Github page for workshop:

https://github.com/margaretc-ho/BCBB_STx_workshop_2024

Spatial Transcriptomics Part 1: Intro to Methods and Concepts for Data Analysis

Margaret Ho, PhD
Bioinformatics and Computational Biosciences Branch (BCBB) at NIAID
margaret.ho@nih.gov

Moderator: Colton McNinch, PhD (BCBB/NIAID)

Image credit to Bo Xia

Bioinformatics and Computational Biosciences Branch (BCBB) at NIAID

BCBB offers the following scientific services & resources such as software development, collaboration & training, and biocomputing resources for the NIAID research community and its collaborators:

Biovisualization and 3D Printing
Clinical Genomics
Data Science and Biostatistics
Imaging
Metagenomics
Non-Human / Microbial Genomics
Structural Biology

More about BCBB:
https://www.niaid.nih.gov/research/bioinformatics-and-computational-biosciences-branch-scientific-services
Looking for bioinformatic and genomics analysis expertise? bioinformatics@niaid.nih.gov

Learning Objectives of Part 1

- understand concepts underlying spatial transcriptomics (STx) methods
- consider important aspects of experimental design depending on biological sample and research question and select best suited STx method for your experiments

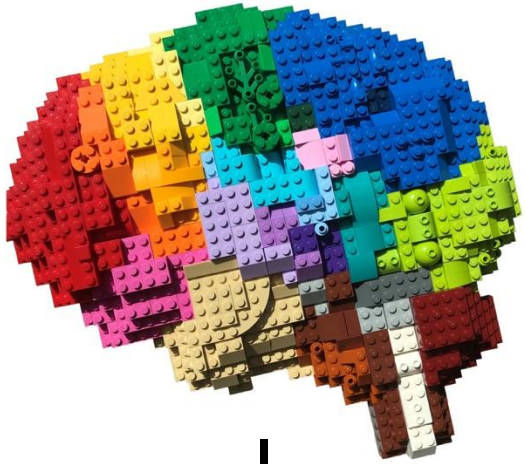
(quick 5 min break)

- understand techniques for processing and analyzing STx data with Seurat and potential downstream analysis

Github page for workshop
https://github.com/margaretc-ho/BCBB_STx_workshop_2024

Followup questions and inquiries are welcome!
margaret.ho@nih.gov

Spatial Transcriptomics Concepts



the original organ

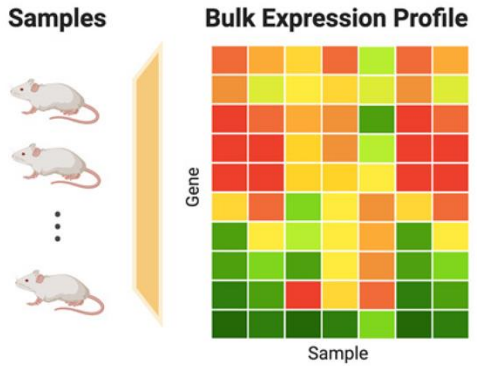
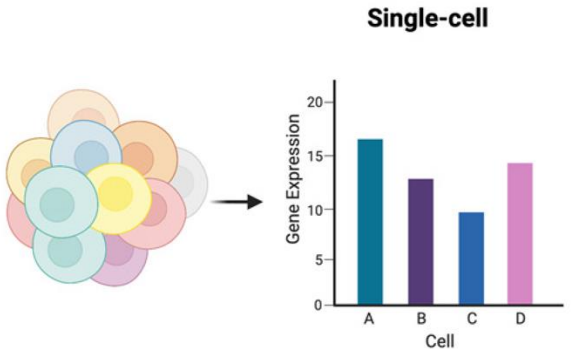
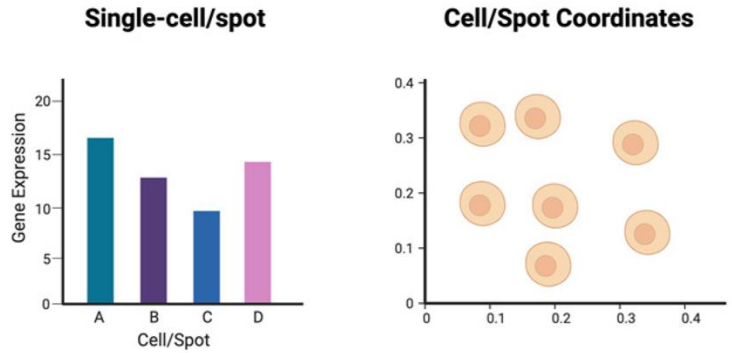
Cell- cell interaction networks
Spatial information preserved

bulk RNA-seq

single-cell RNA-seq (scRNAseq)

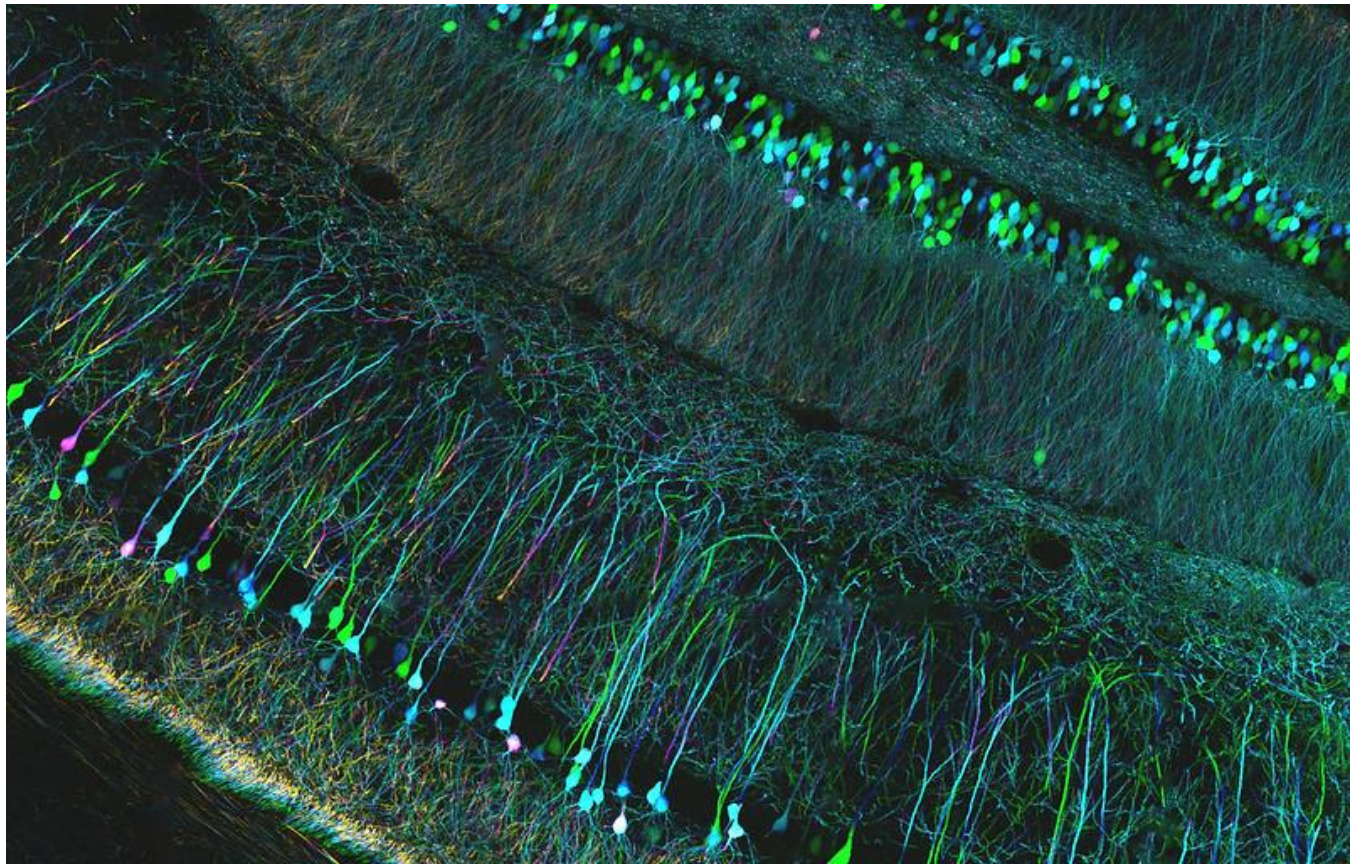
spatial transcriptomics (STx)

Image credit to Bo Xia

	Bulk RNA-Seq	Single-cell RNA-seq	High-throughput Spatial Transcriptomics
Level			
Data Structure	Subject x Gene Expression Count Data	Cell x Gene Expression Count Data	Cell/Spot x Gene Expression Count Data Cell/Spot 2-dimensional Coordinates
Detection Target	Differentially Expressed Genes	Differentially Expressed Genes Cell Sub-populations	Spatially Variable Genes Tissue Architecture Cell-Cell Communication

Jeon, H., Xie, J., Jeon, Y., Jung, K. J., Gupta, A., Chang, W., & Chung, D. (2023). Statistical power analysis for designing bulk, single-cell, and spatial transcriptomics experiments: Review, tutorial, and perspectives. *Biomolecules*, *13*(2), 221. https://doi.org/10.3390/biom13020221

Why spatial transcriptomics?



- Location, location, location!
- Despite the success of scRNA-seq, one needs to liberate viable cells from whole tissue without inducing stress, cell death, or cell aggregation
- **Two major advantages: No need for dissociation** and **preserves the spatial context of cells**
- Subcellular localization of RNA can be very important for function

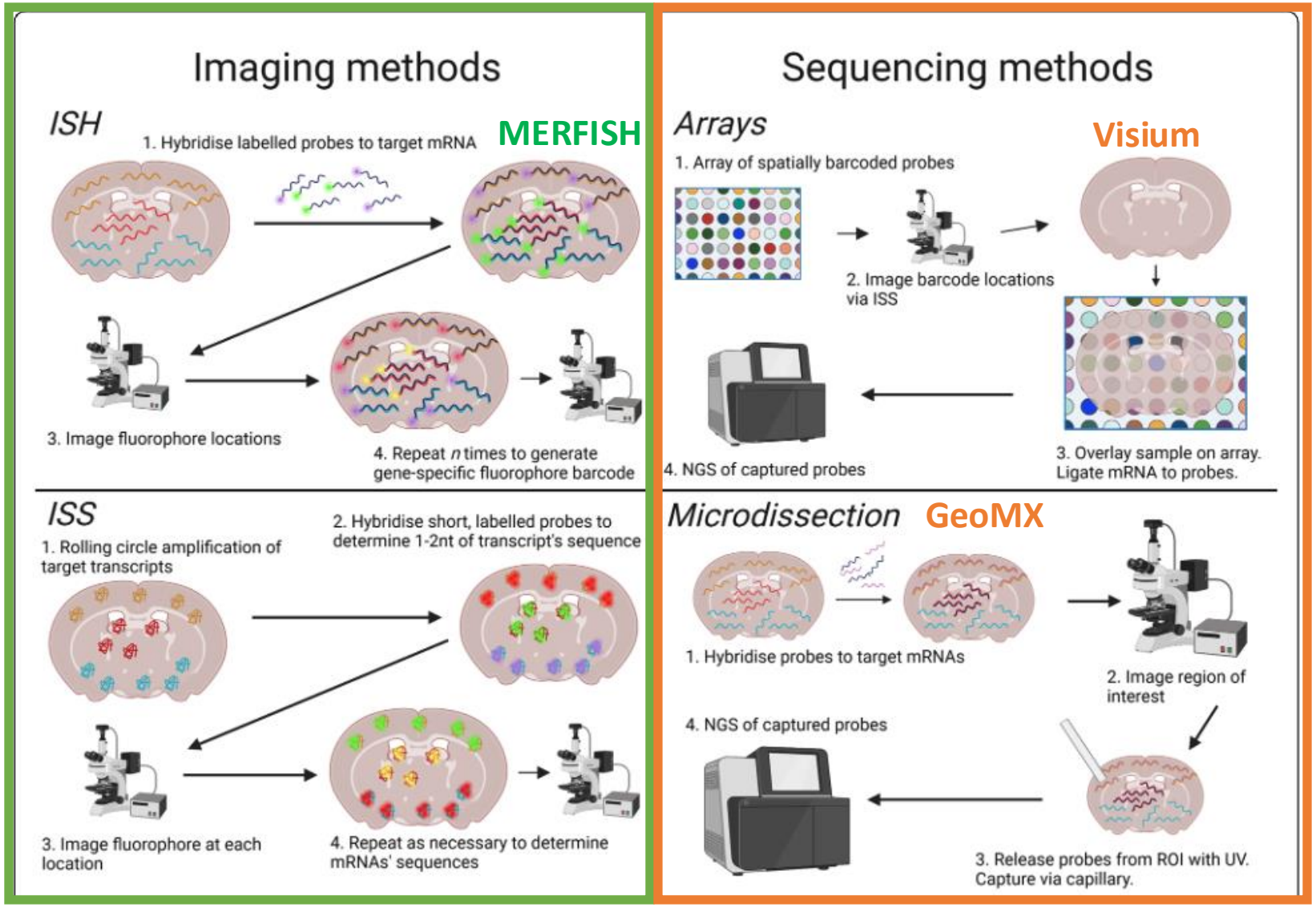
https://www.flickr.com/photos/zeissmicro/10799673016/in/photostream/
Mouse hippocampal neurons (depth coded projection)

Imaging vs Sequencing-based STx methods

In all methods, tissues are stained with other antibodies / histology and imaged for overall spatial orientation

Tradeoffs include:

- Multiplexing (# of transcripts)
- Resolution
- Throughput
- Sensitivity



Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., & Haque, A. (2022). An introduction to spatial transcriptomics for biomedical research. *Genome Medicine*, *14*(1), 68.

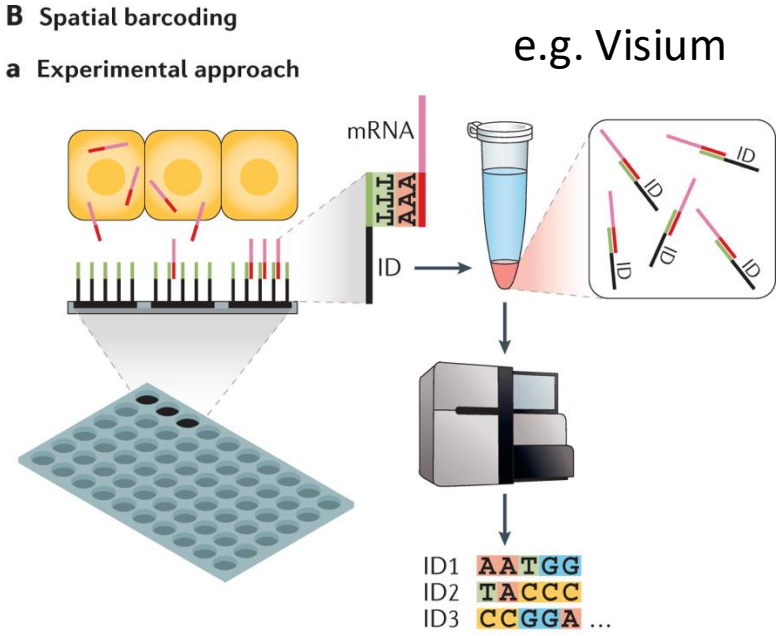
A tradeoff between cellular resolution and # of transcripts

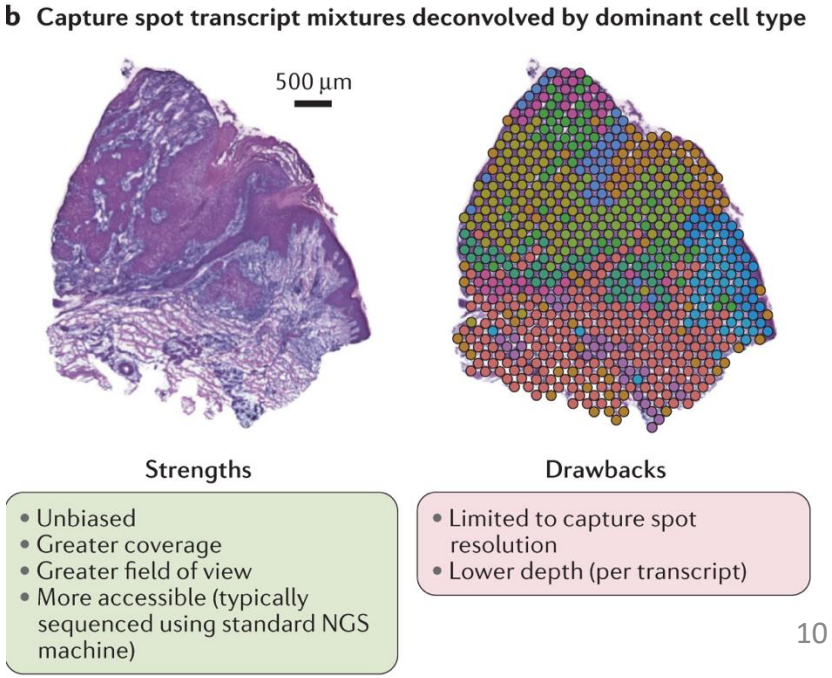
Imaging methods (e.g. MERFISH, Xenium)	Cellular or Subcellular resolution	1uM Sub-cellular resolution		Detects fewer transcripts (panel-based hundreds of genes)
Sequencing-based	Larger than cellular resolution (array-based like Visium) or region of interest (GeoMX or similar)	10uM Single-cell resolution	100uM Multi-cell resolution	Detects more transcripts (de novo or whole transcriptome)

Cell resolution: https://doi.org/10.1038/s41576-023-00586-w

Sequencing-based STx

- **Can profile whole transcriptome**
- **Unbiased** / Less need for *a priori* knowledge
- **Lower spatial resolution**
- Typically lower than cellular resolution depending on tissue – can require **deconvolution** and/or **mapping with scRNA-seq datasets** to analyze
- **More accessible** (ie. standard NGS system with some fluorescence and brightfield imaging to capture spot information or use of Visium CytAssist machine)



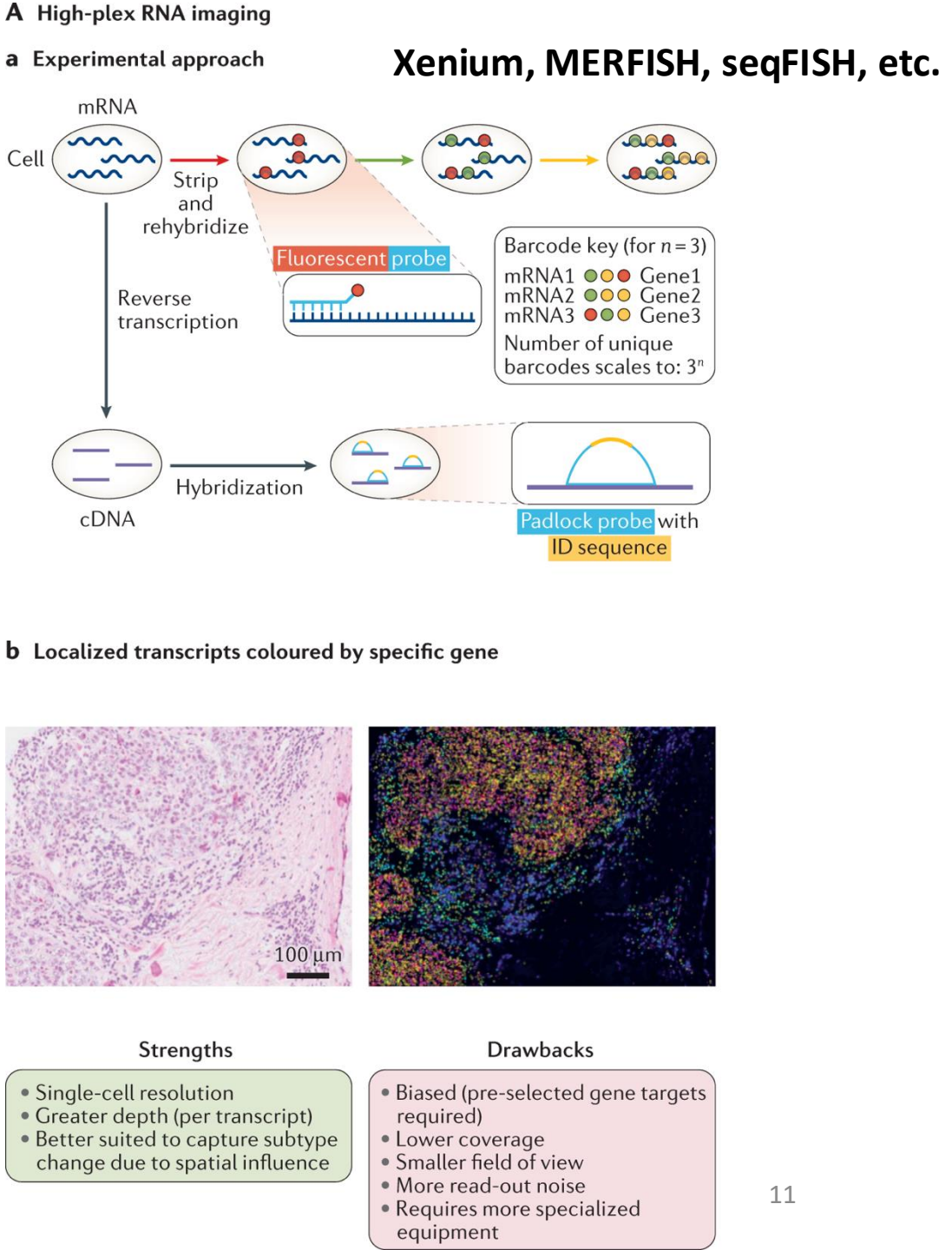


Longo, S. K., Guo, M. G., Ji, A. L., & Khavari, P. A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews. Genetics*, *22*(10), 627–644.

Imaging-based STx

- **Based off of single molecule FISH (smFISH)**
- **Single Cell to Subcellular Resolution** via localization of single mRNA molecules
- Often relies on **cell segmentation** with immunostaining with membrane markers to delineate cell boundaries
- Requires good tissue clearing methods
- ***A priori* knowledge** needed to select genes and design/use existing probe set
- **Typically 500-1000 genes**
- Need to consider issue of molecular crowding

At least one paper gets around this with ExM to do ten thousand genes and looking at ER subcellular localization of transcripts but it has not been implemented in commercial kits -- Xia Fan et al. PNAS 2019 -- https://www.pnas.org/doi/10.1073/pnas.1912459116



Longo, S. K., Guo, M. G., Ji, A. L., & Khavari, P. A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews. Genetics*, *22*(10), 627–644.

Choosing an STx method

Comparison of major commercial STx platforms

	Platform	# of genes profiled	Spatial Resolution	RNA Capture efficiency	Imaging Area	Time Required
Imaging-Based	**MERSCOPE**	**500**	**100 nm**	**95% for cells 80–85% for tissue**	**10 mm × 10 mm**	**28–30 h**
	10X Xenium	**400**	**50 nm**	**Unavailable**	**12 mm × 24 mm**	**2 days**
	CosMX SMI	1000	50 nm	Unavailable	20mm x 15mm	3 days-1 week
Sequencing-Based	**10X Visium HD**	**Whole Transcriptome**	**2-8um**	**Unavailable**	**(6.5 mm × 6.5 mm)x2**	**Few hours**
	GeoMx DSP	Whole Transcriptome	50um/ ROI	Unavailable	35.3 mm × 14.1 mm	Few hours
	Stereo-Seq	Whole Transcriptome	0.5um	12,661/100um^2	10 mm × 10 mm 13.2 cm × 13.2 cm	Few hours

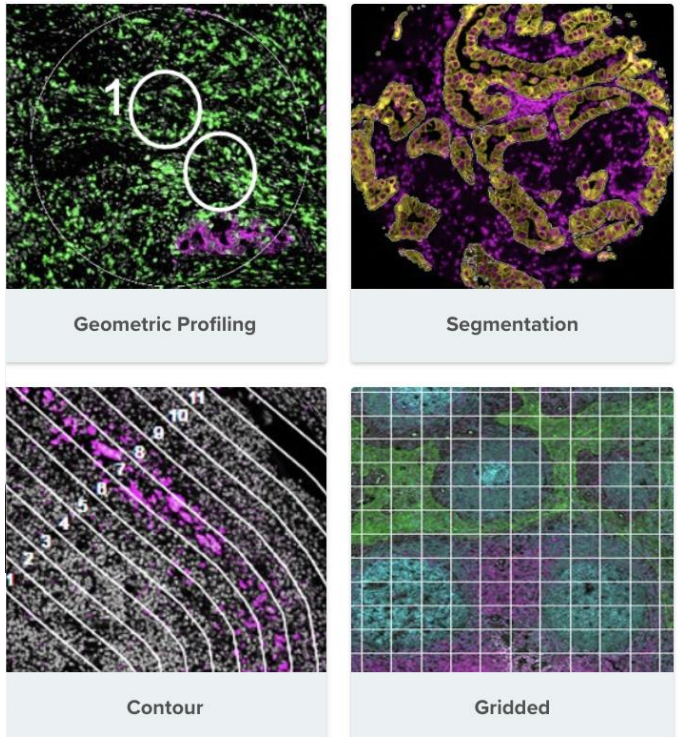
Adapted from Wang, Y., Liu, B., Zhao, G., Lee, Y., Buzdin, A., Mu, X., Zhao, J., Chen, H., & Li, X. (2023). Spatial transcriptomics: Technologies, applications and experimental considerations. *Genomics*, *115*(5), 110671.

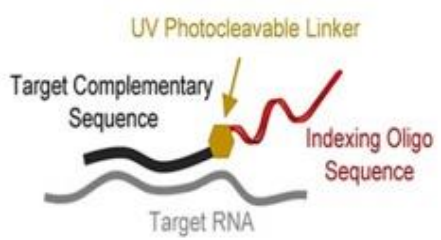
Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
GeoMx DSP	Whole Transcriptome	50um / ROI	35.3 mm × 14.1 mm	Few hours

Launched in 2019

GeoMx DSP

Within a Region of Interest (ROI), Probe with Gene-Specific Barcode Is Released Upon UV Exposure





Each probe is linked with a gene-specific barcode via UV cleavable linker. The barcodes are cleaved from the selected region of interest, and collected for library prep and sequencing. **Resolution of sequencing is limited to ROI, throughput is low (96 ROIs)**

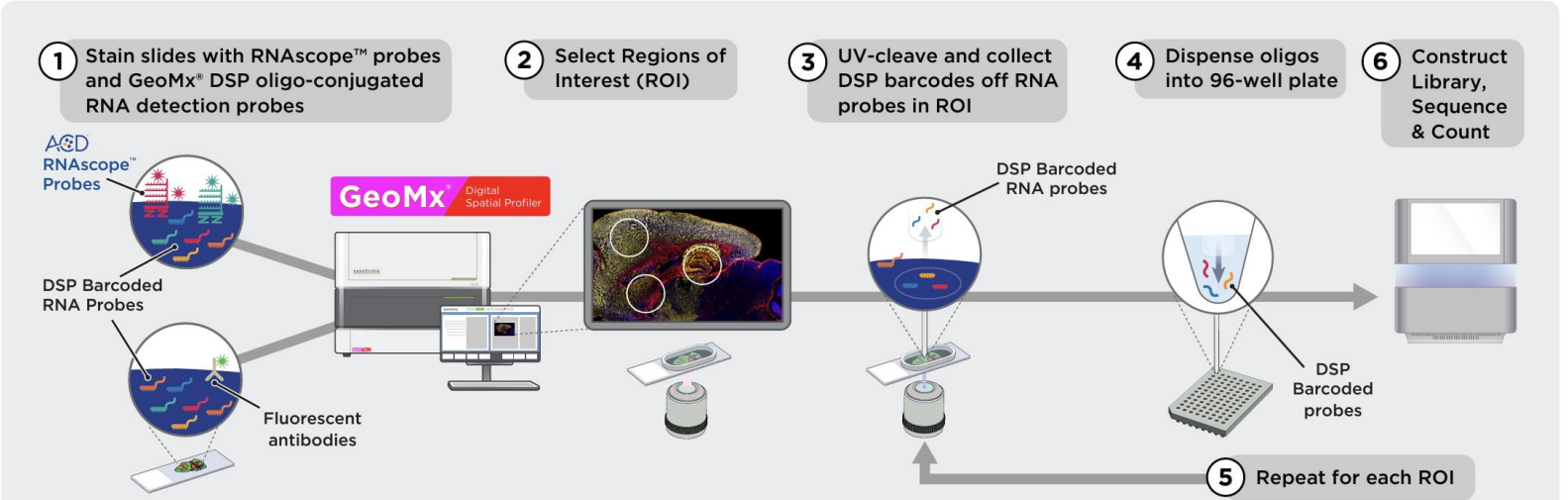


Figure 5 : GeoMx DSP workflow enables the analysis of whole transcriptome data from user-defined regions.

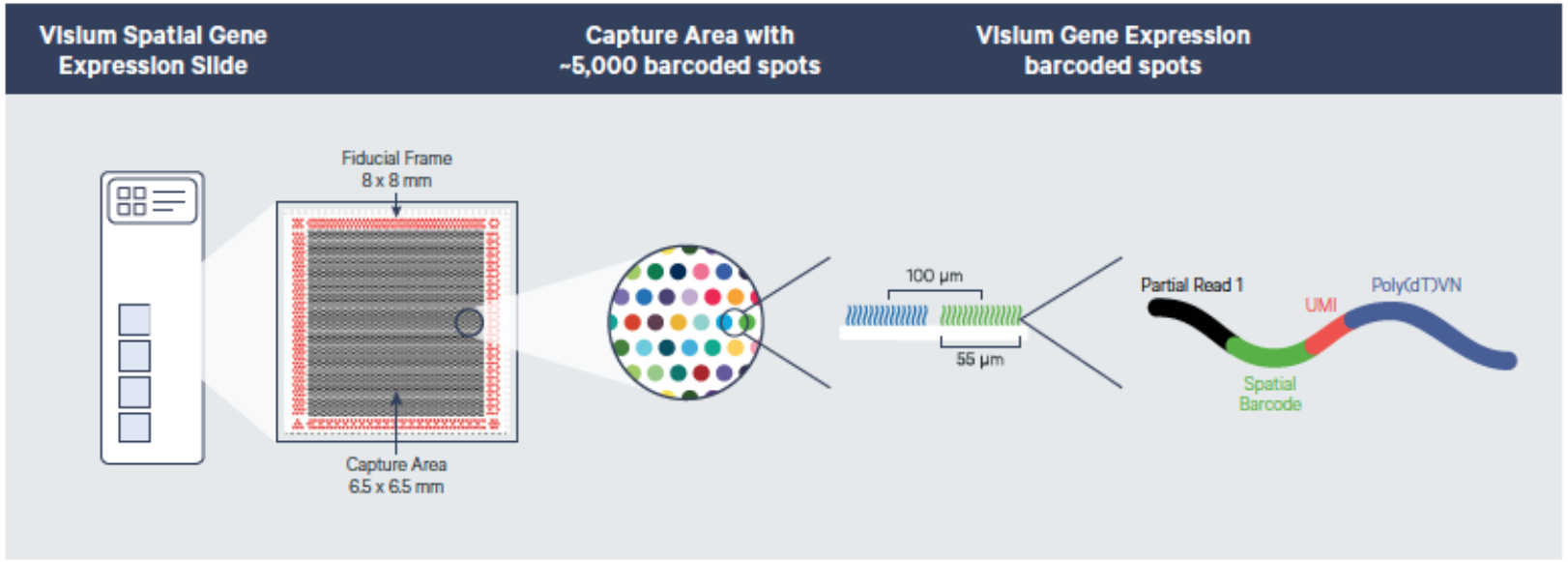
https://nanostring.com/products/geomx-digital-spatial-profiler/geomx-dsp-overview/

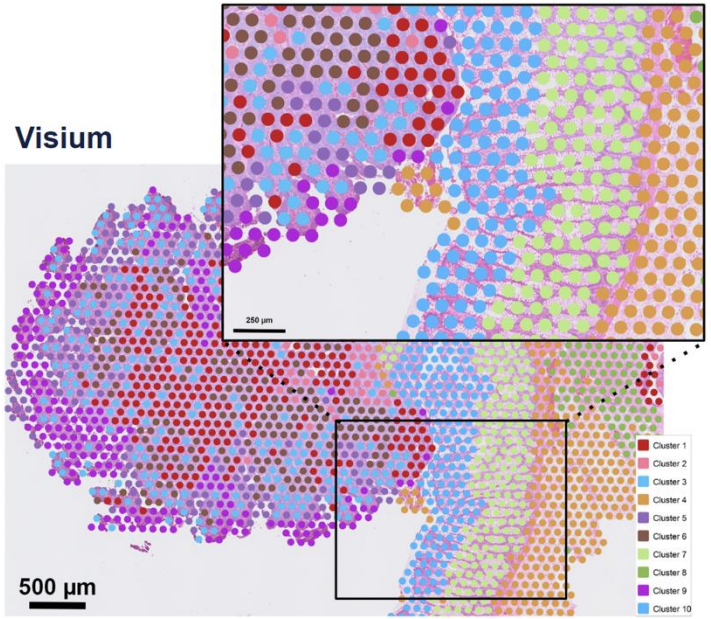
Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
10X Visium	Whole Transcriptome	100um	(6.5 mm × 6.5 mm)X4	Few hours

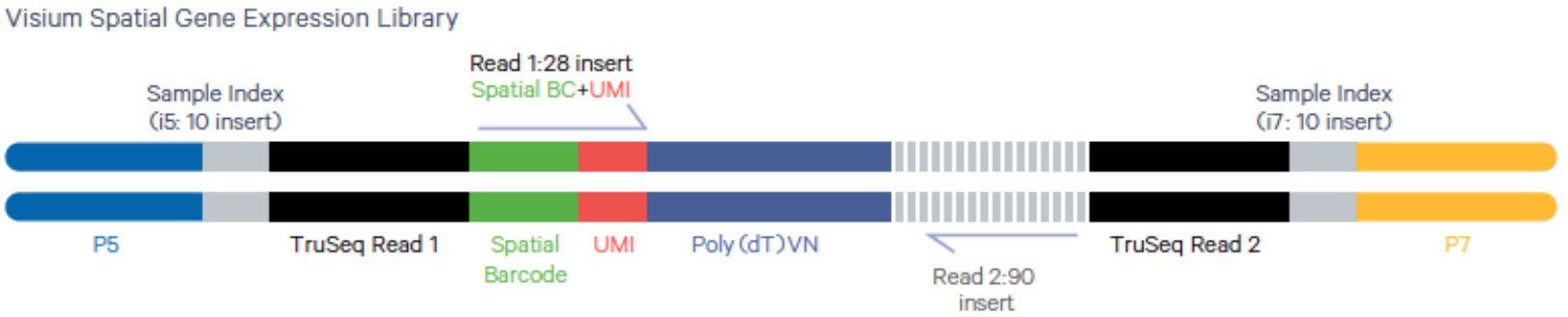
10X Visium

Launched end of 2019

Spot array-based spatial barcoding







4992 total spots per each of 4 capture areas
15k read pairs per spot

1-10 mammalian cells per spot depending on tissue type

Fresh Frozen Only

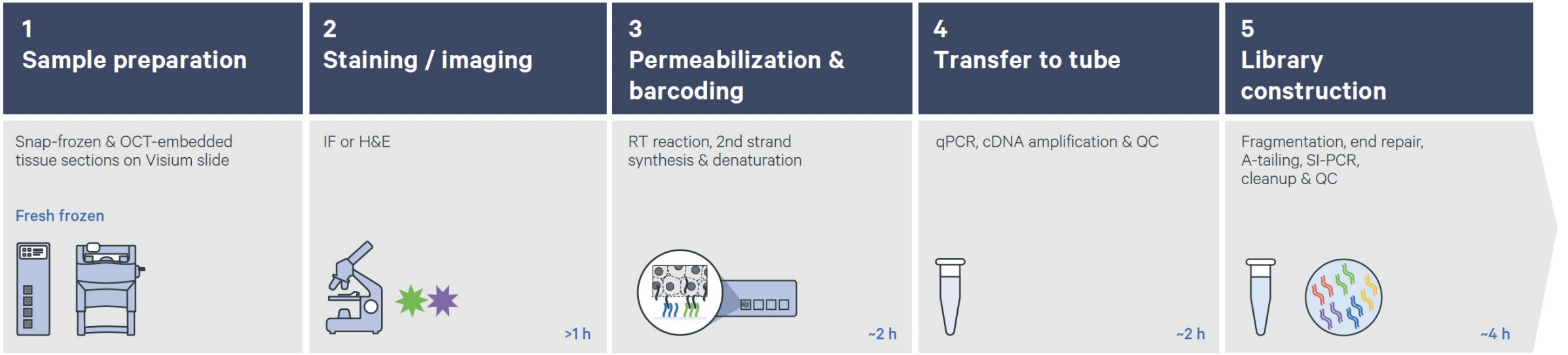
https://kb.10xgenomics.com/hc/en-us/categories/360002304851-Visium

Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
10X Visium	Whole Transcriptome	100um	(6.5 mm × 6.5 mm)X4	Few hours

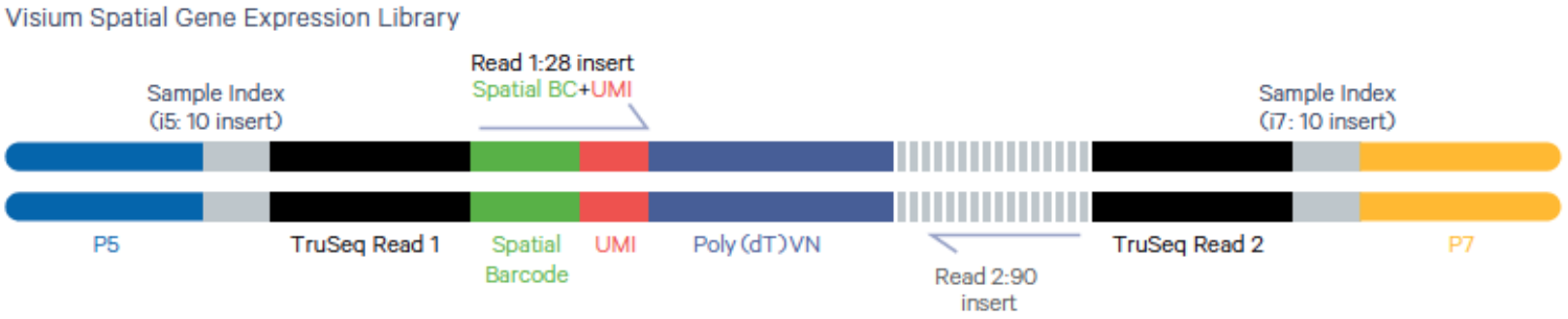
10X Visium

Spot array-based spatial barcoding

Fresh frozen



→ **Standard NGS**



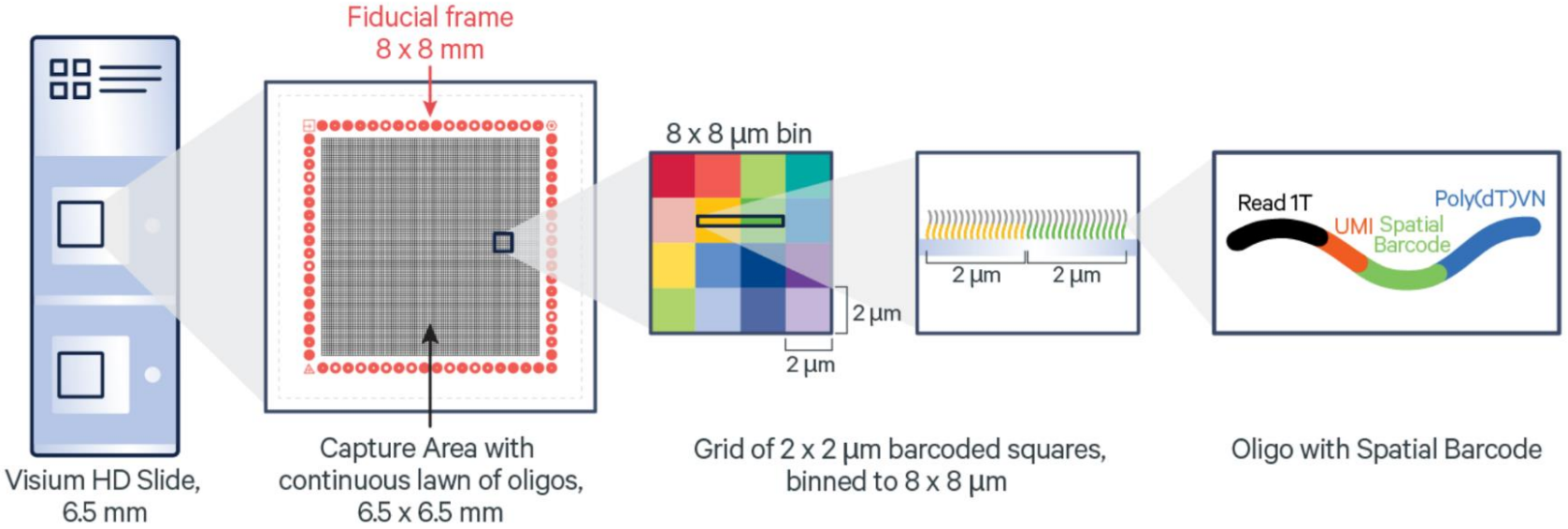
Wang, Y., Liu, B., Zhao, G., Lee, Y., Buzdin, A., Mu, X., Zhao, J., Chen, H., & Li, X. (2023). Spatial transcriptomics: Technologies, applications and experimental considerations. *Genomics*, *115*(5), 110671.
https://kb.10xgenomics.com/hc/en-us/articles/360035999152-What-are-the-imaging-system-requirements-for-running-Visium-for-fresh-frozen

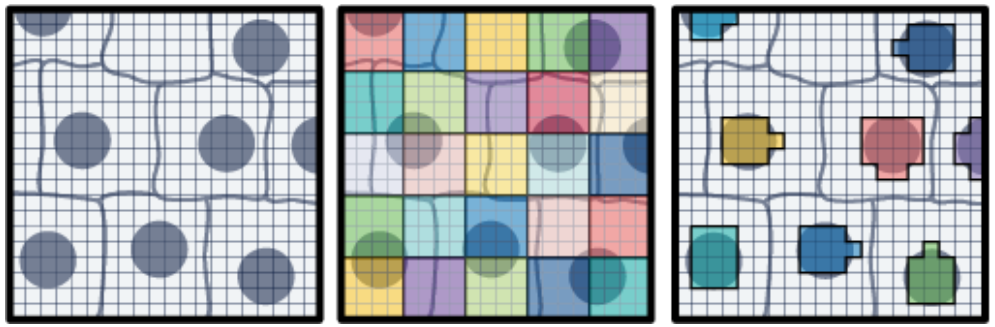
Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
10X Visium HD	Whole Transcriptome (Probe-Capture)	2-8um	(6.5 mm × 6.5 mm)x2	Few hours

10X Visium HD

Grid array-based spatial barcoding

Visium HD launched 2024





Two approaches for binning 2x2 µm barcode squares in Visium HD

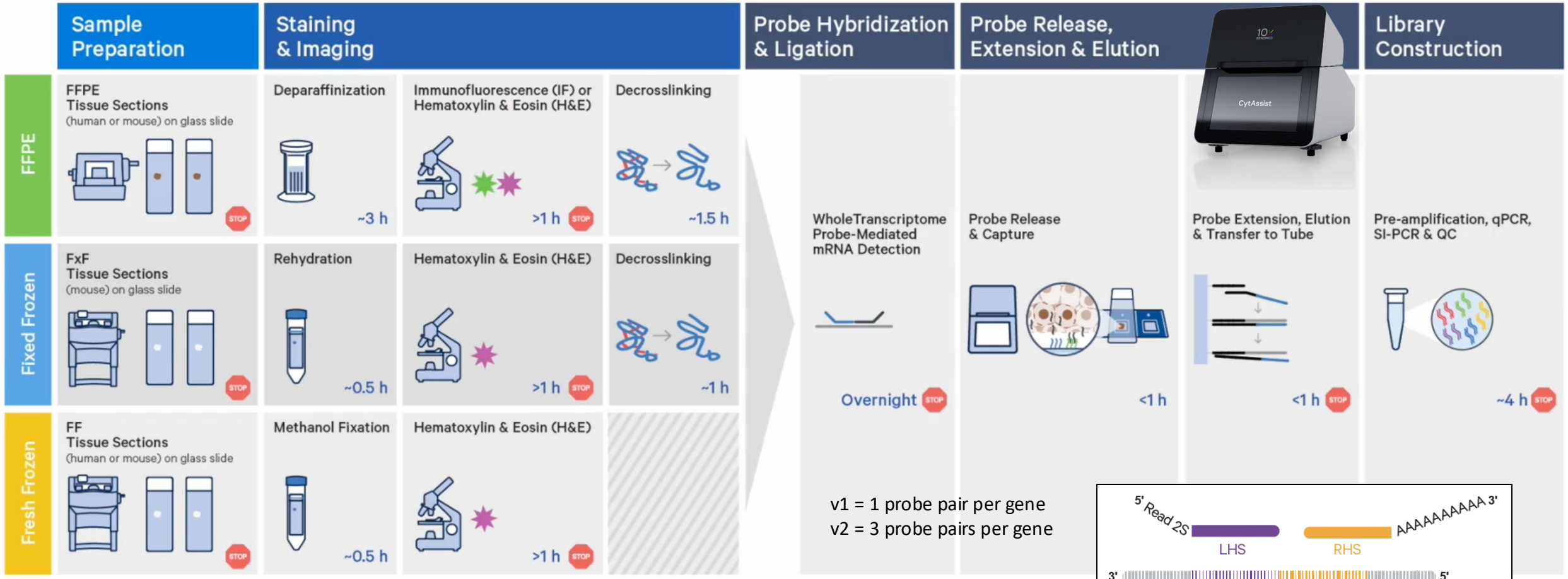
11.2 million 2 x 2 µm barcoded squares without gaps

FF, Fixed Frozen and FFPE

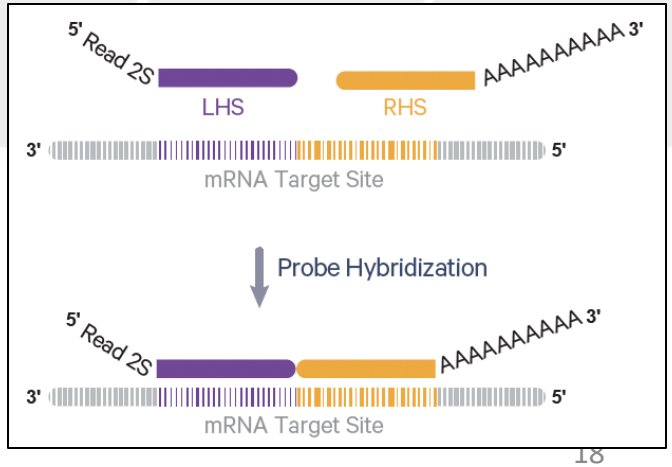
Probe-Capture with Visium Cyt Assist before NGS Sequencing

https://www.10xgenomics.com/blog/your-introduction-to-visium-hd-spatial-biology-in-high-definition
https://www.10xgenomics.com/analysis-guides/segmentation-visium-hd

Important: **Visium HD relies on probe sets to capture whole transcriptome, but output is still NGS**



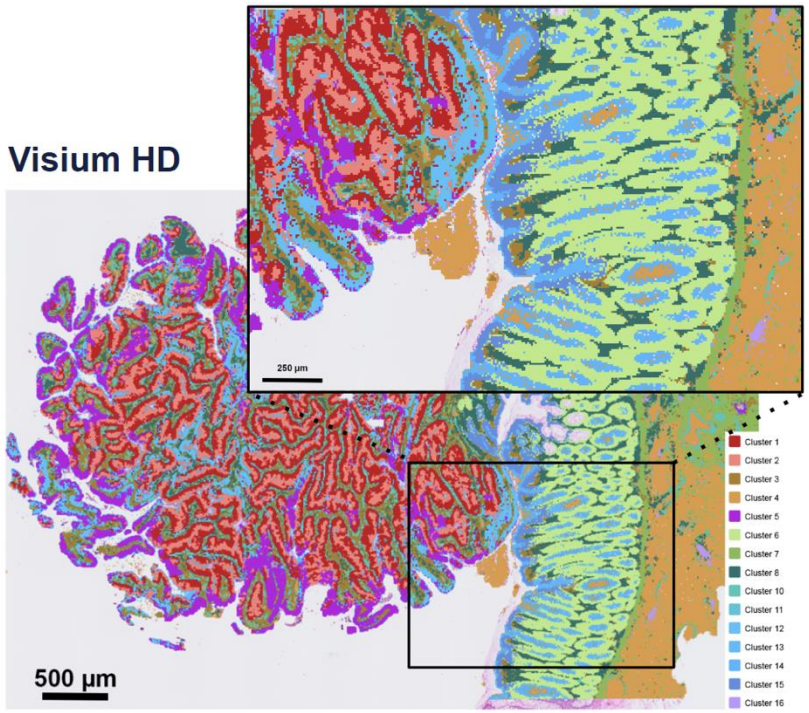
v1 = 1 probe pair per gene
v2 = 3 probe pairs per gene



Currently, whole transcriptome custom probe sets must be designed for species other than human and mouse for Visium HD

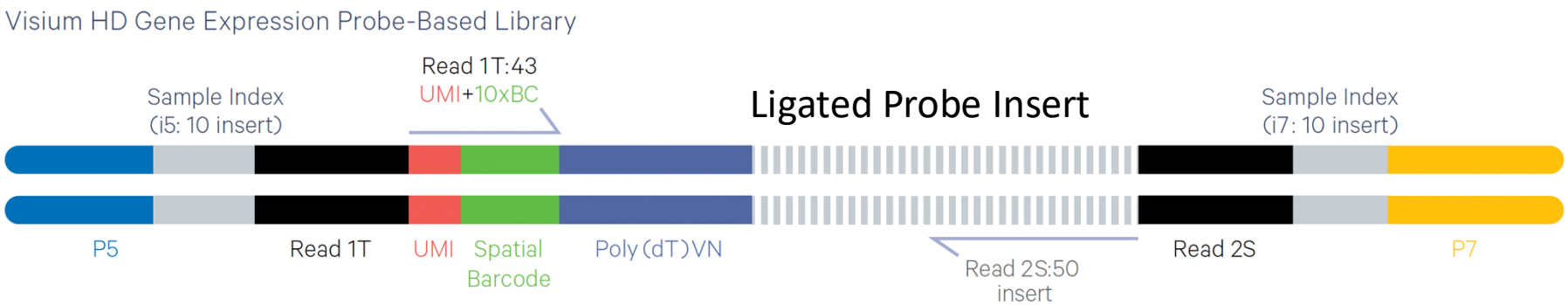
https://www.10xgenomics.com/blog/simplifying-spatial-transcriptomics-protocols-with-visium-cytassist

Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
10X Visium HD	Whole Transcriptome (Probe-Capture)	2-8um	(6.5 mm × 6.5 mm)x2	Few hours



Thickness of 3–10 µm sections (recommend 5um)

FF, Fixed Frozen and FFPE



Gene expression library is sequenced at a recommended min depth of 275 million read pairs for Capture Areas covered fully by tissue

https://www.10xgenomics.com/blog/your-introduction-to-visium-hd-spatial-biology-in-high-definition
https://www.10xgenomics.com/analysis-guides/segmentation-visium-hd

Platform	# of genes profiled	Spatial Resolution	RNA Capture Efficiency	Imaging Area	Time Required
MERFISH	500	100 nm	95% for cells 80–85% for tissue	10 mm × 10 mm	28–30 h

MERFISH 2.0 launched 2024

MERFISH

Multiplex error-robust FISH

MERFISH workflows involve four major steps:

Order will vary based on tissue preservation type (FFPE or FF)

Preparation

Tissue samples are mounted on slides and permeabilized.

Staining for protein co-detection can be added at this stage.

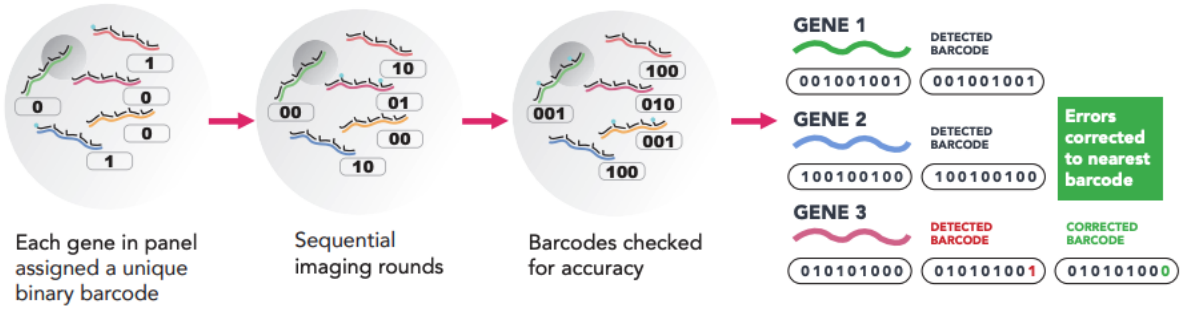
Hybridization

Embedding tens of thousands of unique encoding probes onto the sample.

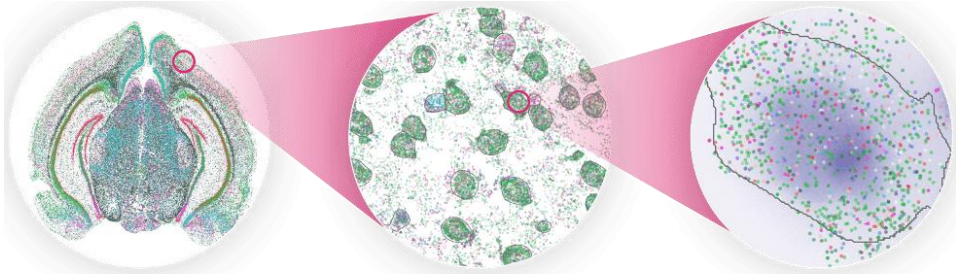
Clearing

Using a gel embedding and clearing process to remove unnecessary components while preserving transcripts and bound probes.

Imaging



(Simplified cartoon from product manual, method details covered in next slide)

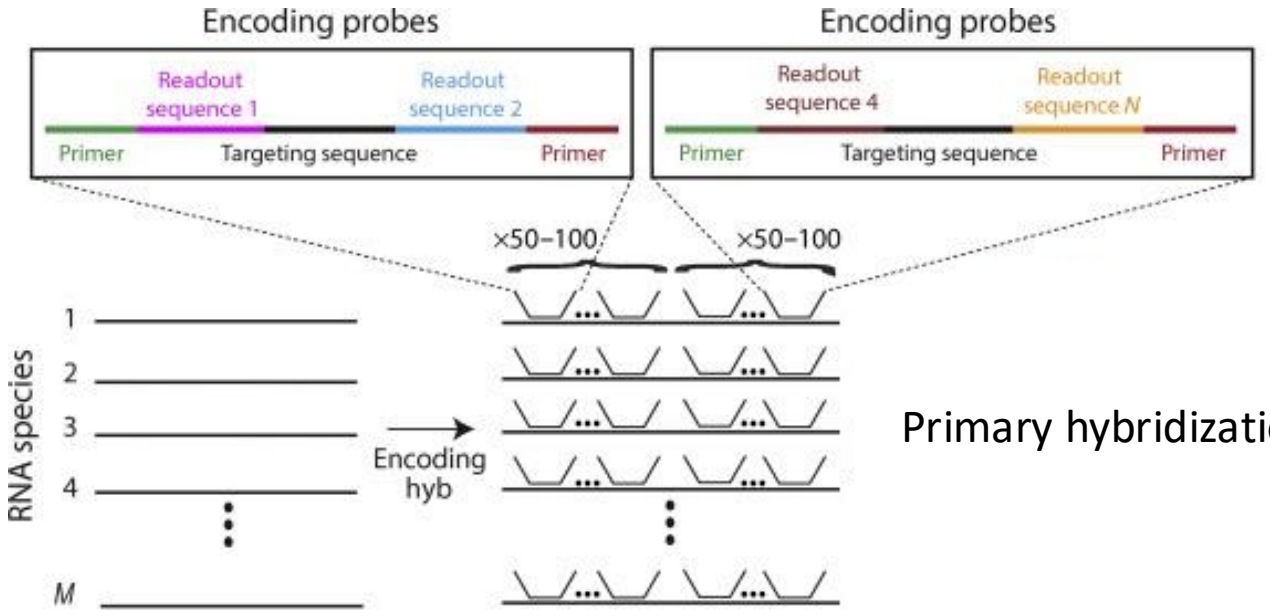


vizgen

FFPE, FF, and fixed frozen tissue

https://vizgen.com/products/

MERFISH Probe Hybridization and Fluorescence Readout



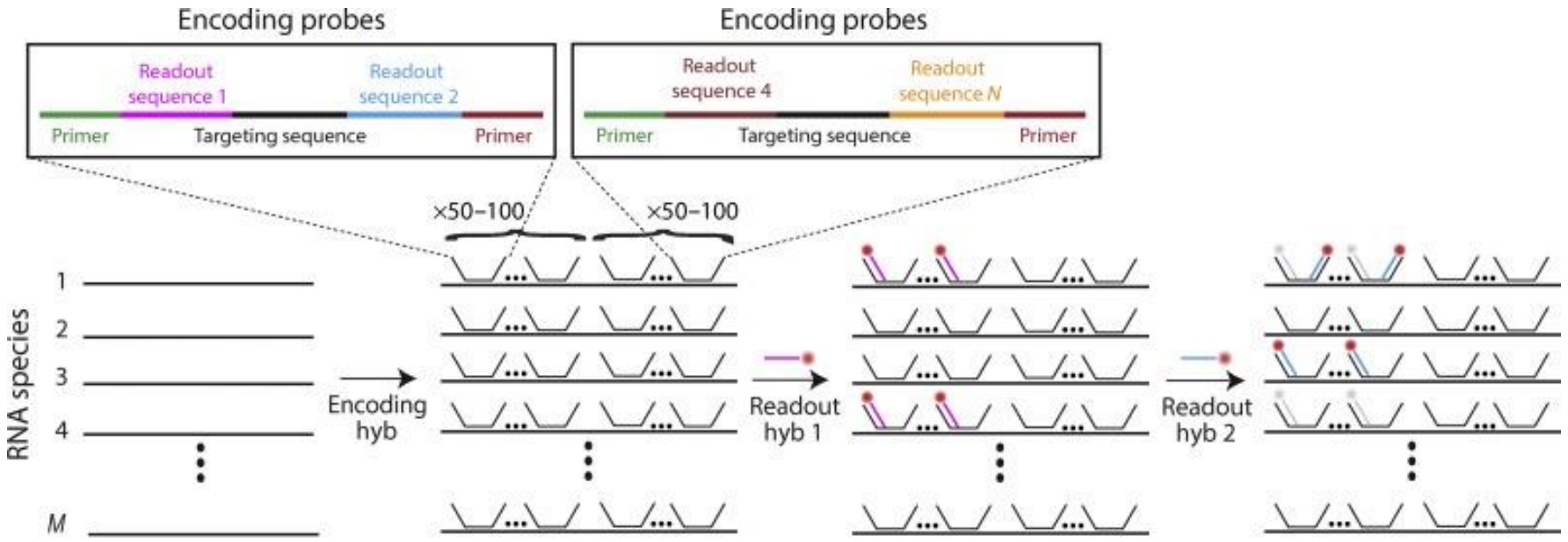
Each encoding probe contains a targeting sequence which directs their binding to specific RNA, as well as **two readout sequences**

To increase the signal from each copy of the RNA, 50-100 encoding probes, each with a different target region, are bound to the same RNA.

Primary hybridization of **encoding probes**

Moffitt, J. R., & Zhuang, X. (2016). RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). *Methods in Enzymology*, *572*, 1–49.

MERFISH Probe Hybridization and Fluorescence Readout



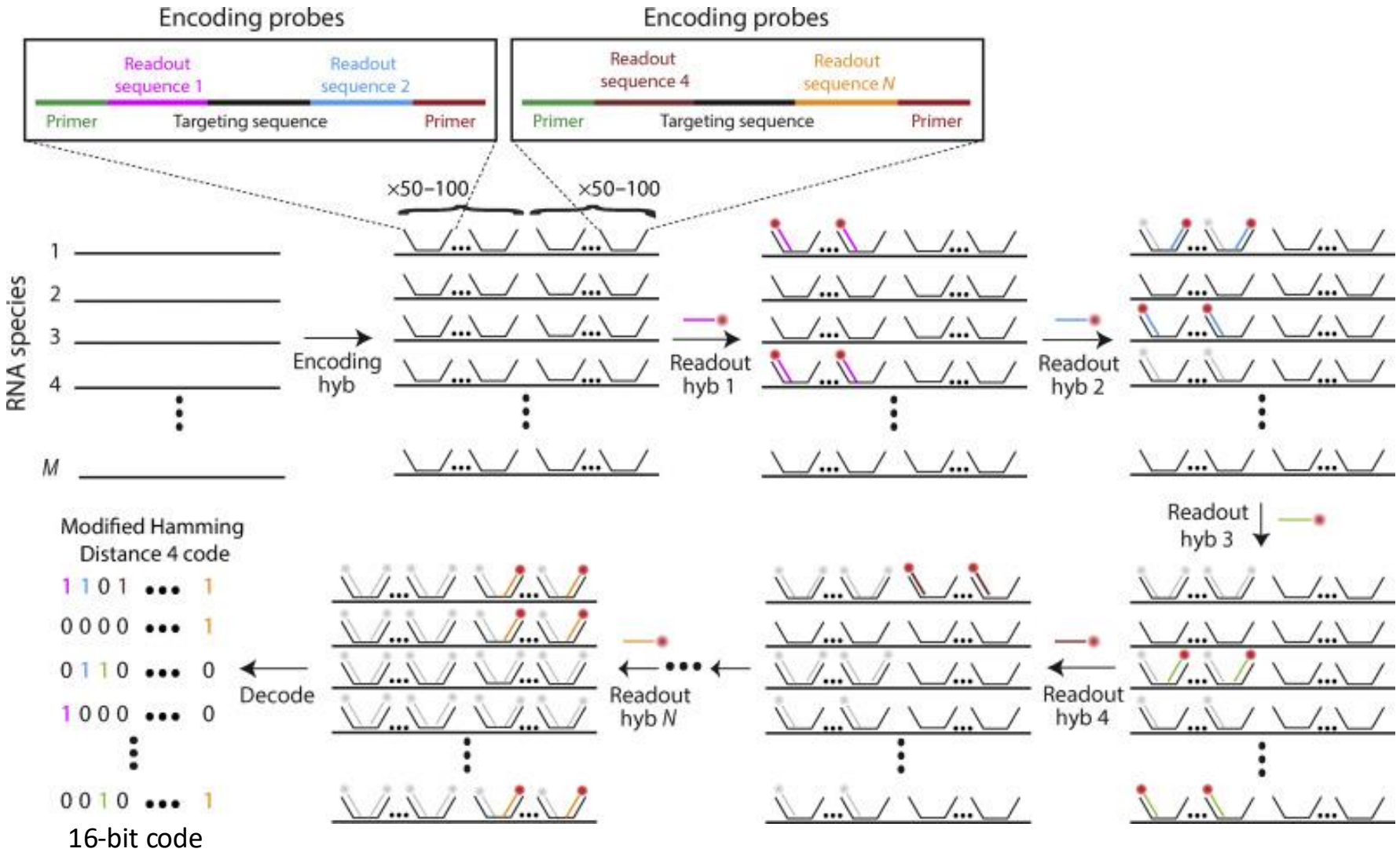
16 rounds of hybridization of **secondary** **fluorescent** **probes**

To identify the readout sequences contained on the encoding probes bound to each RNA, 16 rounds of hybridization and imaging are performed.

Each round uses a unique, **fluorescently** **labeled** **probe** whose sequence is complementary to the readout sequence for that round.

Moffitt, J. R., & Zhuang, X. (2016). RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). *Methods in Enzymology*, *572*, 1–49.

MERFISH Probe Hybridization and Fluorescence Readout



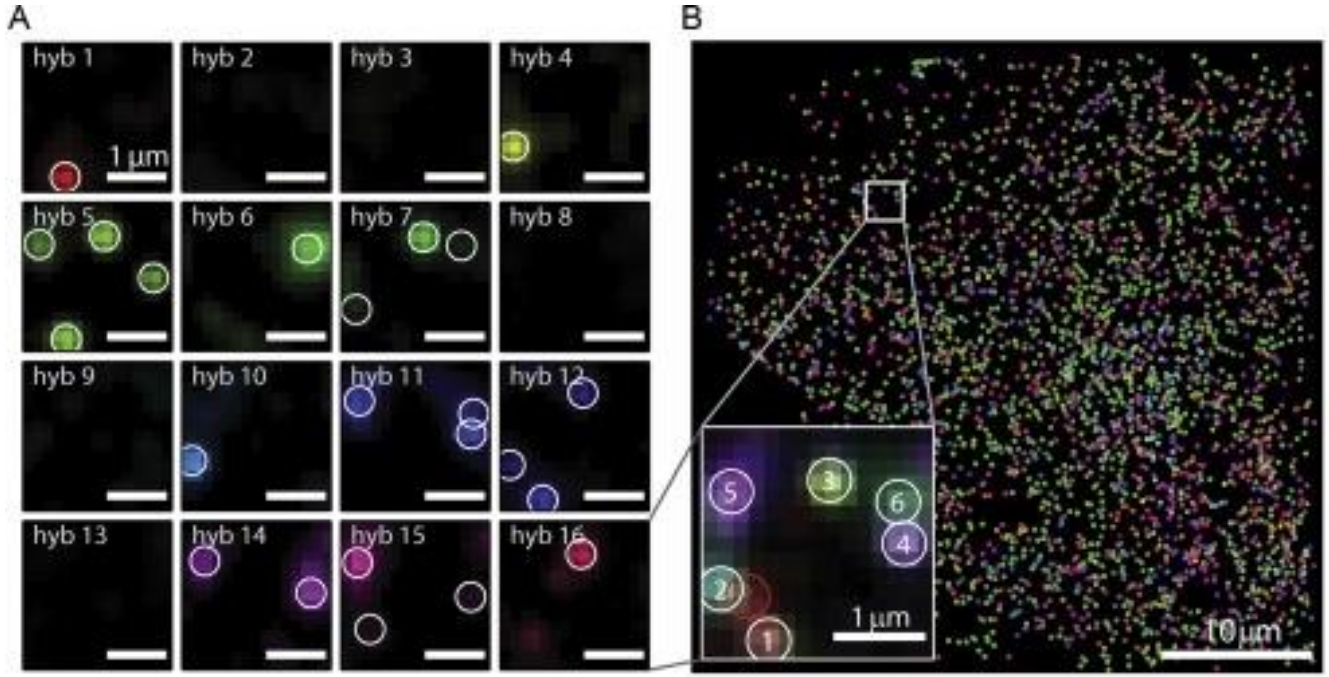
Codebook Design Schema

To identify the readout sequences contained on the encoding probes bound to each RNA, 16 rounds of hybridization and imaging are performed.

Each round uses a unique, **fluorescently labeled probe** whose sequence is complementary to the readout sequence for that round.

Moffitt, J. R., & Zhuang, X. (2016). RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). *Methods in Enzymology*, *572*, 1–49.

MERFISH error robustness



Modified Hamming Distance 4 code

1 1 0 1 ••• 1
0 0 0 0 ••• 1
0 1 1 0 ••• 0
1 0 0 0 ••• 0
⋮
0 0 1 0 ••• 1

C

Hybridization round

Spot number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	1	0	0	0	1	0	0	0	0	0	0	1	0	0/1	0	0	DYNC1H1
2	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	EGFR
3	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	FLNA
4	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	TLN1
5	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	TLN1
6	0	0	0	0	0	1	1	0	0	0	1	0	0	0/1	0	0	LRP1
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	Unidentified

Probes designed with Hamming Distance of 4 to make robust to errors. Errors can then also be corrected to “call” the correct probe

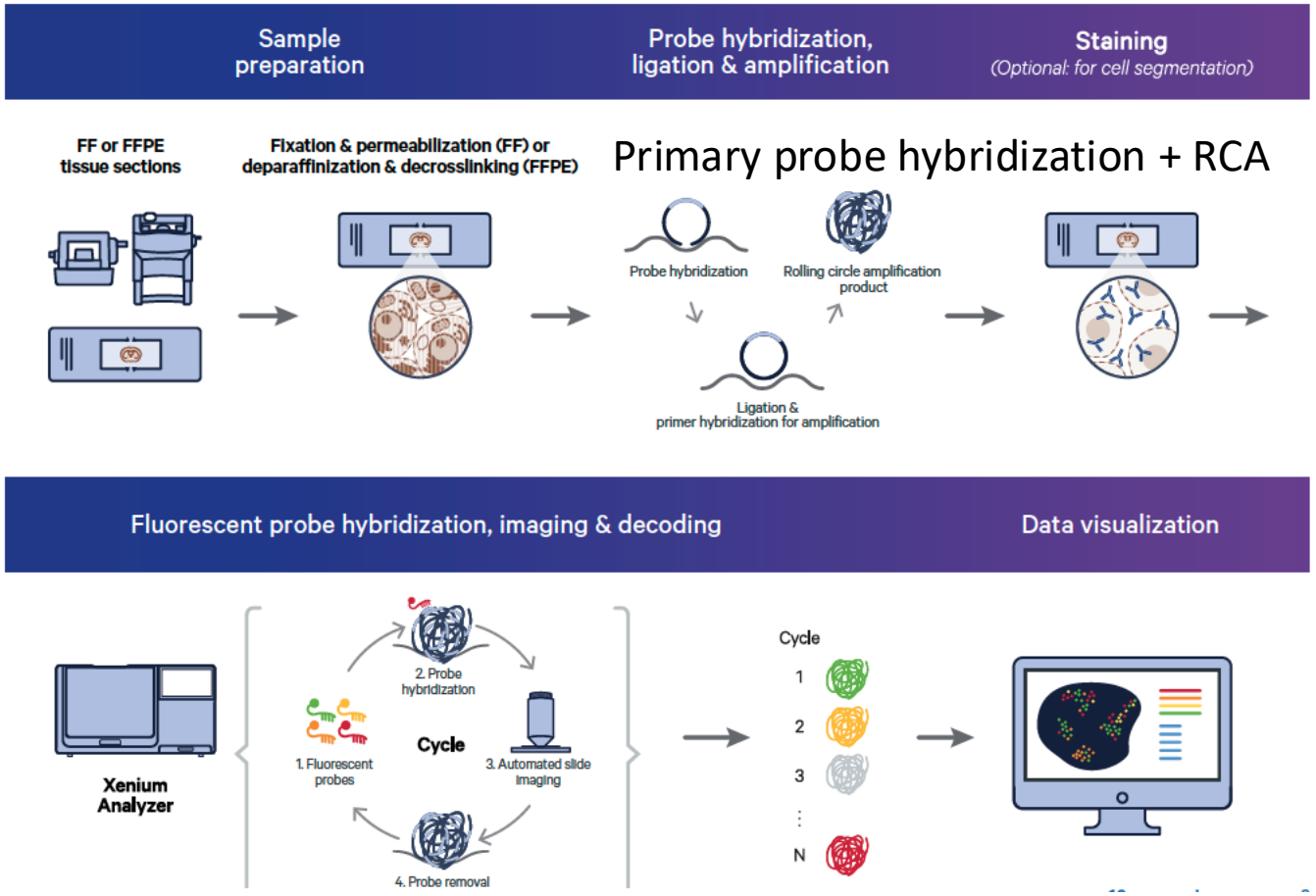
Constant Hamming weight is used to ensure probes have same number of 1’s and 0’s (since rate of reading 1 as 0 is > than reading 0 as 1)

16-bit MHD4 codebook allows for error correction

Moffitt, J. R., & Zhuang, X. (2016). RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). *Methods in Enzymology*, *572*, 1–49.

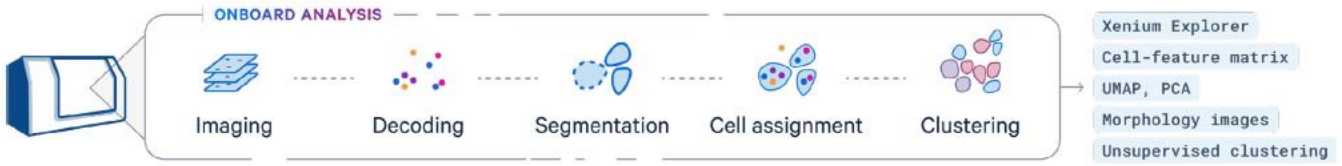
Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
10X Xenium	400	50 nm	12 mm × 24 mm	2 days

Xenium launched 2022



Primary probe hybridization + RCA

Rounds of hybridization of secondary fluorescent probes



FFPE: 5um thickness
FF: 10um thickness

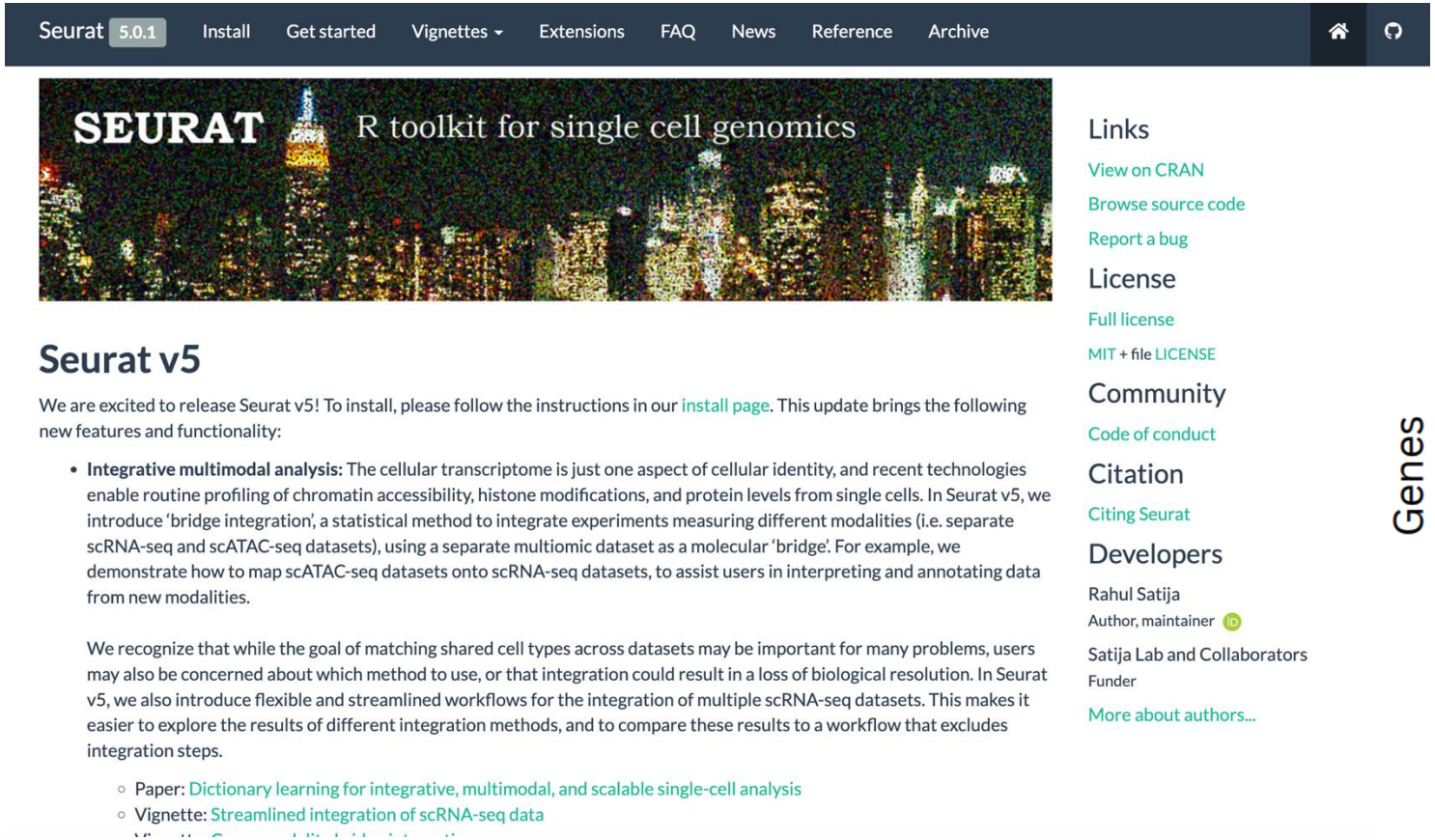
5 min break

Resume for next part on STx data analysis

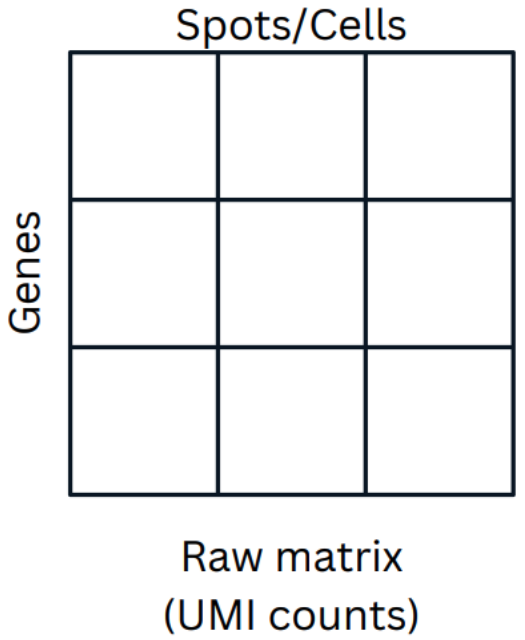
STx Data Analysis

Seurat 5.0 package

Developers: Rahul Satija lab at NYU



Seurat object

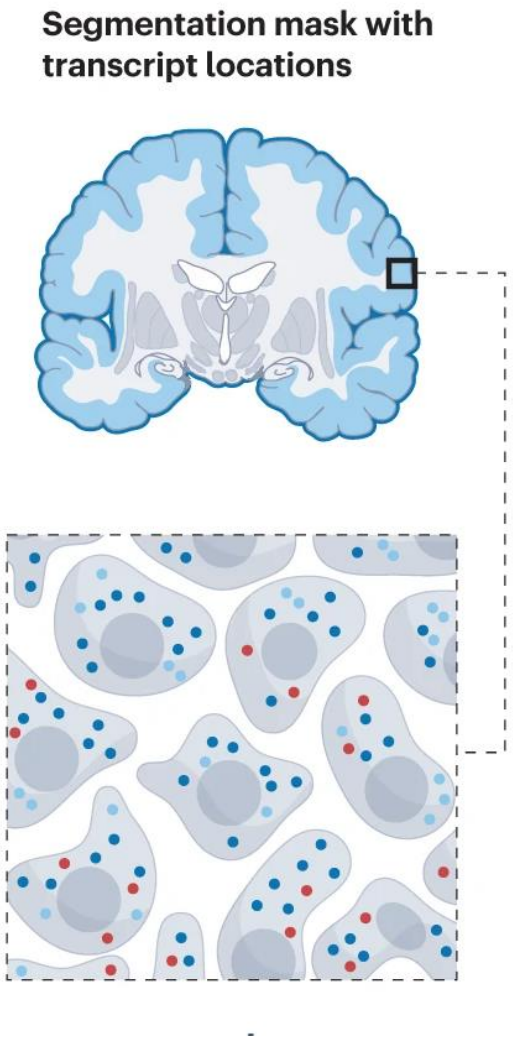


+ associated info:
Metadata such as images,
Normalized counts,
Dimension reduction
Embeddings,
Spatial coordinates
etc.

We will focus on using the Seurat package next session!

Analysis of spatial datasets (Sequencing-based)
Analysis of spatial datasets (Imaging-based)

Image-based spatial transcriptomics: cell segmentation



Coordinates and counts of transcript

	x	y	Count	Cell
gene1	–10	3	15	1
gene2	–15	4	2	1
gene1	2	3	5	2
gene2	4	2	10	2
gene1	–12	10	3	3
gene2	–14	8	1	3
...	...	...	...	...

Cell count matrix and cell coordinates

n_{raw} cells	m_{raw} genes →			x	y
	15	2	...	–12	3
	5	10	...	3	2
	3	1	...	–13	9
	...	...	...	...	...

Cell Segmentation is handled by MERSCOPE and Xenium software, but other methods can also be run post-hoc

Heumos, L., Schaar, A.C., Lance, C. et al. Nat Rev Genet (2023). https://doi.org/10.1038/s41576-023-00586-w

Nuclei: DAPI
Boundary/Membrane: anti-ATPase Ab
Interior RNA: anti-18S rRNA Ab

Multimodal Cell Segmentation

Example from Xenium

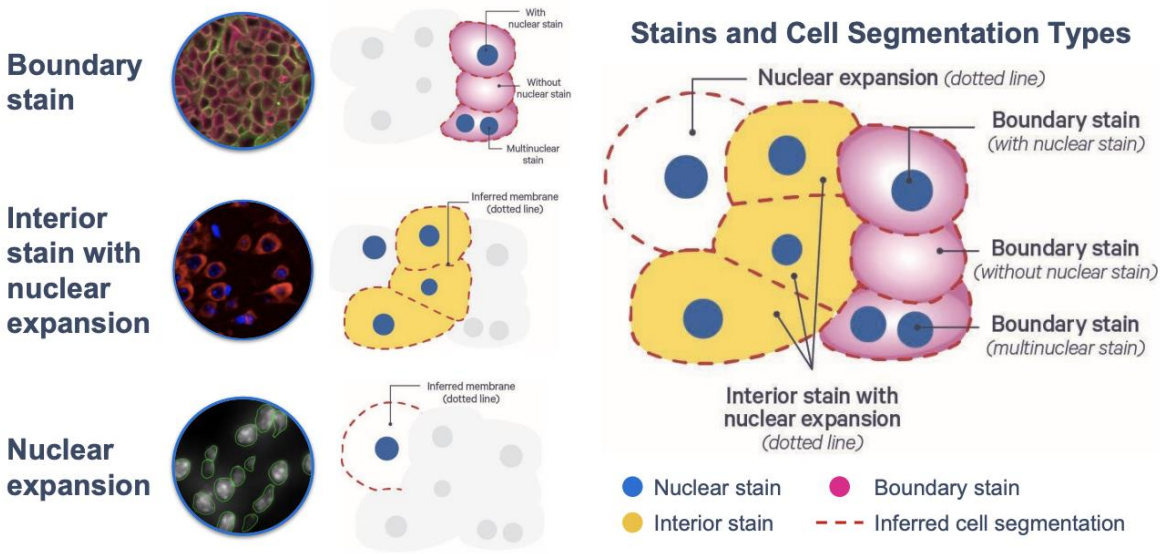
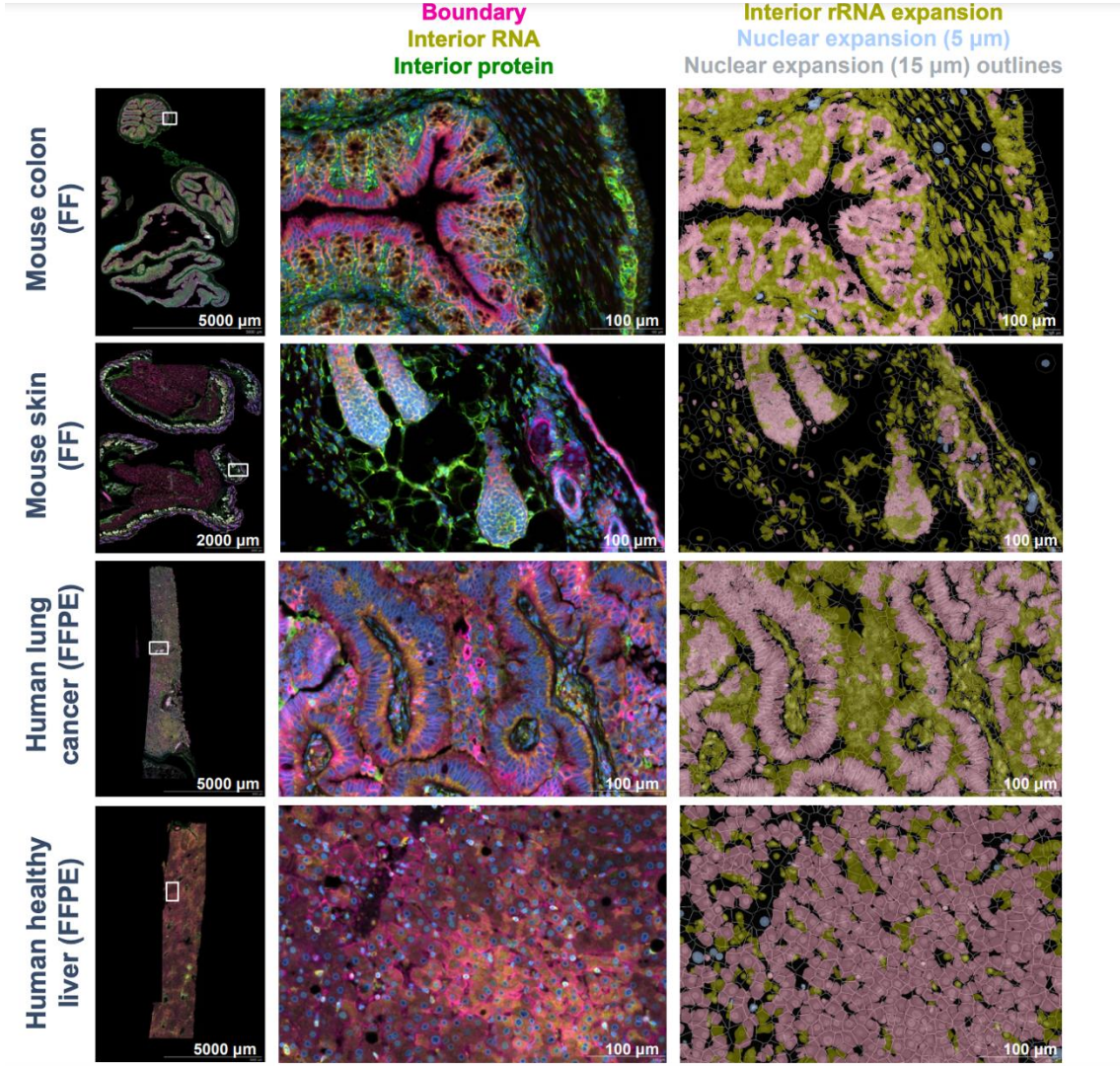


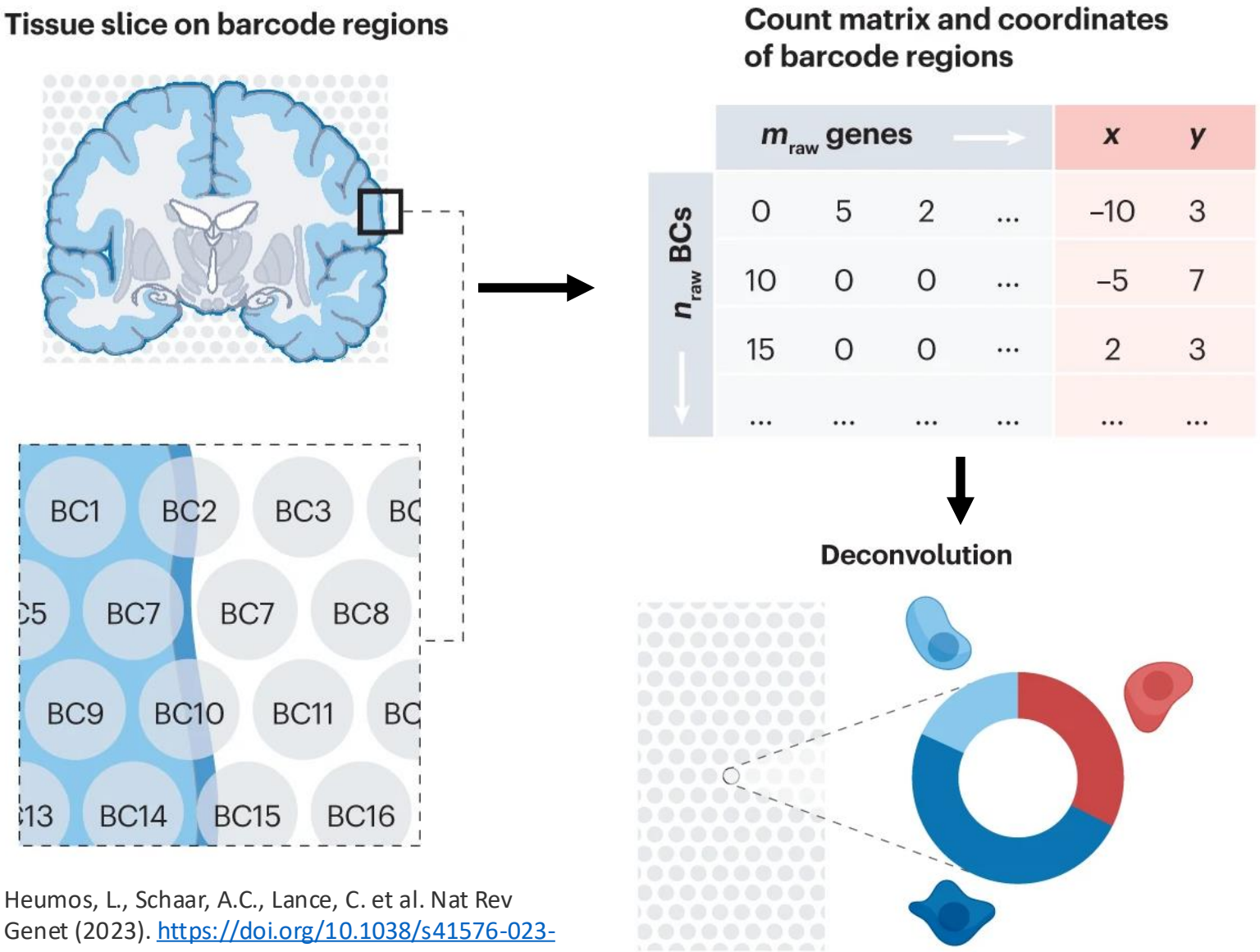
Figure 1. multimodal cell segmentation approach. After nucleus segmentation with DAPI, the algorithm segments each cell with one of three methods applied in a stepwise fashion: boundary segmentation, expansion from the nucleus to the cell interior stain edge, and nuclear expansion. Unlike conventional single-step methods, this multimodal approach effectively addresses certain cell types that cannot provide complete or even partial boundary information due to lack of clear membrane markers.



Since Xenium's boundary segmentation model does not require the presence of a nucleus, they were able to correctly segment multinucleate and occasionally anucleate cells.

Source: https://www.10xgenomics.com/library/ac7cd9 (AGBT 2024 Cell Segmentation Poster from 10X Genomics)

Array-based spatial transcriptomics: deconvolution



Heumos, L., Schaar, A.C., Lance, C. et al. Nat Rev Genet (2023). https://doi.org/10.1038/s41576-023-00586-w

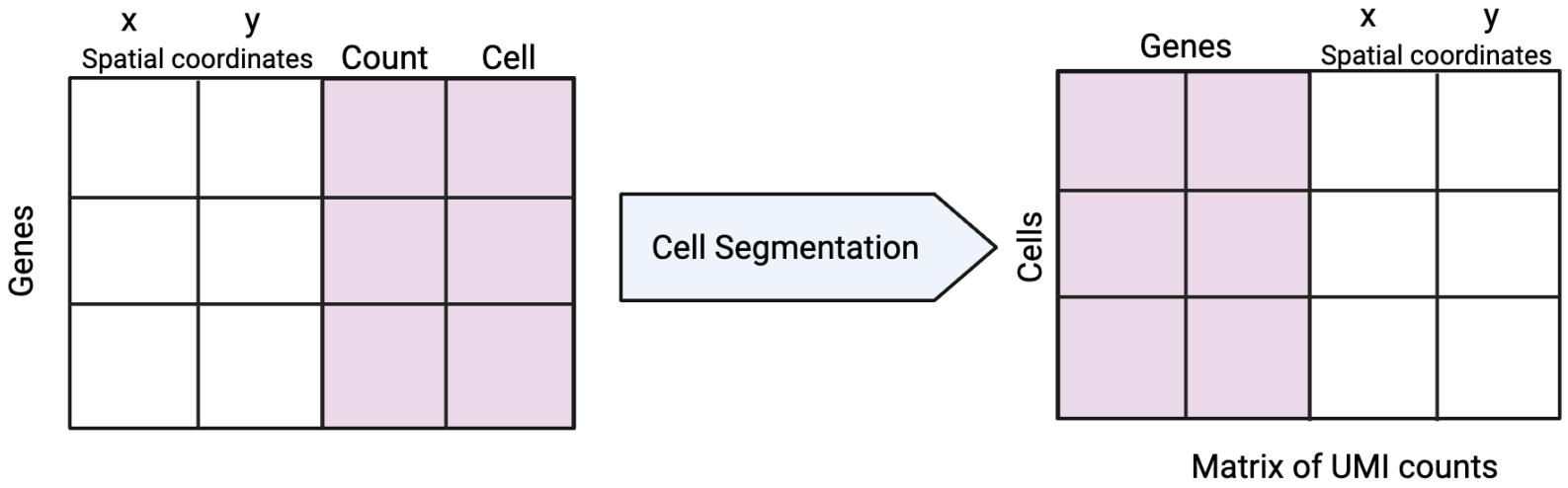
Top Deconvolution Methods

- Cell2location
- SpatialDWLS (Giotto)
- RCTD (supported in Seurat)

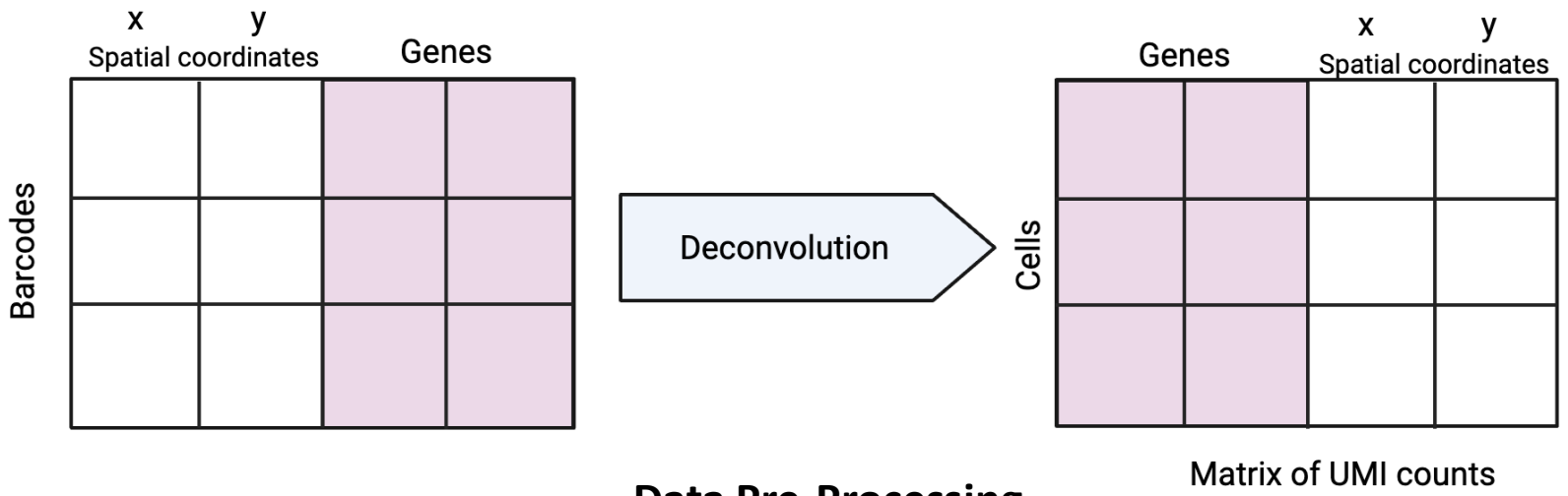
Reference: Li et al. Nat Methods 2022 Benchmarking spatial methods for cell type deconvolution https://www.nature.com/articles/s41592-022-01480-9

Summary: Assigning Genes to Cells and their spatial coordinates

Imaging-based STx needs cell segmentation



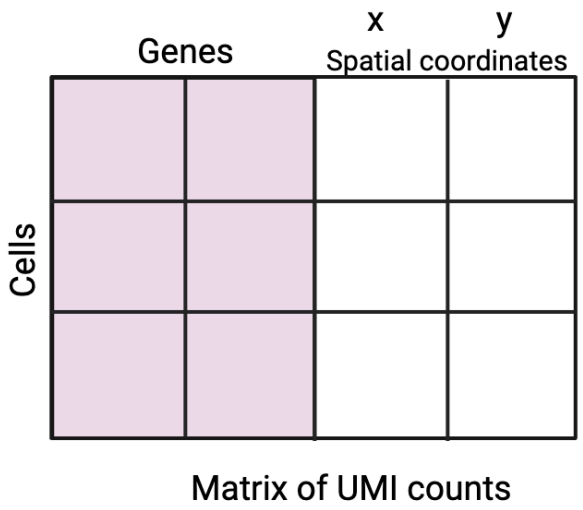
Sequencing Array-based STx needs deconvolution



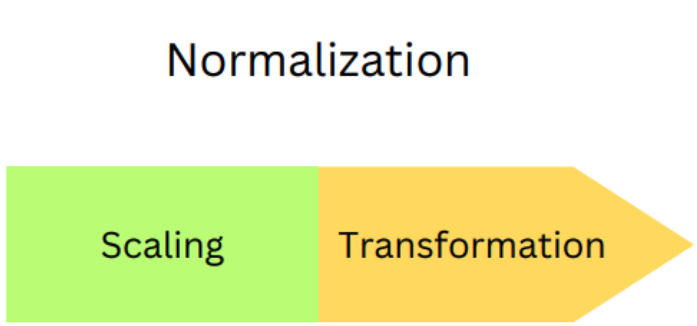
Output:
Gene-Cell Matrix
Location Matrix

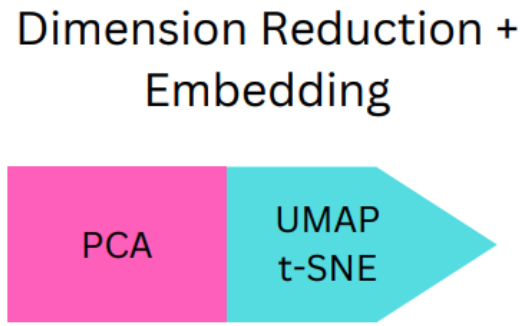
Data Pre-Processing

Count Data Preprocessing



Gene-Cell Matrix
Location Matrix



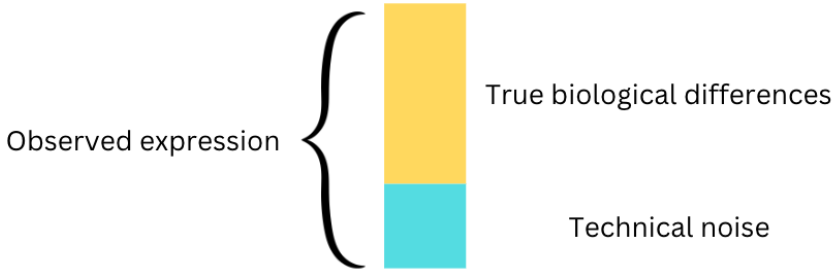


Count Normalization

Rationale: **Gene expression data is often highly overdispersed (greater variance than expected)**
Normalization makes samples more directly comparable and reduces high variance seen in highly expressed genes

Many statistical tests and models require homoskedasticity (constant variance)
Variance stabilization reduces the distortion on plots caused by highly variable genes

Without normalization, the analysis would be dominated by highly expressed genes



1) **Log transformation**
2) **Square root transformation**

$$y_{i,j} = f(x_{i,j})$$

Generalized linear model:

3) Pearson residual transformation (used by Seurat's sctransform)

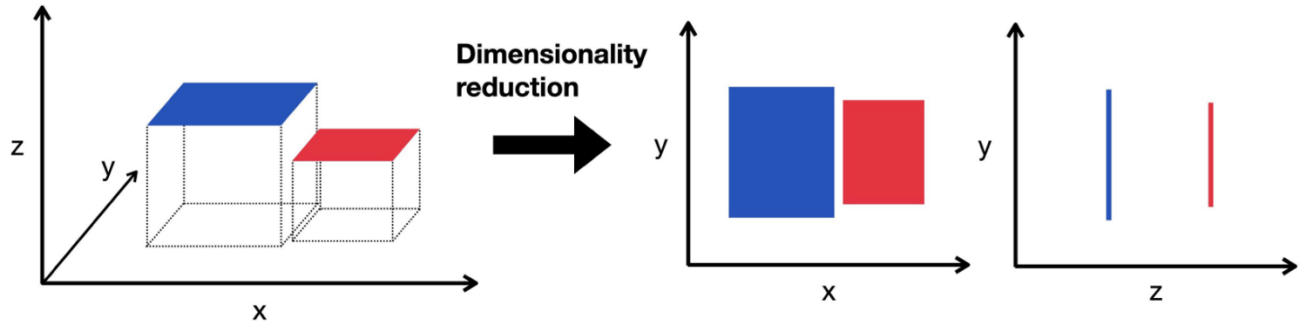
$y_{i,j} = w_j * x_{i,j}$ where $w_i = 1/\sqrt{\mu_i}$

Instead of transforming each measurement individually, Pearson residuals apply a weight w_i to all measurements of a given gene based on observed mean μ_i

Log transformation, square root transform vs pearson transform (scTransform) https://www.youtube.com/watch?v=huxkc2GH4lk&ab_channel=FlorianWagner
Rafa Irizarry's 2024 lecture on Statistical Methods for Single-Cell RNA-Seq Analysis and Spatial Transcriptomics https://bioinformatics.ccr.cancer.gov/btep/classes/rafael-irizarry
scTransform v2 paper https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02584-9

Dimension Reduction

Rationale: **Reducing the dimensions on high dimension data will speed up computation for downstream analysis such as clustering and avoid overfitting** (reduces dataset noise from random variation)
See: Curse of dimensionality



Data reduction is performed by generating a smaller set of predictors that capture a majority of information in the original variables (select most highly variable genes). This has the effect of reducing the correlation of different predictors to one another.

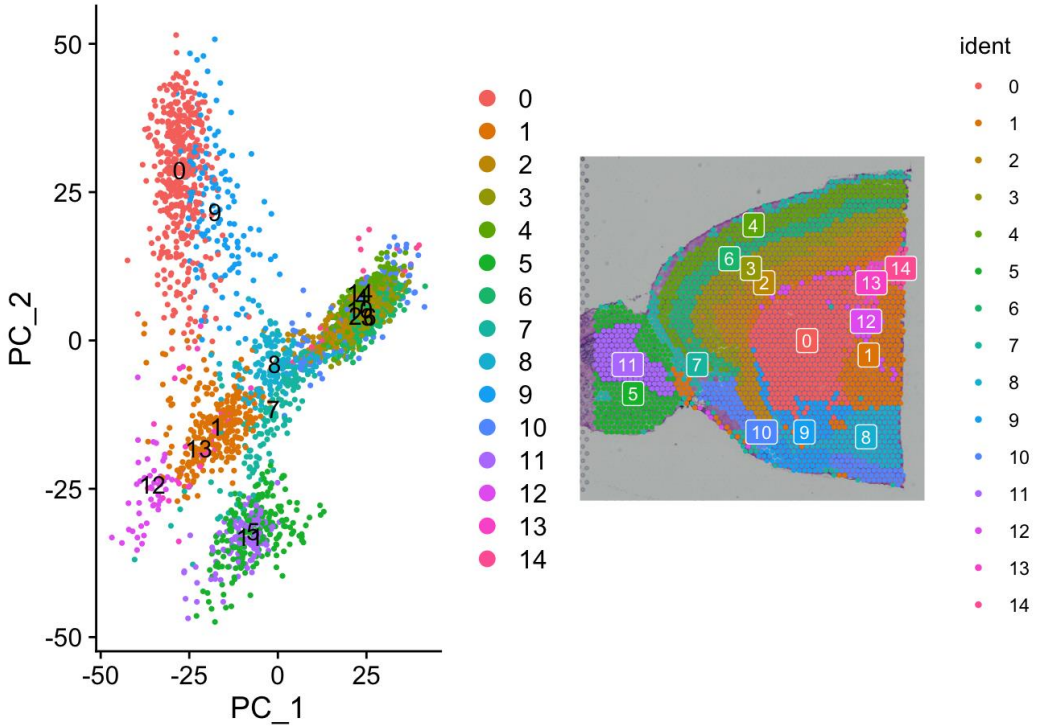
Normalization is required beforehand so that larger scaled variables don't dominate the analysis
For most data reduction techniques, the new predictors are functions of the original predictors.
This class of methods is often called **signal extraction** or **feature extraction** techniques.

Heumos, L., Schaar, A.C., Lance, C. et al. Best practices for single-cell analysis across modalities. Nat Rev Genet (2023). https://doi.org/10.1038/s41576-023-00586-w

Adapted from Kuhn and Johnson. Applied Predictive Modeling 2013

Principal Components Analysis + Clustering

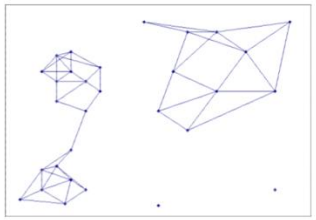
PCA is a commonly used dimension/data reduction technique and seeks to find linear combinations of the predictors, known as principal components (PCs), which capture the most possible variance.
The first PC is defined as the linear combination of the predictors that captures the most variability of all possible linear combinations. Then, subsequent PCs are derived such that these linear combinations capture the most remaining variability while also being uncorrelated with all previous PCs. Mathematically, the jth PC can be written as:



$$\mathrm{PC}_j = (a_{j1} \times \text{Predictor 1}) + (a_{j2} \times \text{Predictor 2}) + \cdots + (a_{jP} \times \text{Predictor } P).$$

P is the number of predictors. The coefficients $a_{j1}, a_{j2}, \ldots, a_{jP}$ are called component weights and help us understand which predictors are most important to each PC.

Because the distance between points from performing PCA are interpretable, the results can be directly used for clustering



Seurat's FindNeighbors computes nearest neighbors graph for a given dataset (k-nearest neighbors, k-NN and shared nearest neighbors, SNN) using PCA space
Seurat's FindClusters performs cluster detection using Louvain (default) or other methods such as Leiden

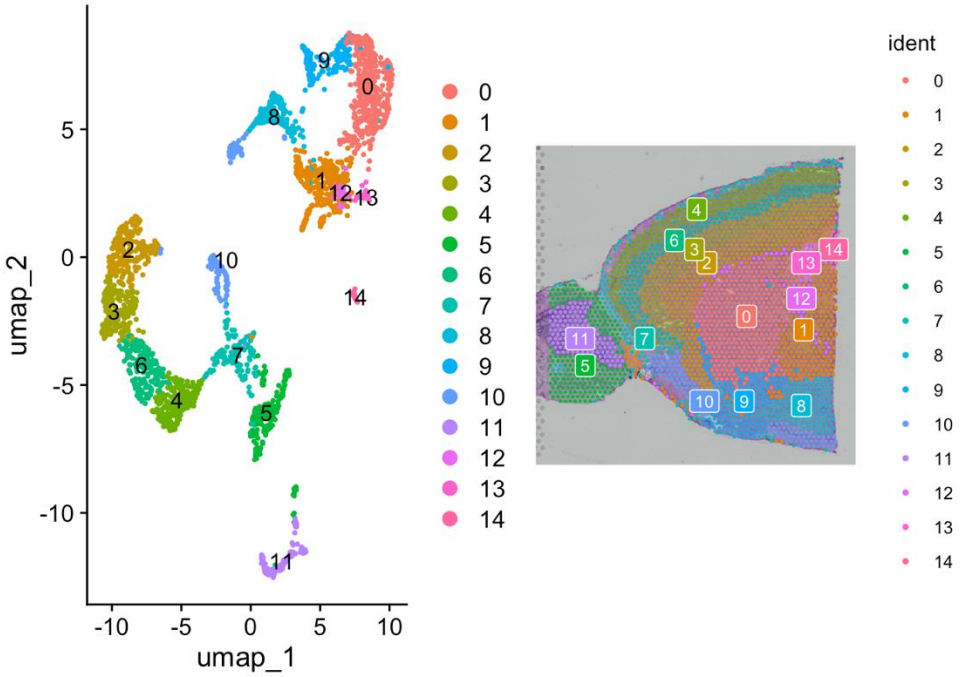
http://mlwiki.org/index.php/SNN_Clustering

Adapted from Kuhn and Johnson. Applied Predictive Modeling 2013

Embedding and Visualization

Following PCA, we can perform embedding and visualization of the clusters

Embedding is the practice of mapping high dimensional data into a lower dimensional space, while preserving the structure and relationships.

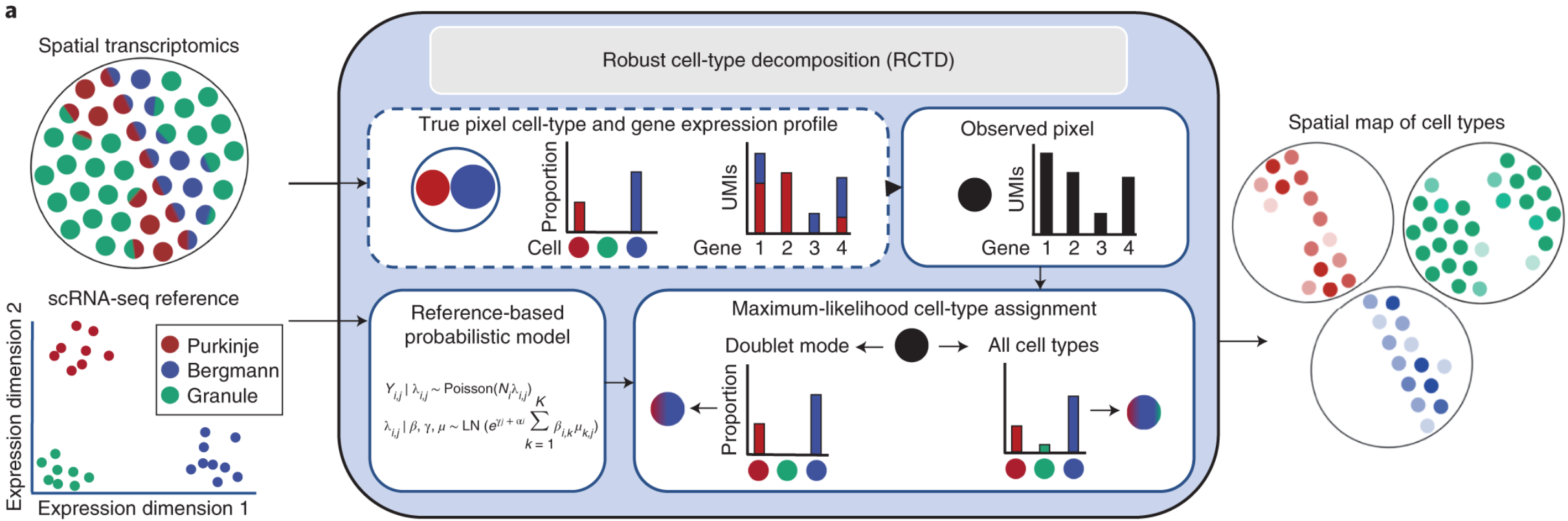


While **Principal Component Analysis** (PCA) assumes the data is linear **Uniform Manifold Approximation and Projection (UMAP)** and **t-Stochastic Neighbor Embedding (t-SNE)** do not assume linearity

The distance between points from performing PCA are interpretable, so they can be used for clustering, whereas those in UMAP and t-SNE embedding are not

More info:
Statquest PCA https://www.youtube.com/watch?v=FgakZw6K1QQ
Statquest UMAP https://www.youtube.com/watch?v=eN0wFzBA4Sc
Statquest t-SNE https://www.youtube.com/watch?v=NEaUSP4YerM
t-SNE subtleties https://distill.pub/2016/misread-tsne/

Annotation: Integration with scRNA-seq data



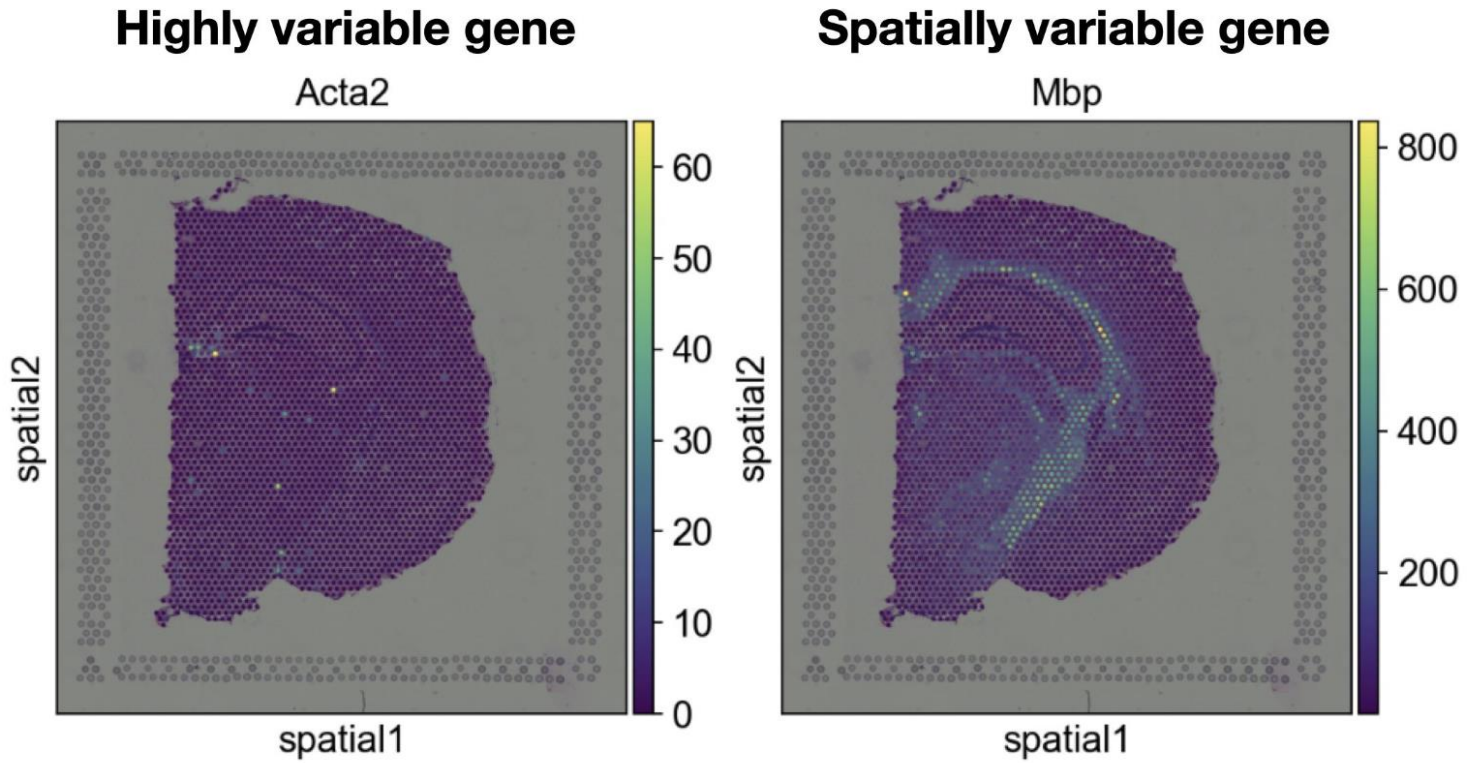
Cable, DM et al. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotech*, *40*(4), 517–526.

Irizarry Lab

Li, B et al. (2022). Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods*, *19*(6), 662–670.
Li, Y et al. (2021). Benchmarking computational integration methods for spatial transcriptomics data. *bioRxiv* https://doi.org/10.1101/2021.08.27.457741

Spatially Variable Genes

In contrast to highly variable genes (genes that differ significantly between cells), spatially variable genes show a distinct spatial pattern

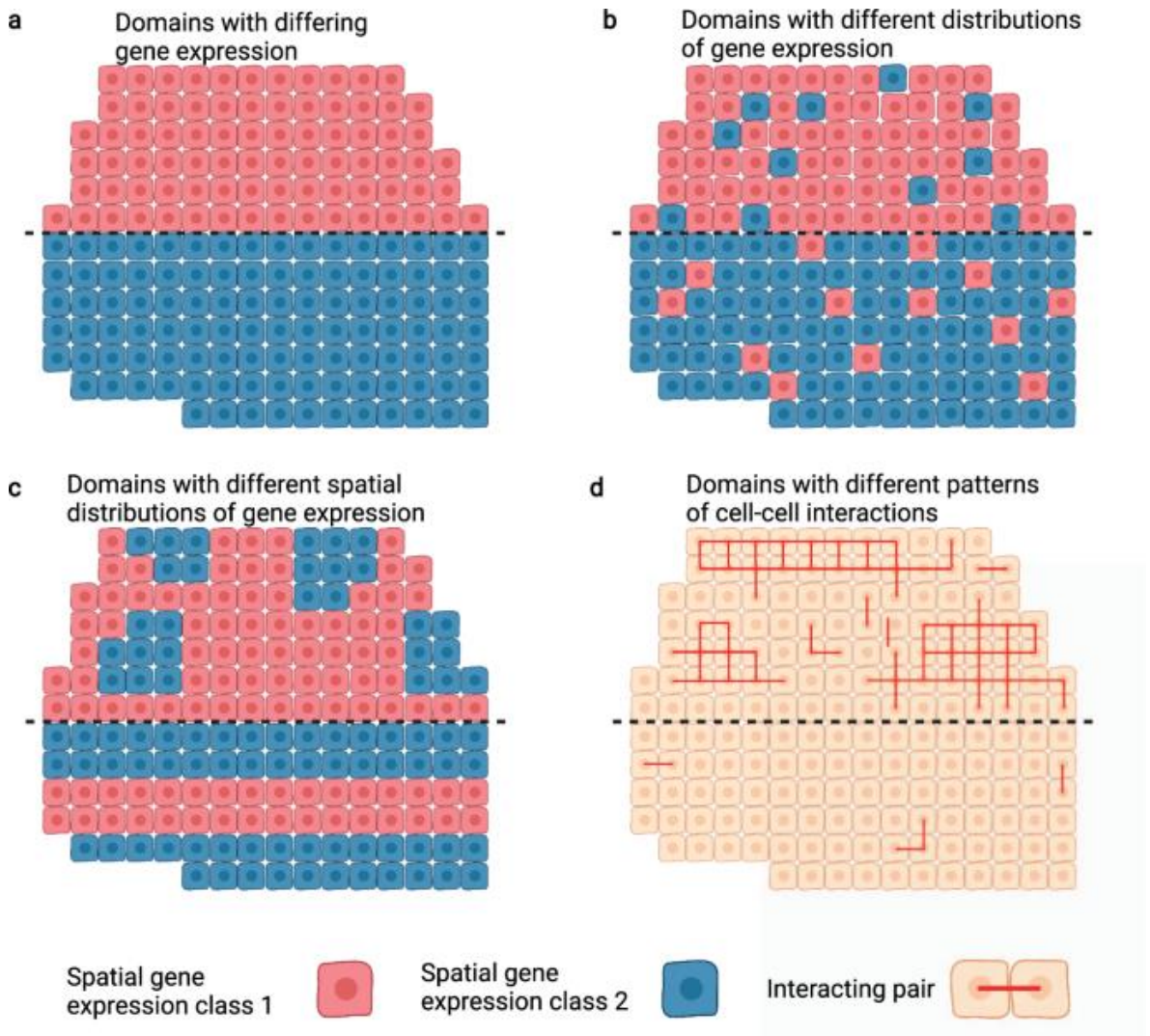


Heumos, L., Schaar, A.C., Lance, C. et al. Best practices for single-cell analysis across modalities. Nat Rev Genet (2023). https://doi.org/10.1038/s41576-023-00586-w

Different scenarios of spatially variable genes

The simplest approach is to look for spatially contiguous regions of cells with maximally similar gene expression (Fig. 2a). This is analogous to the typical clustering analysis in scRNA-seq analysis pipelines, but conscious of spatial position.

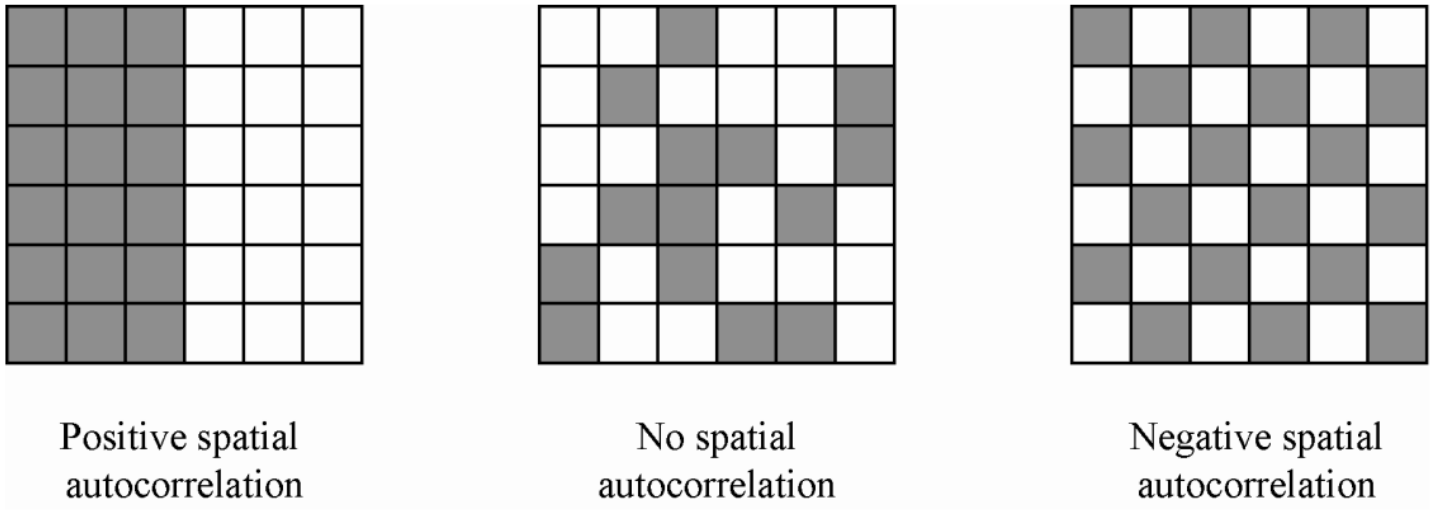
Most currently methods optimize for situations like this, but as methods develop they may have to deal with situations that are more like “salt and pepper” or which have subdomain architecture or take into account specific patterns of cell-cell interactions (CCI)



Walker, B. L., Cang, Z., Ren, H., Bourgain-Chang, E., & Nie, Q. (2022). Deciphering tissue structure and function using spatial transcriptomics. *Communications Biology*, *5*(1), 220.

Spatial Autocorrelation

One simple way to look at spatial relationships is by measuring spatial autocorrelation (i.e. for each gene)



Positive spatial autocorrelation

No spatial autocorrelation

Negative spatial autocorrelation

More info: Moran's I: https://www.youtube.com/watch?v=OJU8GNW9grc

From Radil, S. M. (2011). University of Illinois at Urbana-Champaign

Moran's I

Moran's I is a spatial autocorrelation metric similar to the Pearson correlation coefficient. Range is -1 to 1.

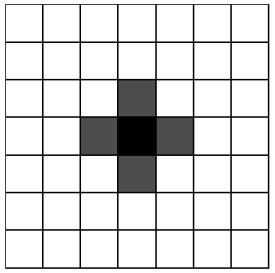
$$\text{Moran's I} = \frac{N}{\sum_{i,j} W_{ij}} \frac{\sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

, where N is the total number of spatial location units indexed by (i, j), and W is a weight matrix to be discussed below. Recall that the Pearson correlation coefficient is

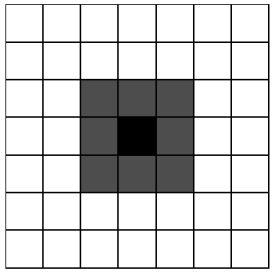
$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$W_{i,j}$ is a weight matrix that can be either contiguity based (first order, second order, etc) or distance-based

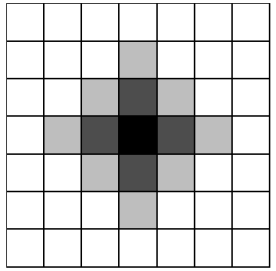
Rook. First order



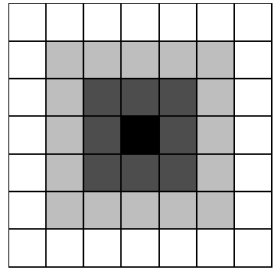
Queen. First order



Rook. Second order



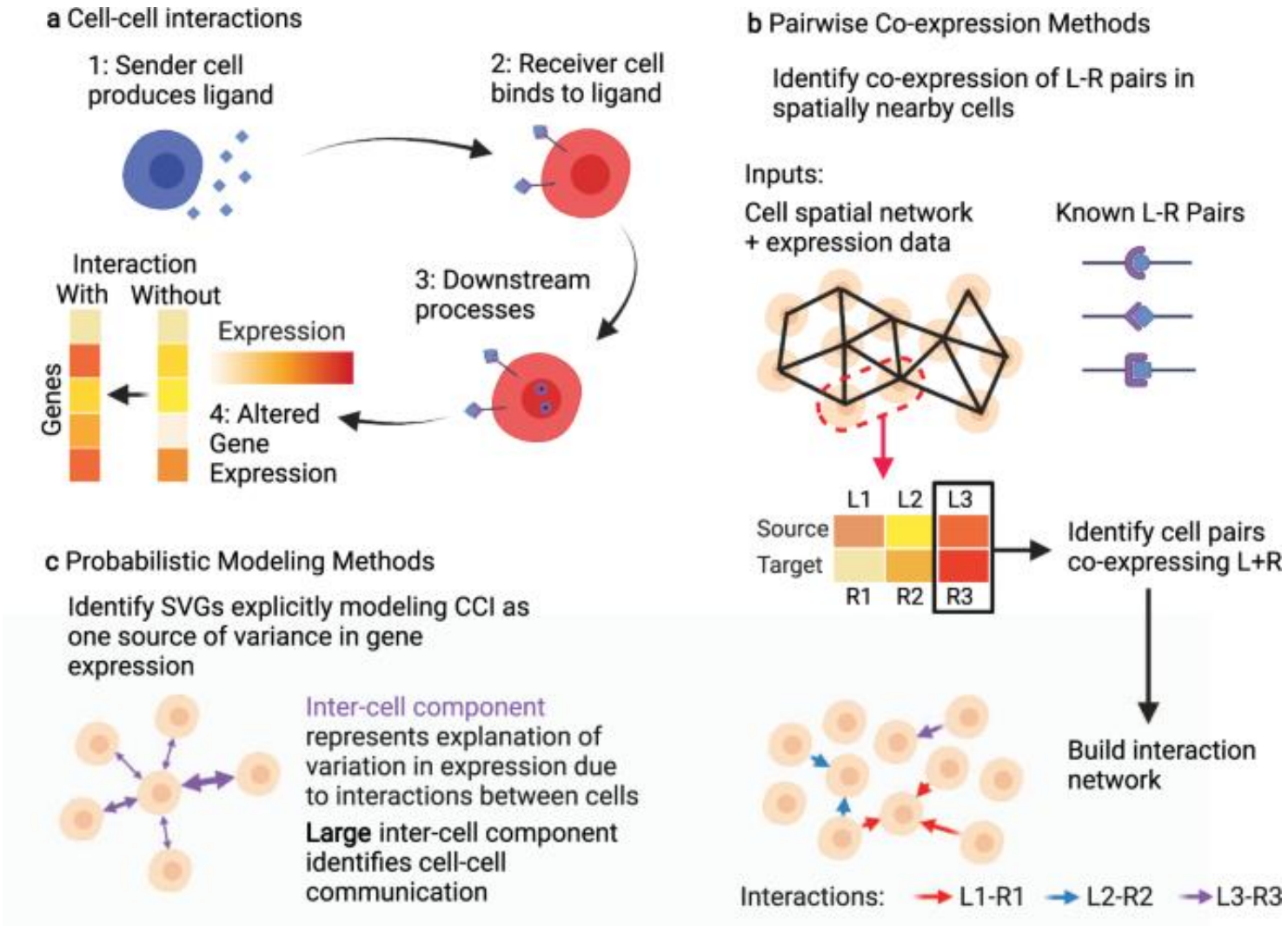
Queen. Second order



Intuitive explanation for understanding and deriving Moran's I:
https://www.youtube.com/watch?v=OJU8GNW9grc&ab_channel=ritvikmath

First and second order nearest neighbors graphic from here
Moran's I definition from here

Methods to identify cell-cell interactions from STx data

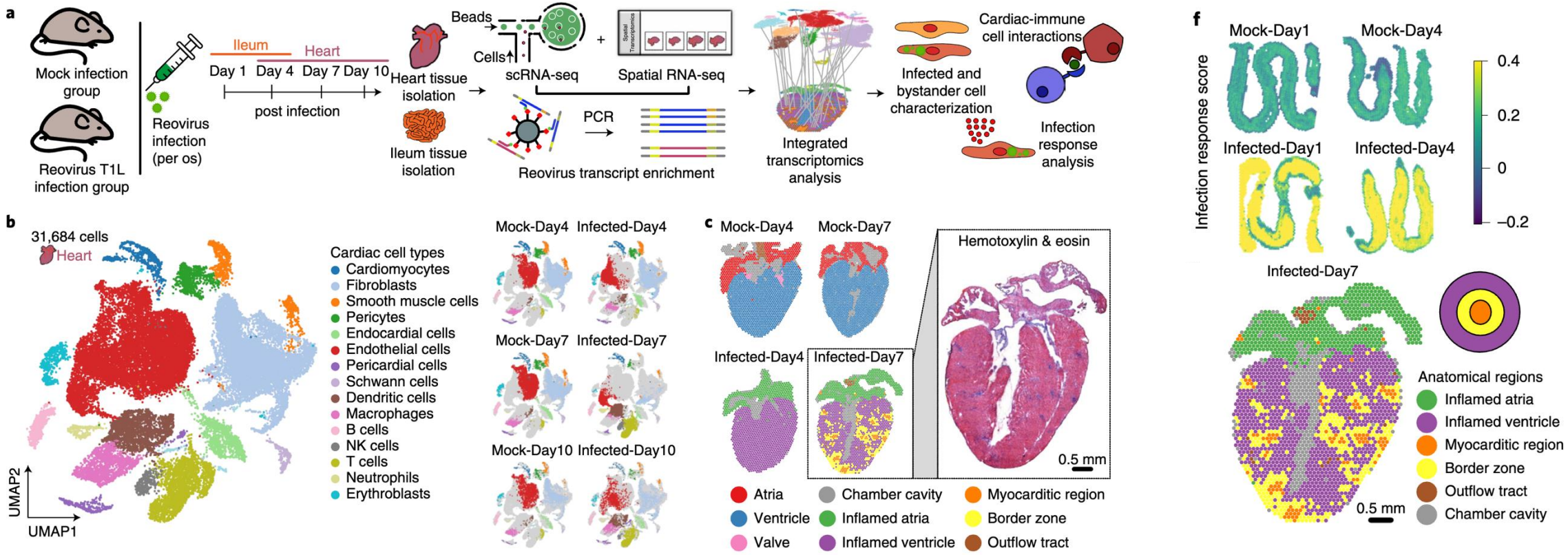


Walker, B. L., Cang, Z., Ren, H., Bourgain-Chang, E., & Nie, Q. (2022). Deciphering tissue structure and function using spatial transcriptomics. *Communications Biology*, *5*(1), 220.

Ongoing challenges in the spatial transcriptomics field

- Experimental technologies still being improve (cost, applicability to all kinds of samples such as FF or FFPE, sensitivity, field of view, depth, etc)
- Methods for data analyses still being developed
- Truly “spatially variable” genes that are not just marker genes
- Spatial DE (example: C-SIDE method using covariate matrix)
- Spatial datasets as 3D maps rather than just 2D coordinates
- Spatial image registration across replicates and samples
- Effective integration with other datasets (scRNA-seq, ATAC-seq, proteomics)

Example Applications 1: Spatial Transcriptomics reveals pathogenesis of viral myocarditis (reovirus on ileum and heart with Visium and scRNA-seq)



-- Identified gene module score (230 genes) sig upregulated in reovirus infection

-- Chronology leading to reovirus-induced myocarditis incl. gut infection, then secondary tissue heart infection

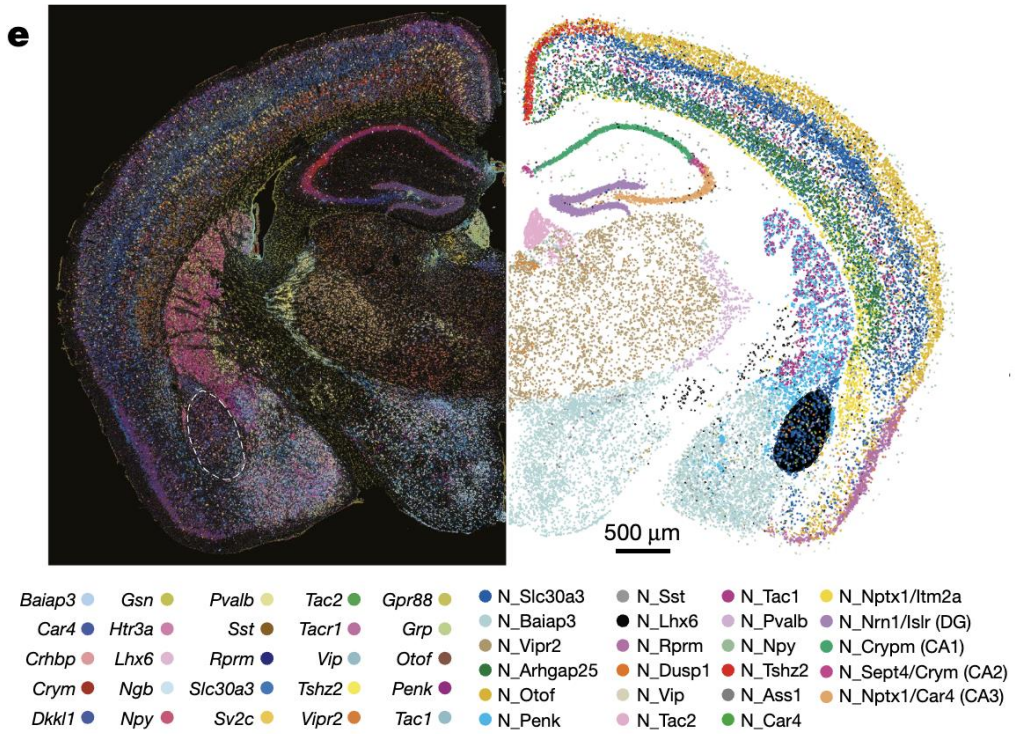
-- Identify role for the basal type-I interferon (innate immune response) in endothelial cells lining cardiac vasculature, which secrete chemokines to recruit cytotoxic T cells, then undergo cell death

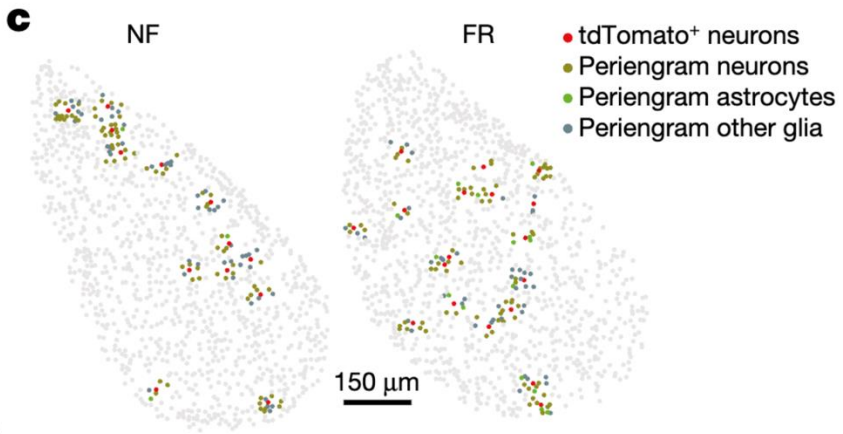
Mantri et al. Nature Cardiovascular Research 2022 (Vlaminck Lab)

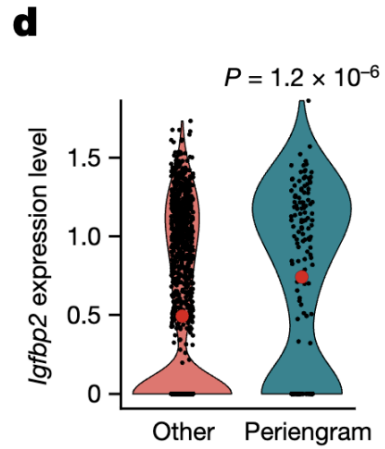
https://doi.org/10.1038/s44161-022-00138-1

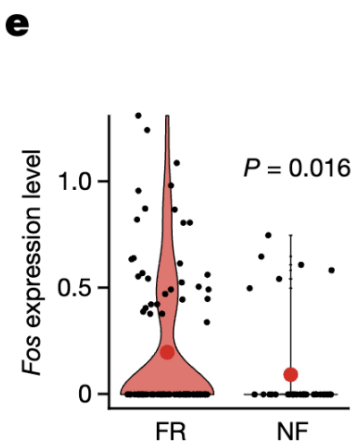
Example Applications 2: Spatial transcriptomics reveal neuron-astrocyte synergy in long term memory

-- Spatially resolved ensemble of engram neurons and locally associated astrocyte subtype using MERFISH
-- Evidence supporting idea of perineuronal nets to contribute to memory stabilization in the BLA

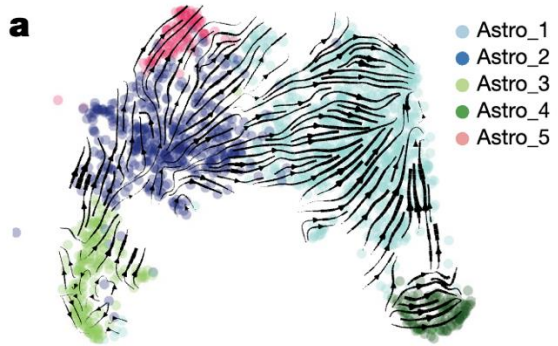








FR = fear training and recall
NF = no fear control



Sun et al. Nature 2024 (Quake Lab)

https://doi.org/10.1038/s41586-023-07011-6

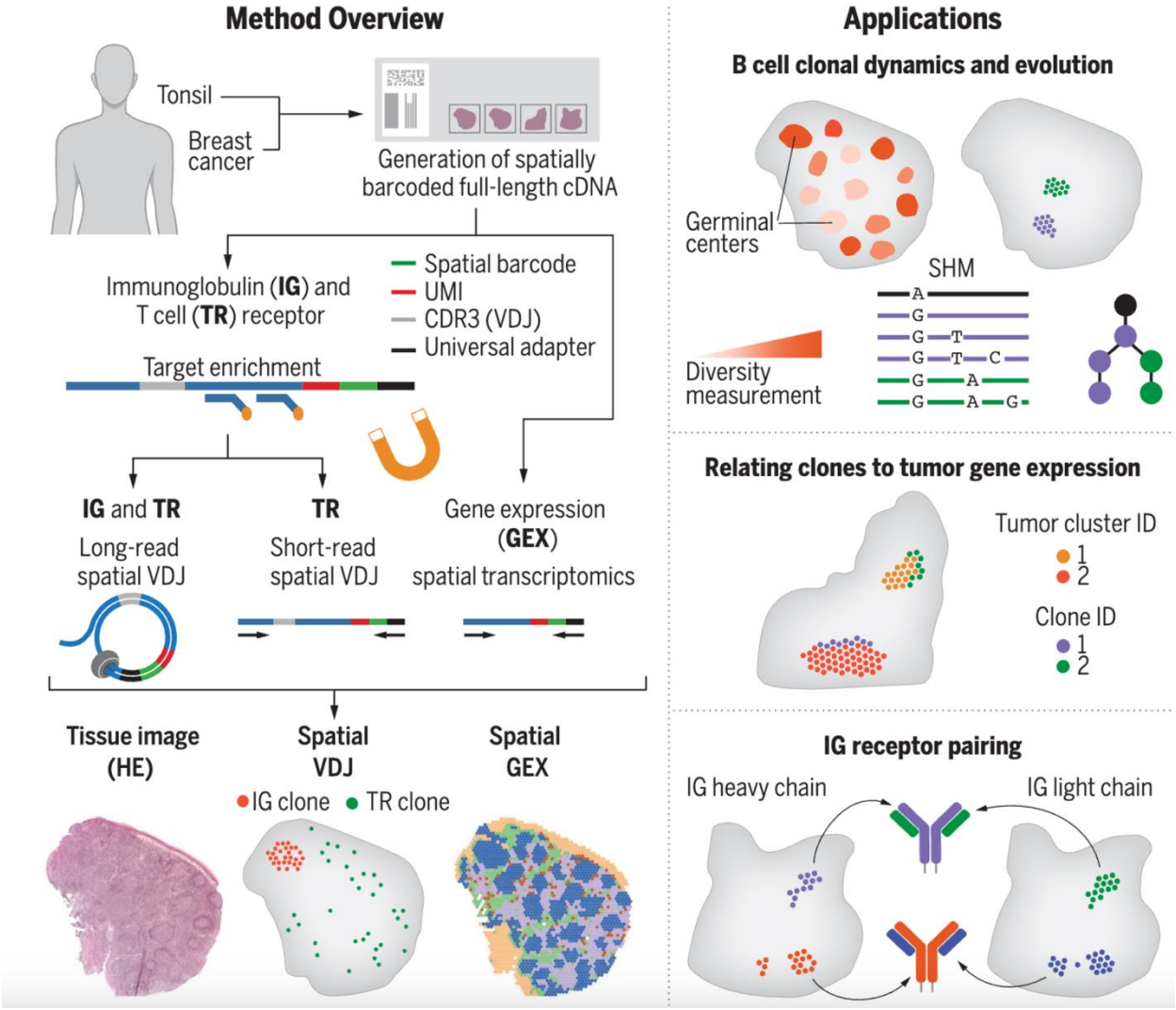
Example Applications 3: Long read Spatial VDJ reveals B and T cell clonal dynamics in both lymphoid and cancer tissue

Visium of tonsils or breast cancer

1) Long read / PacBio Spatial VDJ sequencing of spatially barcoded libraries of full length IG and TR antigen receptor transcripts
2) Short read Spatial VDJ sequencing of TR sequences only of CDR3 regions

Hybridization-capture enrichment of TR and IG constant regions to enrich for antigen receptor transcripts

-- Able to detect B cell clonal dynamics in germinal centers of tonsil tissue and somatic hypermutation
-- Visualize B cell clonality in breast cancer tissue



Engblom et al., Science 2023 (Lundeberg, Frisen labs)

https://doi.org/10.1126/science.adf8486

Spatial Transcriptomics Resources at NIH

- **Spatial Biology Interest Group Listserv / Email List**
  - https://oir.nih.gov/sigs/spatial-biology-interest-group
- **Specialized Instrumentation at NIH:**
  - **NCI** CCR
    - 10X Xenium https://ostr.ccr.cancer.gov/emerging-technologies/spatial-biology/xenium/
    - 10X Visium https://ostr.ccr.cancer.gov/emerging-technologies/spatial-biology/visium/ (FFPE samples can be processed with CytAssist instrument)
    - SCAF https://ostr.ccr.cancer.gov/resources/provider_details/nci-ccr-single-cell-analysis-facility-scaf
    - CosMX and GeoMX at Spatial Imaging Technology Resource (SpITR) https://spitr.ccr.cancer.gov/
  - Vizgen said that **NIA** (Baltimore) and **NEI** both each have a MERSCOPE instrument
  - Xenium and Visium CytAssist are also at **NIAMS, NINDS**. **NIAID VRC** also acquiring both this year
  - **NIAID** RTB (Research Technologies Branch): https://rtb.nih.gov/section/STR have Visium and GeoMX DSP
- **Software on Biowulf:**
  - GeoMX NGS Pipeline: https://hpc.nih.gov/apps/geomx_ngs_pipeline.html
  - Xenium Ranger: https://hpc.nih.gov/apps/xeniumranger.html
  - Space Ranger (for Visium): https://hpc.nih.gov/apps/spaceranger.html

- **Looking for bioinformatic and genomics analysis expertise? Contact us at bioinformatics@niaid.nih.gov**

Prepare for Part 2 next week!

Please make sure you install and can run R and 

either locally or on Biowulf

1) Either download and install Rstudio locally:
https://posit.co/download/rstudio-desktop/

Or use these instructions for running R/R Studio Server on Biowulf using tunneling:
https://hpc.nih.gov/apps/rstudio-server.html

2) Download Quarto Markdown document .qmd (contains R code we'll run) from Github

https://github.com/margaretc-ho/BCBB_STx_workshop_2024

3**) Inside Quarto document in red is code for setup (please run before Part 2 of workshop):**
Install packages (**Seurat**, **Seurat-data** which includes Visium dataset etc) and check that you can load the libraries
Download Allen cortex scRNA-seq data

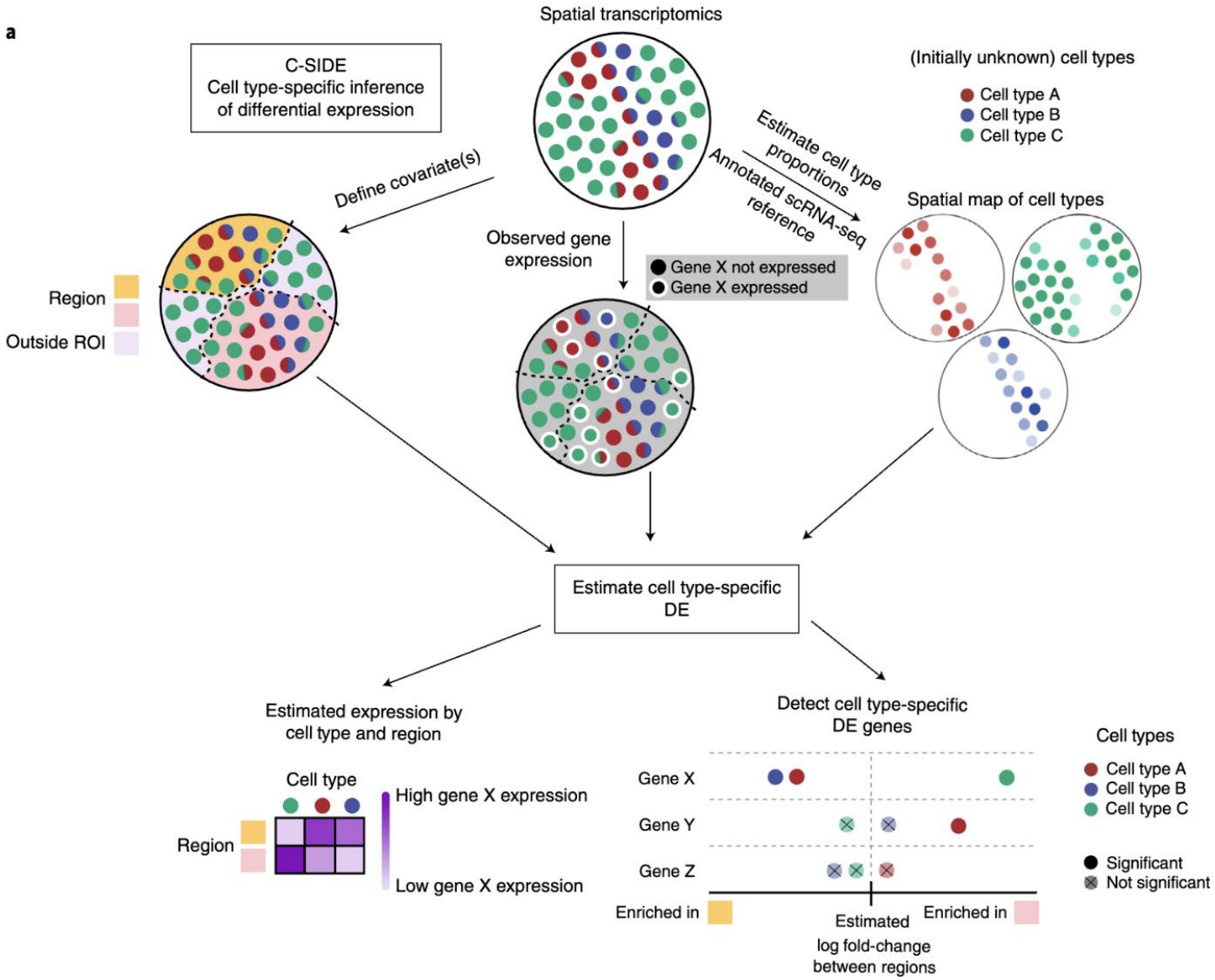
Followup questions and inquiries are welcome!
margaret.ho@nih.gov

Further Reading / Resources

- Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., & Haque, A. (2022). **An introduction to spatial transcriptomics for biomedical research**. *Genome Medicine*, *14*(1), 68. [REVIEW]
- Longo, S. K., Guo, M. G., Ji, A. L., & Khavari, P. A. (2021). **Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics**. *Nature Reviews. Genetics*, *22*(10), 627–644.
- Walker, B. L., Cang, Z., Ren, H., Bourgain-Chang, E., & Nie, Q. (2022). **Deciphering tissue structure and function using spatial transcriptomics**. *Communications Biology*, *5*(1), 220.
- Heumos, L., Schaar, A.C., Lance, C. et al. **Best practices for single-cell analysis across modalities**. Nat Rev Genet (2023)
- **Rafa Irizarry 2024 lecture on Statistical Methods for Single-Cell RNA-Seq Analysis and Spatial Transcriptomics** https://bioinformatics.ccr.cancer.gov/btep/classes/rafael-irizarry
- Li, B et al. (2022). **Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution**. *Nat Methods*, *19*(6), 662–670
- Jeon, H., Xie, J., Jeon, Y., Jung, K. J., Gupta, A., Chang, W., & Chung, D. (2023). **Statistical power analysis for designing bulk, single-cell, and spatial transcriptomics experiments: Review, tutorial, and perspectives**. *Biomolecules*, *13*(2), 221.
- Cable, DM et al. (2022). **Robust decomposition of cell type mixtures in spatial transcriptomics**. *Nat Biotech*, *40*(4), 517–526.
- Choudhary, S., & Satija, R. (2022). **Comparison and evaluation of statistical error models for scRNA-seq**. Genome Biology, 23(1), 27.
- Moffitt, J. R., & Zhuang, X. (2016). **RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH).** *Methods in Enzymology*, *572*, 1–49.
- Wang, Y., Liu, B., Zhao, G., Lee, Y., Buzdin, A., Mu, X., Zhao, J., Chen, H., & Li, X. (2023). **Spatial transcriptomics: Technologies, applications and experimental considerations**. *Genomics*, *115*(5), 110671.

Spatial Differential Gene Expression analyses

C-SIDE as part of spacexr Package (Irizarry Lab)



Spatial Differential Gene Expression analyses

C-SIDE as part of spacexr Package (Irizarry Lab) will perform DE across a covariate (e.g. proximity to pathology or discrete regions)