



2023 제19회 KOGO 동계워크샵

Spatially Resolved Transcriptomics (SRT) Principles

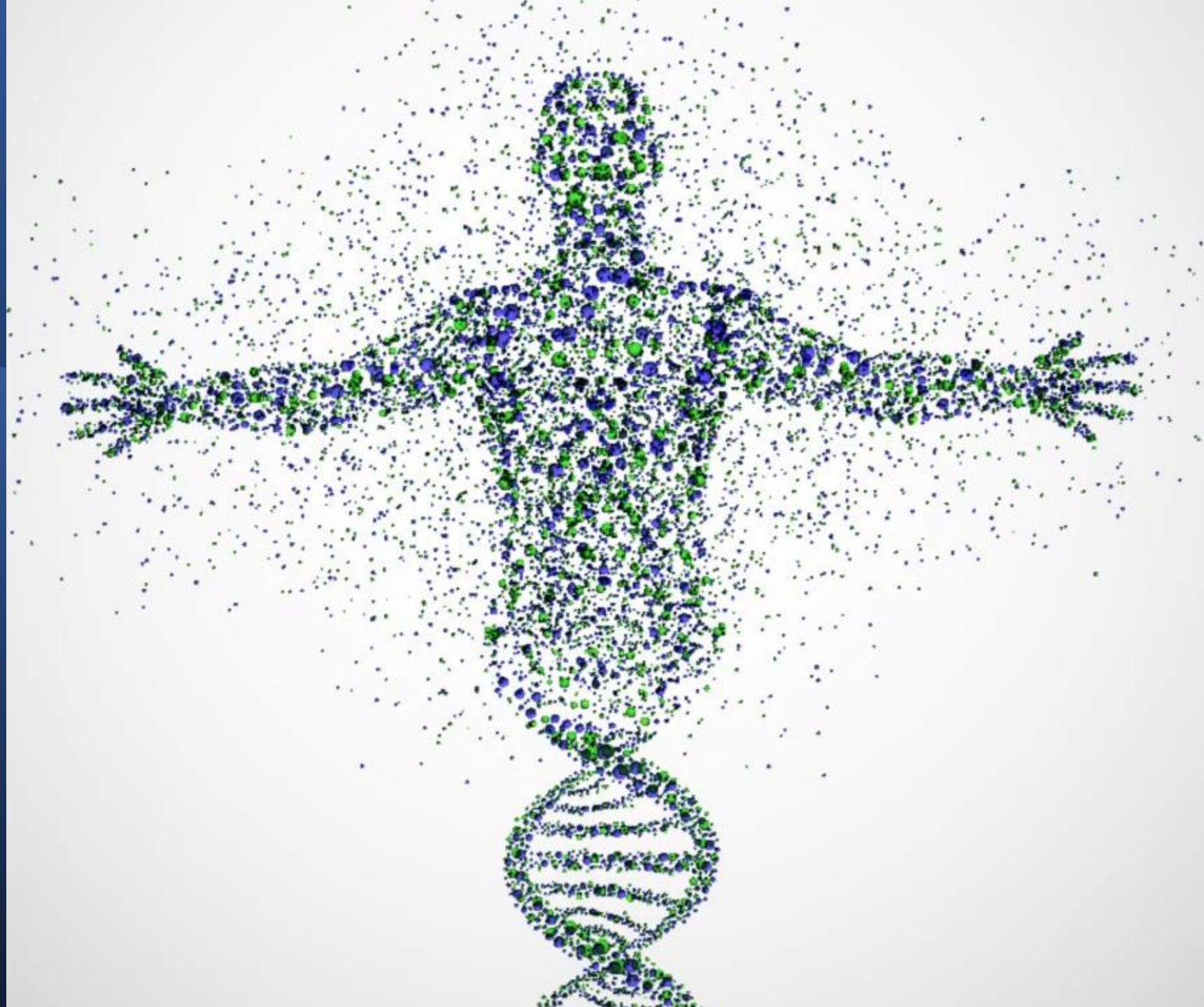
Jungmin Choi, Ph.D.

Department of Biomedical Sciences
Korea University, College of Medicine

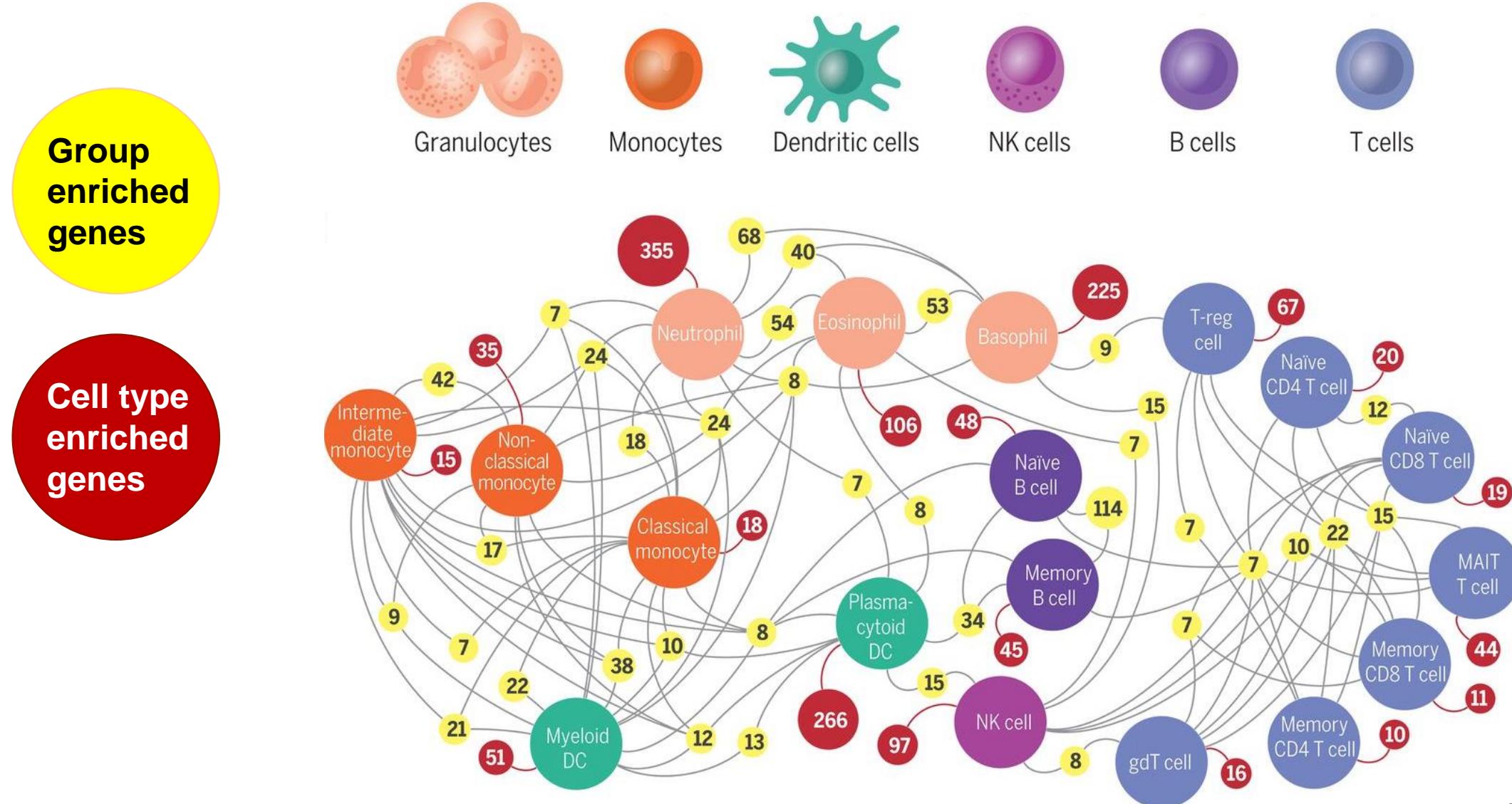
Contents covered today

- **Introduction**
 - A broad overview of single-cell data and experimental spatially resolved techniques
- **Computational methodology and frameworks**
 - Different flavors of currently available spatially resolved data analysis methods

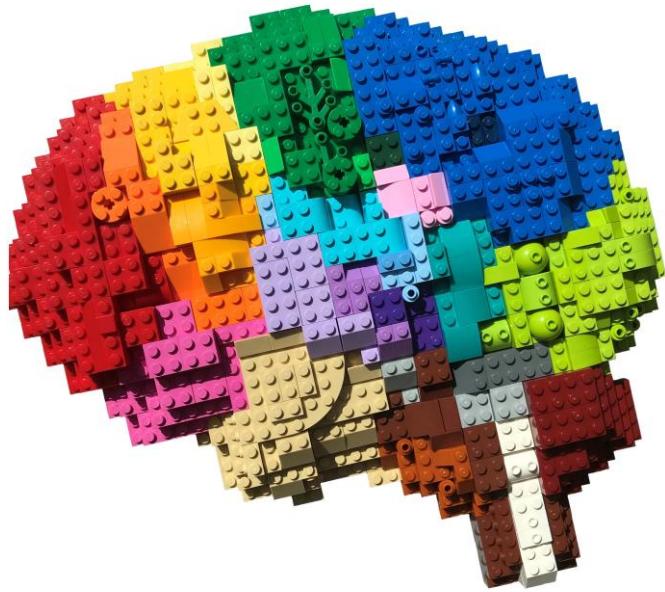
**Human Body
is Made of
37 Trillions
of Cells**



Each cell expresses a set of different genes that enable them to carry out specialized functions



Single cells help gain insight into complexity of biology



Brain tissue



Bulk genomics

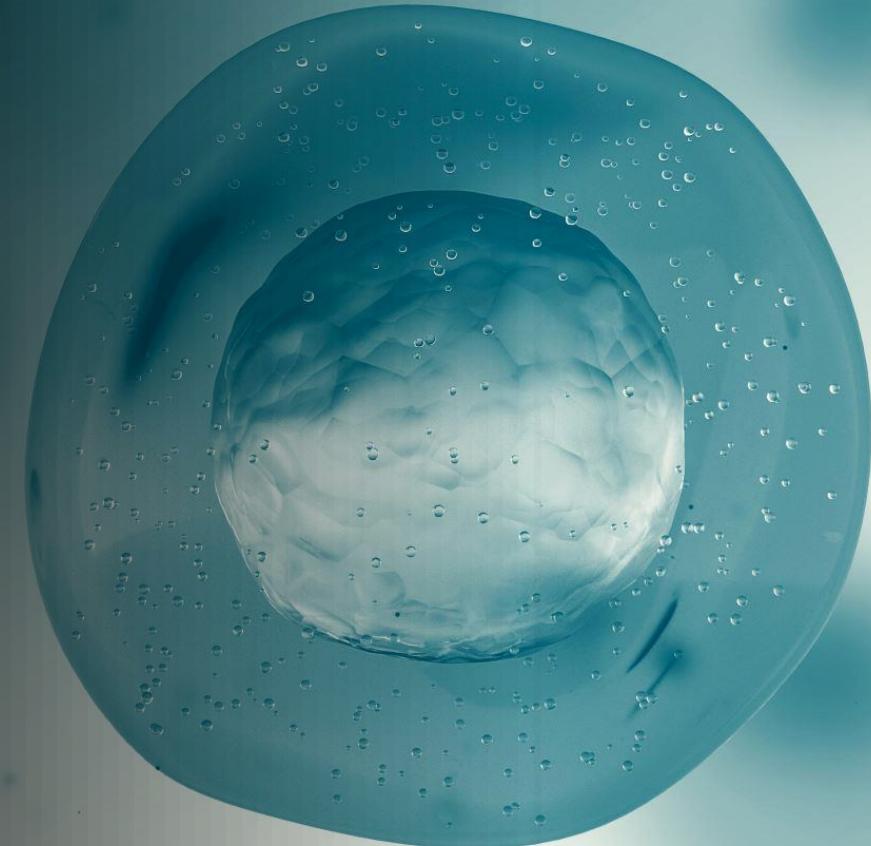


Single-cell genomics

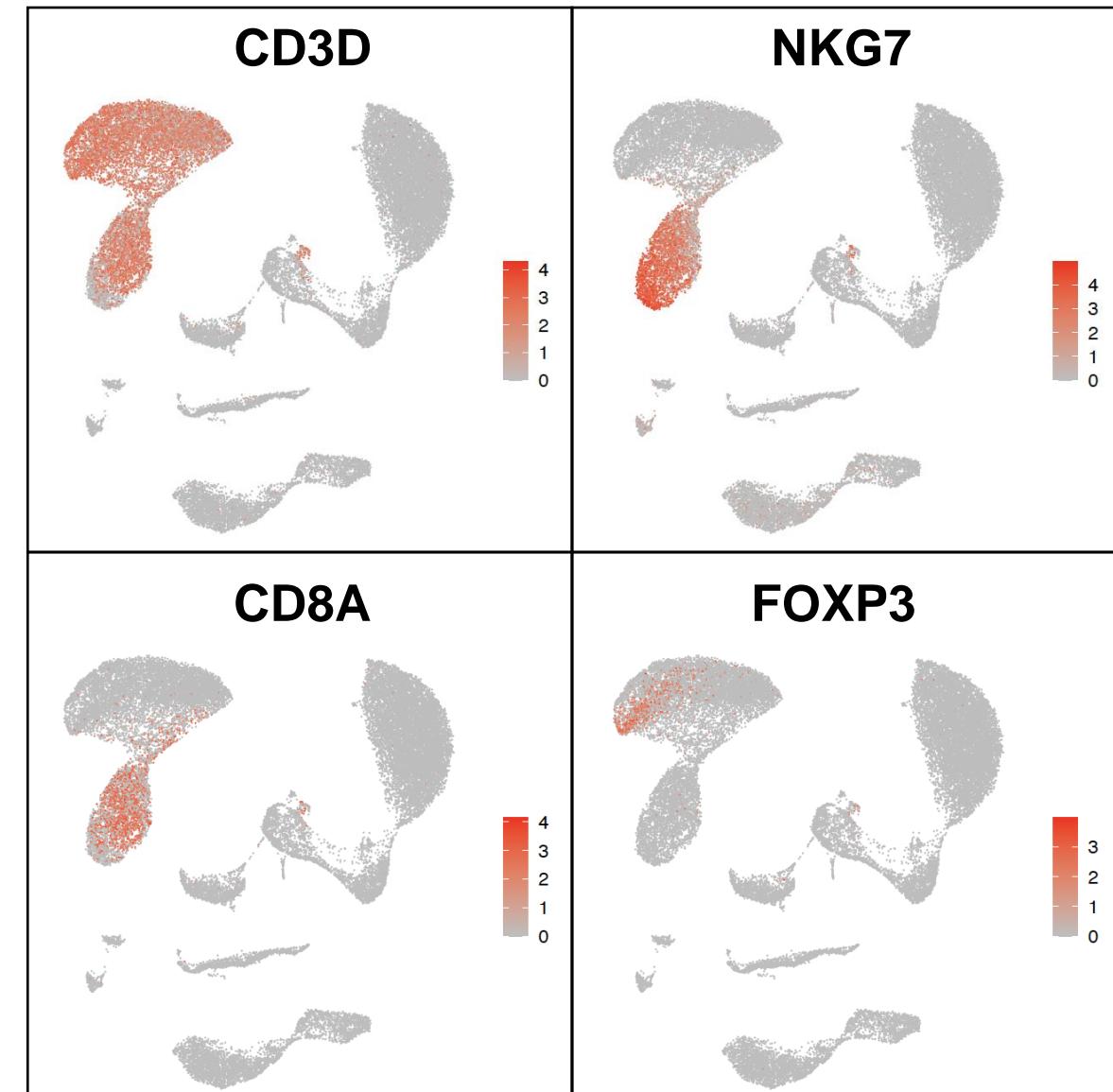
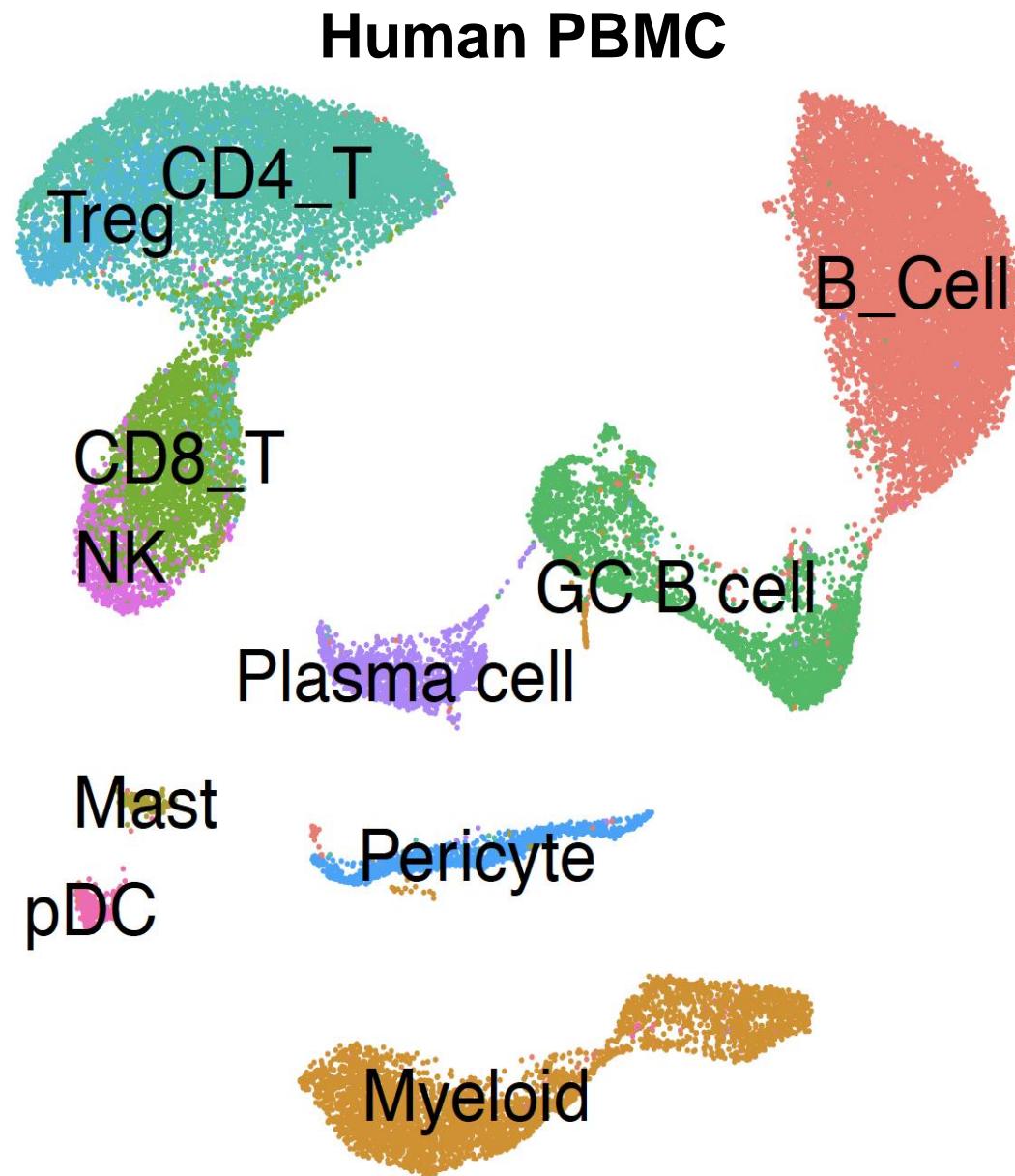
The average behavior measured in millions of cells (bulk genomics)
does not necessarily reflect the behavior in individual cells



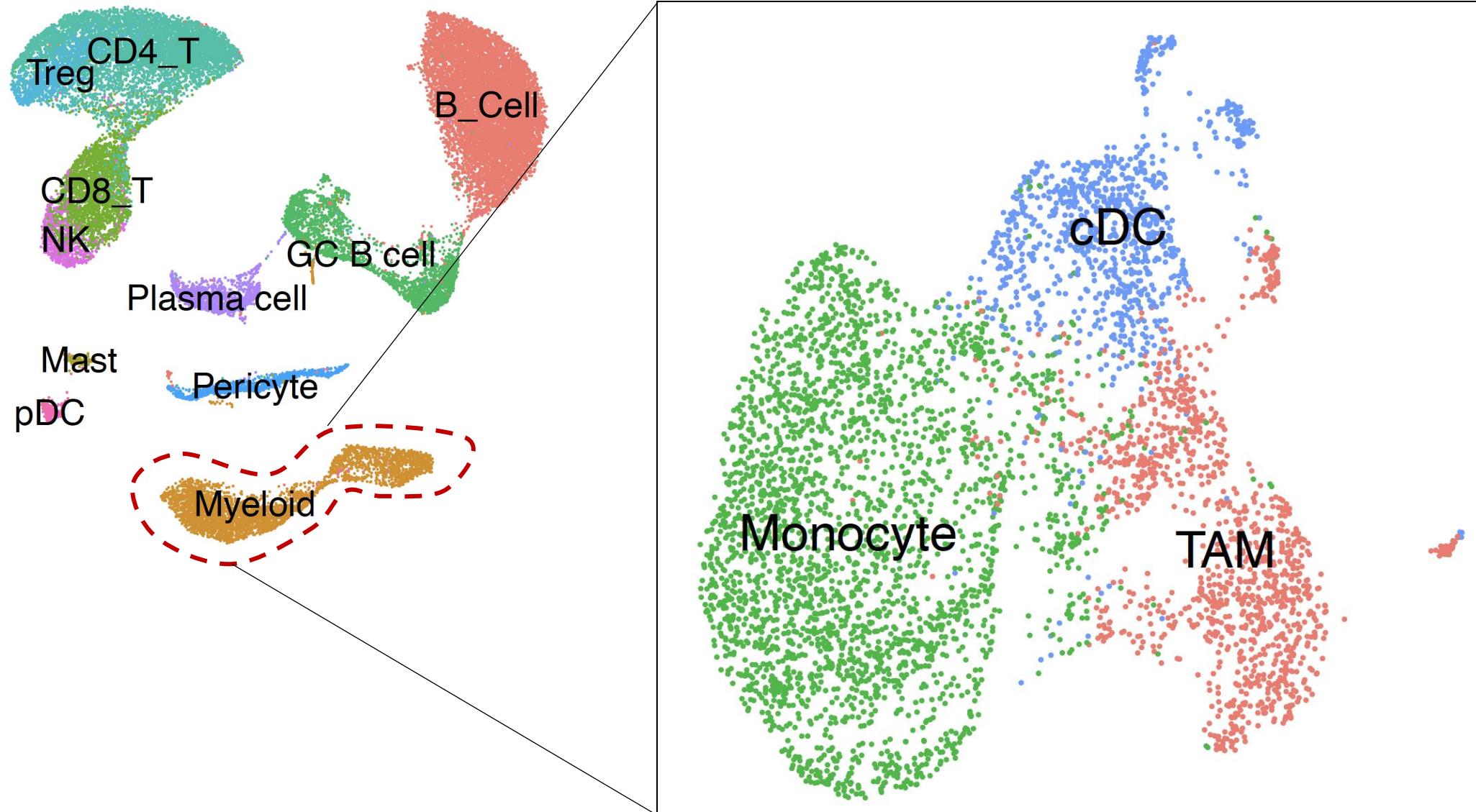
Why care single cells?



1. Single cells can identify and characterize cell types

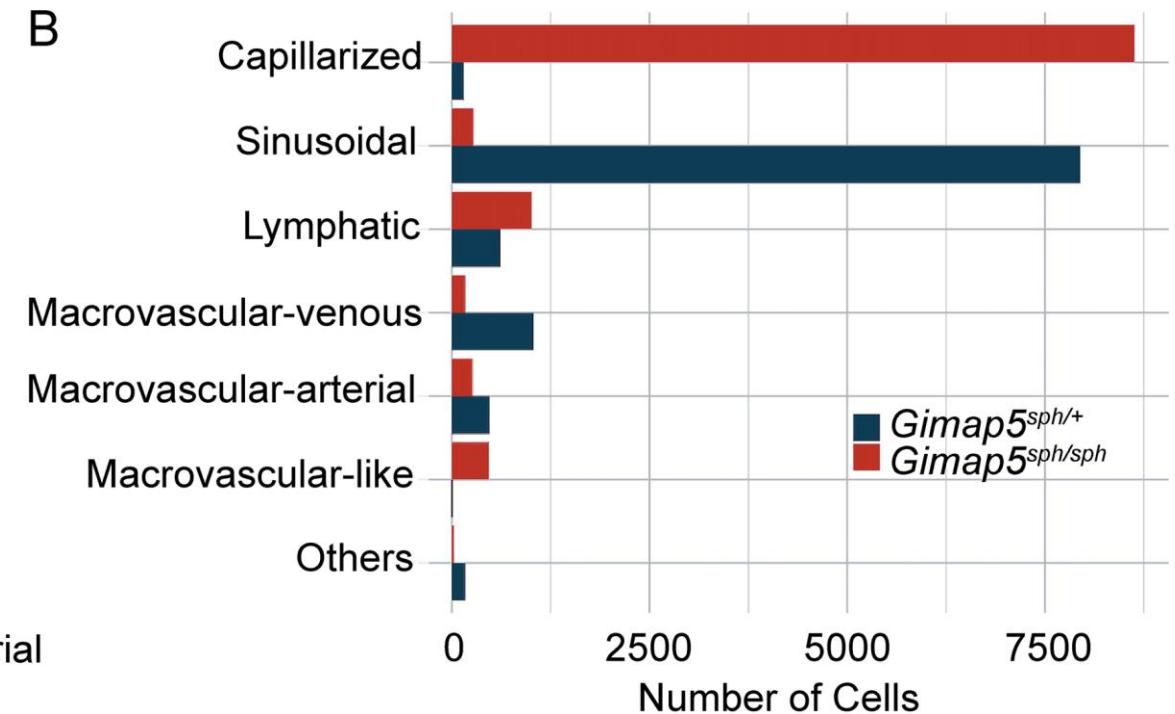
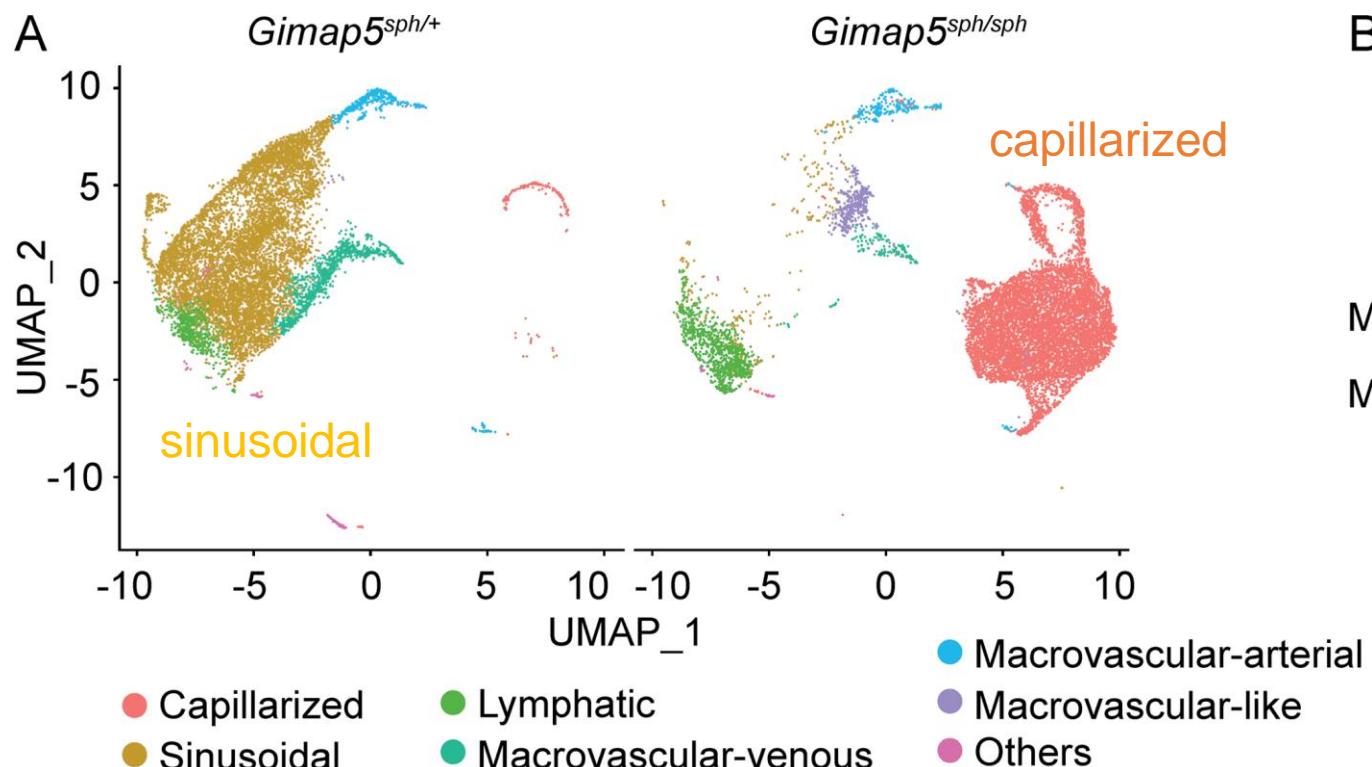


1. Cell clusters of interest can be even further explored



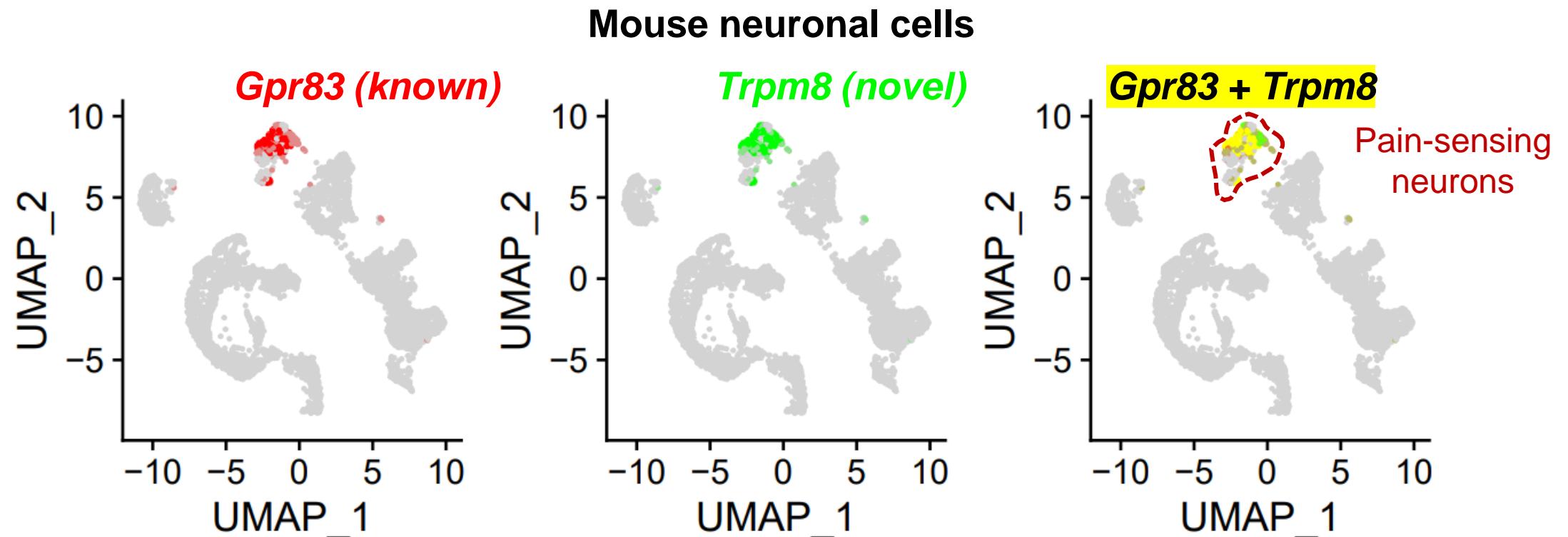
2. Single cell data can delineate cellular composition and their dynamics of cell state

Mouse endothelial cells

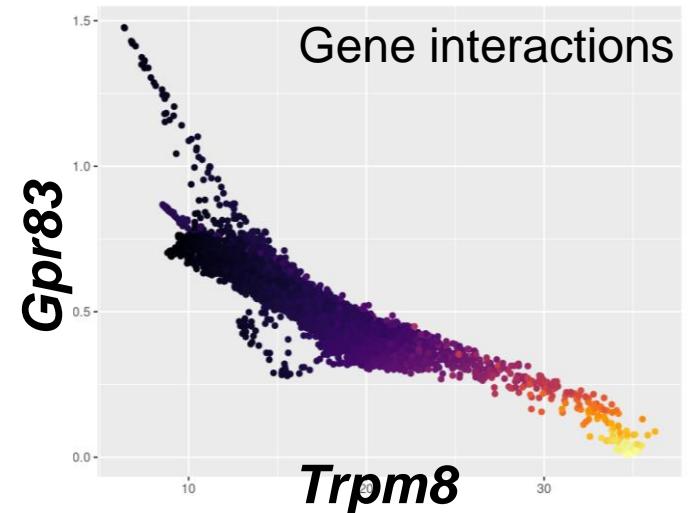


The homozygous deletion of *Gimap5* leads to the accumulation of a specific type of endothelial cells, resulting in portal hypertension

3. Single cells can isolate strongly associated genes

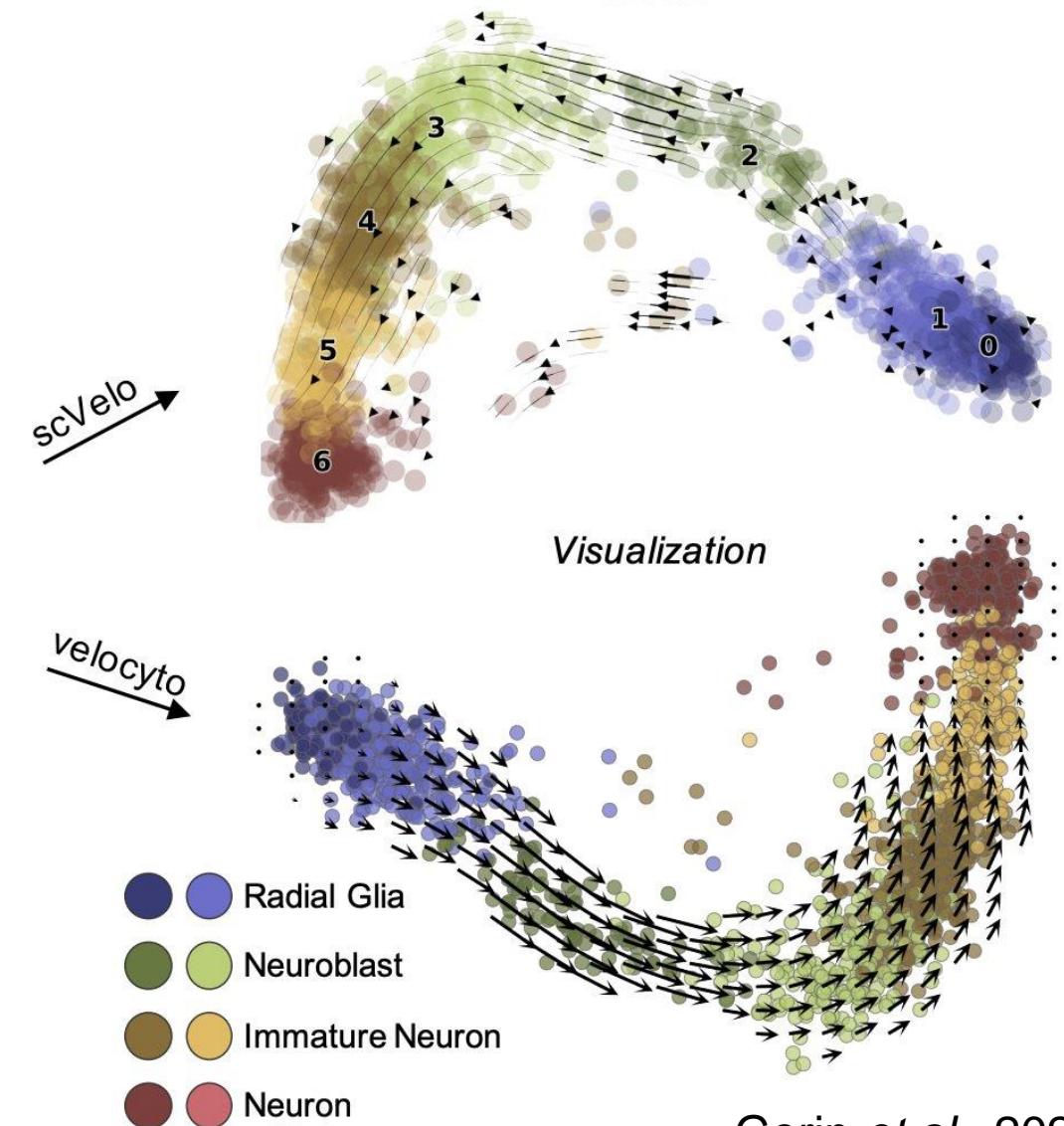
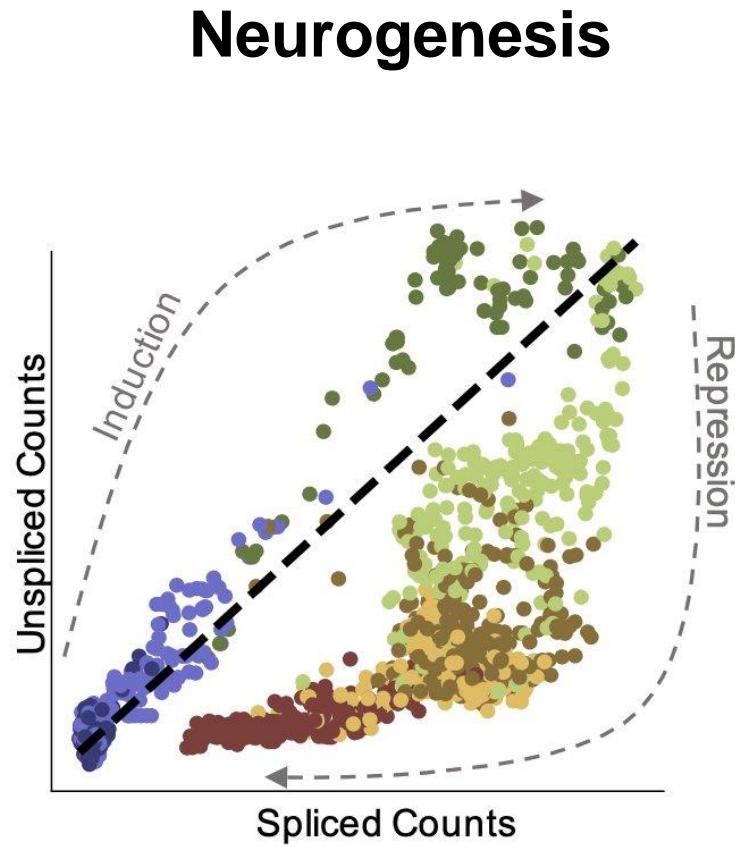


Gpr83 and Trpm8 interaction is observed in pain-sensing neurons exclusively, suggesting important roles



4. Single cell data can predict a continuum from static snapshots of cell state

A sequential trajectory
of cliff diving



Single cell genomics was pioneered by James Eberwine (Upenn) in the early 1990

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1663–1667, March 1990
Biochemistry

Amplified RNA synthesized from limited quantities of heterogeneous cDNA

(cerebellum/guanine nucleotide-binding protein/T7 RNA polymerase/Purkinje cell)

RUSSELL N. VAN GELDER*,†, MARK E. VON ZASTROW‡, ANDREA YOOL§, WILLIAM C. DEMENT*,
JACK D. BARCHAS‡, AND JAMES H. EBERWINE‡¶

*Nancy Pritzker Laboratory of Behavioral Neurochemistry, *Sleep Research Laboratory, Department of Psychiatry, and §Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305

Communicated by Seymour Benzer, November 13, 1989

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 3010–3014, April 1992
Neurobiology

Analysis of gene expression in single live neurons

(amplified, antisense RNA/expression profile/mRNA complexity/pyramidal cell)

JAMES EBERWINE*†‡, HERMES YEH§, KEVIN MIYASHIRO*, YANXIANG CAO*, SURESH NAIR*,
RICHARD FINNELL*¶, MARTHA ZETTEL§, AND PAUL COLEMAN§

Departments of *Pharmacology and *Psychiatry, University of Pennsylvania Medical School, Philadelphia, PA 19104; and Department of §Neurobiology and Anatomy, University of Rochester Medical Center, Rochester, NY 14642

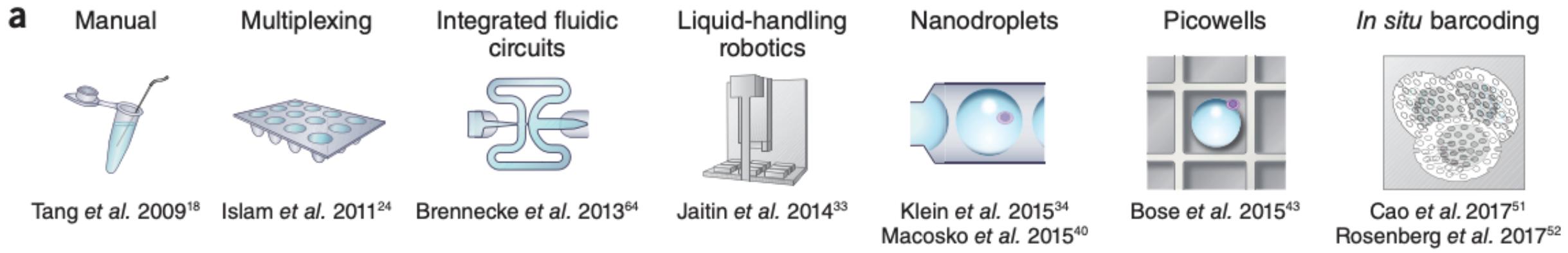
Communicated by George Koelle, December 2, 1991

The very first CSHL Single Cell Workshop in 2009

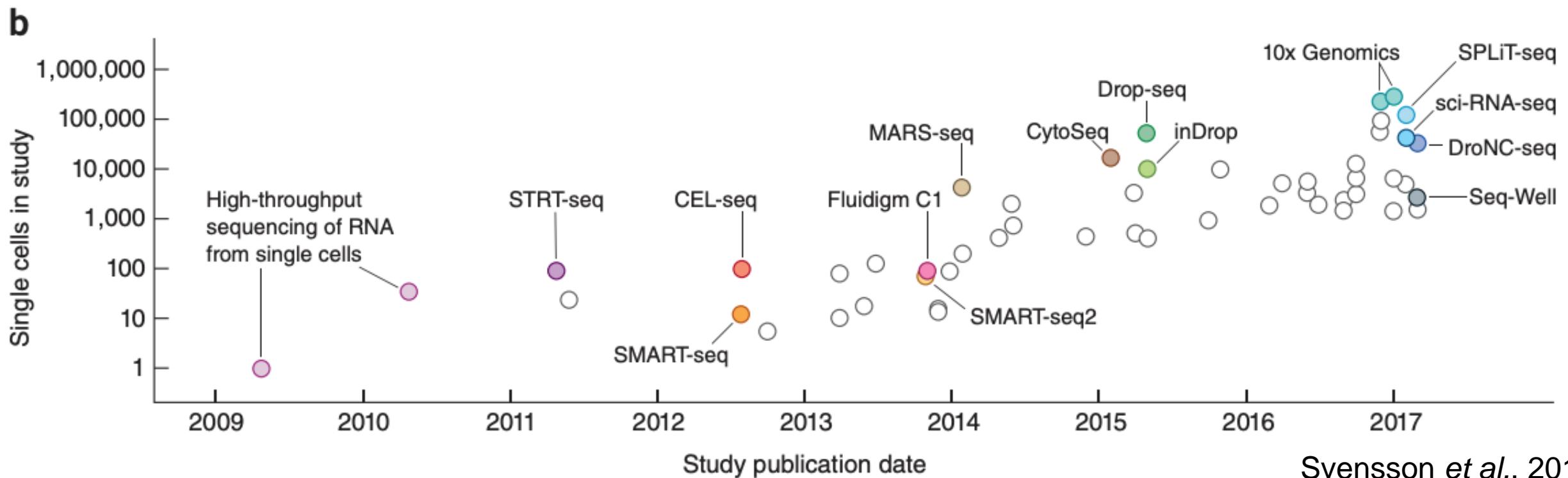


<https://pbs.twimg.com/media/EJRjv6sUwAARHO8?format=jpg&name=4096x4096>

Exponential scaling of single cell sequencing tech.

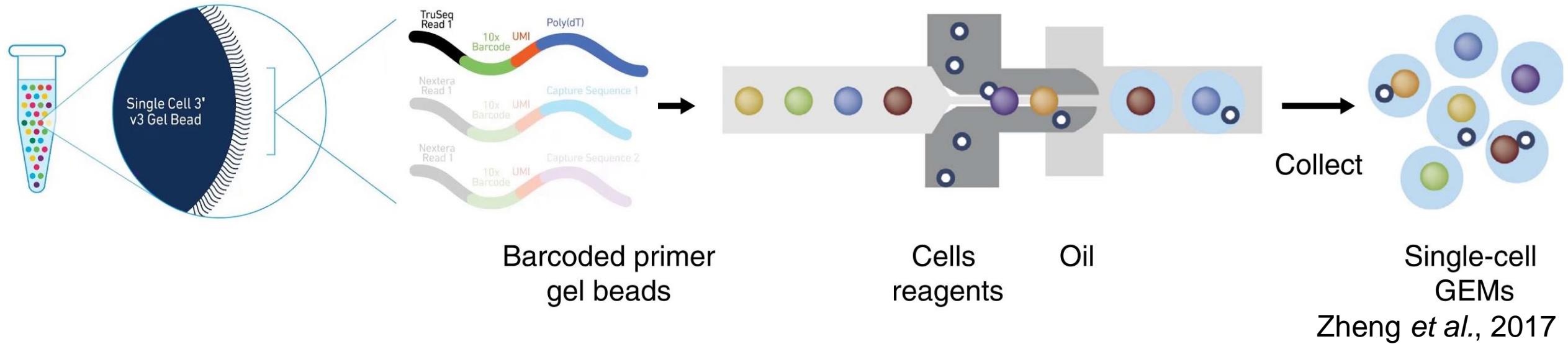


Few cells < 100s cells << 10,000 cells <<<<< 100,000 cells

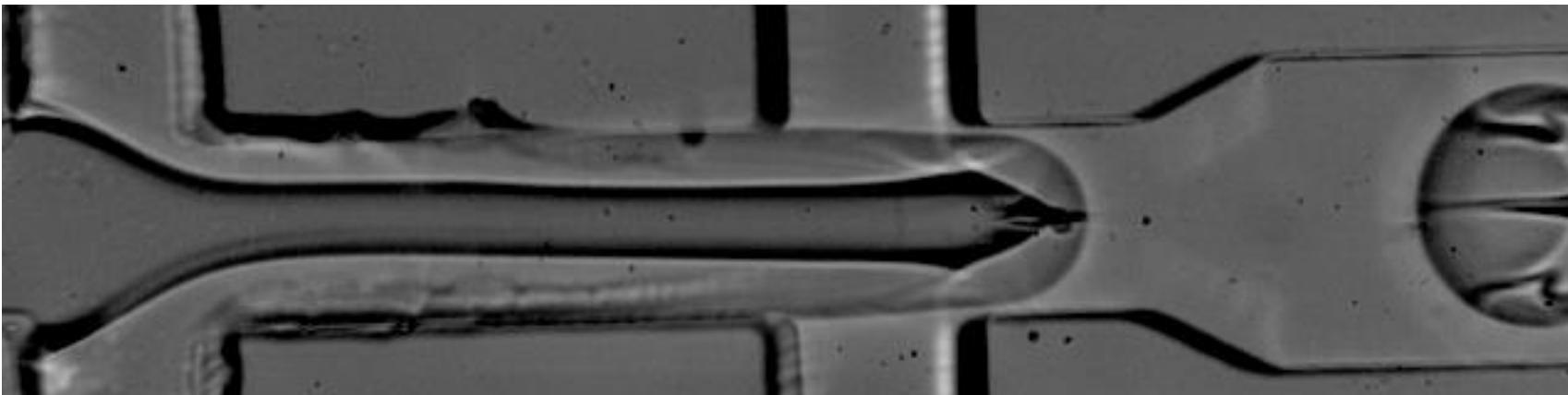


GAME-CHANGER

Microfluidic droplet-based cell isolation

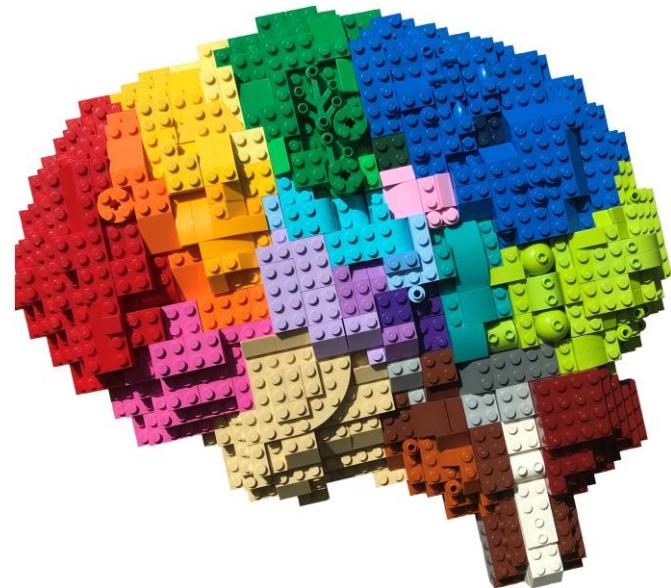


Zheng *et al.*, 2017



A gif image from dropseq.org

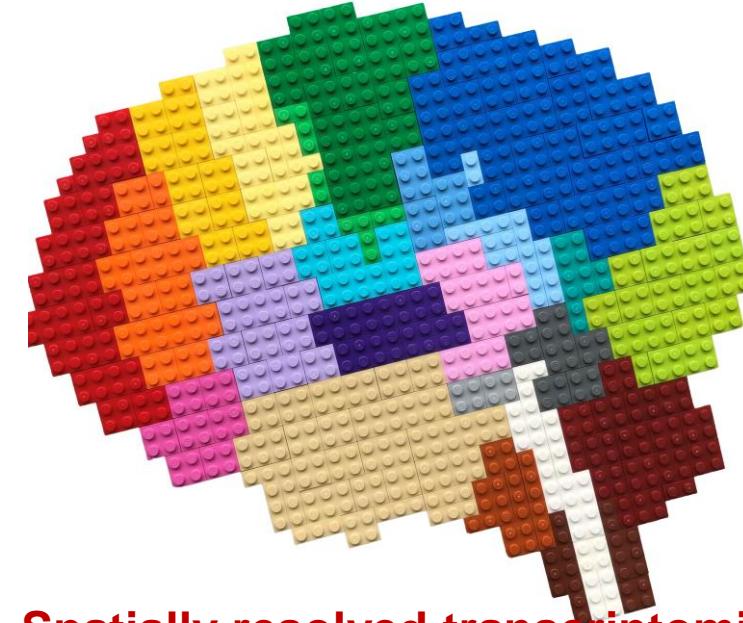
Unfortunately, single cell transcriptomics loses context



The original brain



single-cell RNA Sequencing



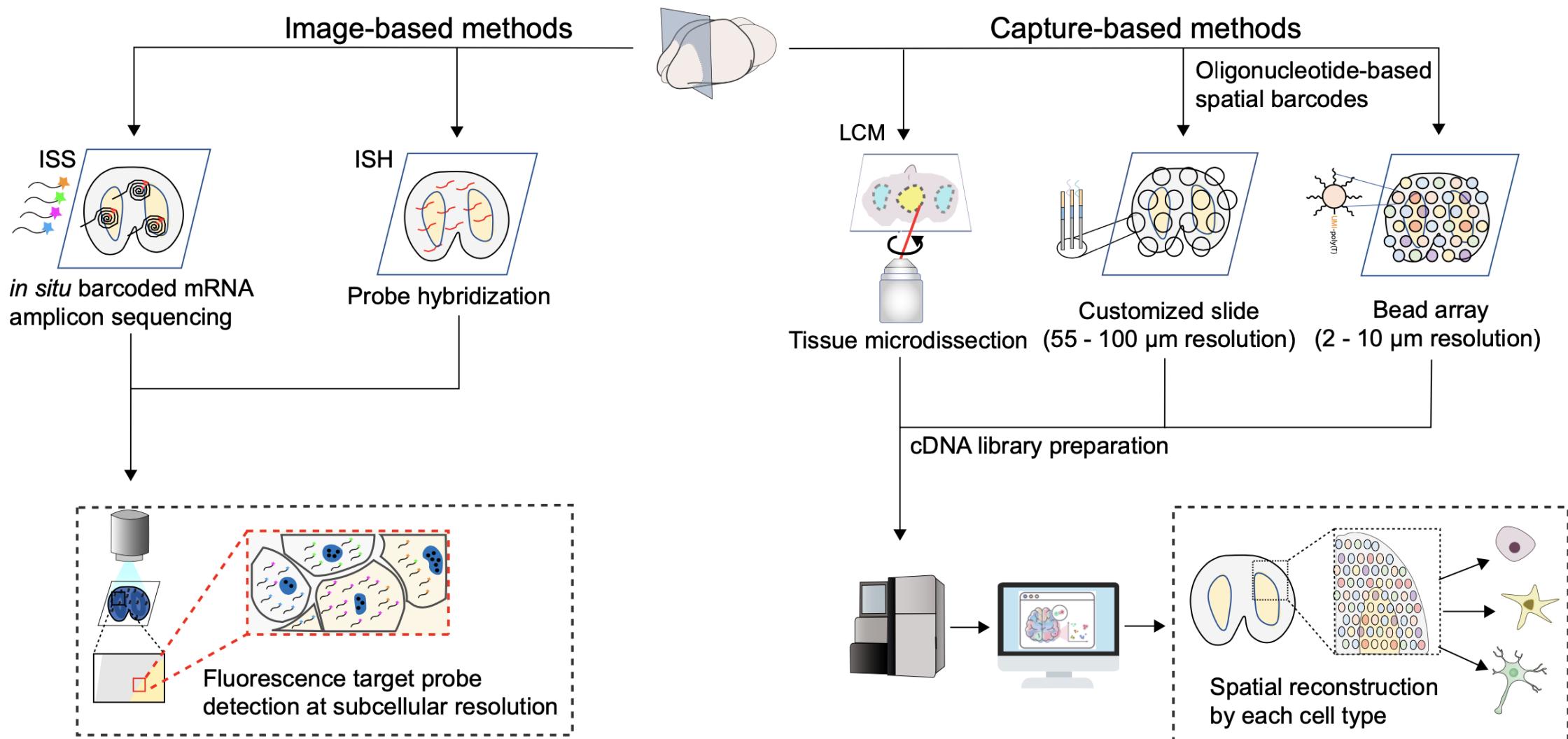
Spatially resolved transcriptomics
(SRT)

The first law of geography

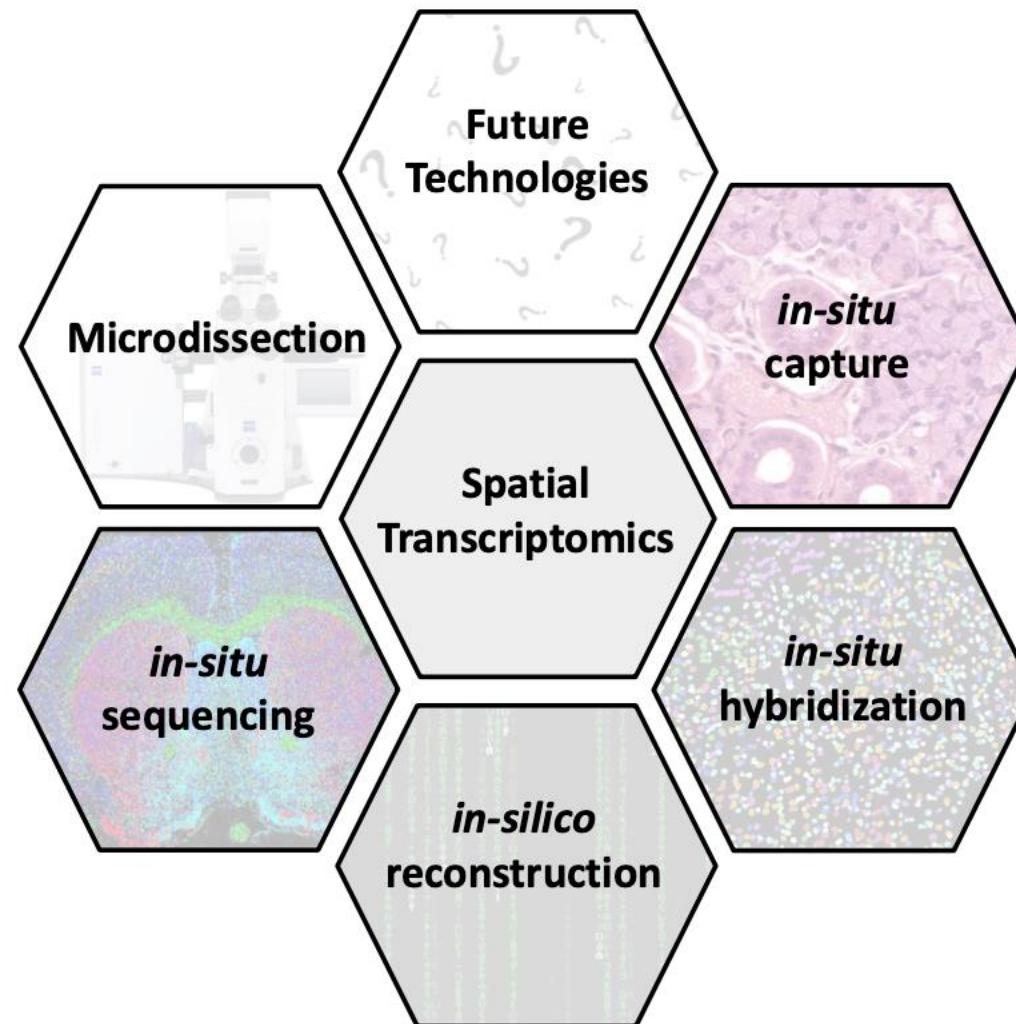
*Everything is related to everything else,
but near things are more related than distant things*

Image credit to Bo Xia

There are two types of Spatially Resolved Transcriptomics (SRT) methods

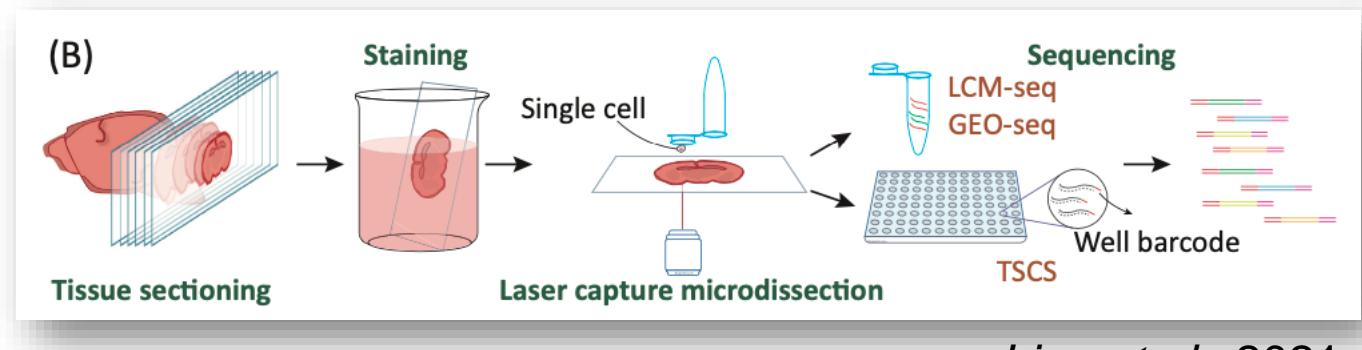
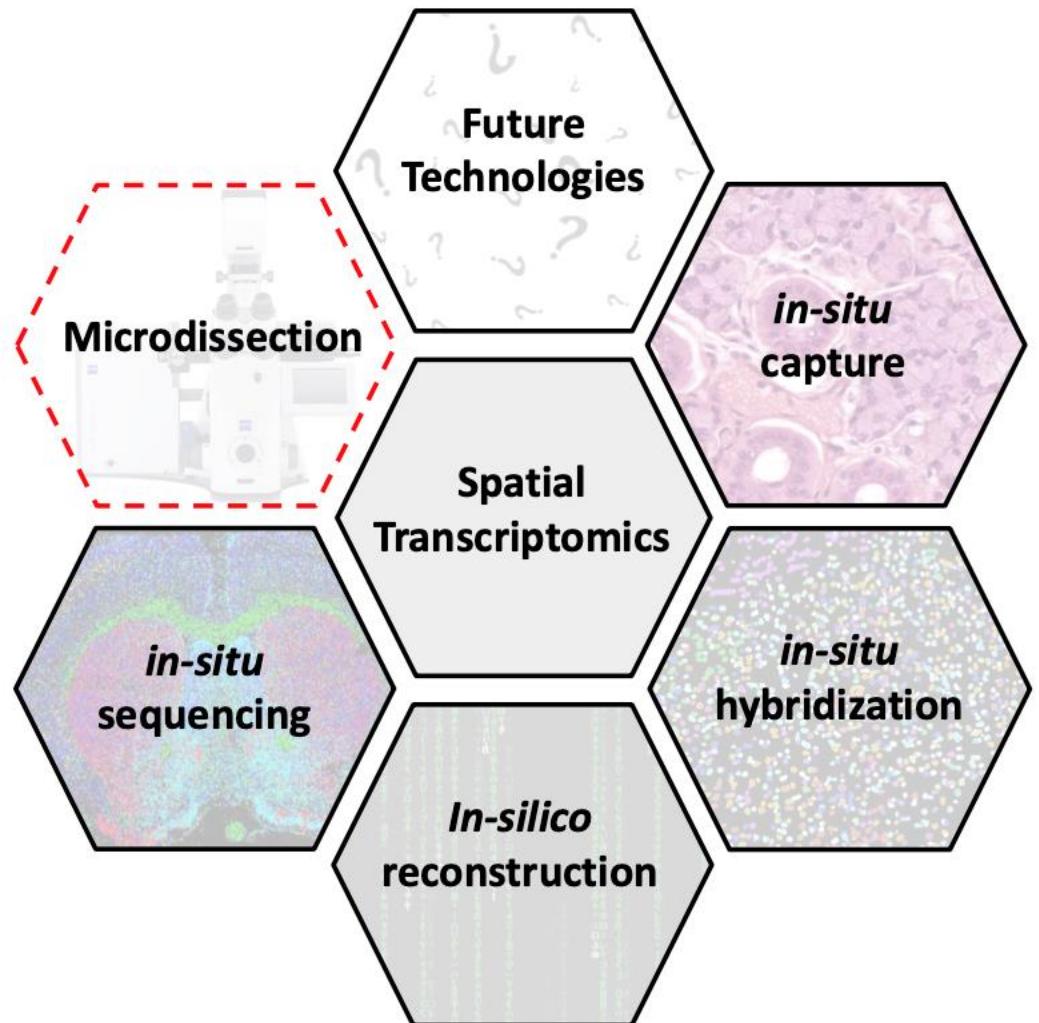


Various experimental spatially resolved techniques



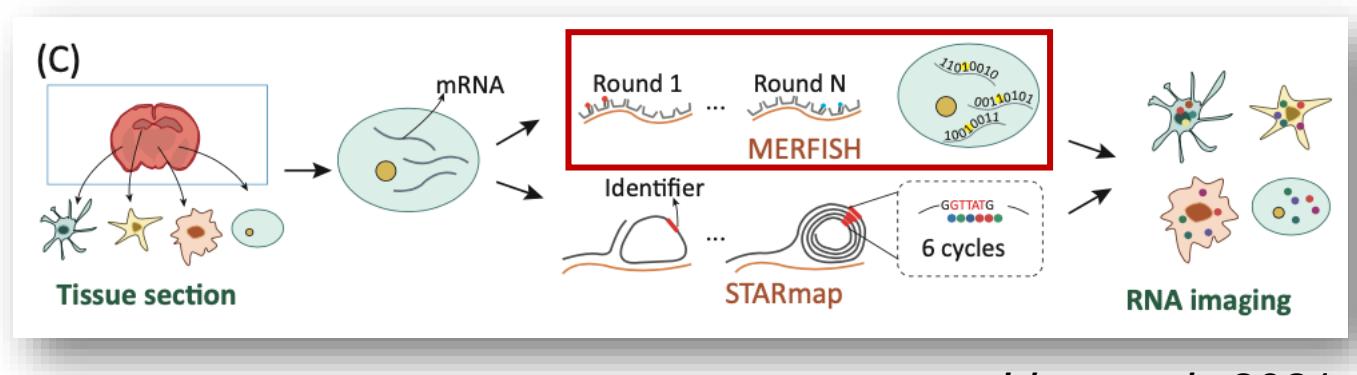
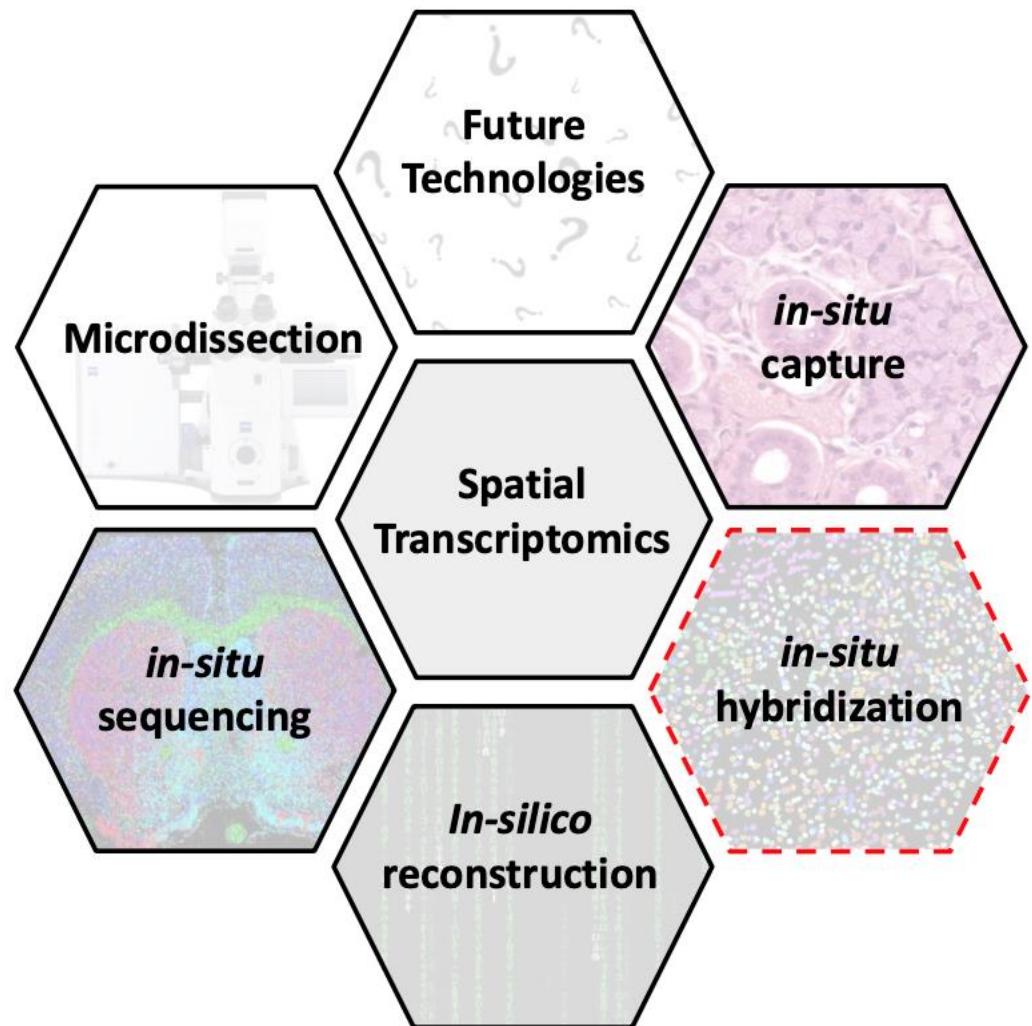
Asp *et al.*, 2020

1. Microdissection-based technology



- Dissect a region of interest (single cell), place isolate in separate wells, and sequence
- Examples: Laser Capture Microdissection (LCM), Tomo-seq, TIVA, ProximID, and Niche-seq

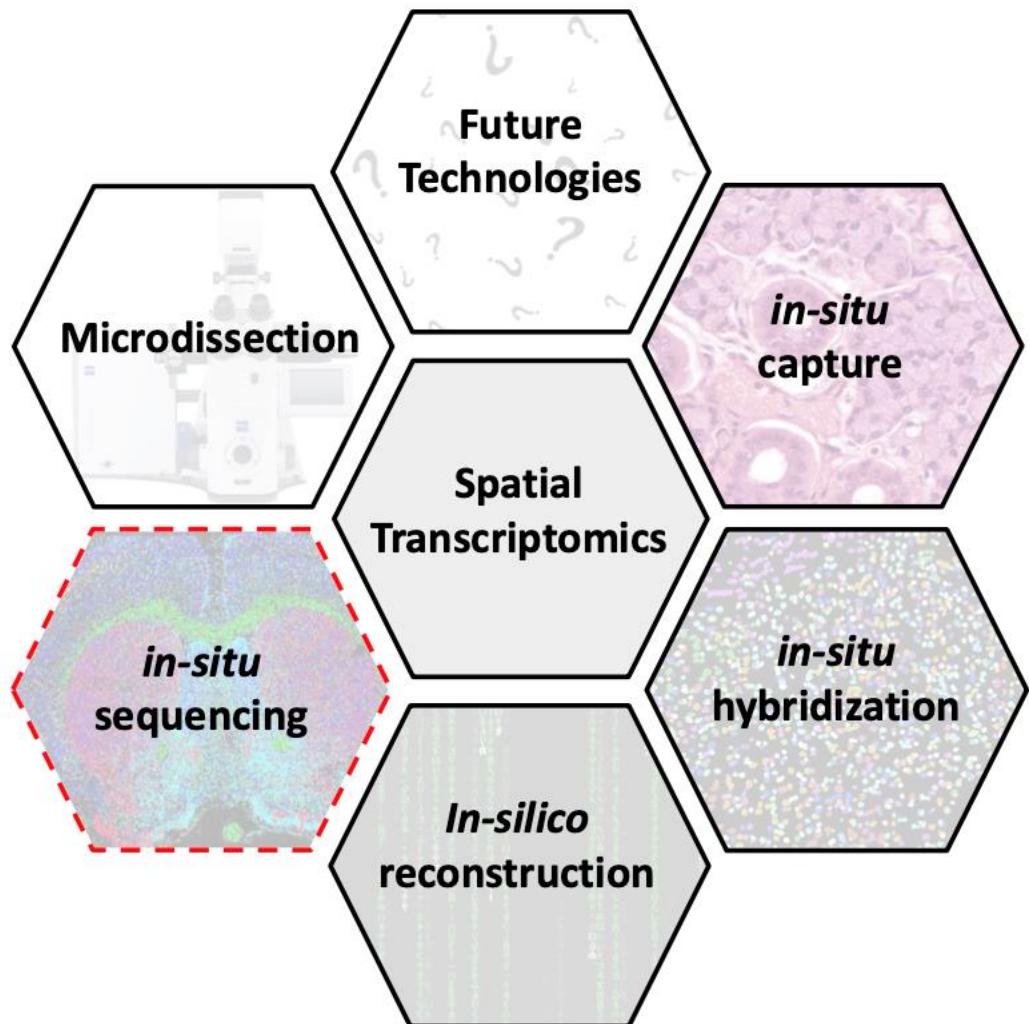
2. *in situ* hybridization-based technology



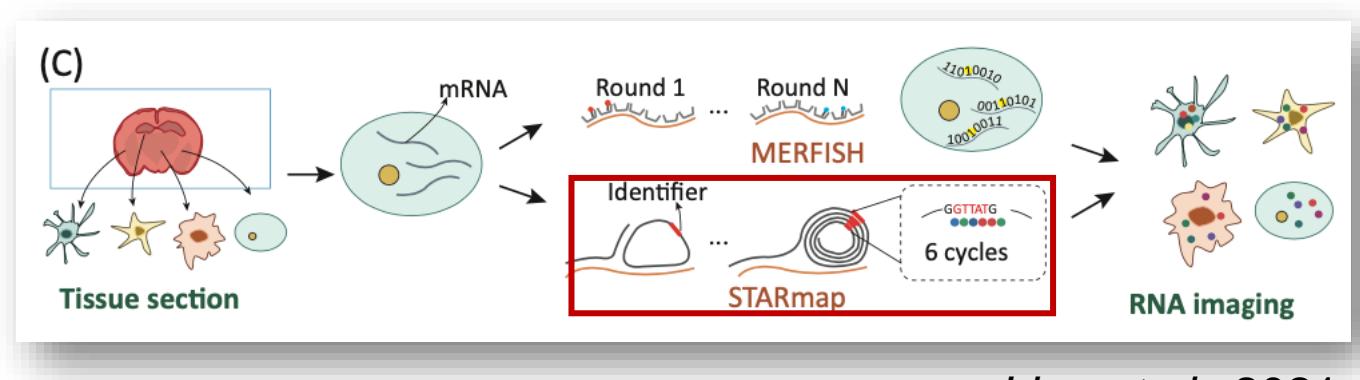
- Label probes for specific targets, hybridize in place
- Require “*a priori*” defined targets (100~1,000)
- Offer sub-cellular resolution.
- Examples) smFISH, seqFISH, **MERFISH**, seqFISH+, osmFISH, and RNA Scope etc.

* FISH – Fluorescence *In Situ* Hybridization

3. *in situ* sequencing-based technology



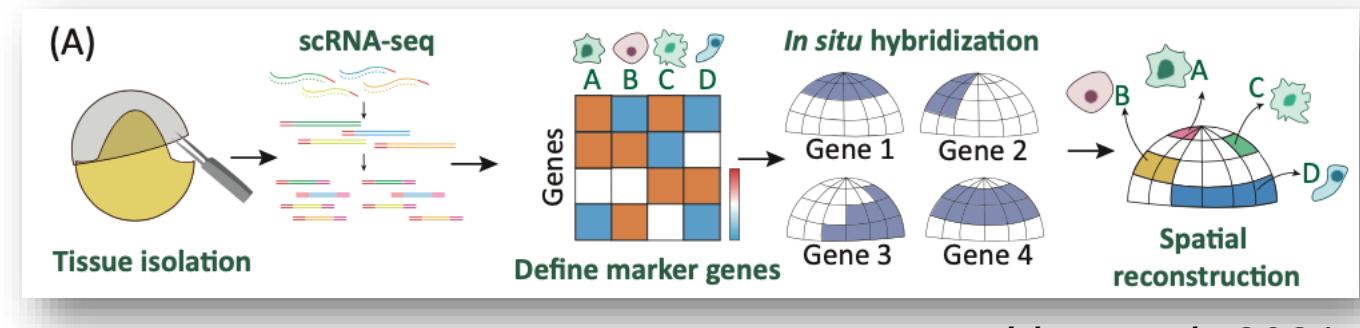
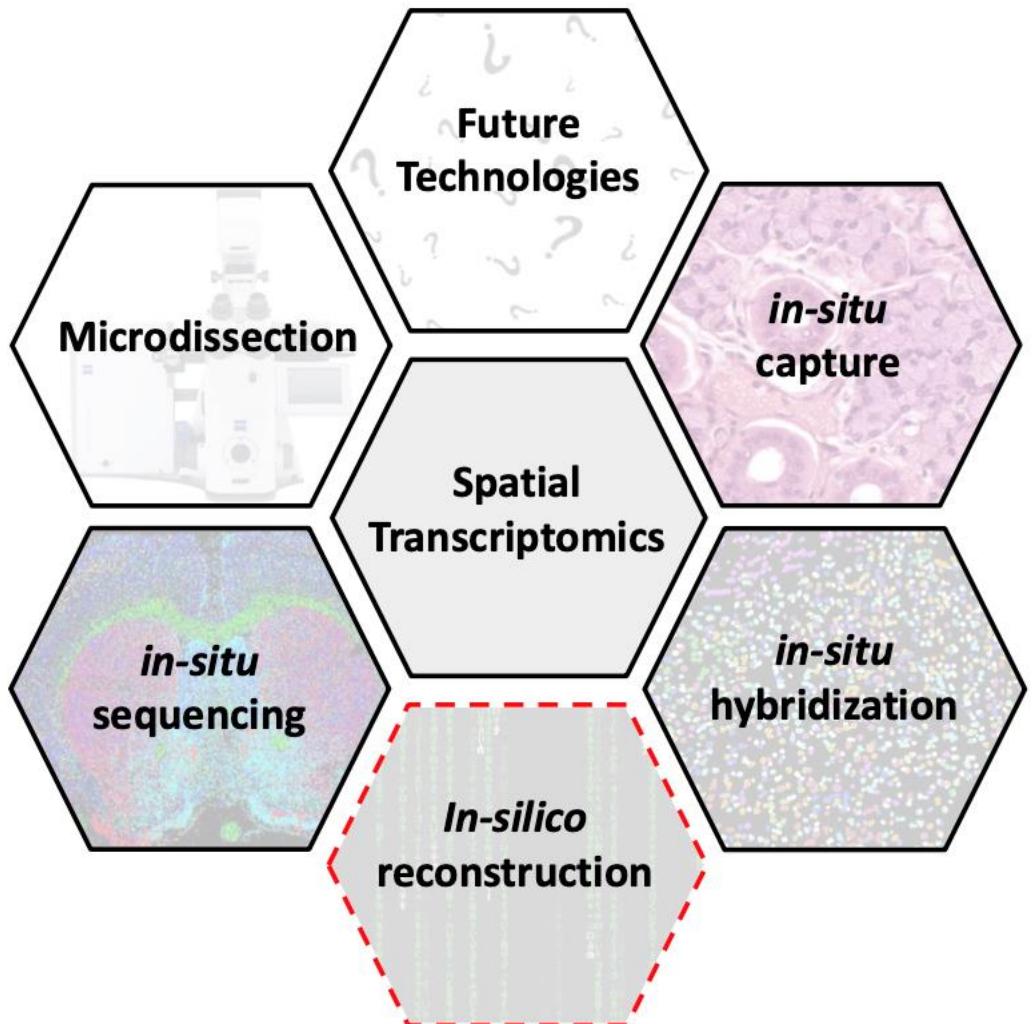
Asp *et al.*, 2020



Liao *et al.*, 2021

- Sequence the transcripts in place after amplification using a probe (e.g., SNAIL probe in STARmap)
- Offer sub-cellular resolution. Some rely on “*a priori*” defined targets
- Examples: ISS/Cartana, BaristaSeq, **STARmap**, and FISSEQ

4. *in silico* reconstruction technology

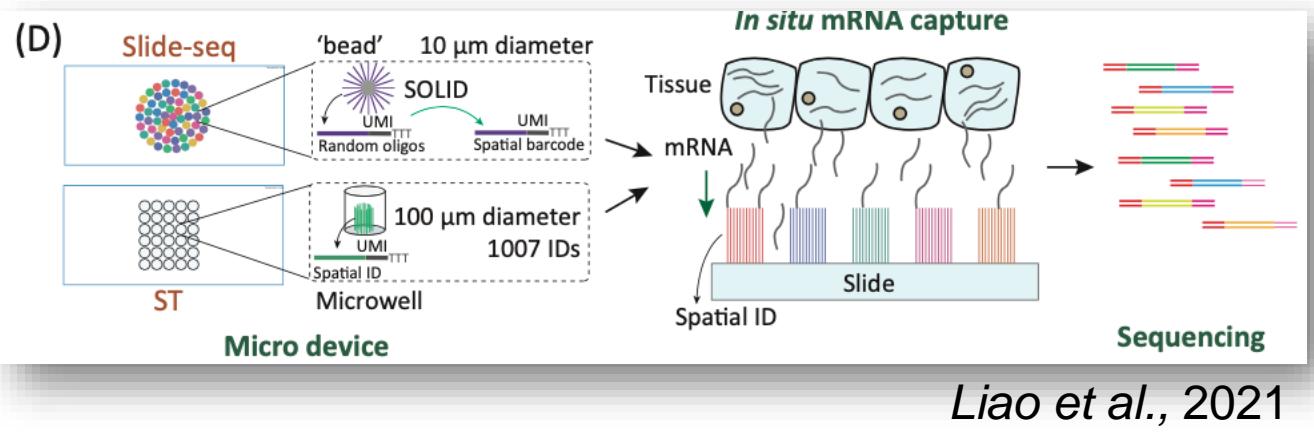
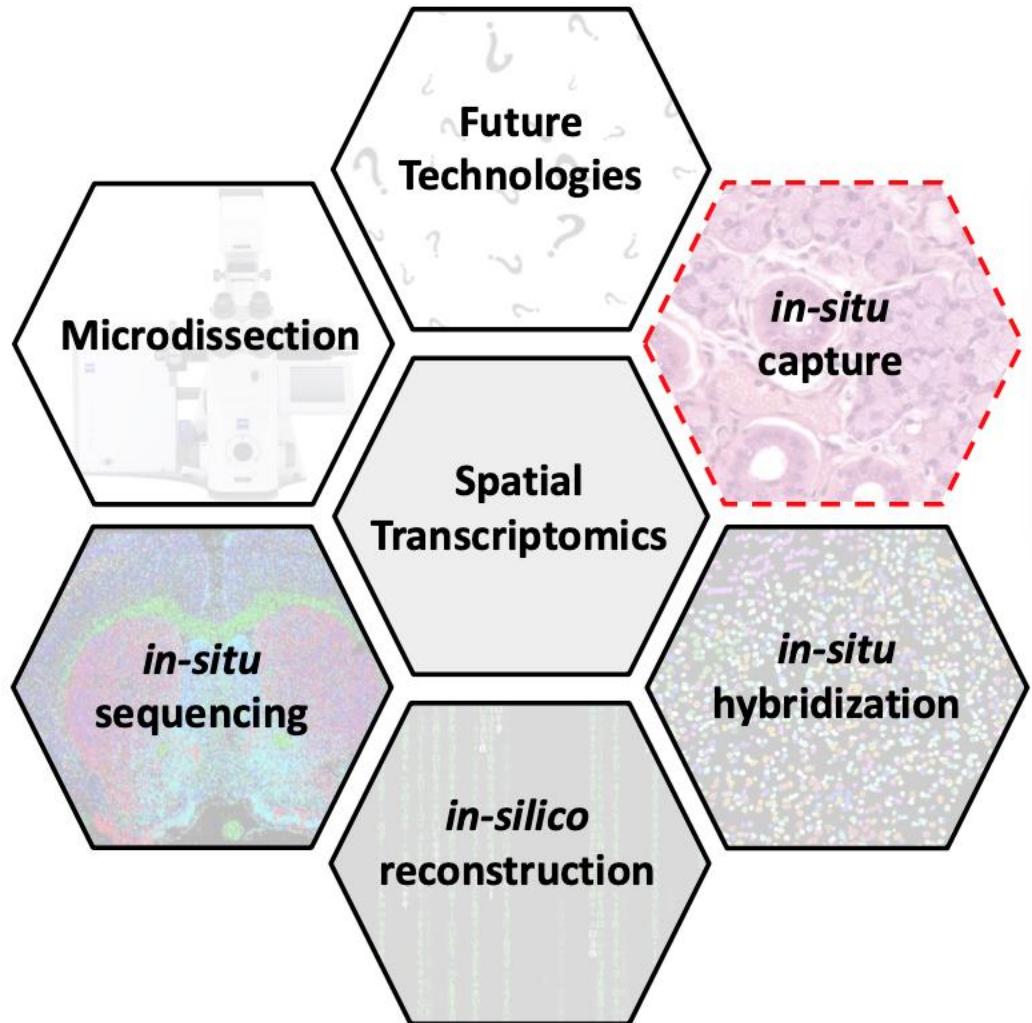


Liao et al., 2021

- Infer and reconstruct spatial structure from non-spatial data like single-cell transcriptomes
- At the core of this framework lies a structural correspondence hypothesis that cells in physical proximity share similar gene expression profiles
- Examples: novoSpaRc, CSOmap, and Seurat v3

Asp et al., 2020

5. *in situ* capture-based technologies

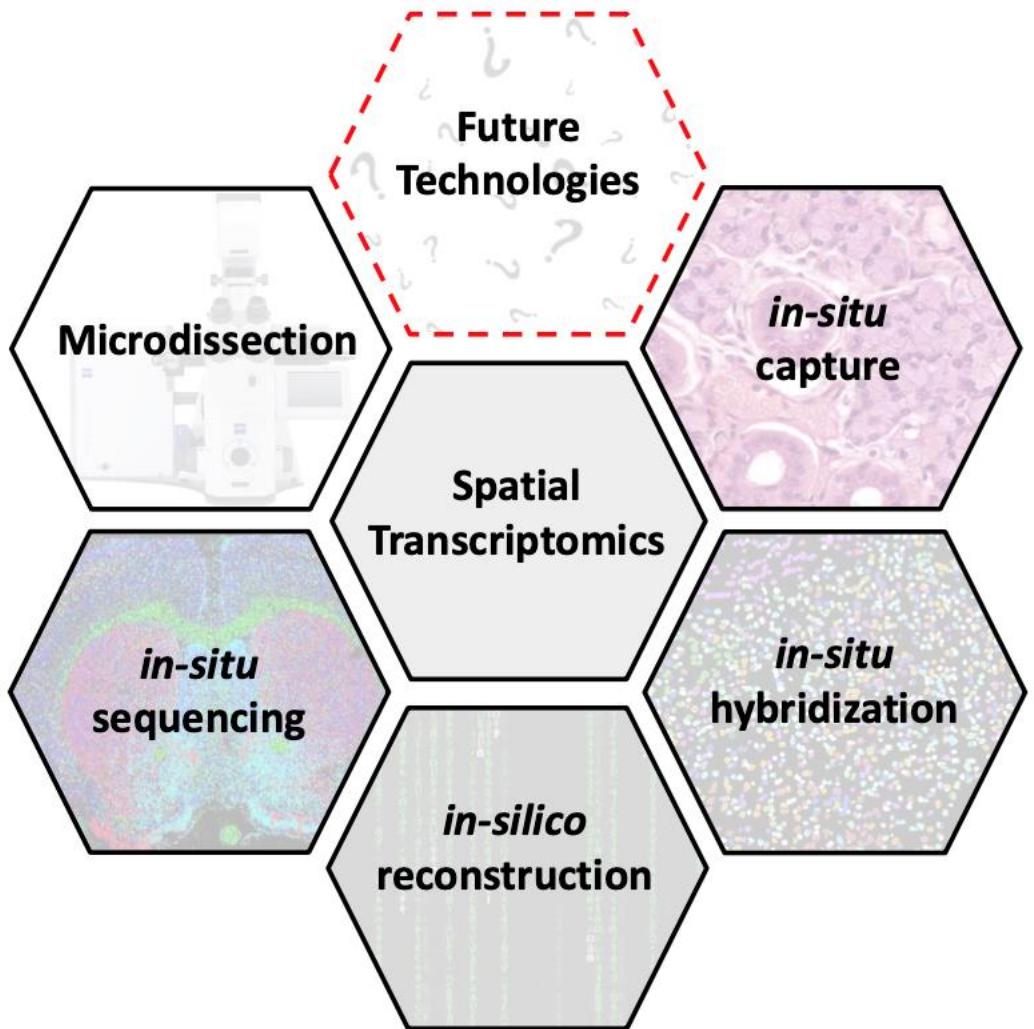


Liao et al., 2021

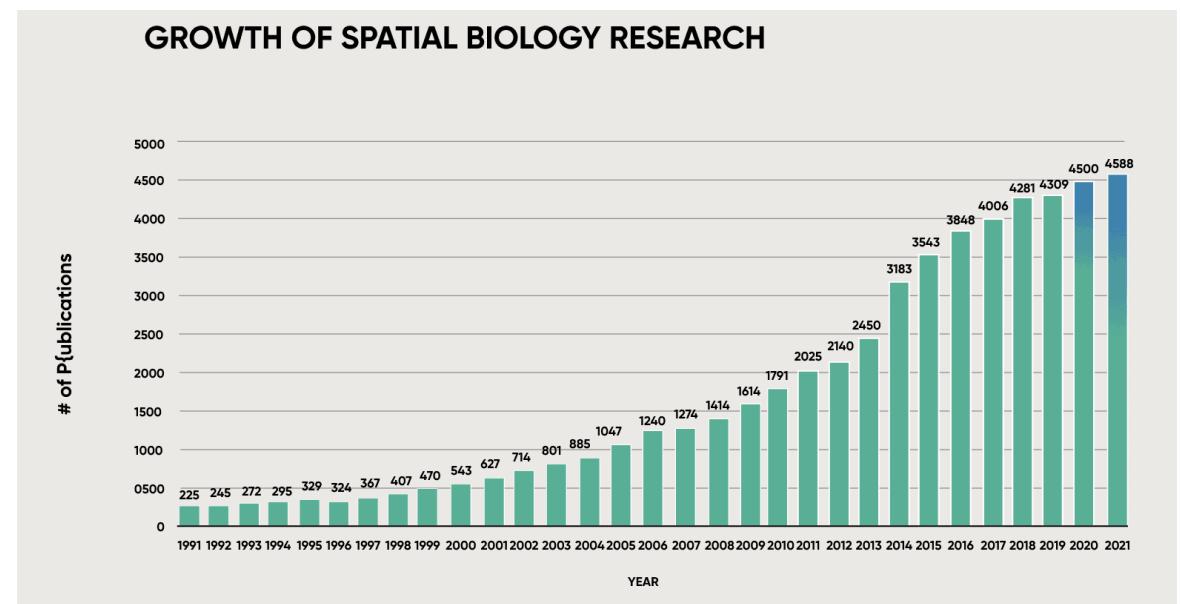
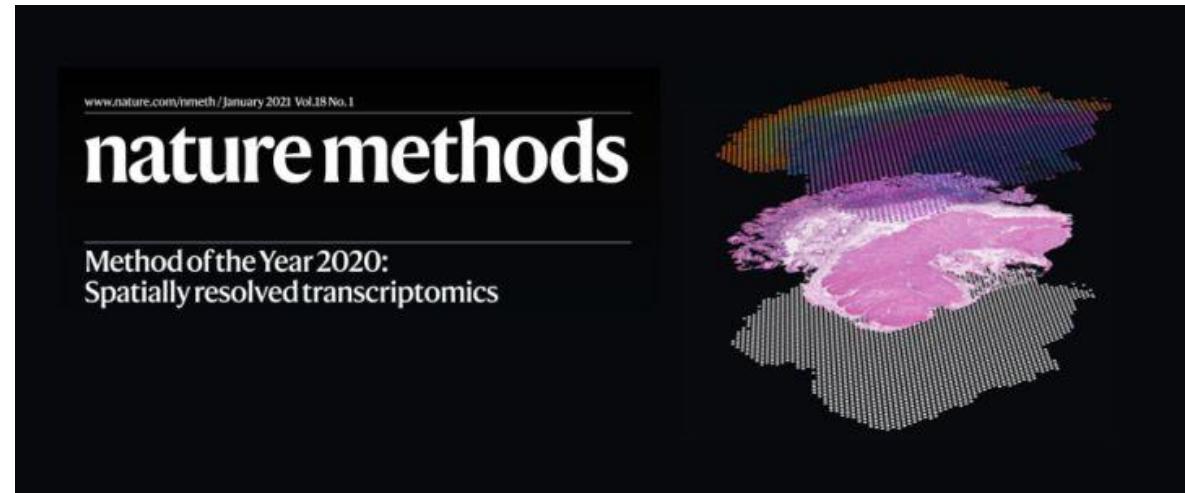
- Capture the transcripts *in situ* but sequence *ex situ*
- Usually less dependent on prior selection of targets
- Examples) Visium, **Spatial Transcriptomics (ST)**, **Slide-seq v2**, HDST, GeoMX, APEX-Seq, and Stereo-seq

Asp et al., 2020

SRT is an emerging class of high-throughput technologies and evolves rapidly



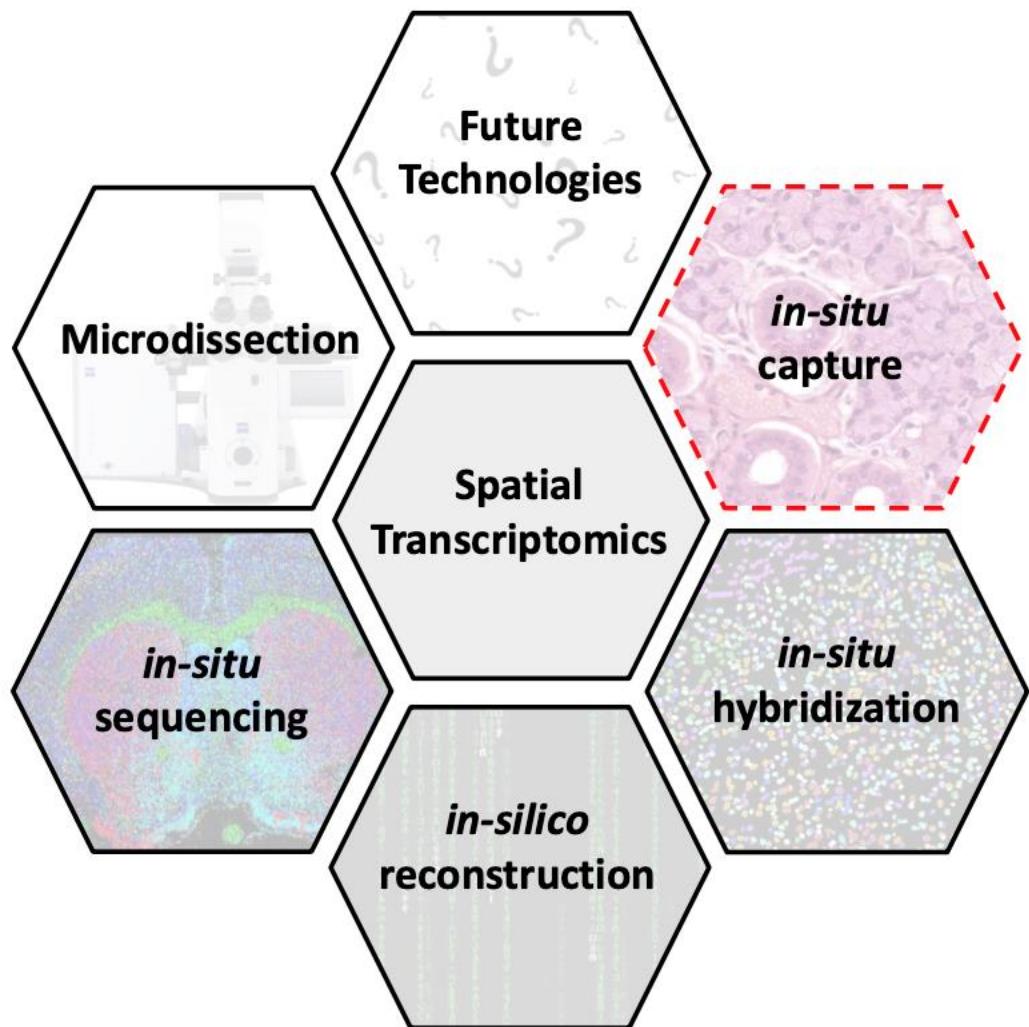
Asp et al., 2020



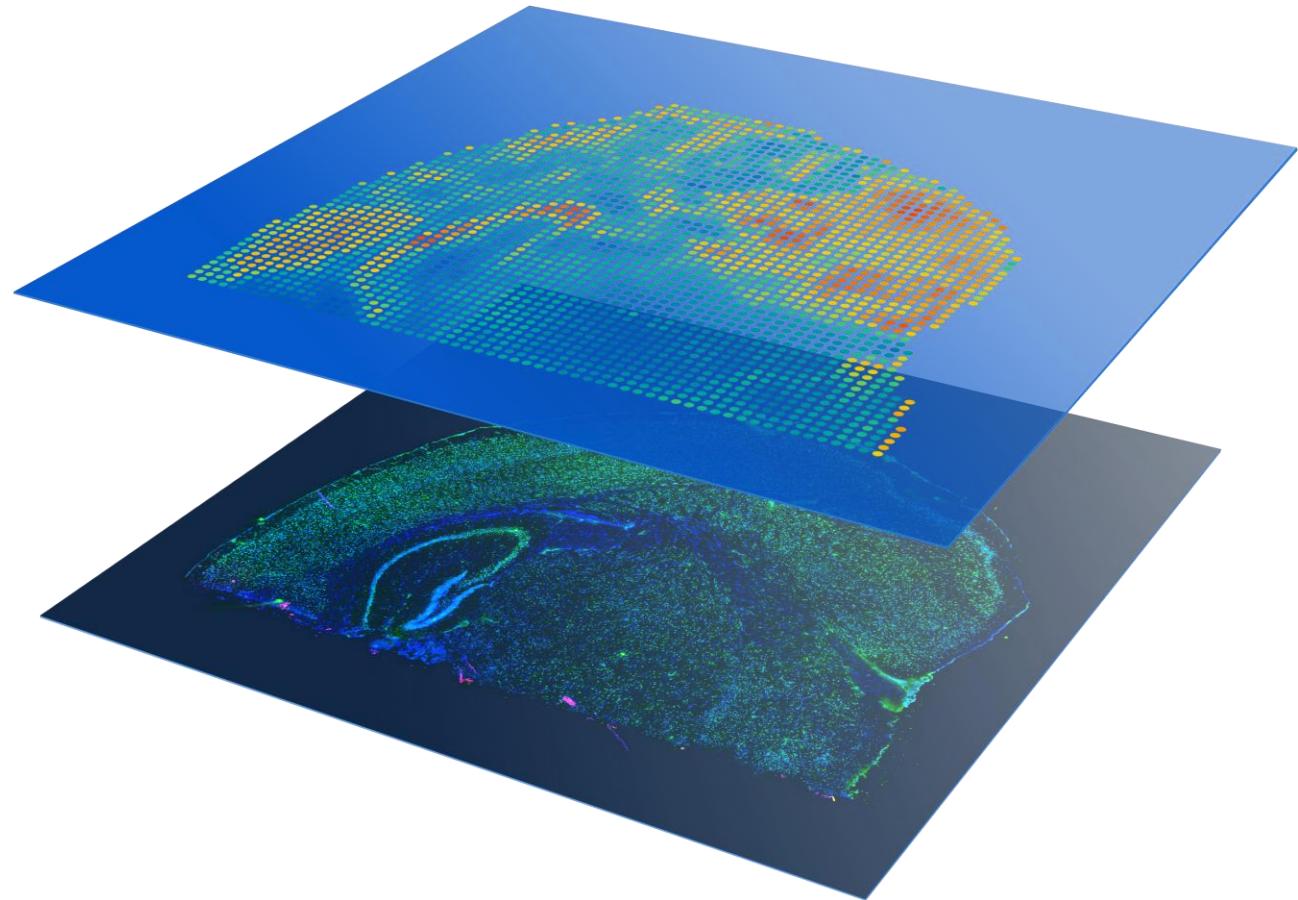
Recommended review literature on SRT

- Rao A, Barkley D, França GS, Yanai I. **Exploring tissue architecture using spatial transcriptomics.** Nature. 2021 Aug;596(7871):211-220. doi: 10.1038/s41586-021-03634-9. Epub 2021 Aug 11. PMID: 34381231; PMCID: PMC8475179.
- Longo SK, Guo MG, Ji AL, Khavari PA. **Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics.** Nat Rev Genet. 2021 Oct;22(10):627-644. doi: 10.1038/s41576-021-00370-8. Epub 2021 Jun 18. PMID: 34145435.
- Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. **An introduction to spatial transcriptomics for biomedical research.** Genome Med. 2022 Jun 27;14(1):68. doi: 10.1186/s13073-022-01075-1. PMID: 35761361; PMCID: PMC9238181.
- Moses L, Pachter L. **Museum of spatial transcriptomics.** Nat Methods. 2022 May;19(5):534-546. doi: 10.1038/s41592-022-01409-2. Epub 2022 Mar 10. Erratum in: Nat Methods. 2022 Apr 19;; PMID: 35273392.
- Lee J, Yoo M, Choi J. **Recent advances in spatially resolved transcriptomics: challenges and opportunities.** BMB Rep. 2022 Mar;55(3):113-124. doi: 10.5483/BMBRep.2022.55.3.014. PMID: 35168703; PMCID: PMC8972138.

More in detail for *in situ* capture-based SRT

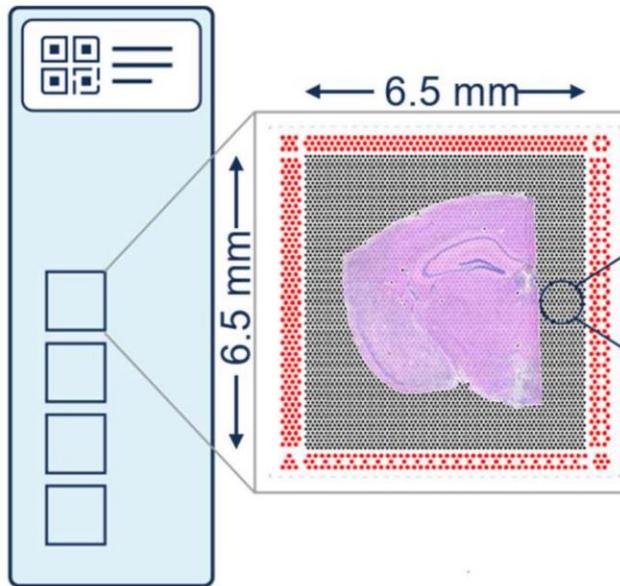


10X Visium



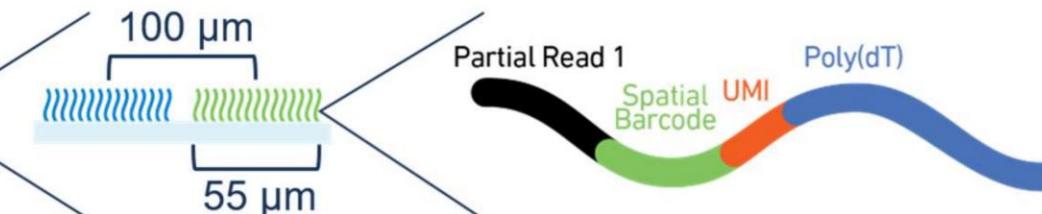
Capture-based SRT pertains spatial context :: Visium

Visium Spatial
Gene Expression
Slide



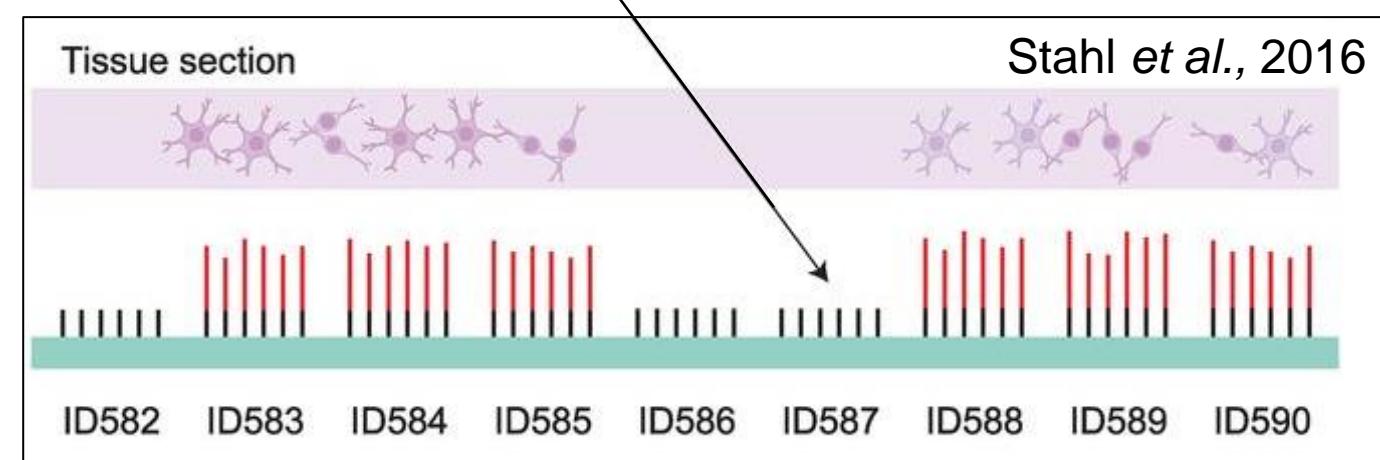
Capture Area with
~5000 Barcoded
Spots

Visium Gene
Expression Barcoded
Spots



<https://images.app.goo.gl/2k36WDczUaUcBFTYA>

1. Array-based technique (4 arrays per slide)
2. 6.5mm x 6.5mm area to put the sample on
3. 4,992 spots arranged in a hexagonal grid
4. Array specs:
 - Spot diameter: 55μm
 - Center-to-center distance: 100 μm

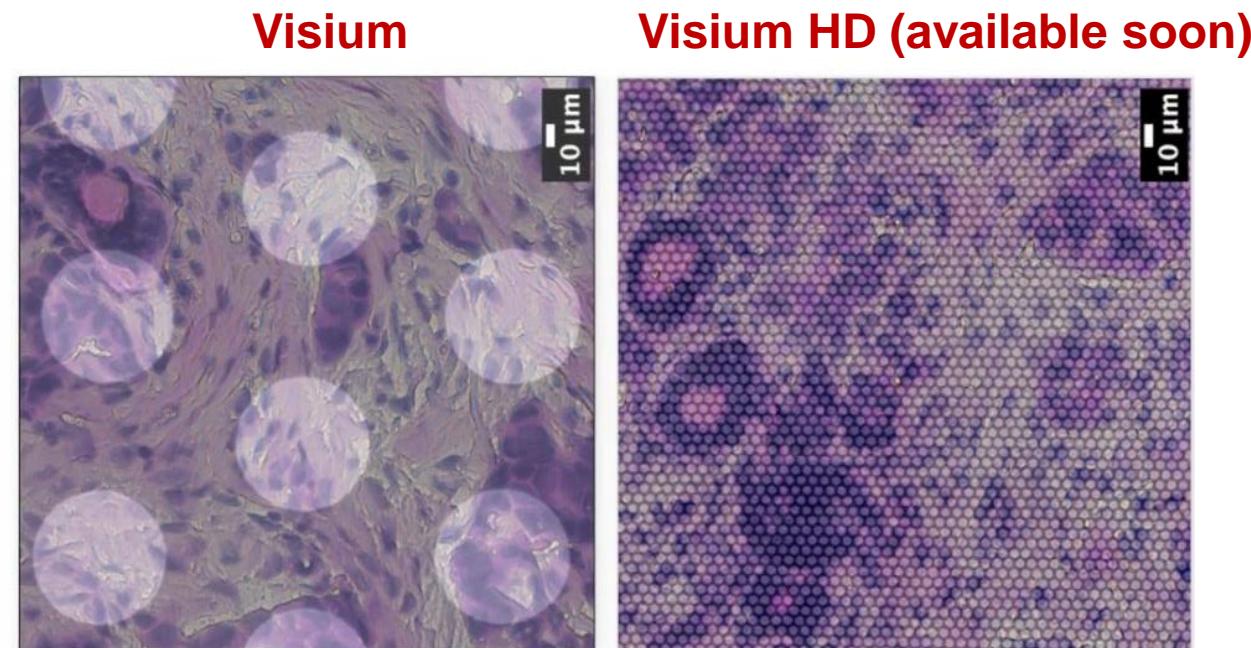


We are witnessing an evolution of Visium last decade



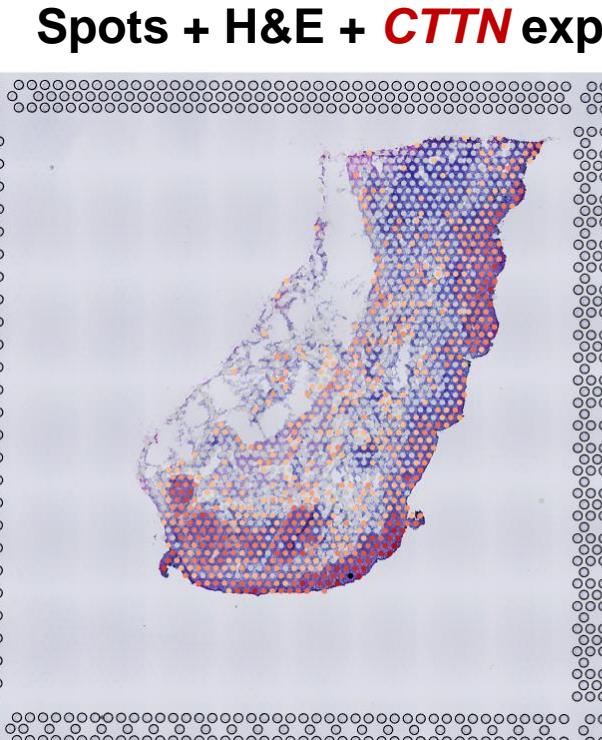
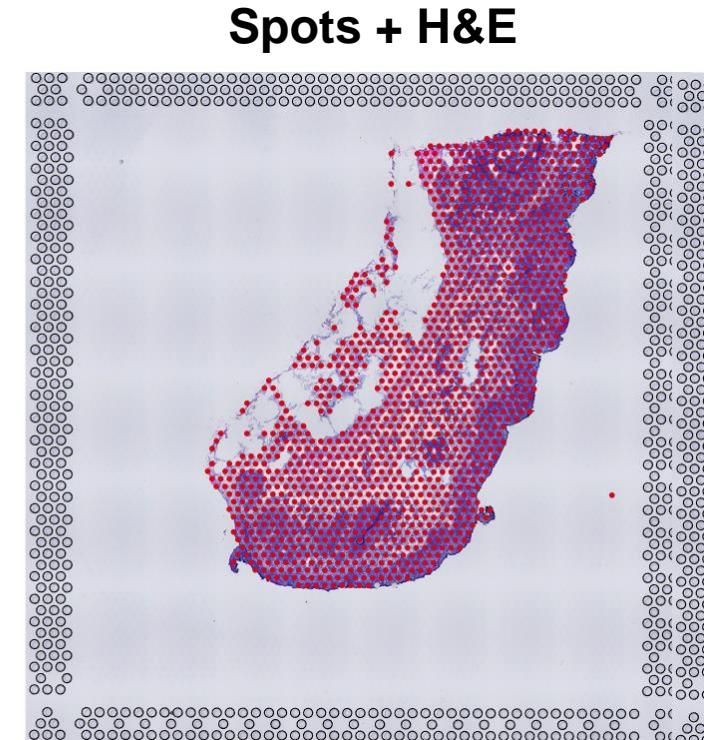
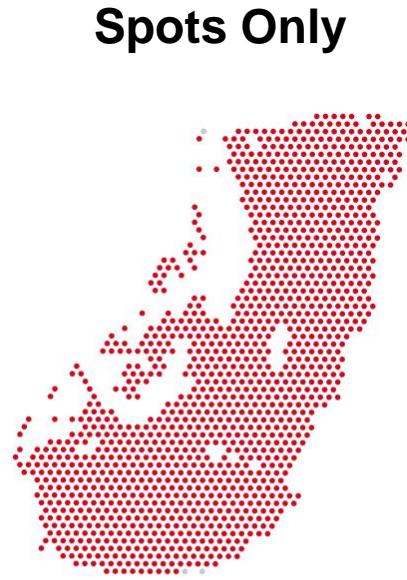
<https://images.app.goo.gl/5jH4HrQoE5VBPksZ9>

- Successor to Spatial Transcriptomics (ST)
- Approx. 1-10 cells contribute to each spot
Not a single-cell resolution!
- Data represented as [spot] x [gene] matrix
- You also get HE images of the same tissue



<https://twitter.com/AlbertVilella/status/1367028429300916225/photo/1>

Human esophageal cancer 10X Visium data example



Face color intensity proportional to
gene expression value

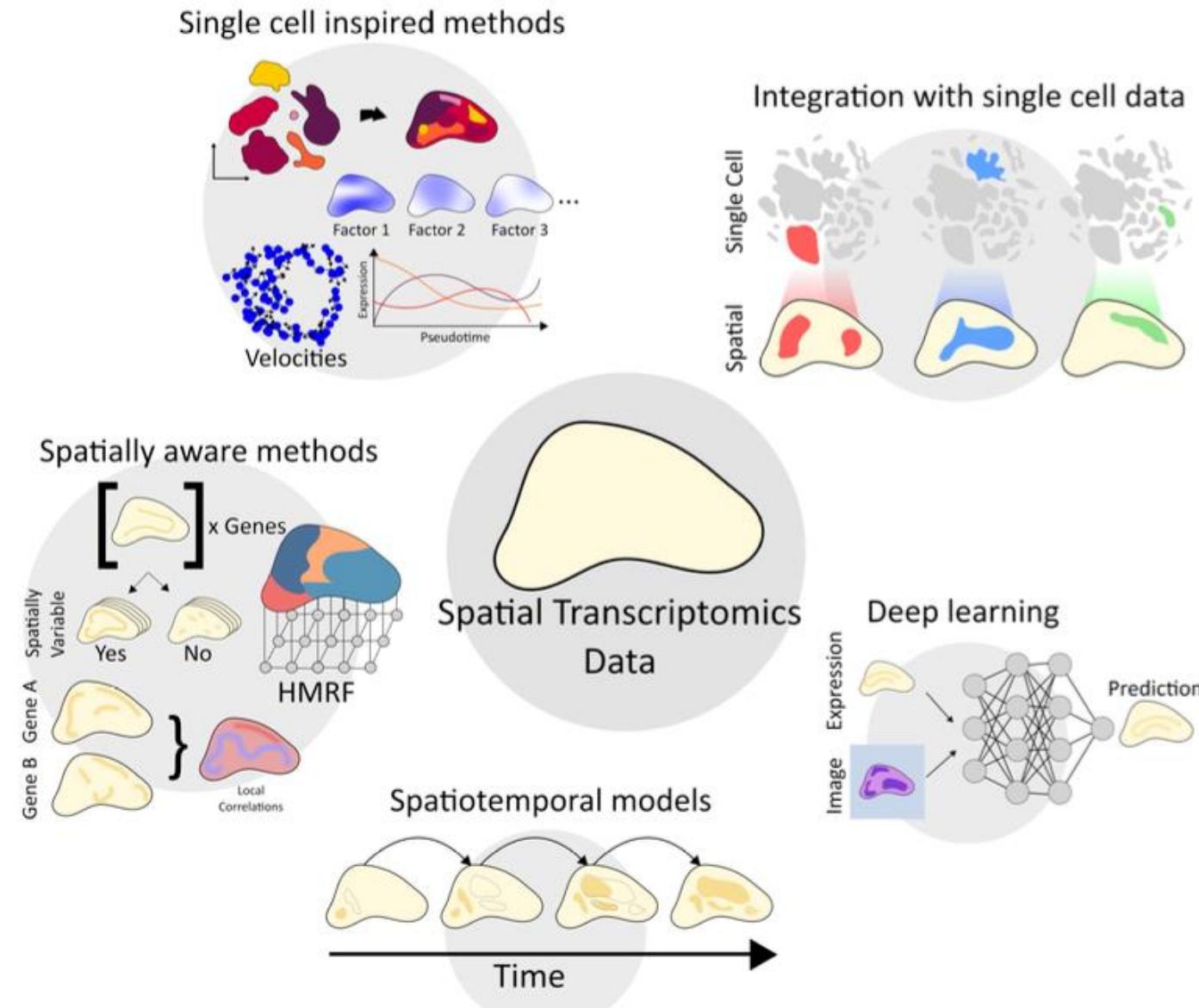
Will be performing hands-on exercise today

Computational data analysis methodologies for SRT

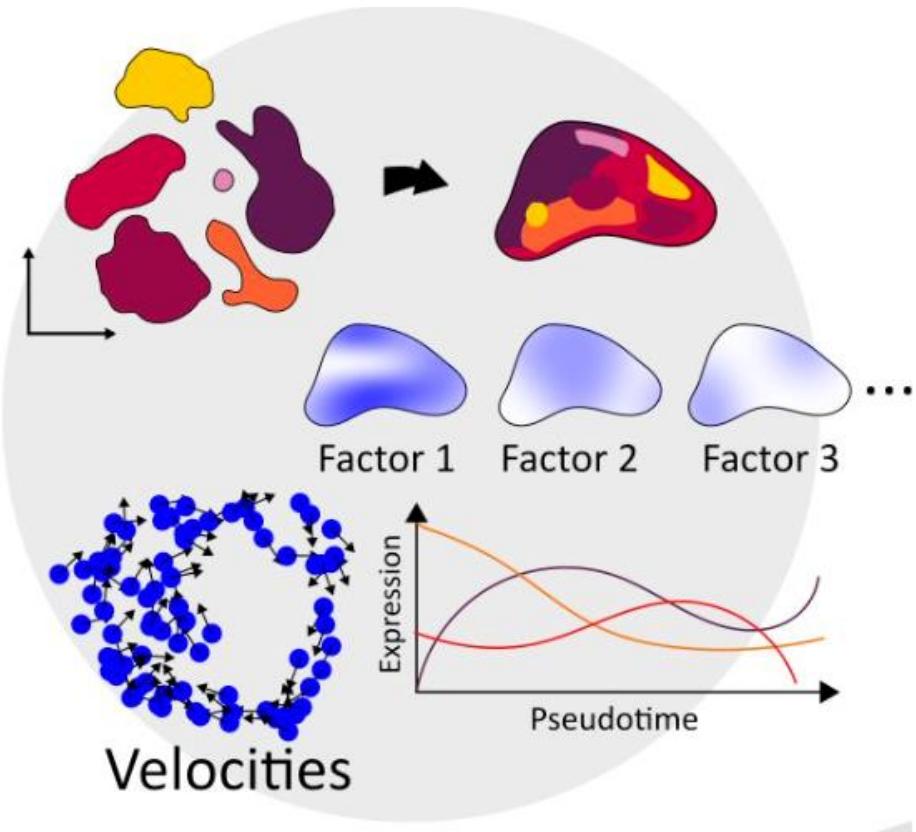
Generalized toolkits for spatial analysis

- R-based tools
 - ✓ **Seurat**
 - ✓ **STUtility** (extended spatial function for Seurat)
 - ✓ **Giotto** (greater variety of built-in tools for spatial analysis)
 - ✓ **SpatialExperiment**
- Python-based tools
 - ✓ **scanpy**
 - ✓ **squidpy** (extended spatial functions for scanpy)
 - ✓ **stLearn** (integrates spatial distance, tissue morphology and gene expression from spatial data)

Diverse technologies for SRT



1. Single cell inspired methods



Apply existing algorithms developed for single-cell data on spatial data

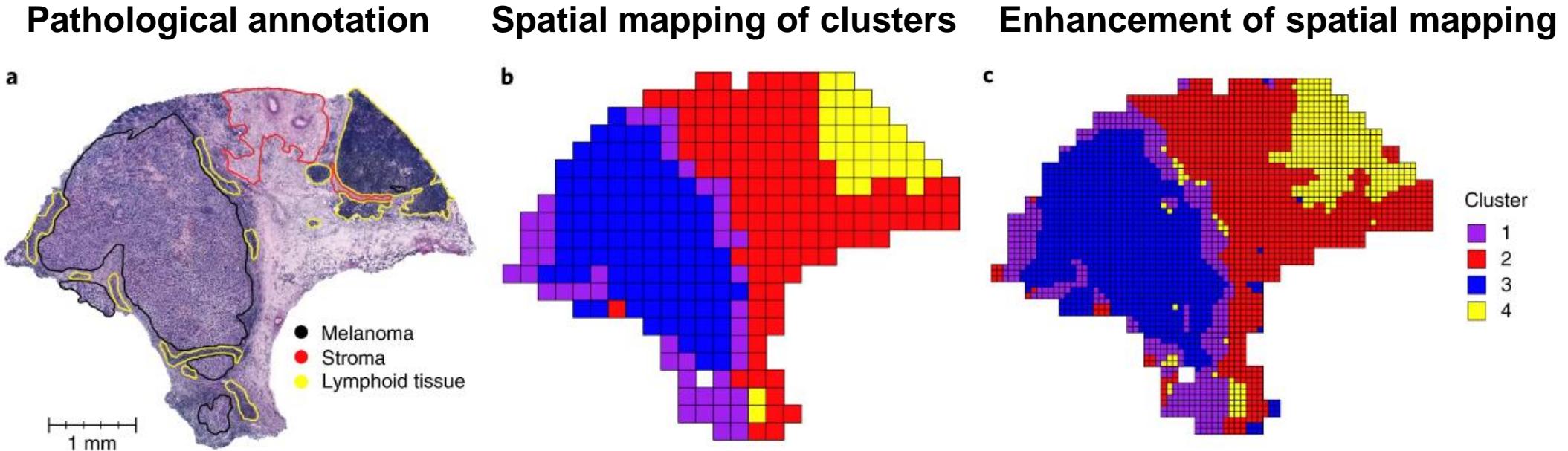
Example:

- Cluster spatial data and show clusters in space
- Factor models for data decomposition
- Trajectory inference

Available algorithms:

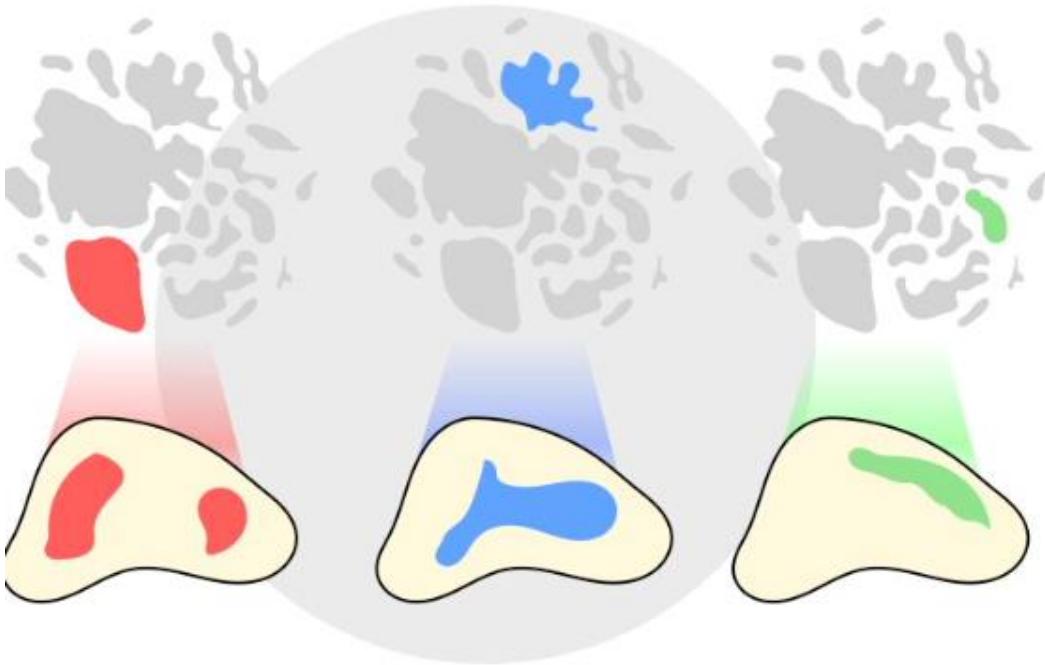
- ✓ Seurat
- ✓ scanpy
- ✓ STUtility
- ✓ stLearn
- ✓ BayesSpace
- ✓ SpatialExperiment (similar to SingleCellExperiment) etc.

1. Single cell inspired methods :: BayesSpace



Zhao *et al.*, 2021

2. Integration with single cell data



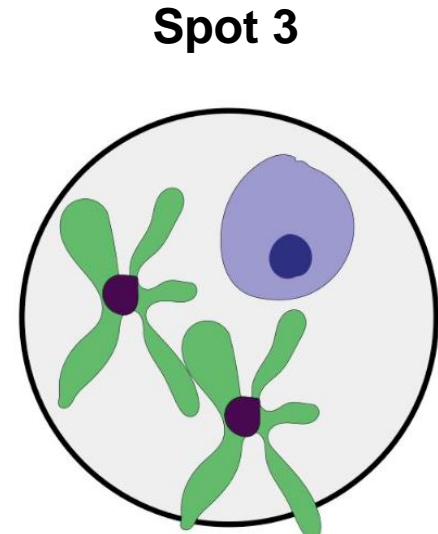
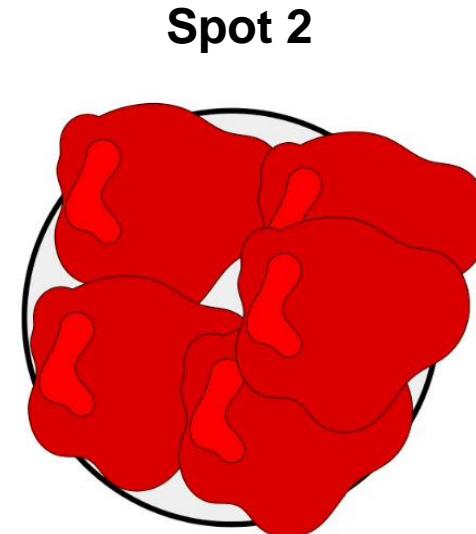
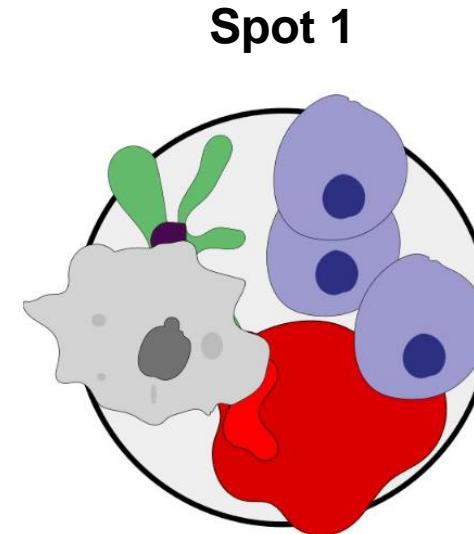
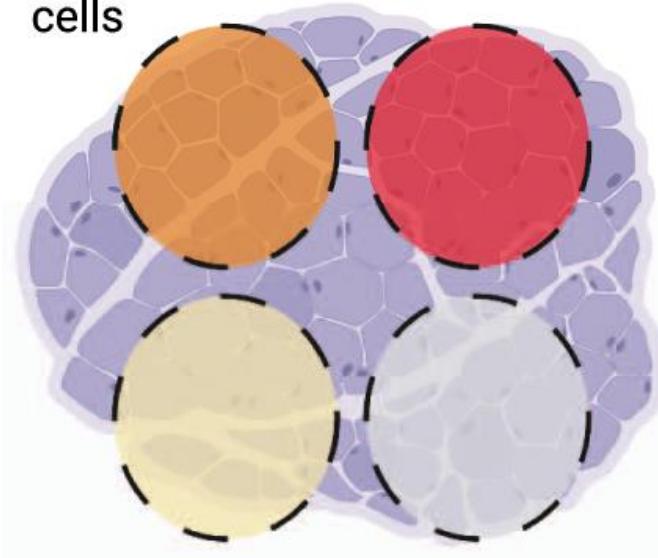
Use single-cell data as a reference when working with spatial data

Why integrate spatial data with single cells?

1. Efficient use of resources: Leverage extensive annotation already done for single-cell data
2. The problem of low resolution: Mixed cells in Visium spots

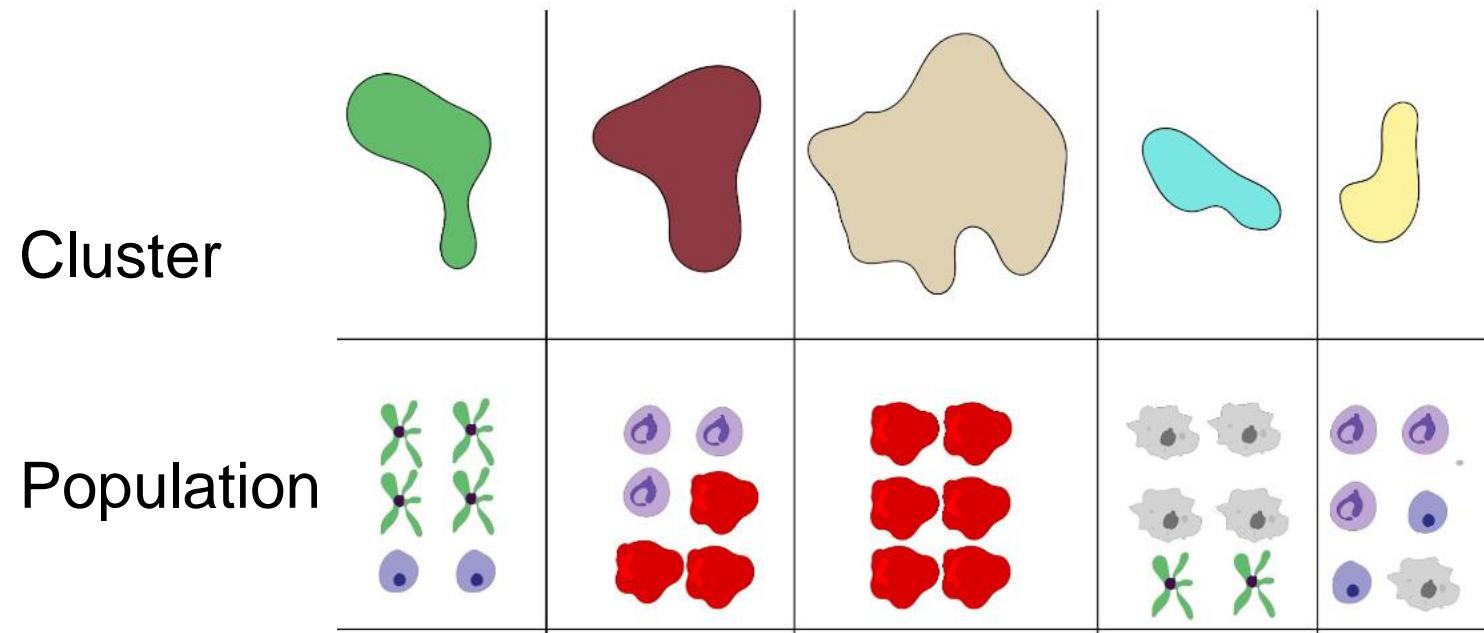
2. Integration with single cell data: mixed cell population

Each spot collects
expression among several
cells



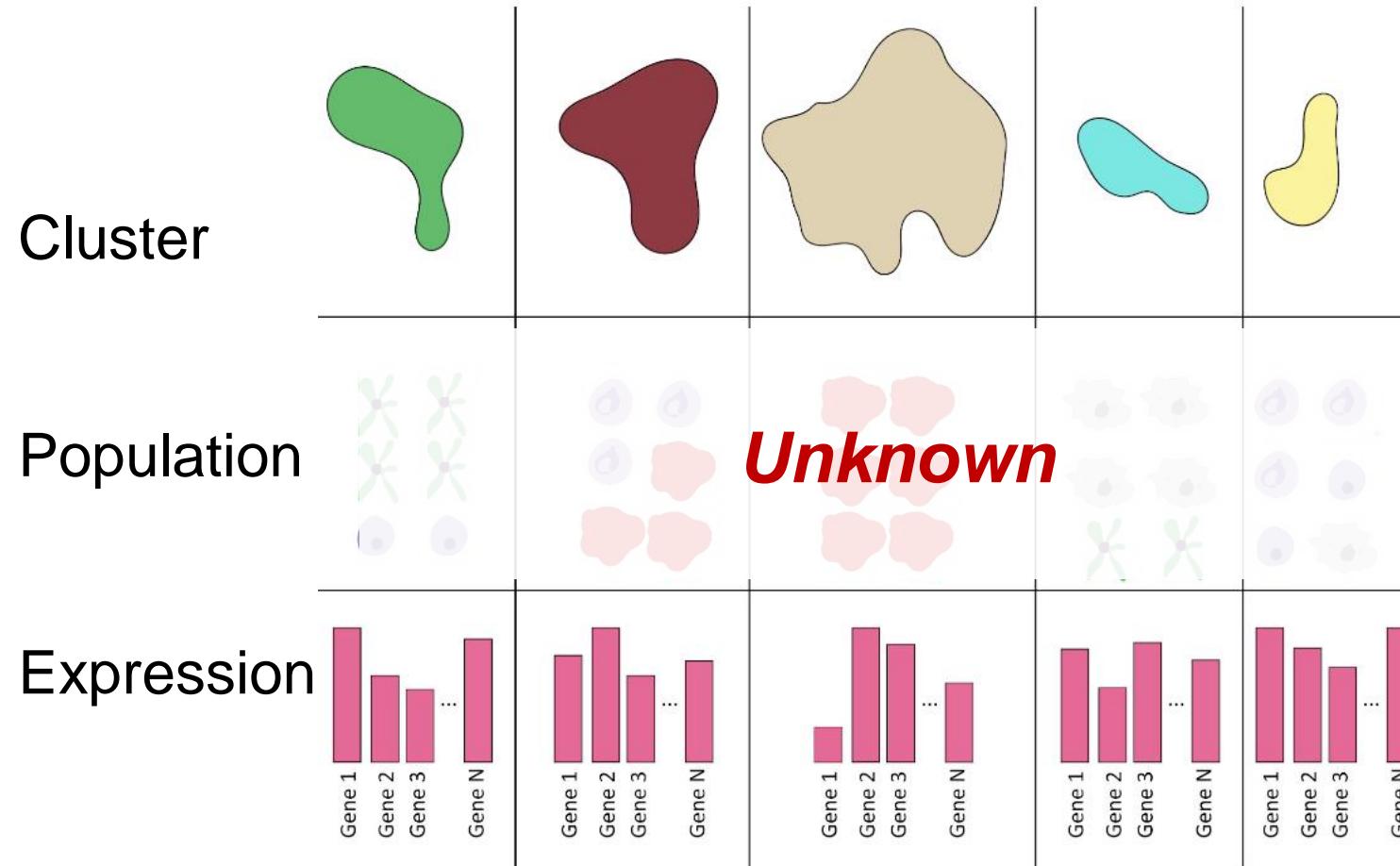
In several capture-based techniques (e.g., Visium and Slide-seq), observed expression values are a combination of multiple cell types and not all necessarily the same cell type.

2. Clusters do not represent single-cell type



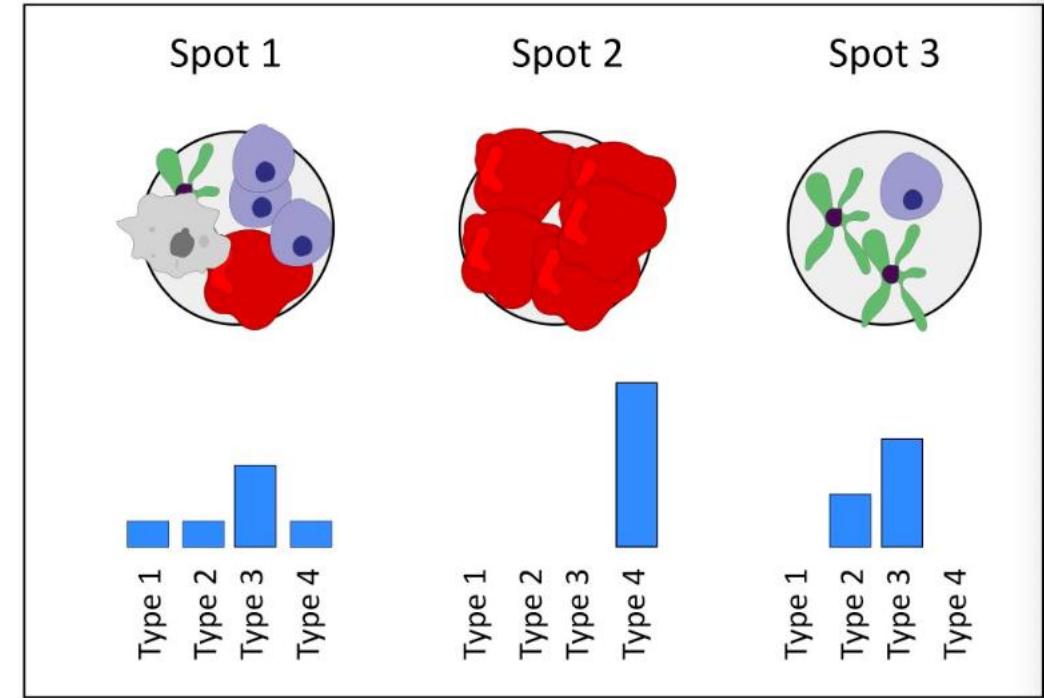
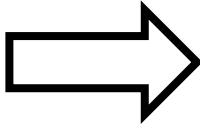
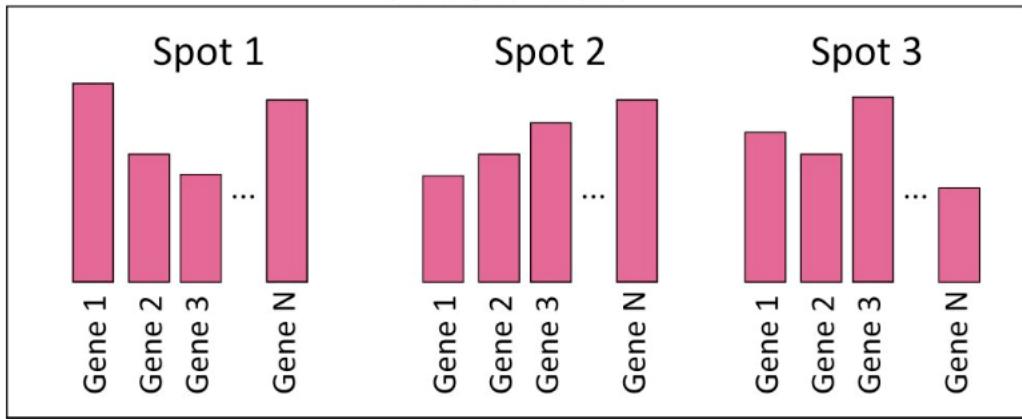
- Rather, clusters are more an assembly of spots with a similar composition of cell types
- We do not know what the cell type population looks like

We do not know what the cell type population look like



- Clusters do not represent single-cell type
- Rather, clusters are more an assembly of spots with a similar composition of cell types
- We only have observed expression

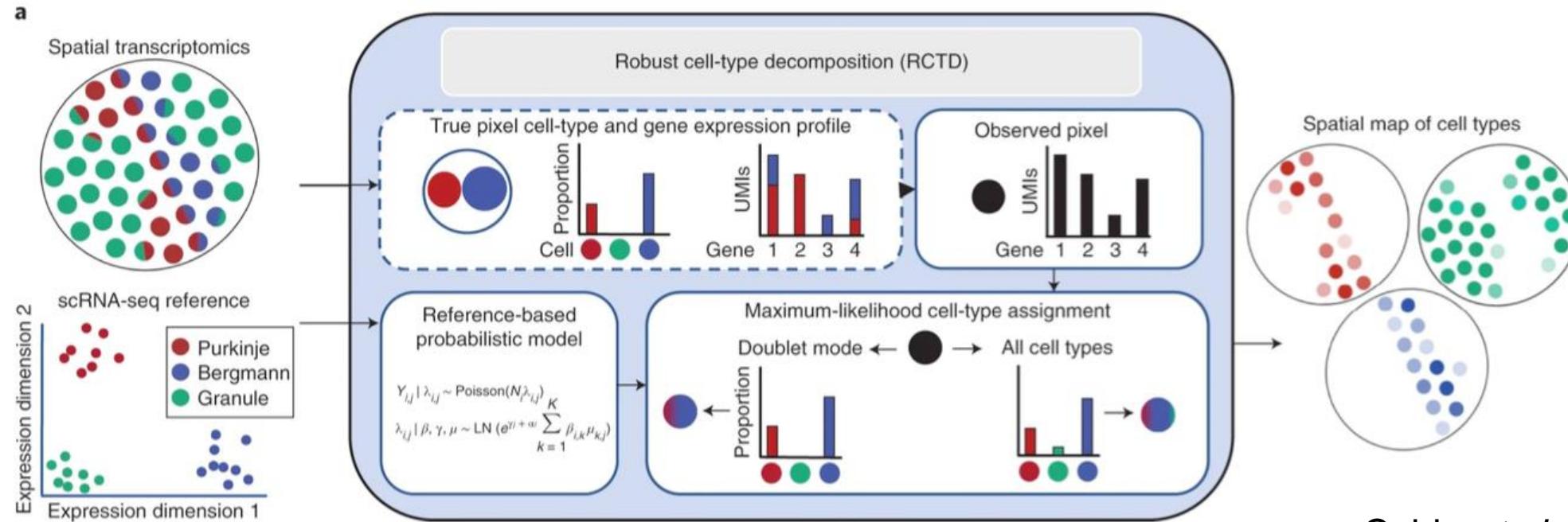
Estimate cell type abundance from observed gene expression profile



2. Different methods for integration with single cells

Marker gene	Anchor	Probabilistic Model	Optimization
Extract marker genes for each cell type from single cell data	Find anchors between single cell and spatial data. Create correction vectors based on expression differences	Assume gene expression follows a certain statistical distribution. Joint model for single cell and spatial data.	Find spatial location where each cell most likely to reside
Compute enrichment score for each set of marker genes in spatial locations	Use correction vectors to integrate two data sets. Transfer labels of single cells to spatial data	Learn cell type parameters from single cell data and use them to deconvolve spatial data	Simultaneously optimize terms such as: <ul style="list-style-type: none">• Cell density• UMI distribution• Gene distribution
Moncada <i>et al.</i> , 2020	Seurat	Stereoscope, RCTD, cell2location	Tangram

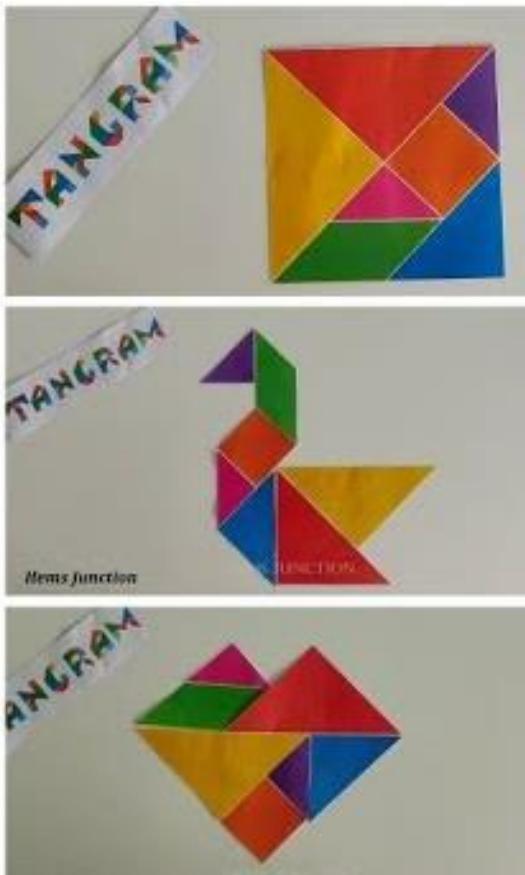
2-1. Integration with single cell data :: RCTD



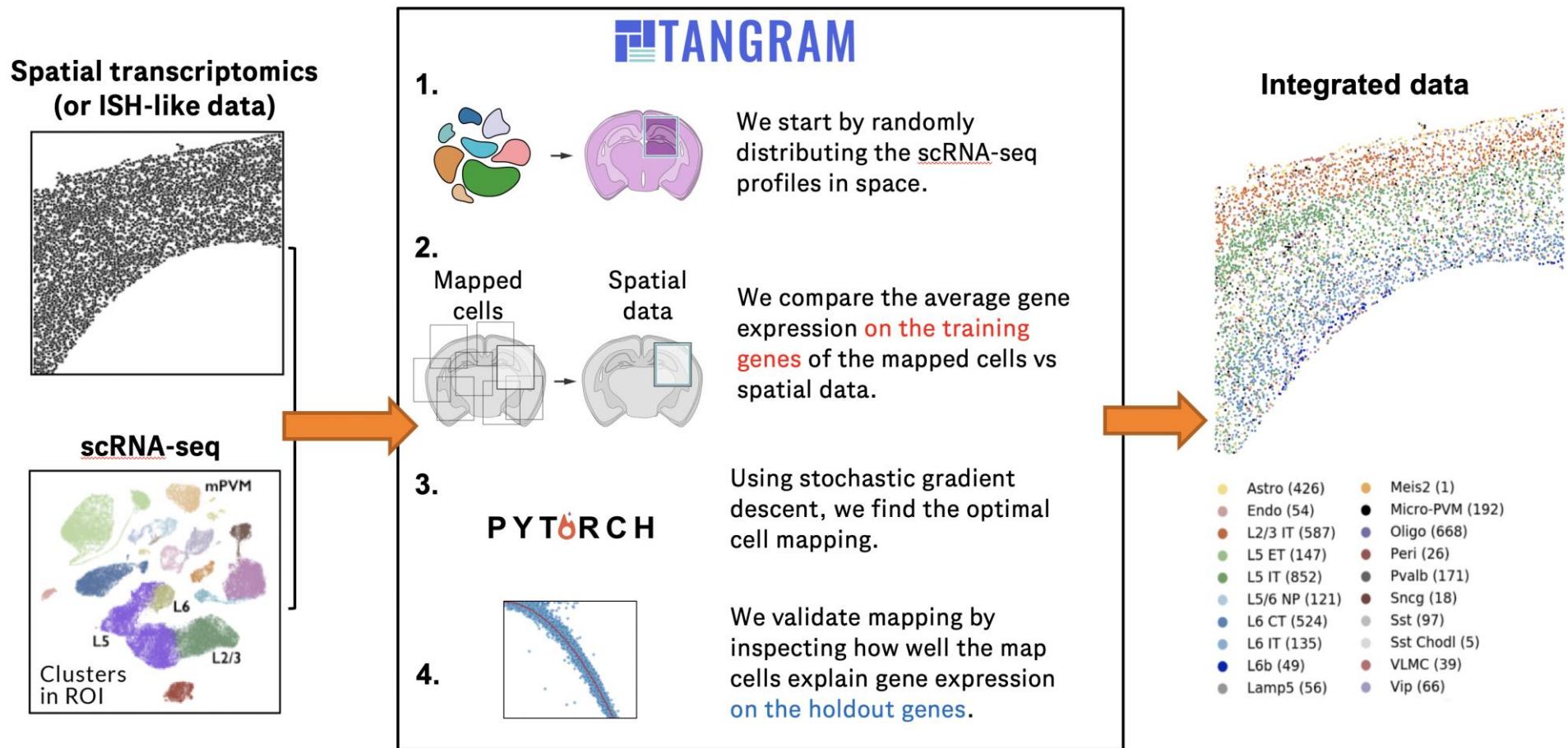
Cable *et al.*, 2022

- A probabilistic model for inferring cell types in spatial transcriptomics data, supervised with a labeled single-cell RNA-seq reference
- Robust cell-type decomposition (RCTD) uses **maximum likelihood estimation** to identify cell types present on each spatial transcriptomics spot, in addition to estimating cell type proportions
- Robust decomposition of cell type mixture in spatial transcriptomics

2-2. Integration with single cell data :: Tangram



<https://images.app.goo.gl/grbf8WGbls6wZcxS6>

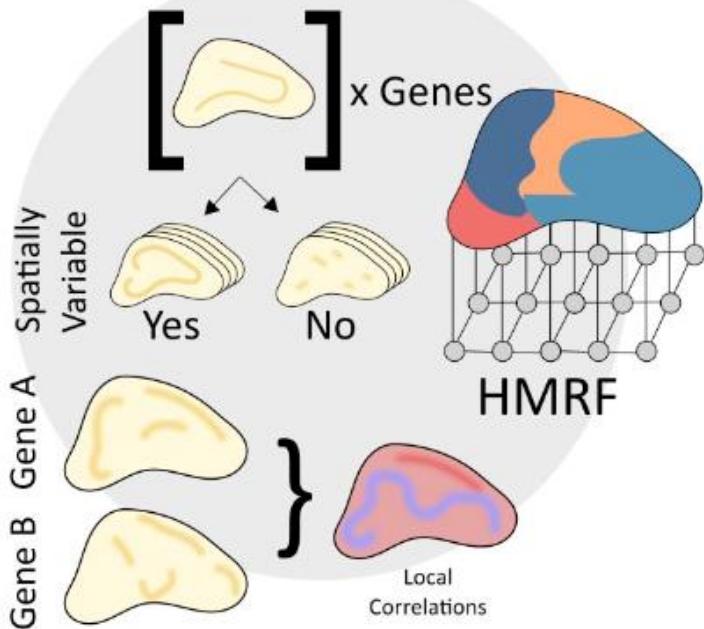


Biancalani et al., 2022

Tangram integrates scRNA-seq and SRT data via maximization of a spatial correlation function using stochastic gradient descent (SGD) optimization

3. Spatially aware methods

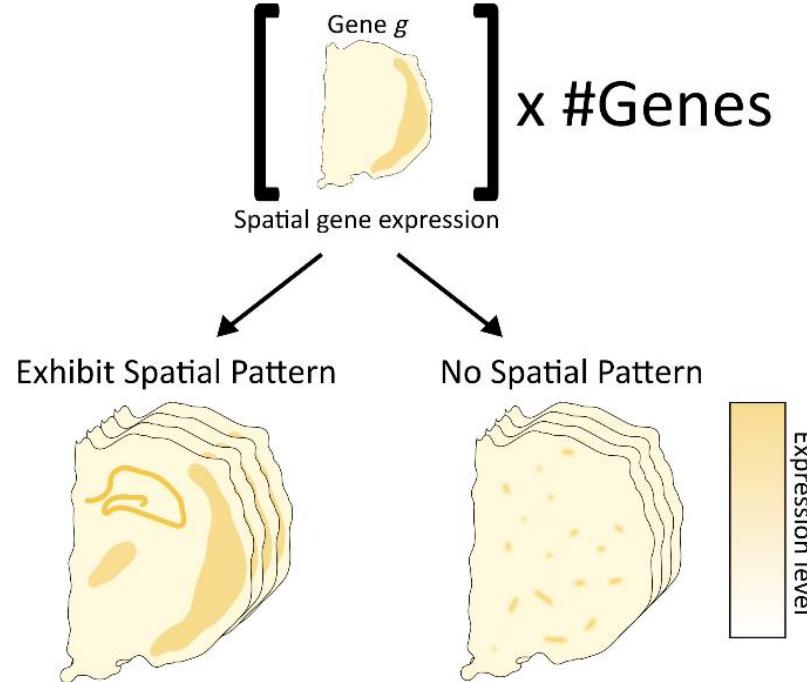
Spatially aware methods



Attempts to include knowledge of spatial structure in the analysis, not only to visualize results

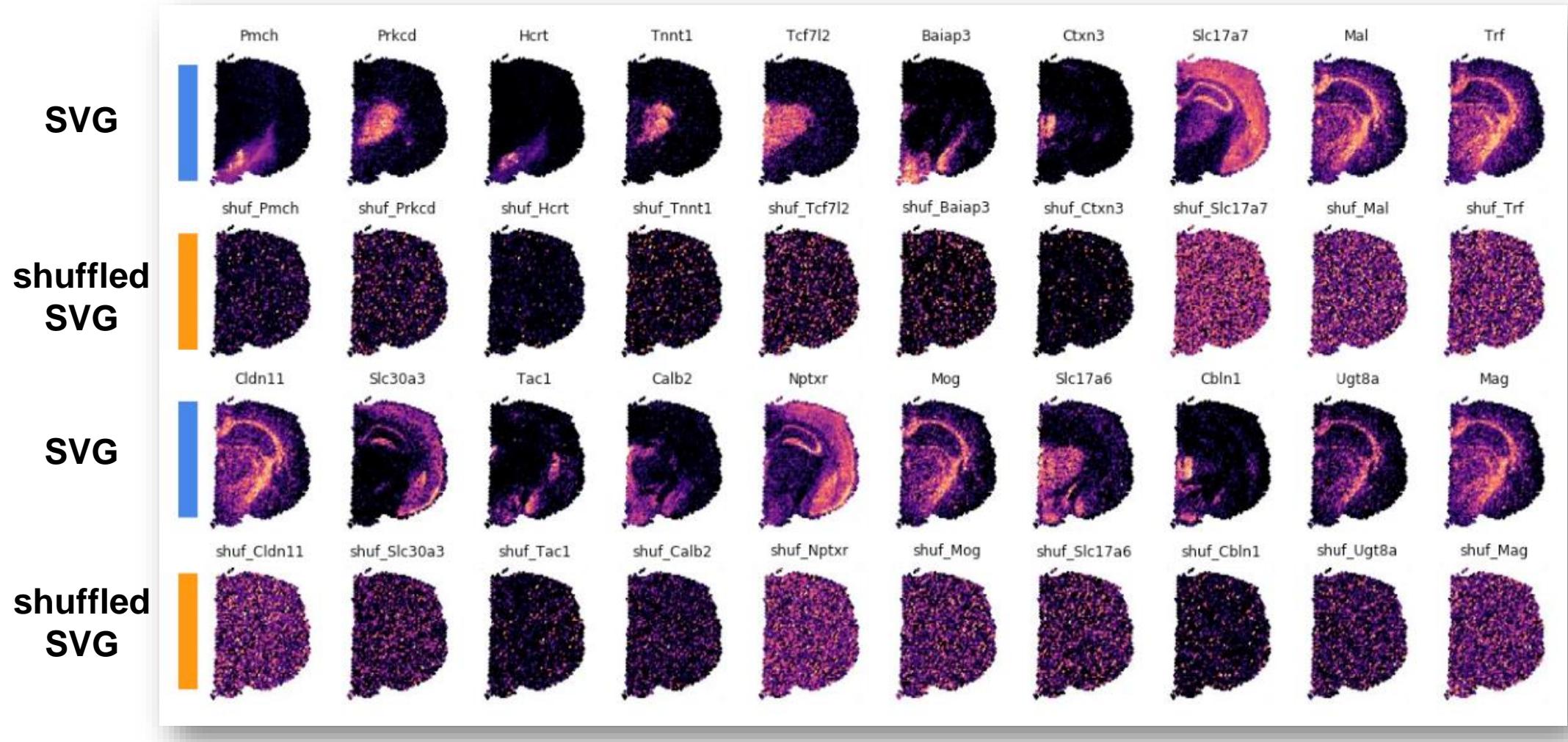
- Identifying spatially variable genes and features
- Finding spatially coherent expression domains
- Leveraging spatial proximity to increase the robustness of inference
- Finding local correlations between features

3. Spatially aware methods :: spatially variable genes



- Spatially variable genes (SVGs) are genes with a highly spatially correlated pattern of expression, which varies along with the spatial distribution of a tissue structure of interest
- Standard statistical measures such as Moran's I or Geary's C can be used to rank genes
- SpatialDE, SVCA, and SPARK use probabilistic models
- Essentially, test whether a "spatial" term in the covariance function significantly increases model's ability to explain data

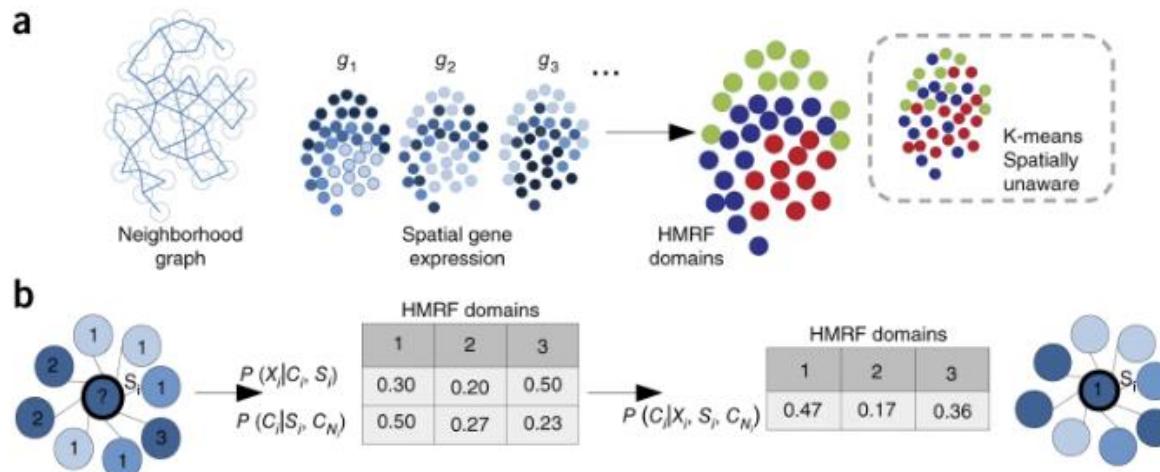
3. Spatially aware methods :: spatially variable genes



- 20 gene expression profiles from mouse brain
- Shuffle spots to get random expression profiles (with “shuf” prefix)

3. Spatially aware methods :: spatial domain patterns

HMRF (Hidden Markov Random Field)



- Standard clustering mainly focus on gene expression
- Leverage spatial information to find spatially coherent clusters (domains)

Zhu et al., 2018

gene expression (x_i) and neighborhood configuration (c_{Ni})

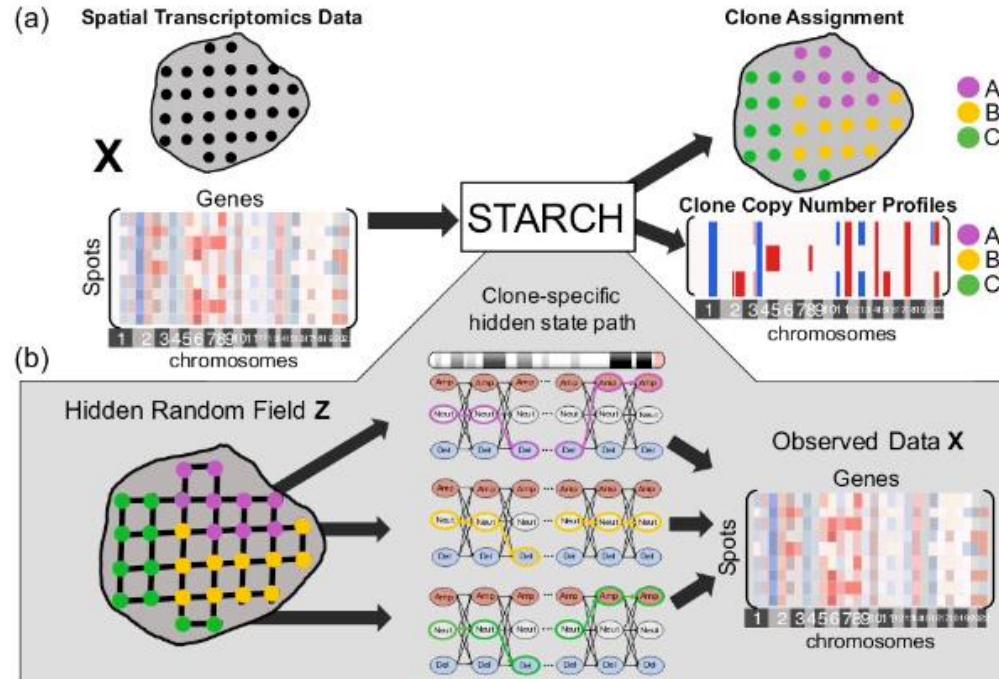
- **Spatially aware methods normally utilize HMRF:**

Construct a graph based on spatial proximity

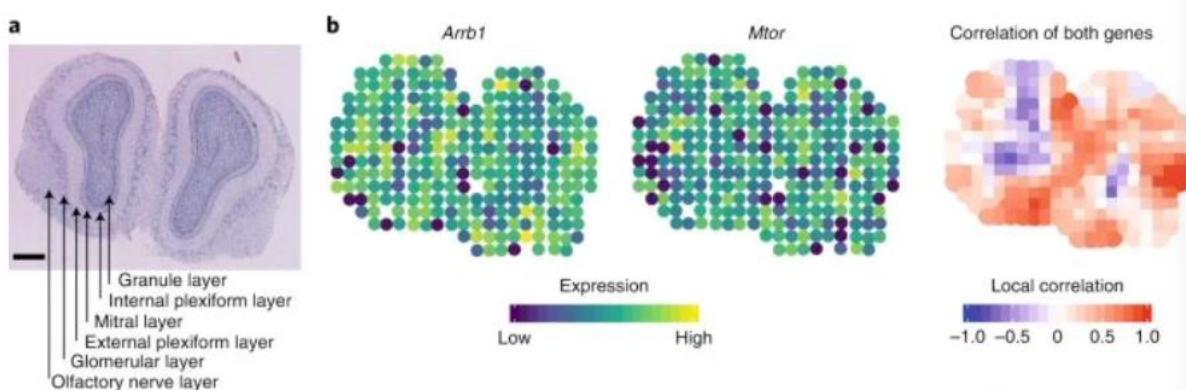
The probability of a node (spot) belonging to a specific domain depends on:

- Agreement with a domain expression profile
- Coherence with neighbors

3. Spatially aware methods :: spatial domain patterns



Elyanow *et al.*, 2021



Ghazanfar *et al.*, 2020

STARCH

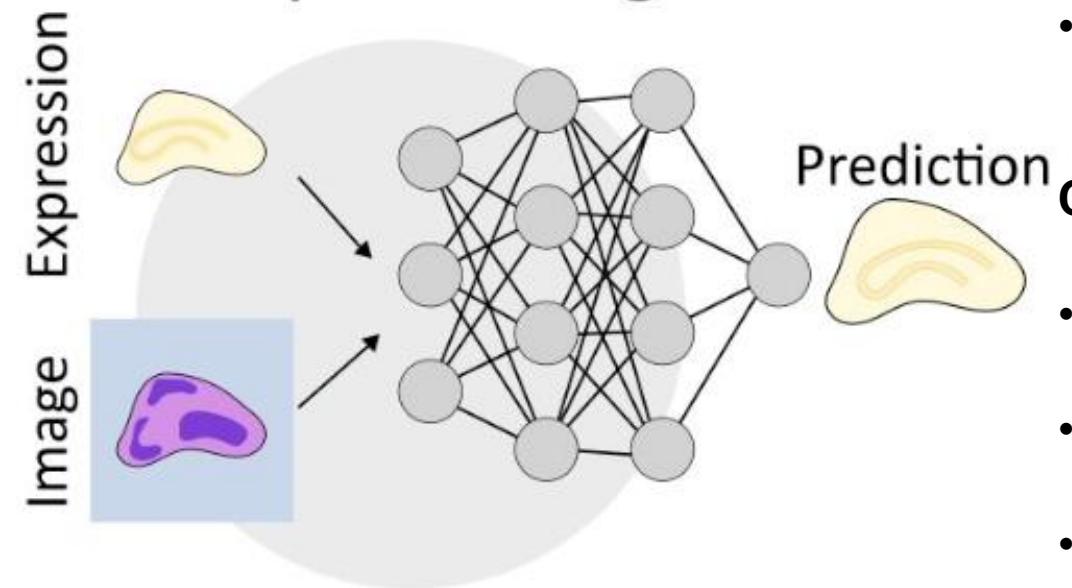
- Infers Copy Number Aberrations (CNA) from spatial transcriptomics data
- Increases robustness of inference by aggregating data in the same domains (similar profiles)
- It uses Hidden Markov Random Fields (HMRF)

scHOT

- Computes (spatially) weighted correlations to find local correlations
- It also uses Hidden Markov Random Fields (HMRF)

4. Deep learning

Deep learning



Apply deep learning to spatial data (very broad)

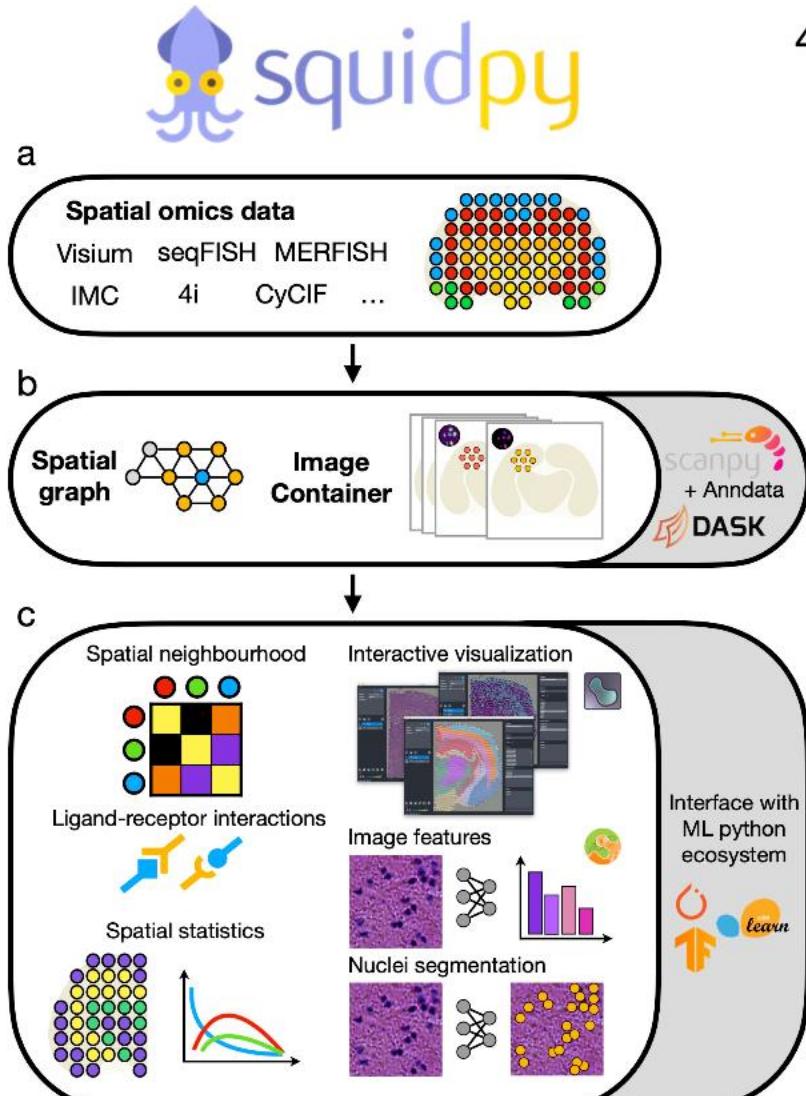
- Relatively few examples. Limited amount of high-quality available data.

Current examples:

- **XFuse**: “superresolution” (pixel) of gene expression by learning joint representation of image and expression data.
- **stPlus**: Uses scRNA-seq data and autoencoders to enhance spatial transcriptomics data
- **SpaGCN**: simultaneous domain and SVG detection using graph convolution layers
- **RESEPT**: Uses graph convolutional network to embed spatial data in RGB space, then uses a CNN to segment data into spatially coherent tissue domains

5. Computational suites: squidpy

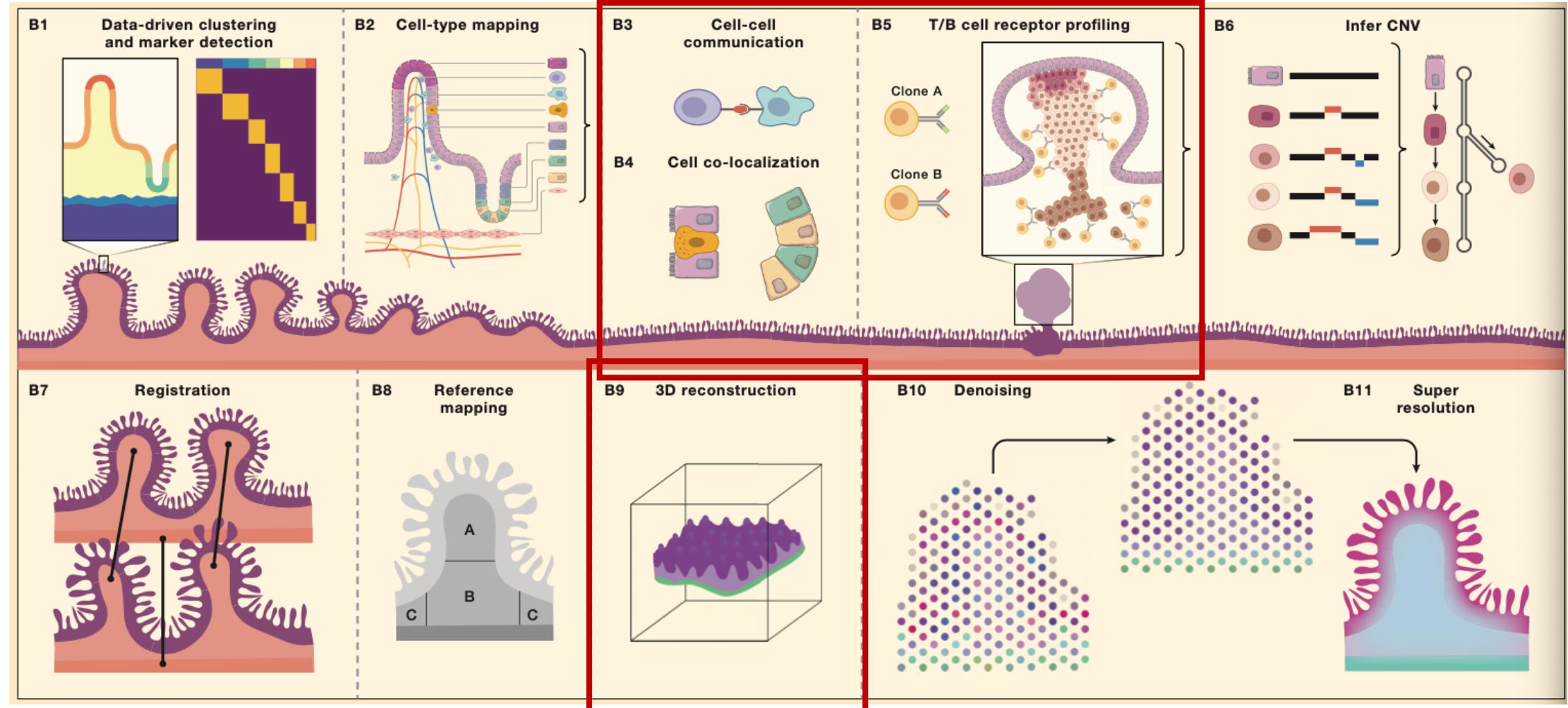
4



“One framework to rule them all, one framework to find them...”

- It builds on top of scanpy and anndata, from which it inherits modularity and scalability.
- Tailored towards spatial data with support for multiple different experimental platforms (not only Visium)
- Easy to construct spatial graphs and perform graph operations
- Has excellent interface with ML ecosystems such as PyTorch, TensorFlow and sklearn

Snapshots of spatial transcriptomics applications



Larsson et al., 2023

Take home messages

- ✓ There are tons of spatial techniques out there
- ✓ In addition, an ever-increasing repertoire of computational methods!
 - A lot of tools out there, but sometimes beneficial to construct custom solutions
- ✓ Spatial-omics data is already improving our understanding of human health and disease in research, diagnostic, and therapeutic setting

Questions?

Have fun with the hands-on SRT exercise



Jihyun Kim



Jooyoung Park



Harim Chun