with segmentation and clustering
- Can import or export lists of cells.
- Memory requirement: 10X recommends 32GB of RAM
- 10X analysis summary example

Xenium QC



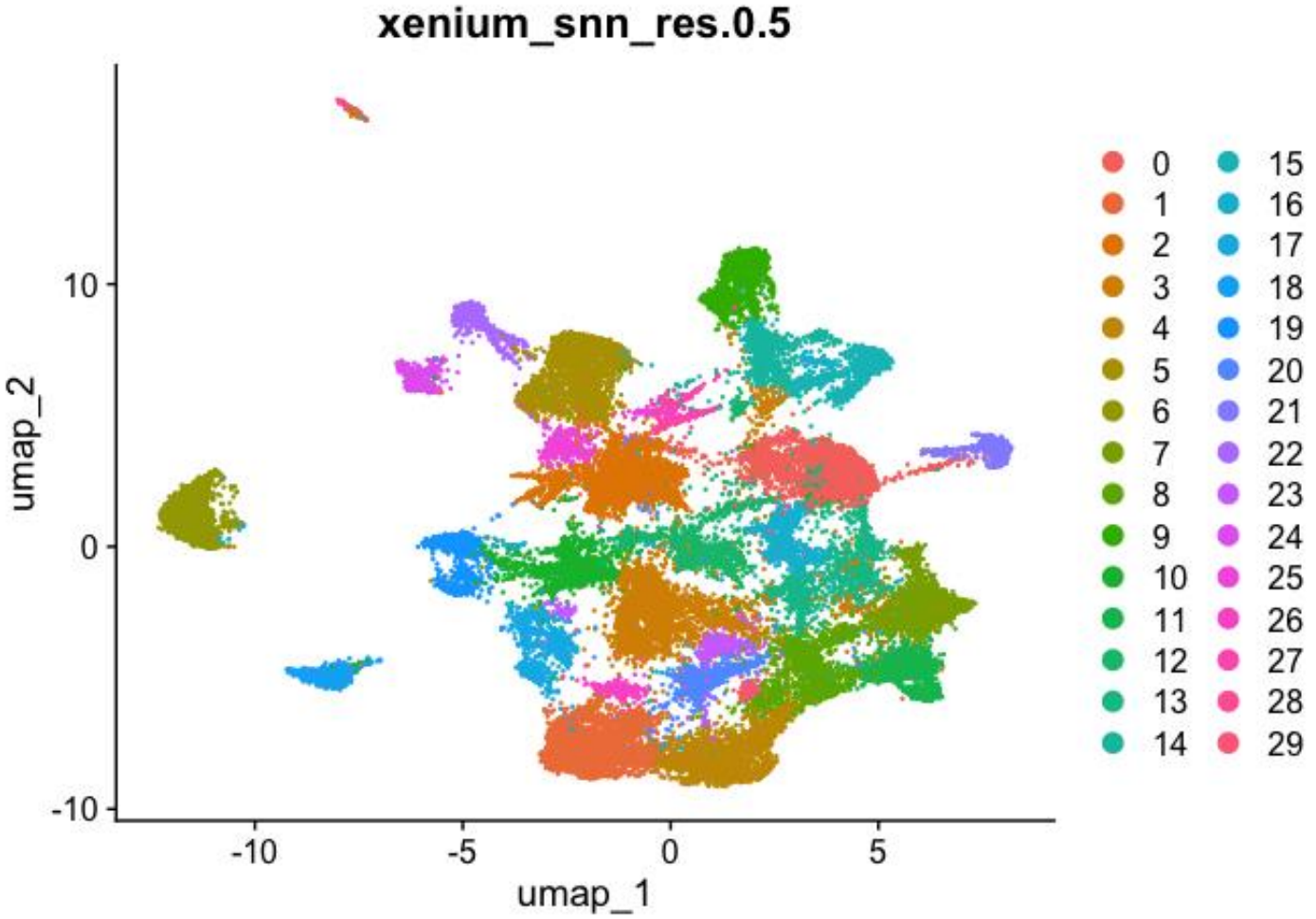


With Xenium data, we will apply similar count filters as with scRNA data. Rather than use fixed cutoffs, it's a good idea to visualize the distribution for the sample. One recommendation is to filter out the lower or upper 2%, but it's also a good idea to check where these cells appear in the tissue before filtering.

Xenium segmentation

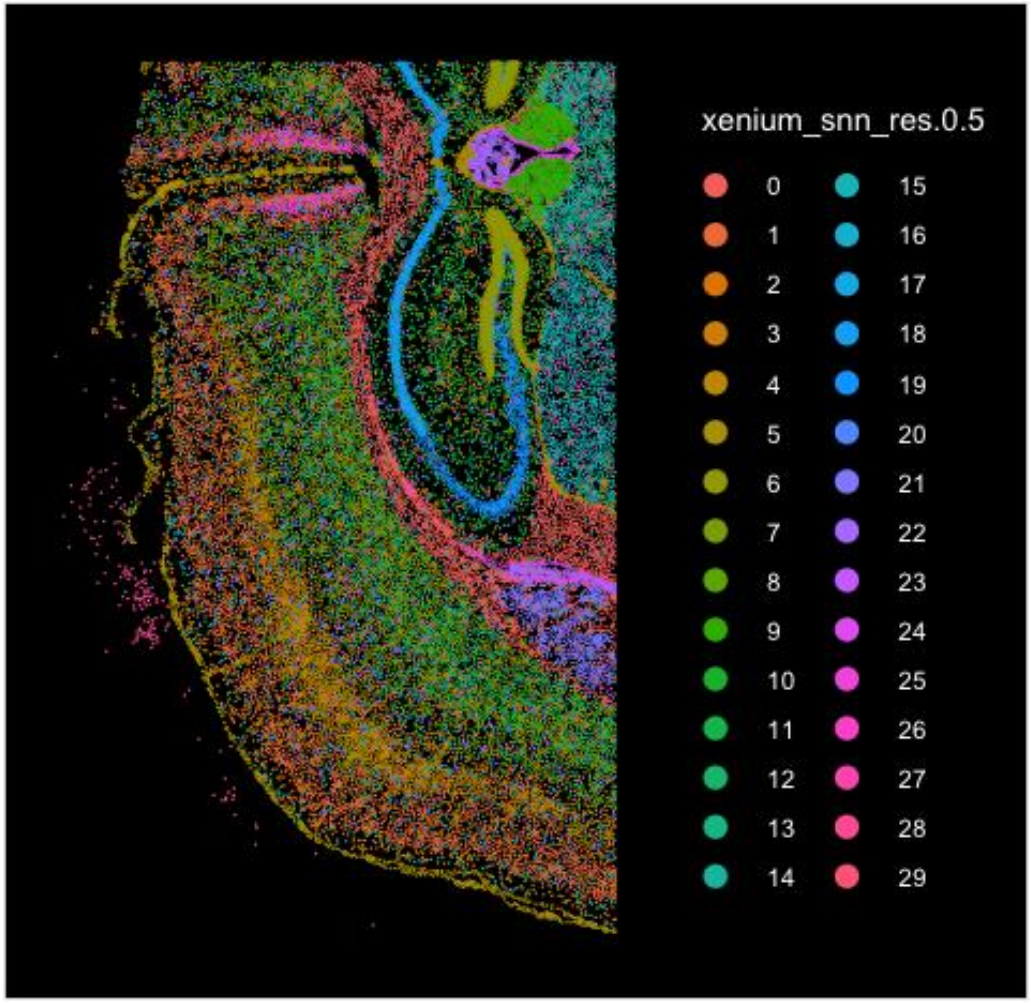
- 10X results provide segmentation by cell or by nucleus
- Segmentation is based on detected cell boundaries in the imaging and tends to be more accurate than Visium HD segmentation.
- Packages for Xenium: Segger, Bering, Baysor, Cellpose (and others)

Xenium Clustering



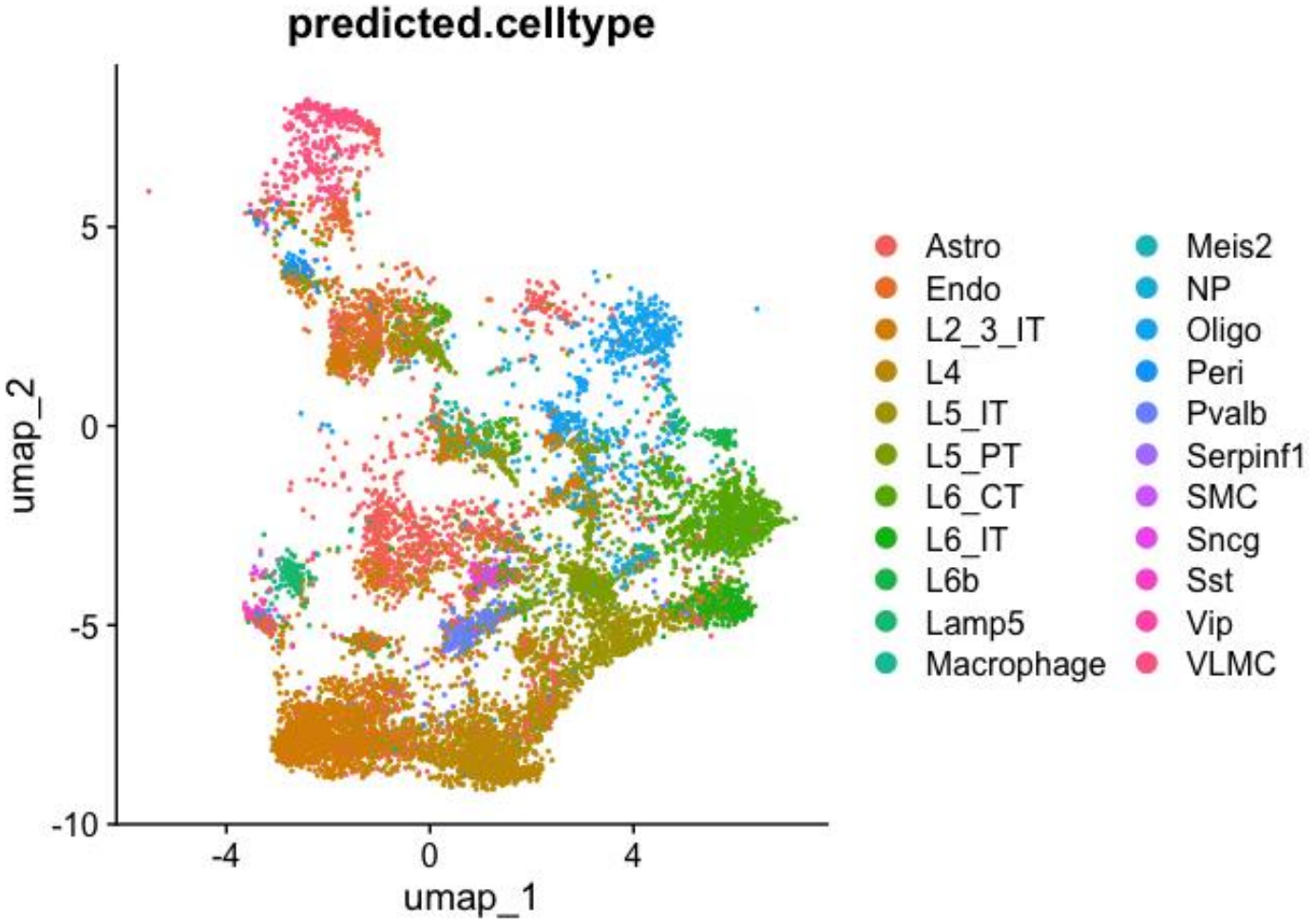
The reduction looks much more like scRNA data, and we detect many more clusters than with Visium. (The clusters above are from Seurat.)

Xenium Clustering



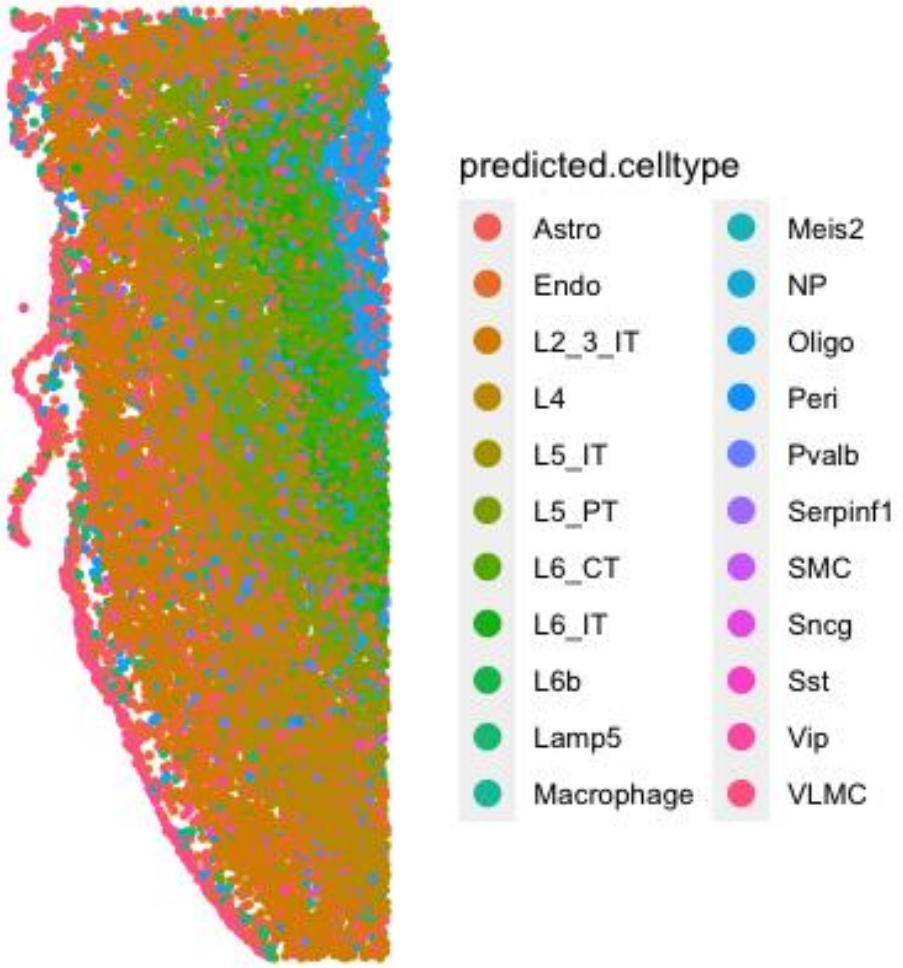
Clusters map well to the tissue, but we can some noise in different regions. Alternative tools (e.g. MEcell) are designed for spatial clustering with Xenium.

Xenium Deconvolution/Annotation



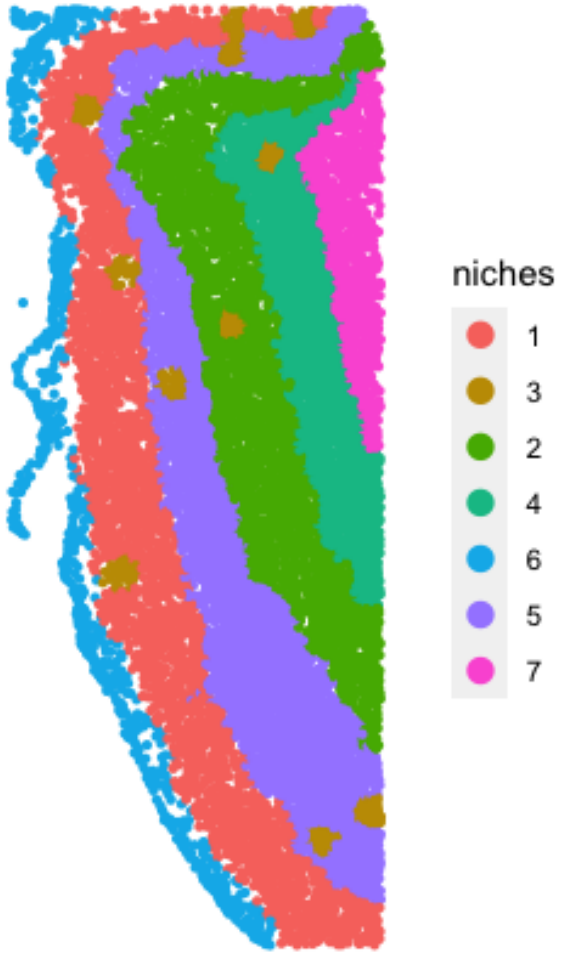
Here, we apply RCTD to a cortex subset of the Xenium data, similar to how we handled Visium. The result is closer to annotation than deconvolution, but some clusters might include doublets or segmentation issues.

Xenium Deconvolution/Annotation



When mapped to the tissues, we can see more noise in the annotations than we might expect from the UMAP. We might want to update the annotation to broader cell types or work on the segmentation.

Xenium Niche Analysis



Seurat has a function called BuildNicheAssay that will take the cell type predictions (or any other grouping like clusters) and build “niches” that represent broader sets of spatially similar cells and interactions.

Xenium in Python

- Due to the large number of cells, it is often recommended to work with Xenium data in Python.
- 10X has an example analysis in a Jupyter notebook that is a good starting point.
- Standard analyses rely on tools from the scverse community, including: scanpy, squidpy, and spatialdata.
- The workflow is similar to this example from the Biocore's previous workshop on CosMx analysis.

Xenium Tutorial Code

- Example code is available here
- Thank you!