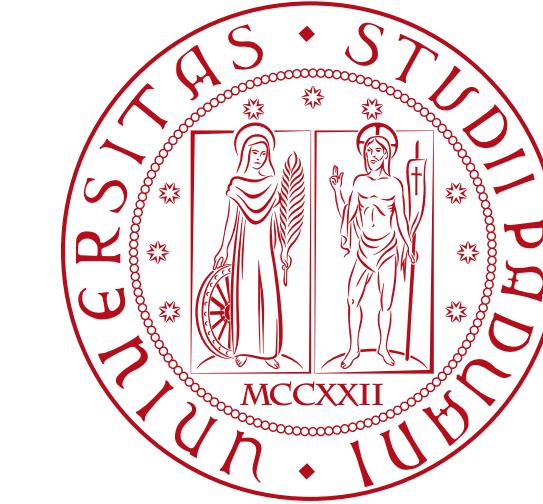




800¹²²²⁻²⁰²²
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

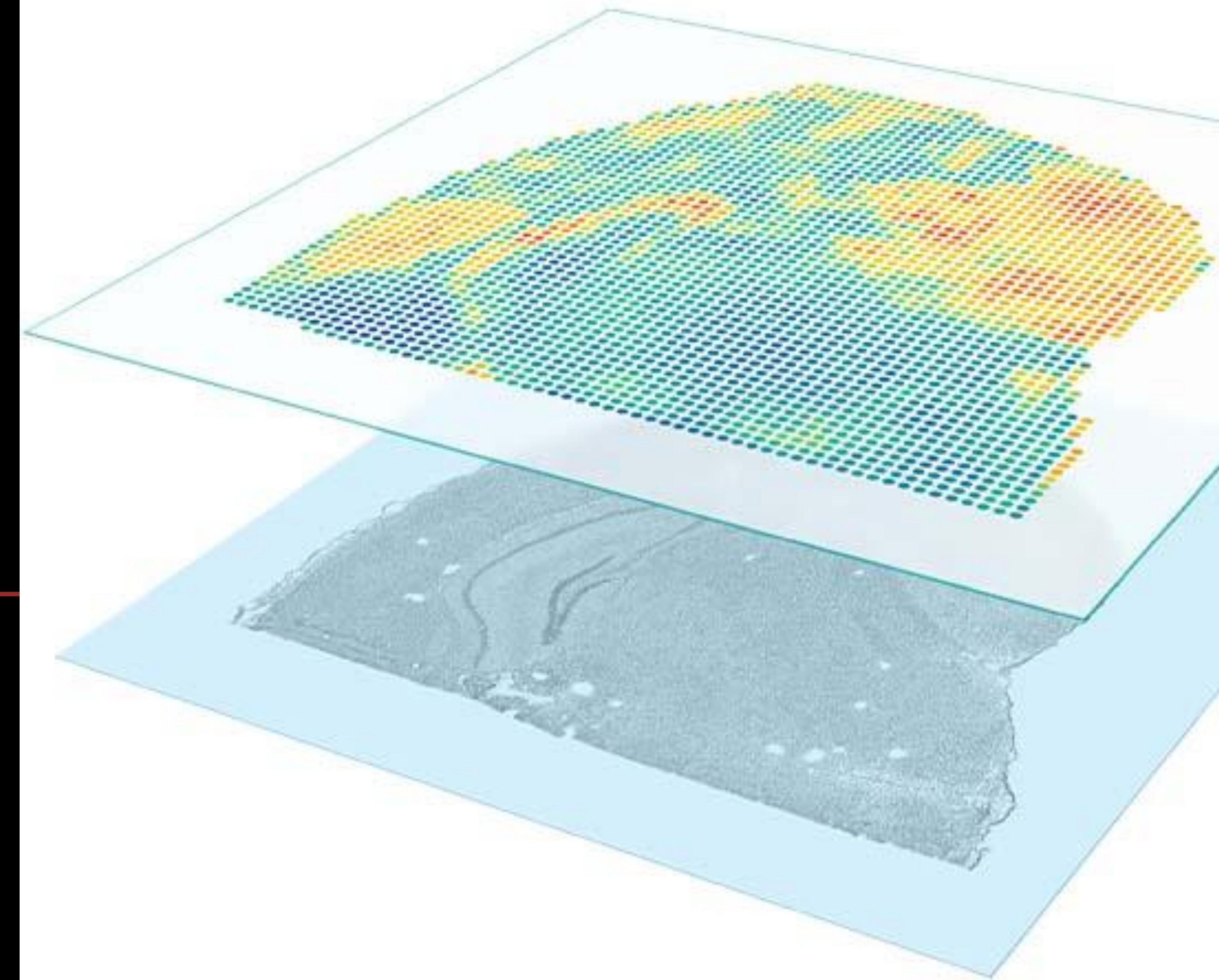
CSAMA 2023 - BRIXEN/BRESSANONE

SPATIAL (TRANSCRIPT)OMICS

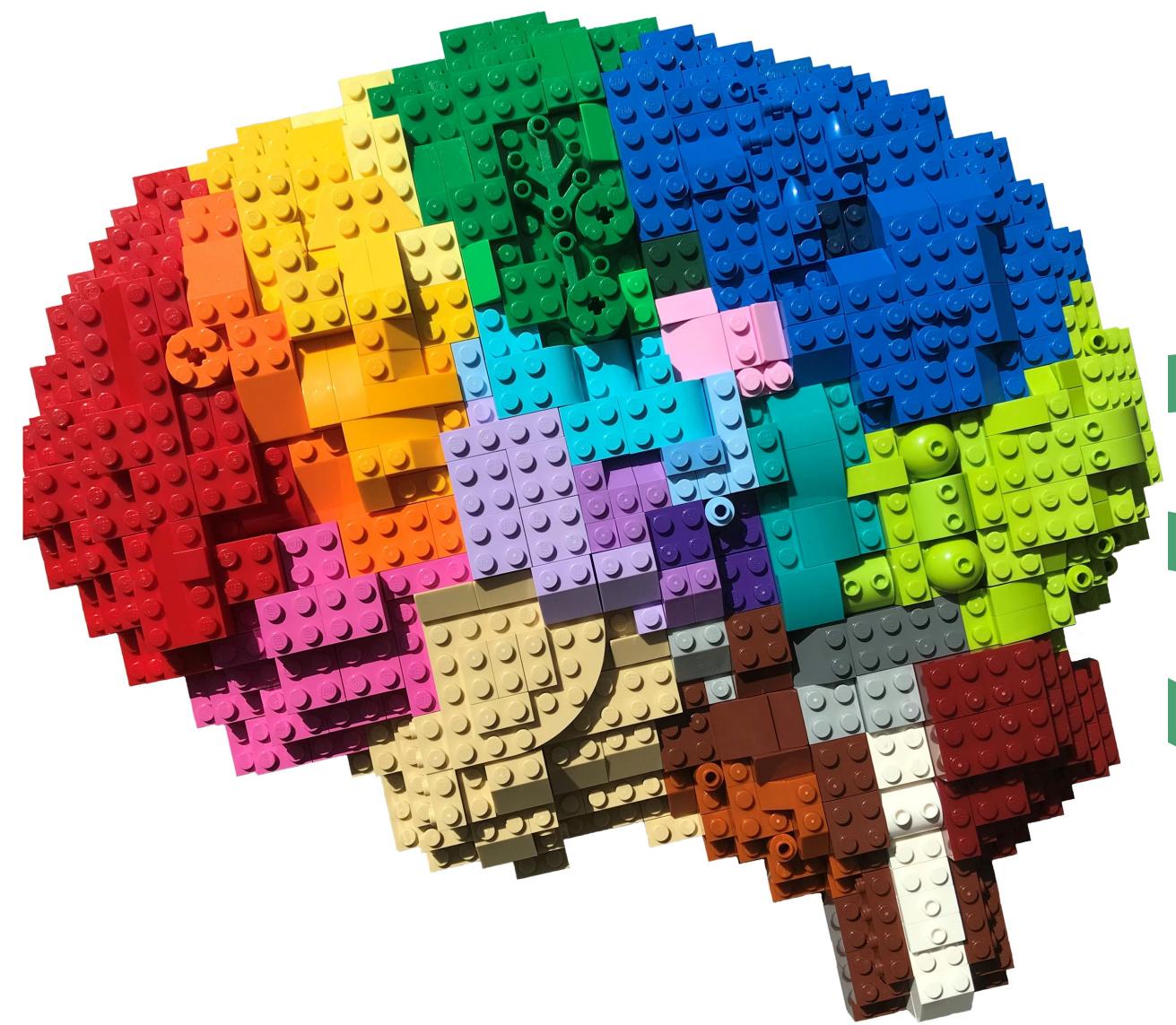


Davide Risso
@drisso1893
@drisso@genomic.social
@drisso

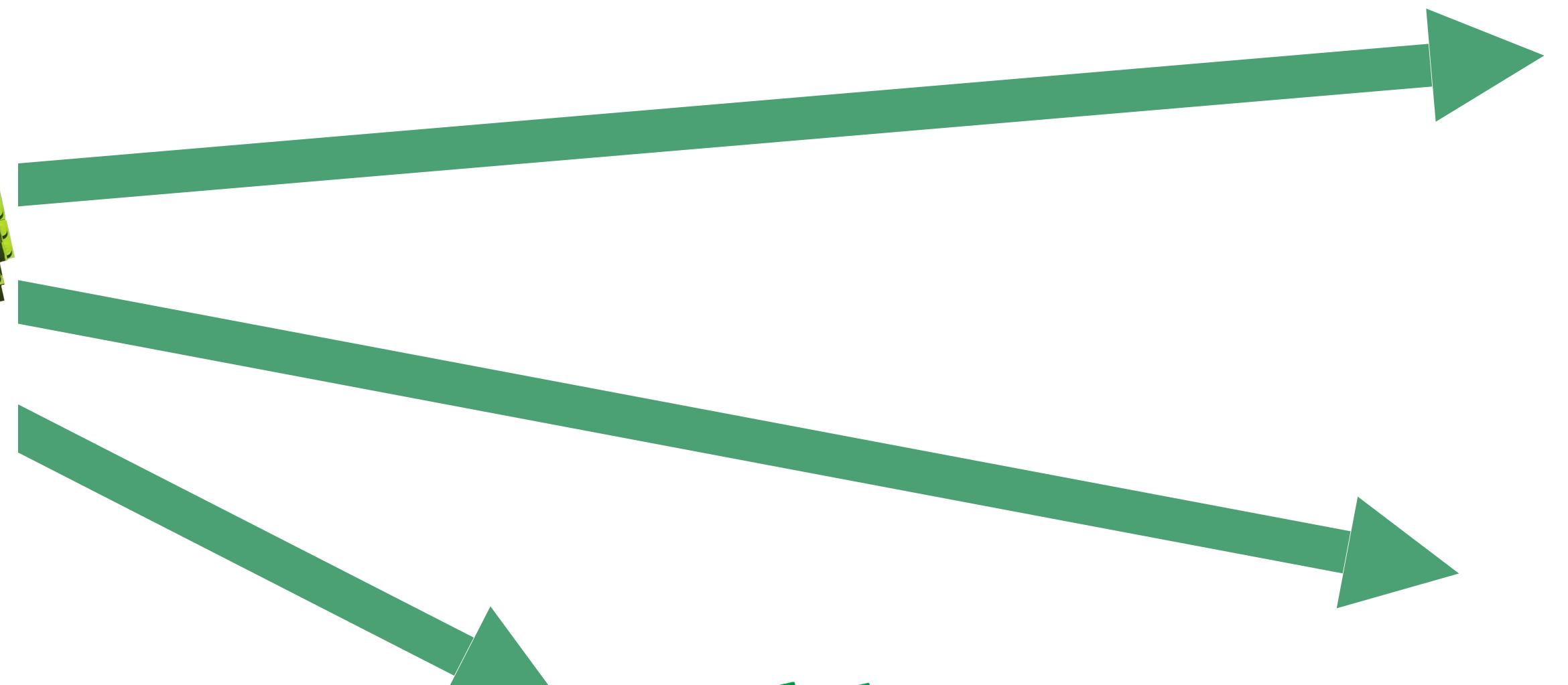
Introduction to Spatial Transcriptomics



The evolution of gene expression measurements



Original organ



Spatial
transcriptomics



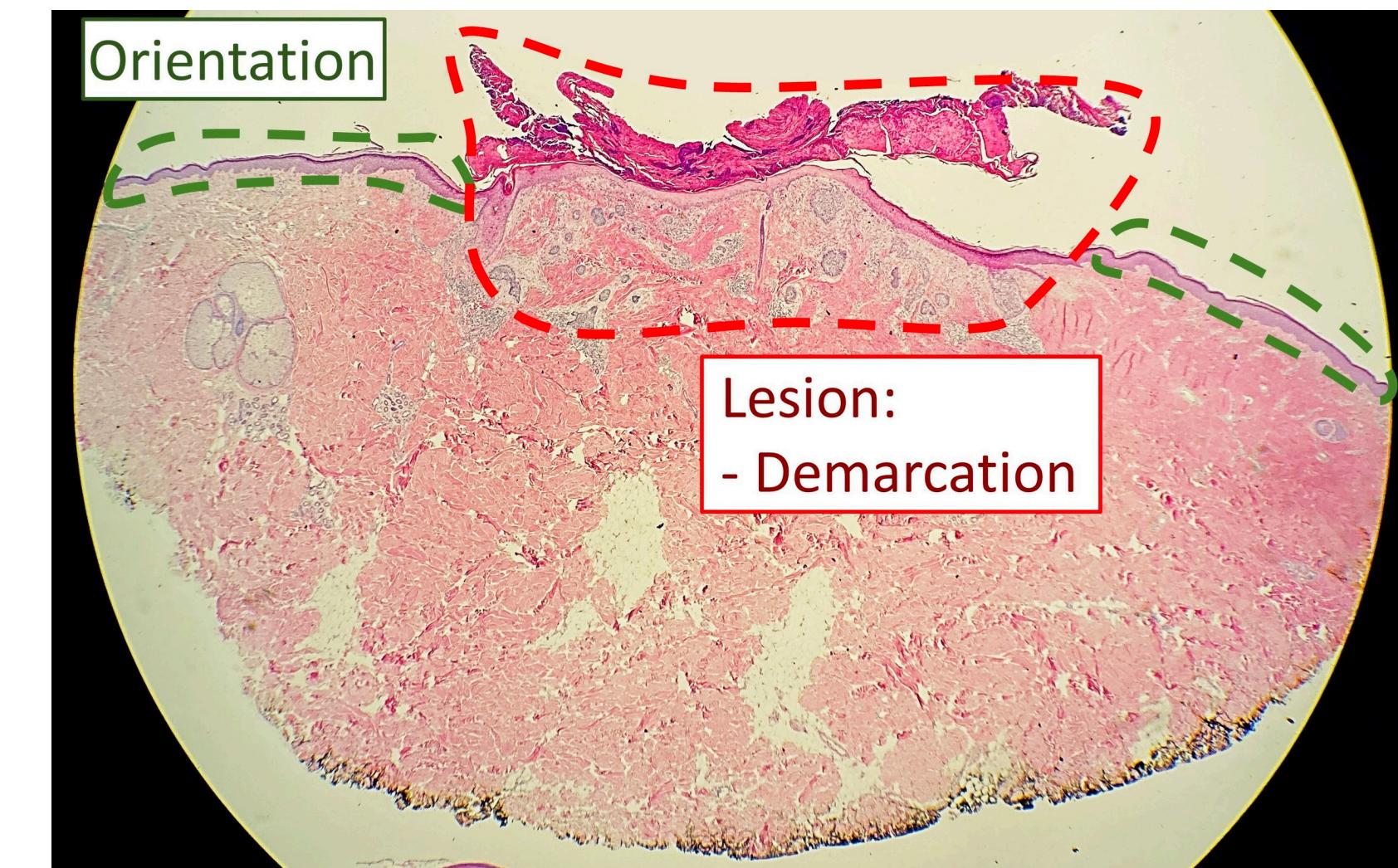
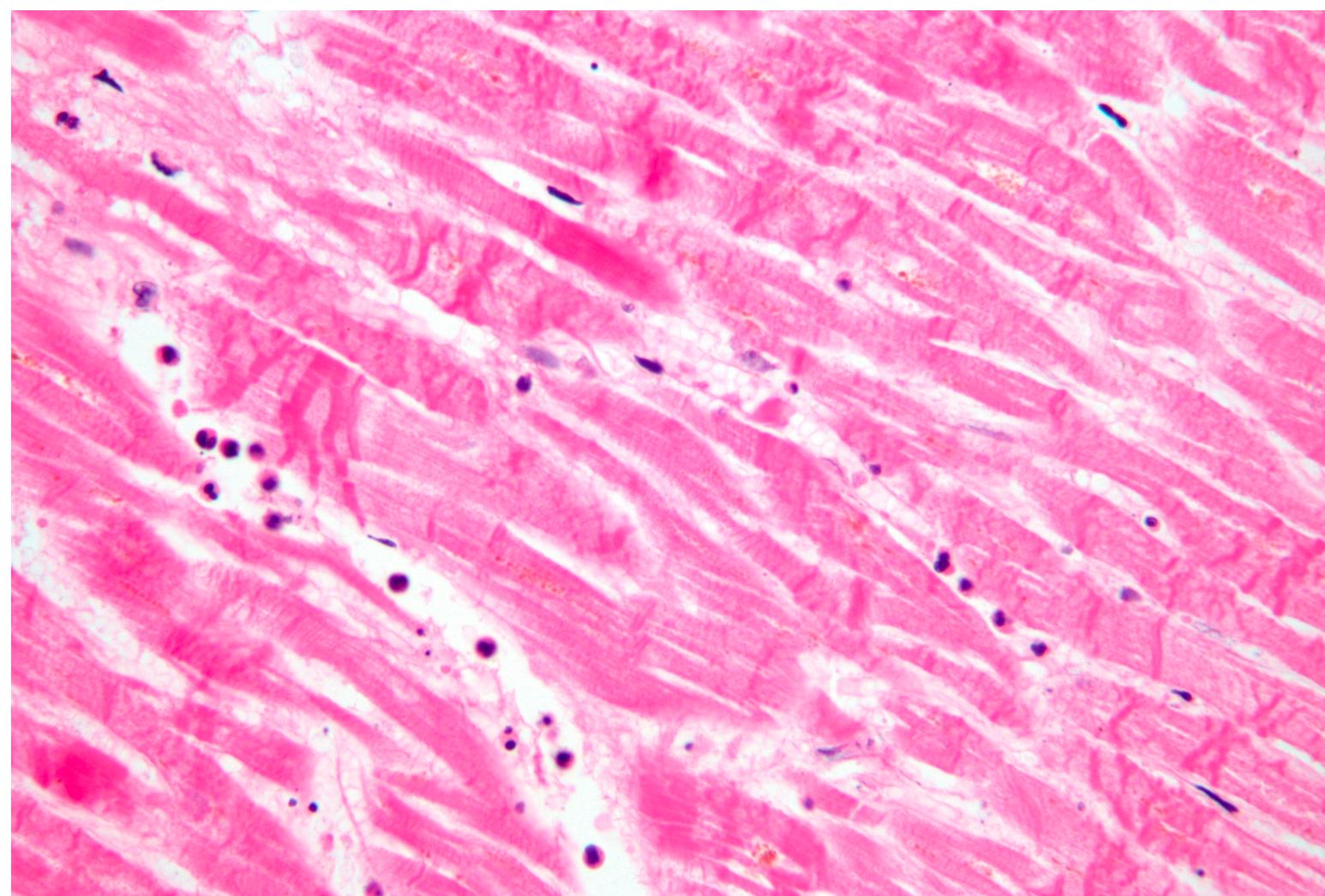
Bulk RNA-seq



Single-cell RNA-seq

Why spatial?

- Key biological processes depend on the physical proximity of cells and on the spatial organization of tissues.
- Several diseases are characterized by abnormal spatial organization within tissues.
- In clinical settings, histopathology is a widely used diagnostic tool.
- In situ hybridization (ISH) and immunohistochemistry are the main techniques.



What is spatial transcriptomics?

By “spatial transcriptomics”, we mean attempts to quantify mRNA expression of **large numbers of genes** within the spatial context of tissues and cells.

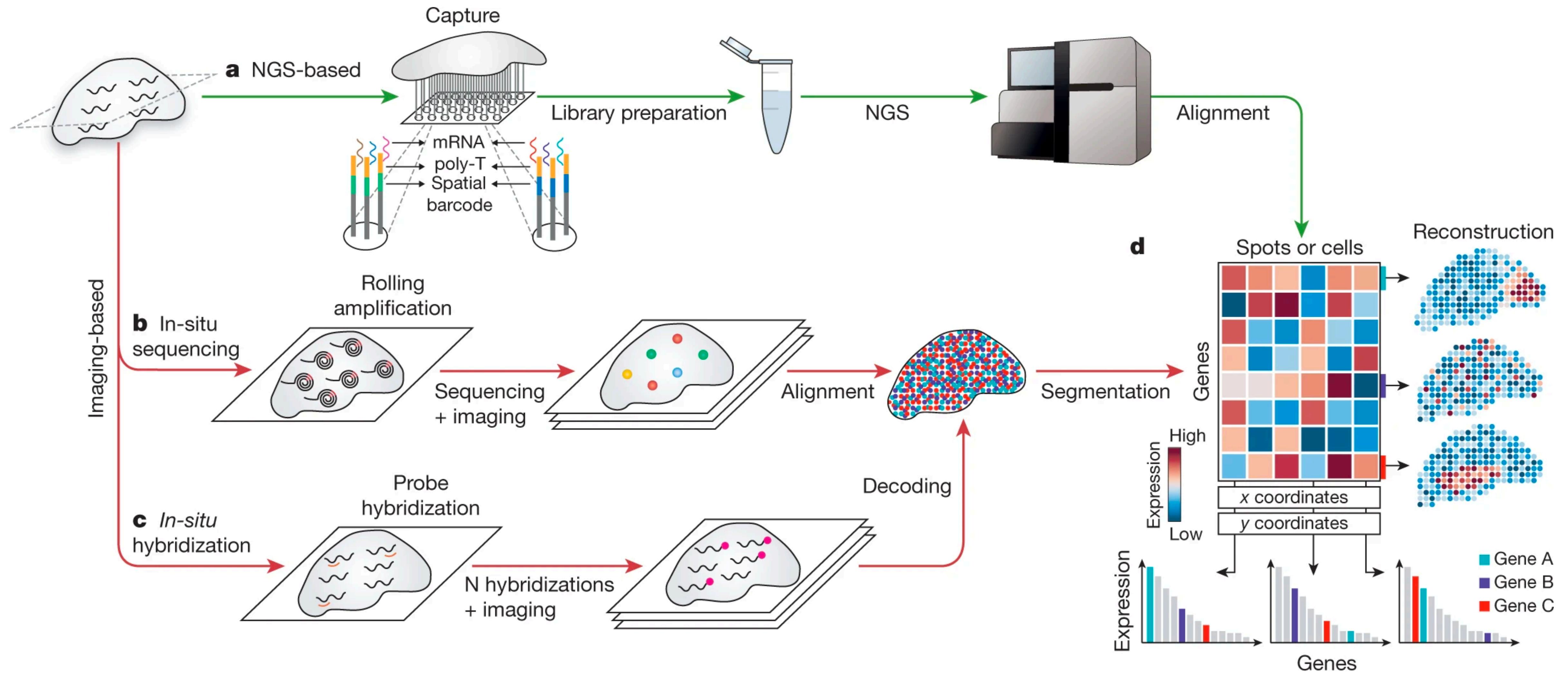
Moses & Pachter (2022)

What is spatial transcriptomics?

- ISH allows the visualization of biological processes by mapping DNA, RNA, and proteins within tissues.
- However, they are limited to only a handful of genes/proteins.
- Over the past decade, technologies have emerged that bridge the gap between traditional spatial approaches and whole transcriptomics/proteomics approaches.

We may define spatial transcriptomics as the set of techniques that allow to query (close to) the entire transcriptome while retaining the tissue's spatial information.

Spatial Transcriptomics technologies



NGS-based methods

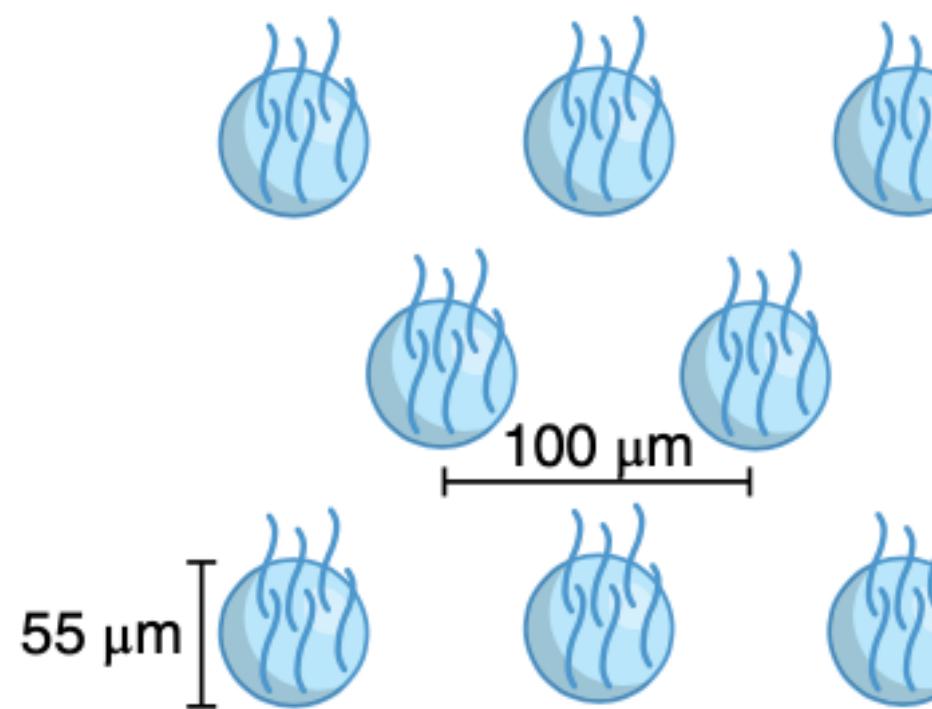
- Built starting from single-cell RNA-seq methods + the addition of a spatial barcode before library preparation.
- Main advantages:
 - “Unbiased”: there is no need to select which genes to measure prior to the experiment
 - Familiarity with NGS data and their biases
 - Close to whole transcriptome
 - Cheap(er) and commercially available
- Main drawbacks:
 - Lower sensitivity (inferior to scRNA-seq)
 - Lower resolution (supra-cellular in most cases)
 - Standardized area size (currently 6 mm² for 10X Genomics Visium)

NGS-based methods

10X Genomics Visium

Technology initially proposed
in Stahl et al. (2016)

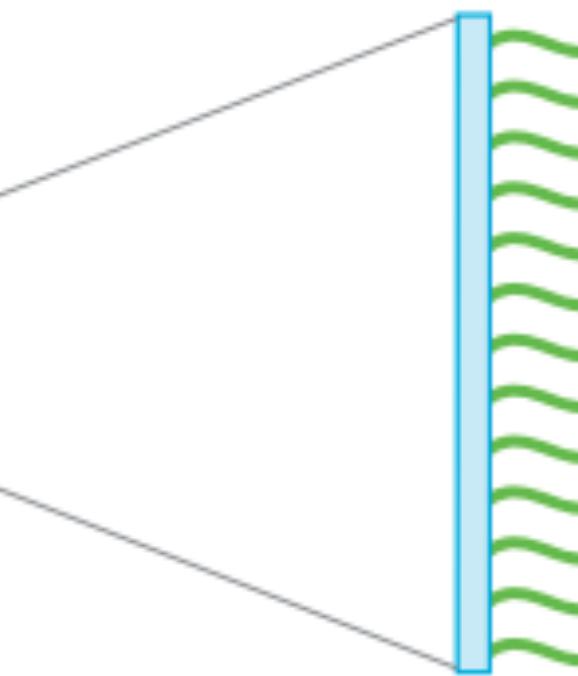
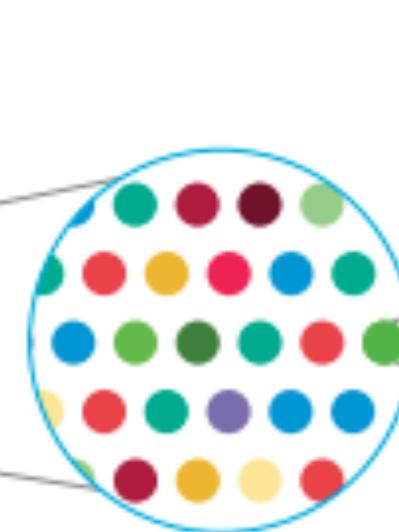
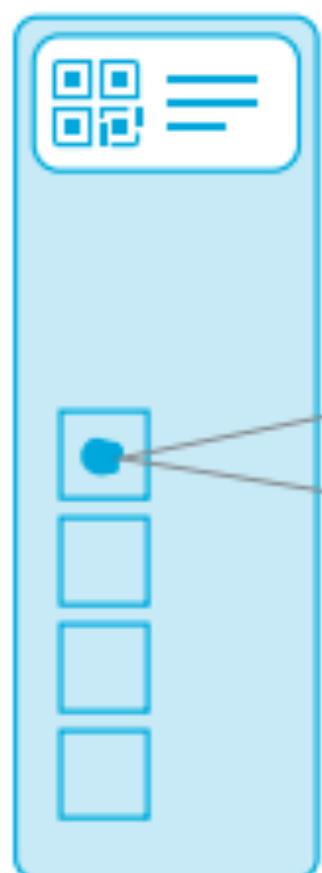
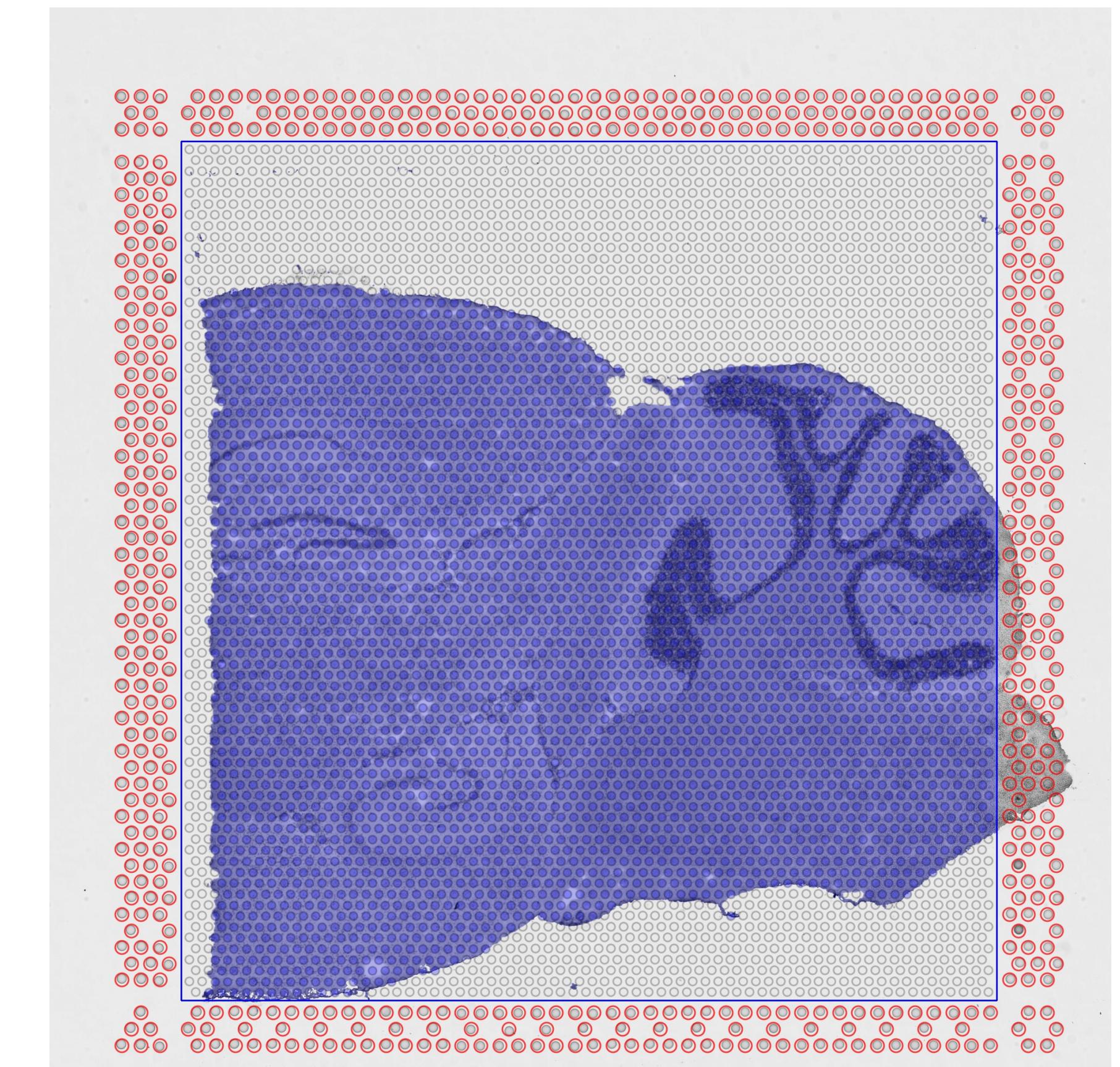
Improved and
commercialized by 10X
Genomics



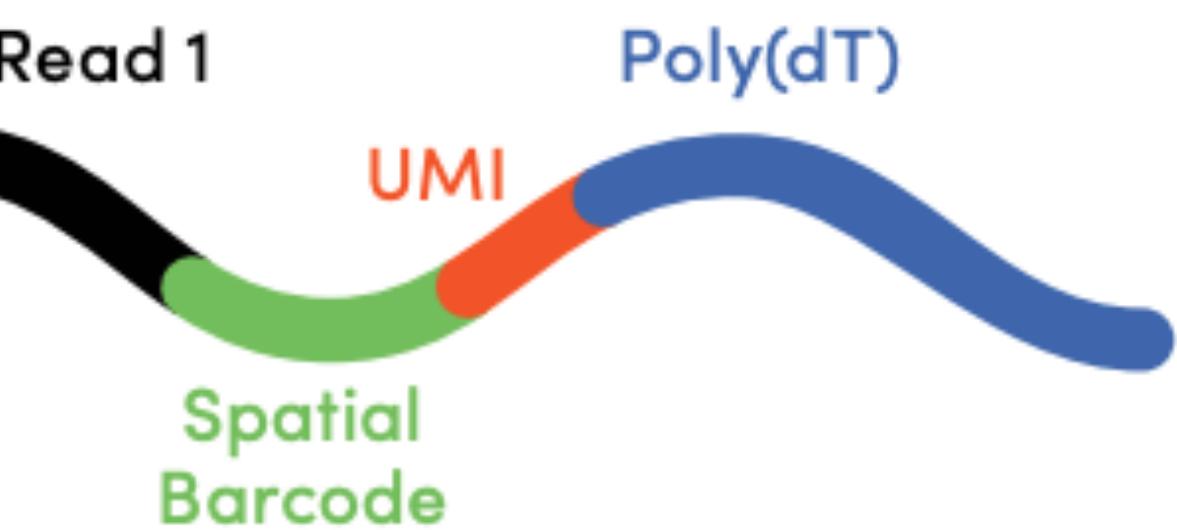
Visium Spatial
Gene Expression Slide

Capture Area with
5000 Barcoded Spots

Visium Gene Expression
Barcoded Spots

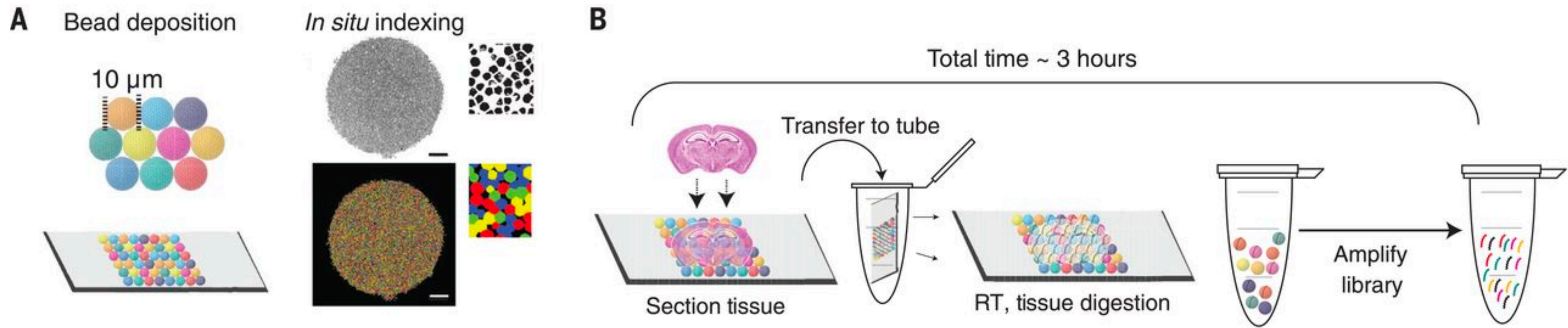


Partial Read 1



NGS-based methods

Slide-Seq



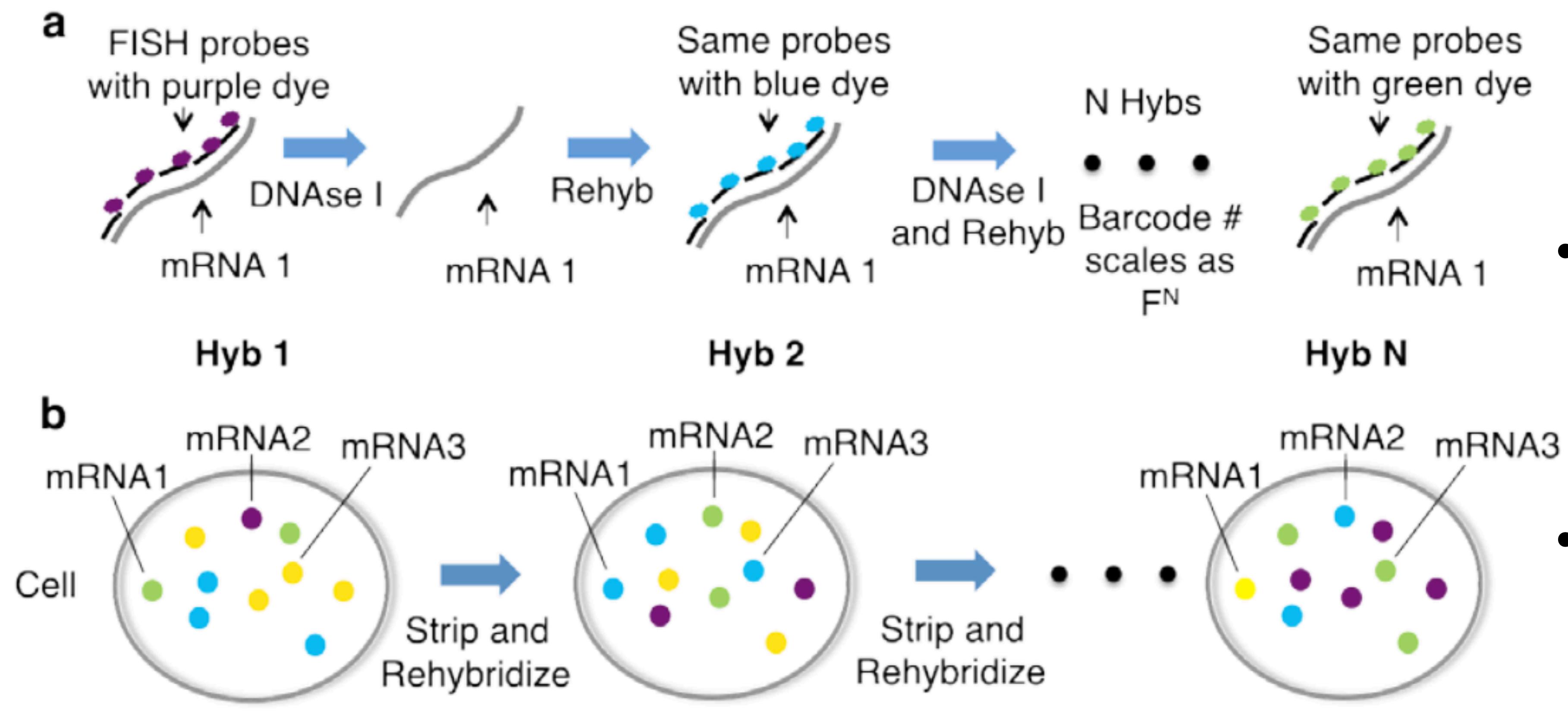
Many other improvements and protocols, including HDST, DBiT-seq, Stereo-seq, Seq-scope, PIXEL-seq. All attempts at increasing resolution.

Image-based methods

- In-Situ Sequencing (ISS), in which transcripts are amplified and sequenced in the tissue.
- ISH-based methods, in which imaging probes are sequentially hybridized in the tissue.
- Main advantages:
 - High sensitivity (up to 80% of detection efficiency of gold standard)
 - High resolution (sub-cellular) — limited only by the optical diffraction limit
 - Can span a wide range of area sizes — highly customizable
- Main drawbacks:
 - ISH-based methods do not report the sequence of the transcripts (cf. mutations, isoforms, ...)
 - Lower gene throughput (100-1000 genes/proteins — approaching 10,000)
 - Less well studied biases and systematic effects

In-situ methods

seqFISH

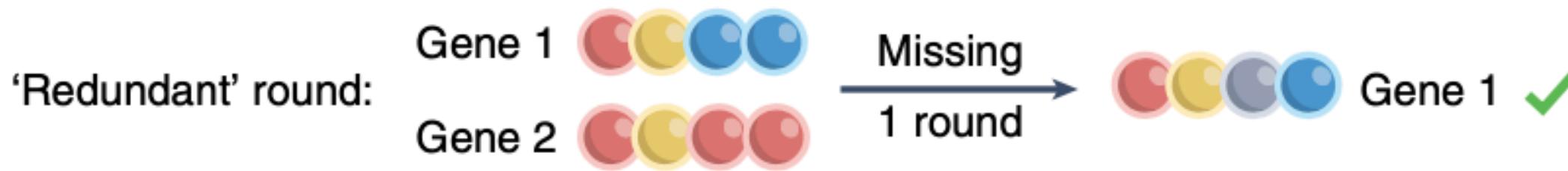
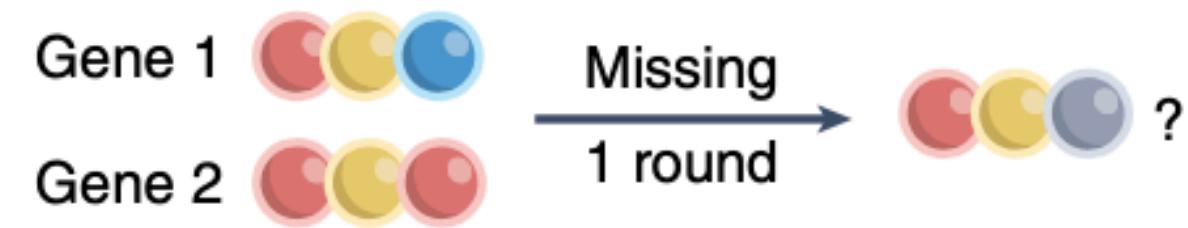
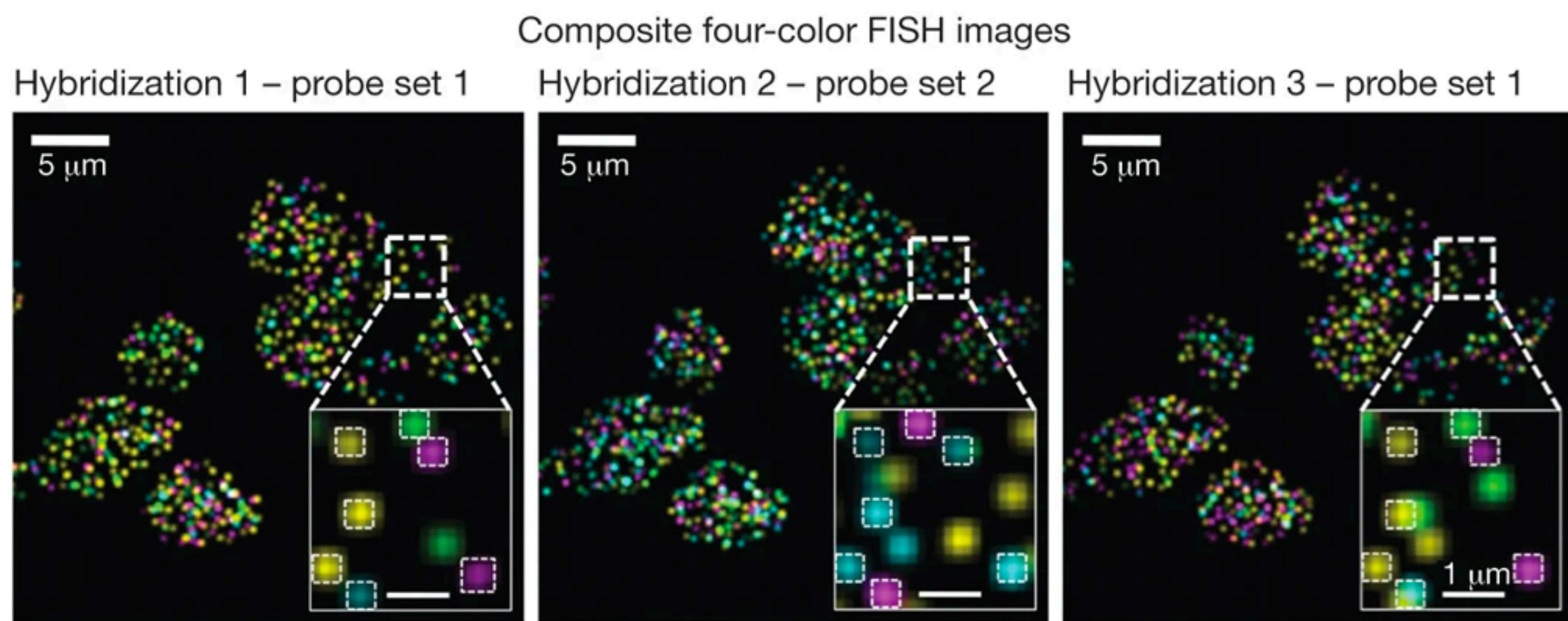


- The mRNAs in cells are barcoded by sequential rounds of hybridization, imaging and probe stripping.
- As the transcripts are fixed, the fluorescent spots can be aligned.
- This sequential barcode is designed to uniquely identify an mRNA.

In-situ methods

seqFISH

b

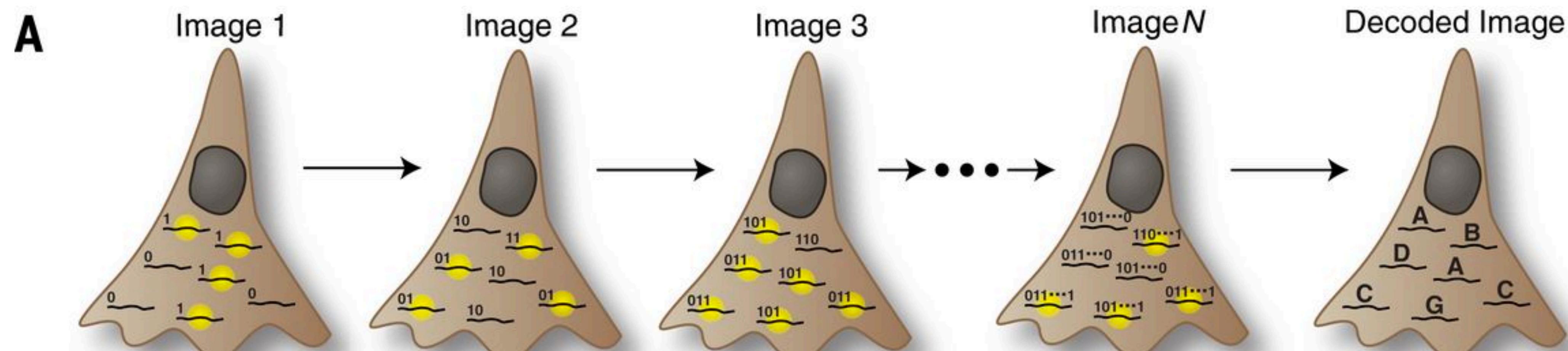


- With 4 dyes, 8 rounds of hybridization can cover the entire transcriptome ($4^8 = 65,536$)
- In practice, an error-correcting round of hybridization is performed, so that genes can still be distinguished if signal from one round of hybridization is missing

In-situ methods

Multiplexed Error-Robust FISH (MERFISH)

- Each gene is encoded by a N-bit binary word
- During each round of imaging, only the subset of RNAs that have “1” in the corresponding bit emit signal

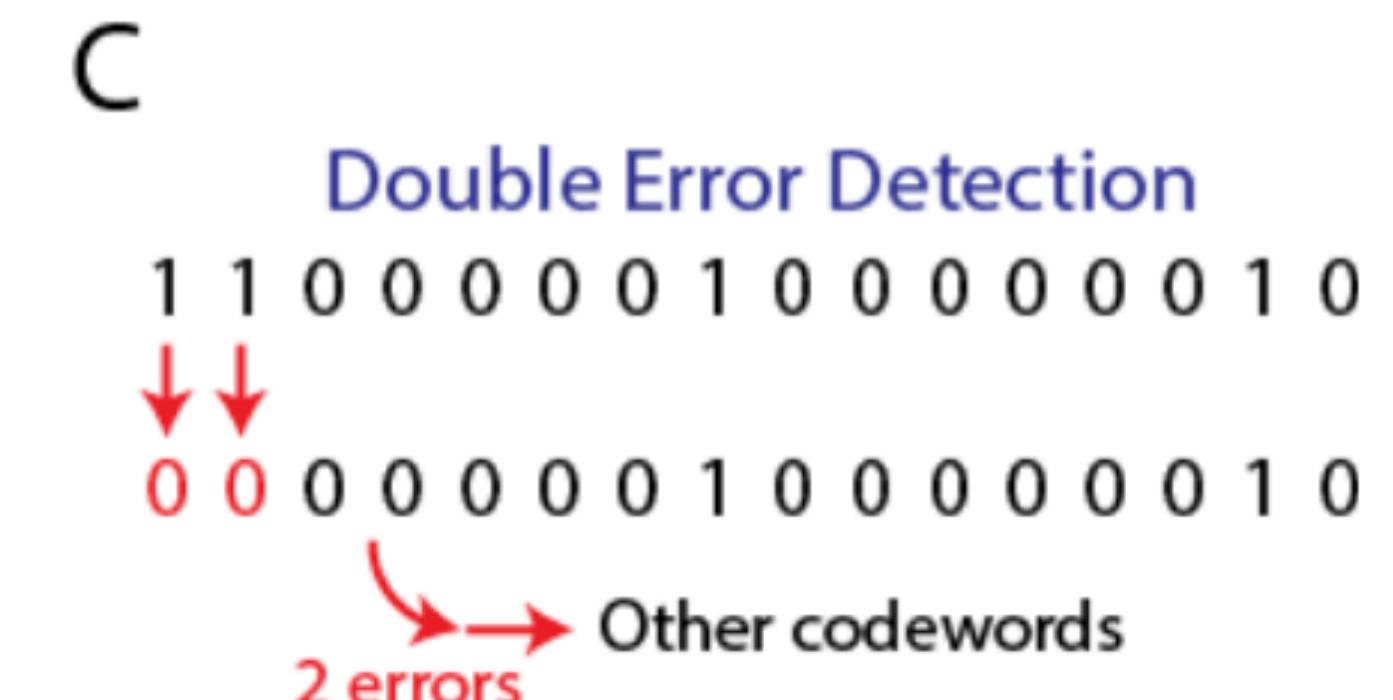
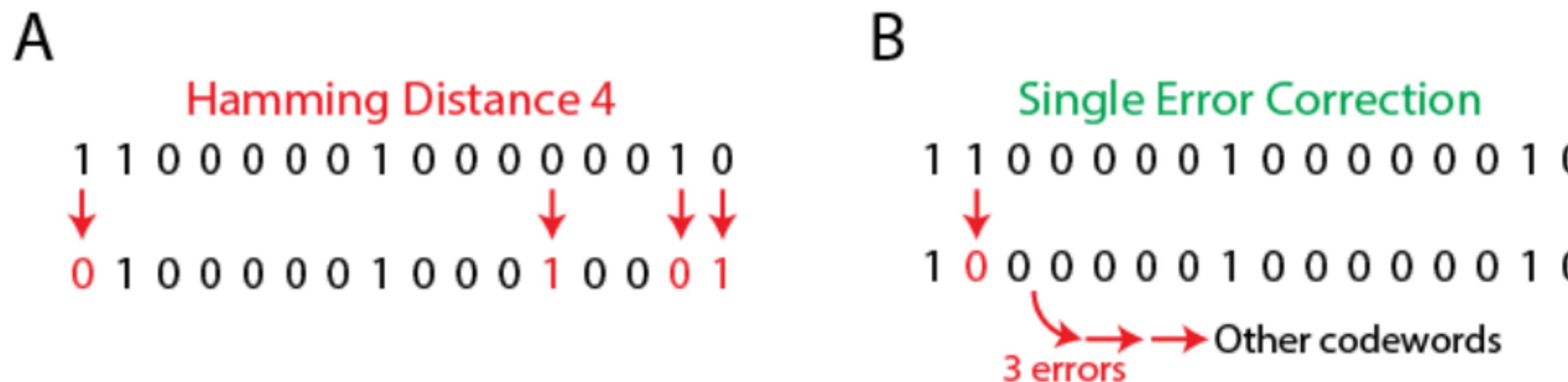


Chen et al (2015)
Moffitt et al (2016)

In-situ methods

Multiplexed Error-Robust FISH (MERFISH)

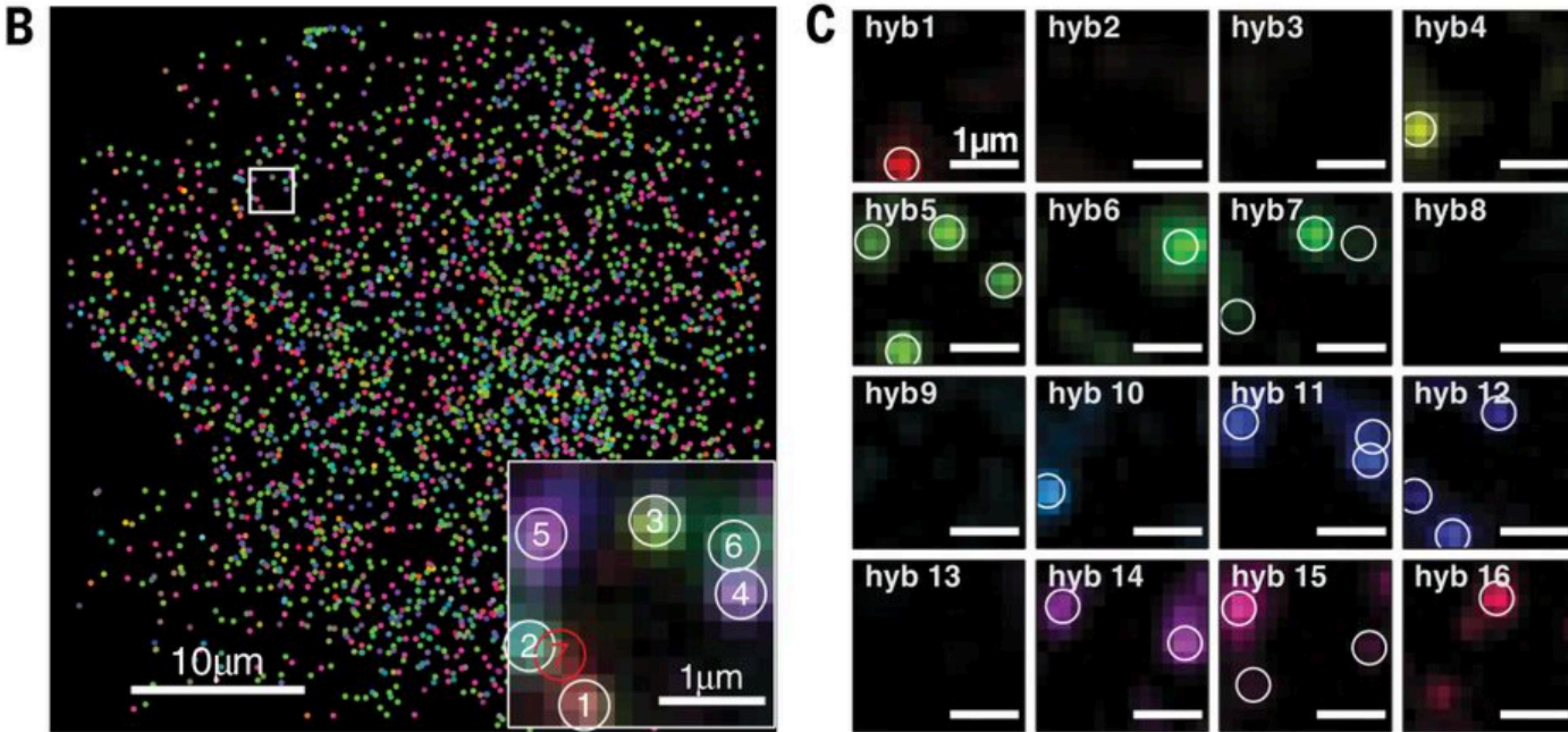
- To correct for errors, only a subset of words separated by a Hamming distance of at least 4 (HD4 code) are used
- Every single-bit error produces a word that is exclusively close to a single RNA-encoding word, allowing such errors to be detected and corrected



Chen et al (2015)
Moffitt et al (2016)

In-situ methods

Multiplexed Error-Robust FISH (MERFISH)



Chen et al (2015)
Moffitt et al (2016)

In-situ methods

Multiplexed Error-Robust FISH (MERFISH)

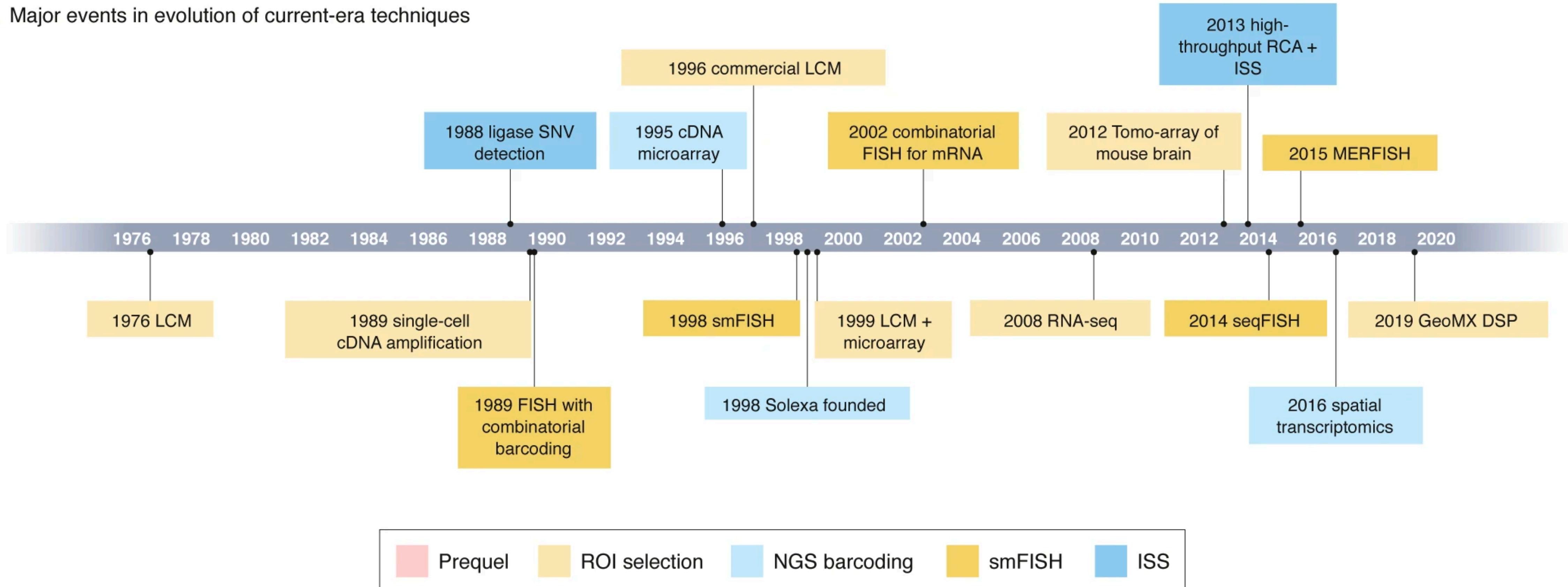
D

Spot number	Hybridization round																Gene
	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	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	LRP1
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	unidentified

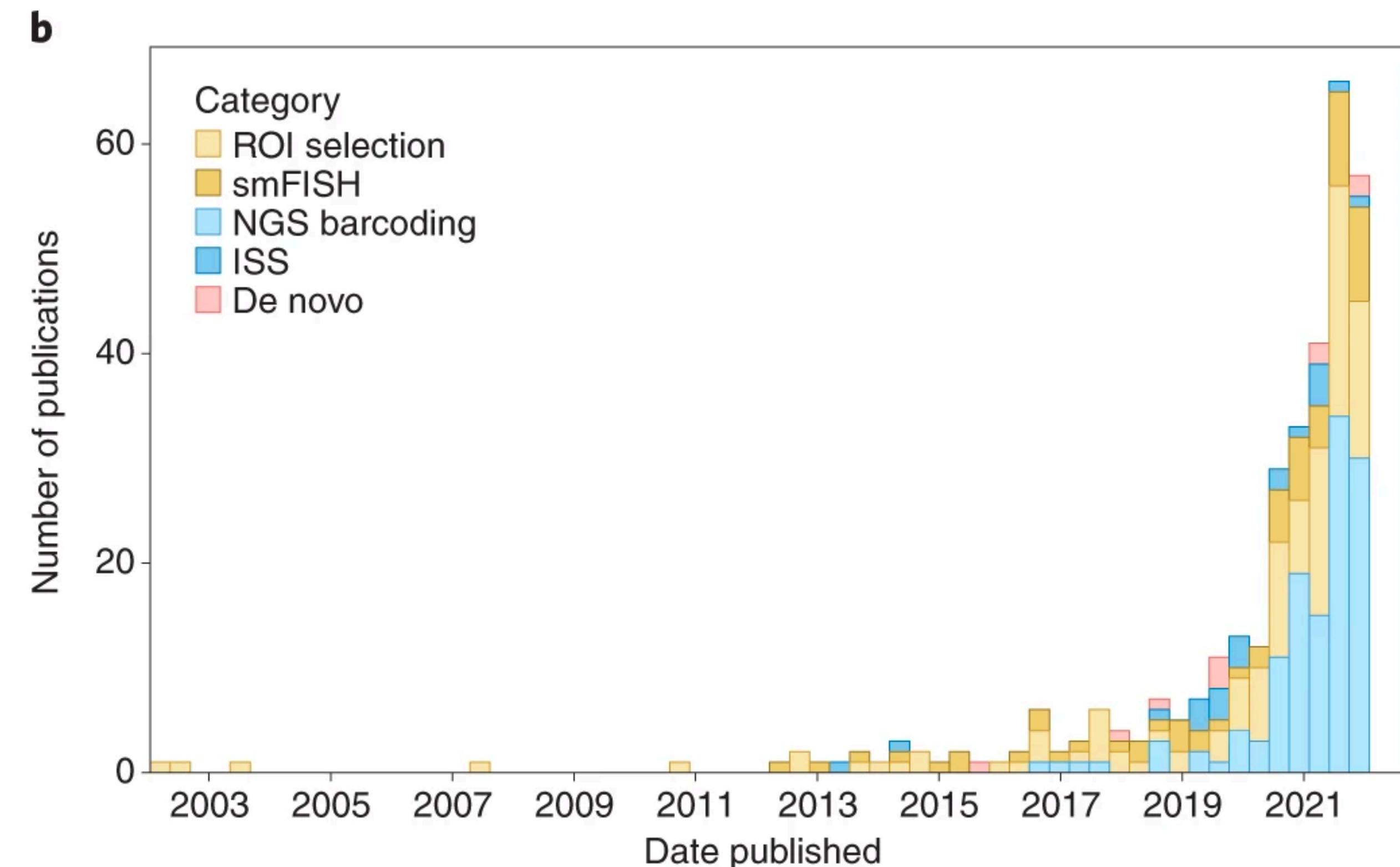
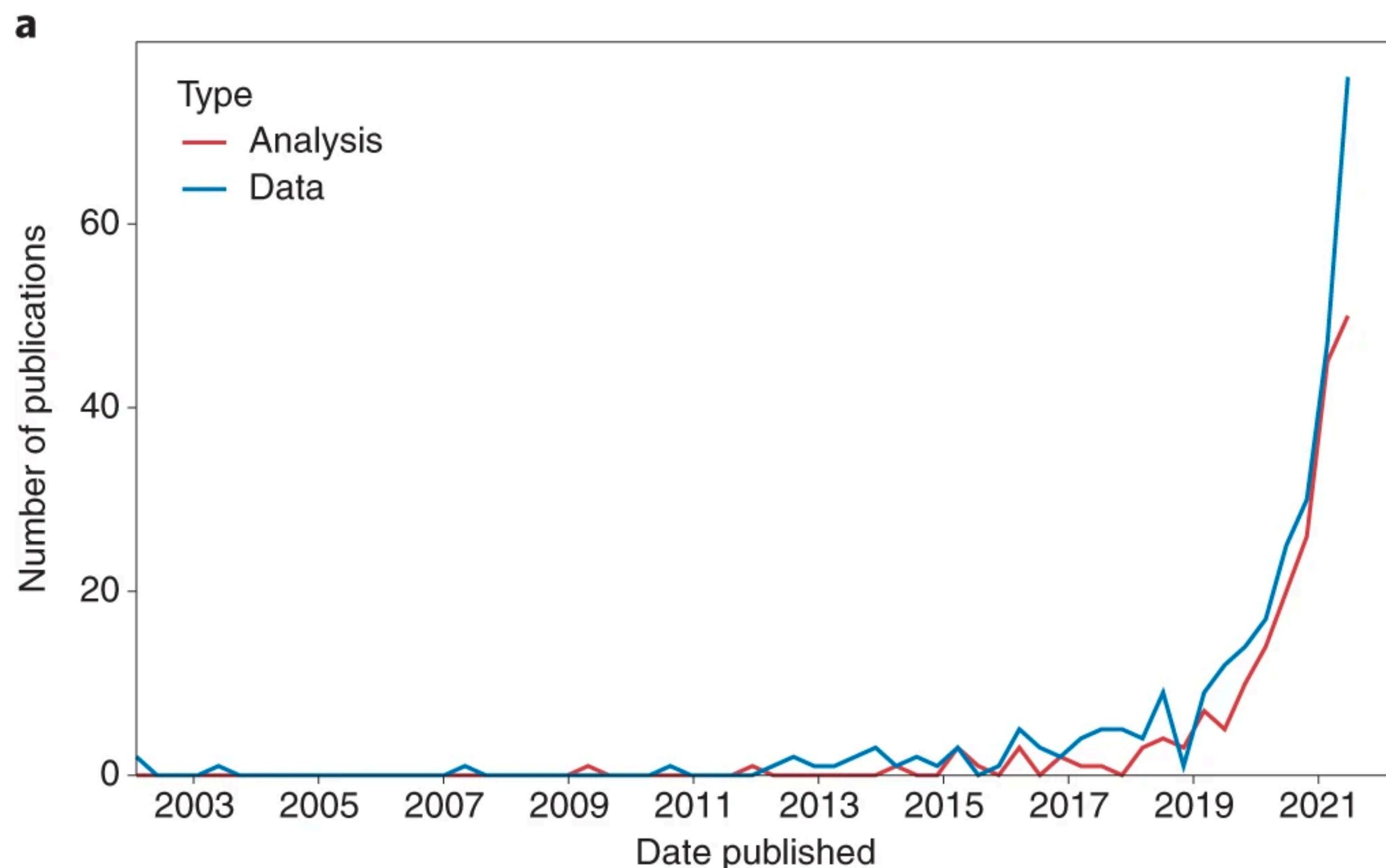
Chen et al (2015)
Moffitt et al (2016)

Timeline

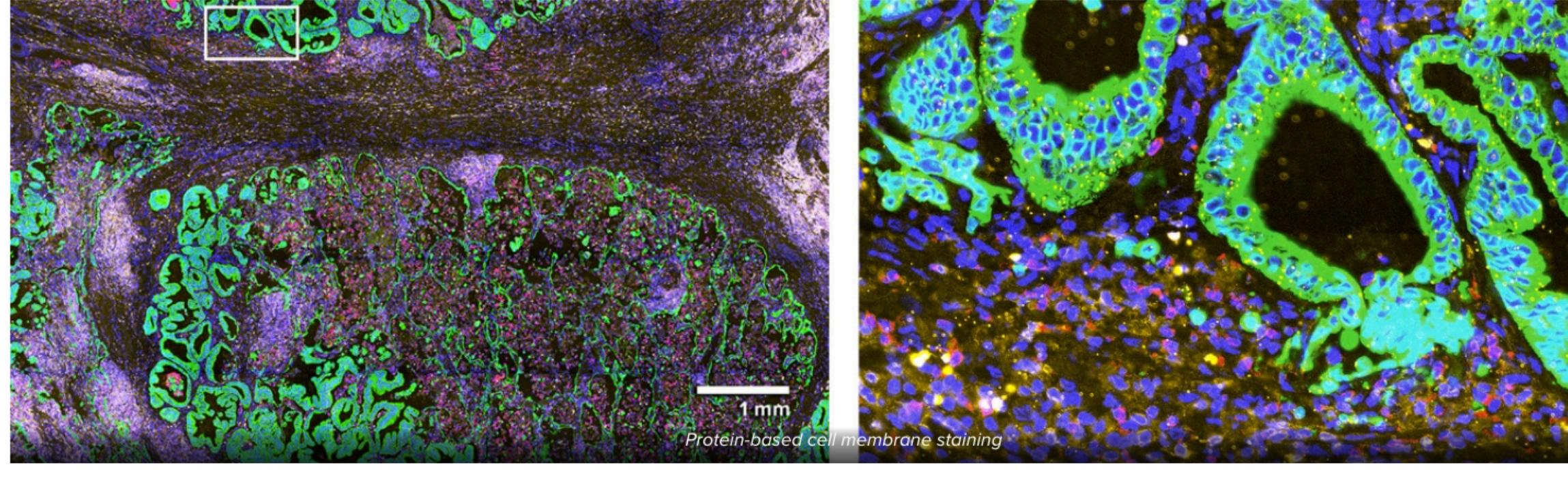
c Major events in evolution of current-era techniques



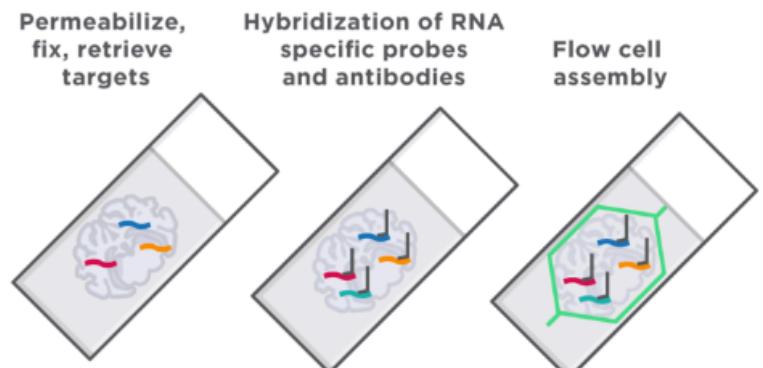
Trends



Outline: commercial FISH

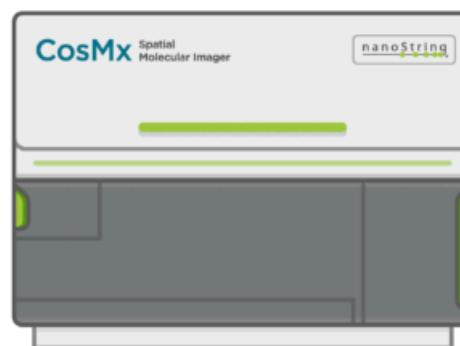


1 SAMPLE PREPARATION > 2 INTEGRATED READOUT > 3 INTERACTIVE DATA ANALYSIS

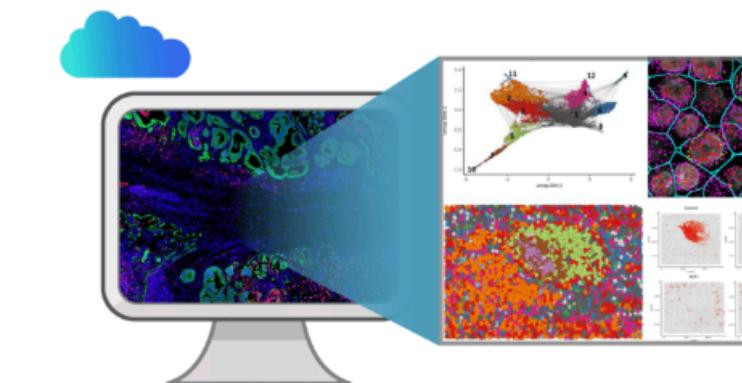


Compatible with formalin-fixed paraffin embedded (FFPE) and fresh frozen (FF) tissues

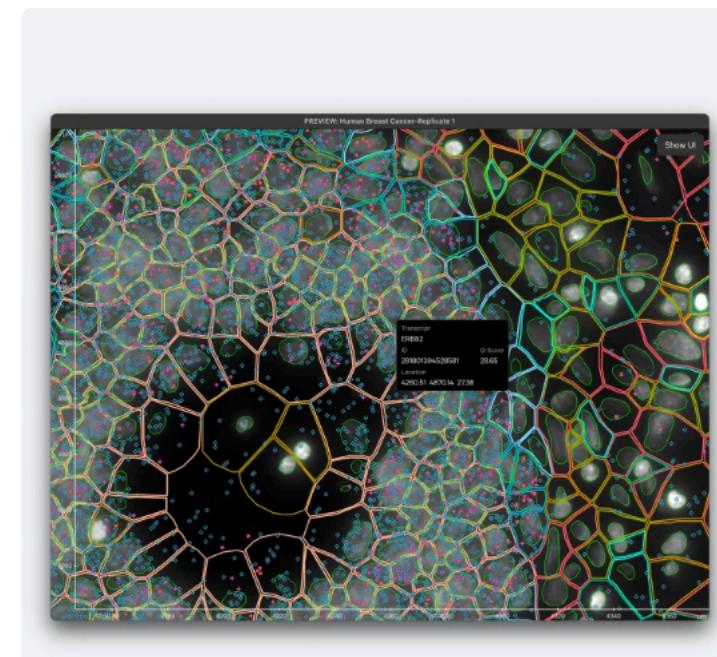
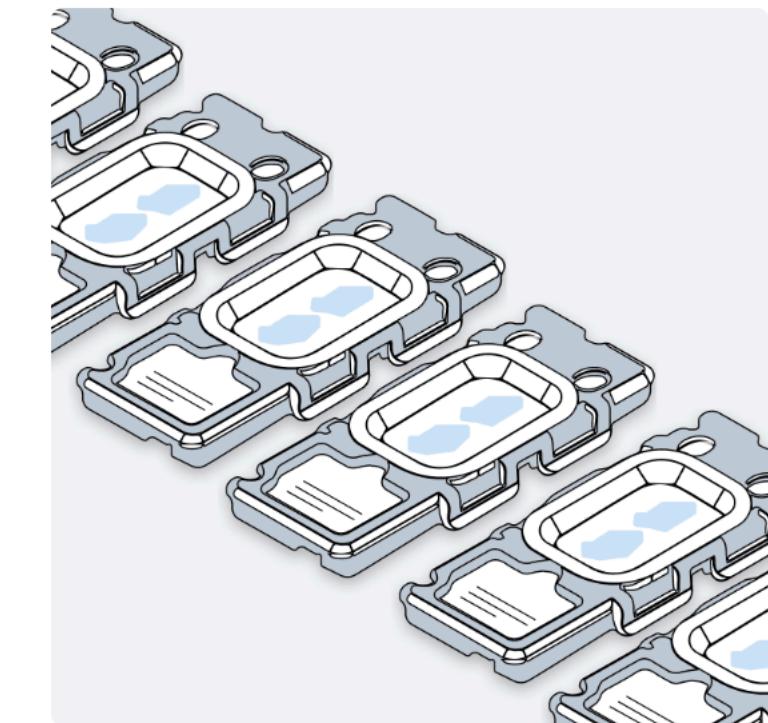
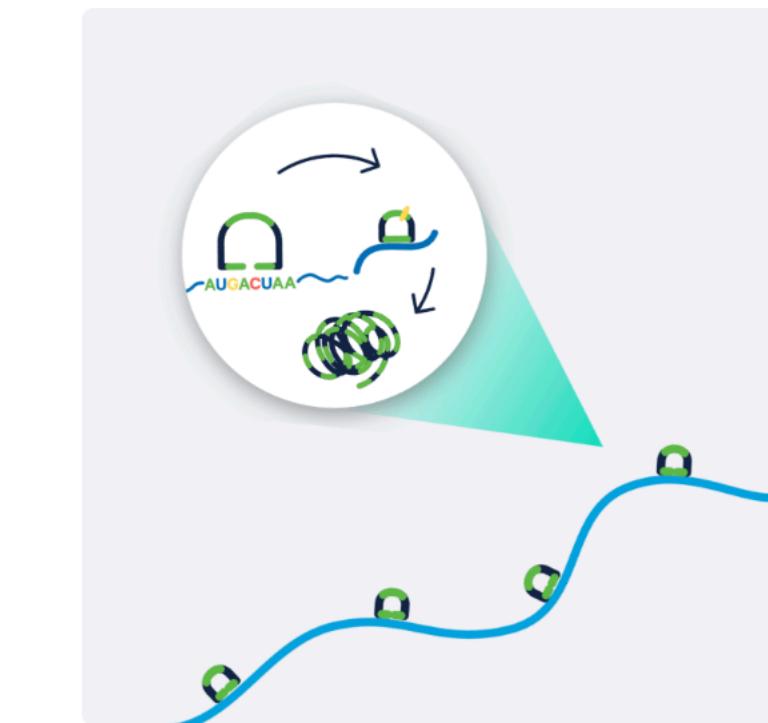
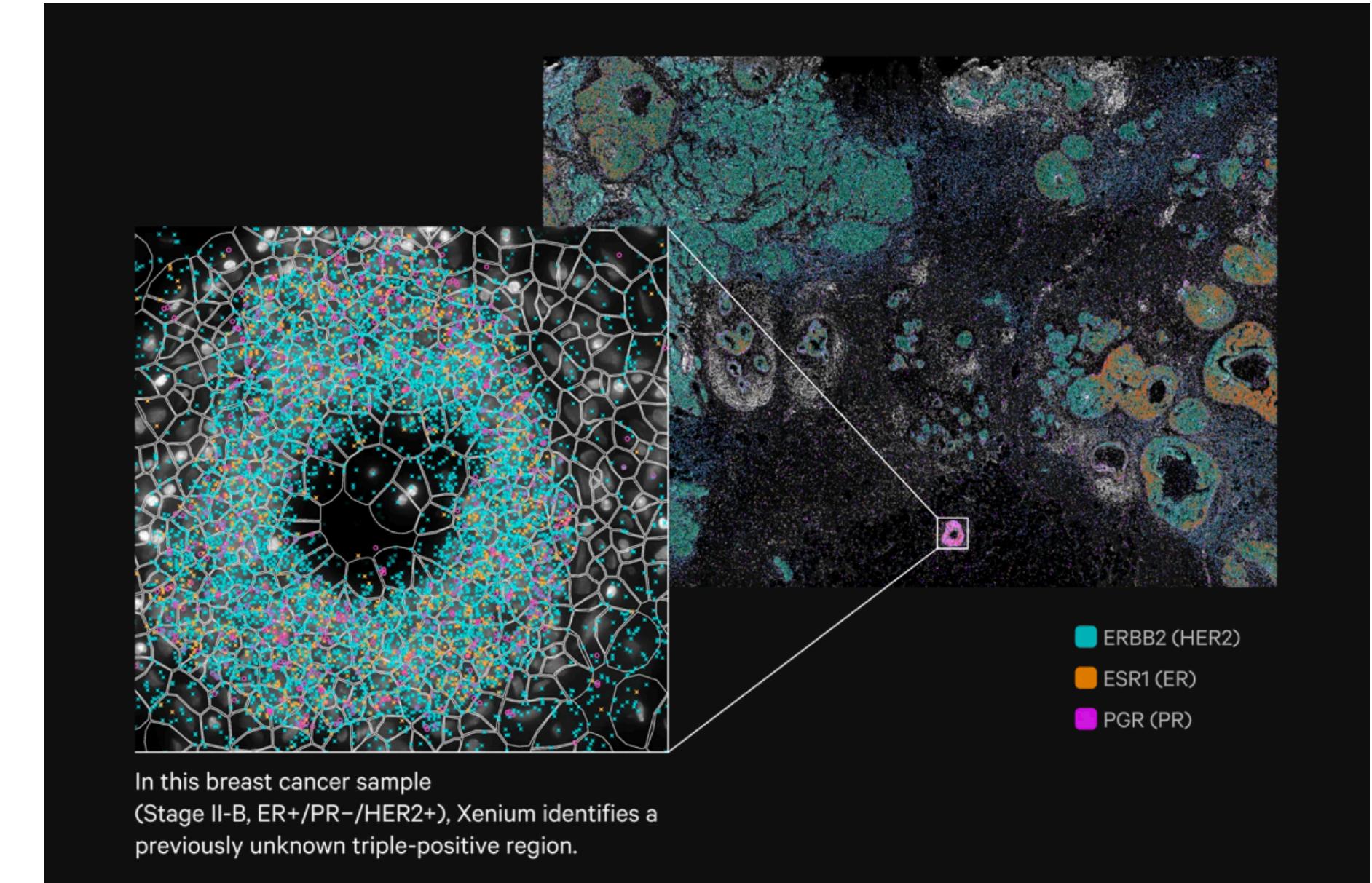
CosMX SMI for single-cell imaging delivers a comprehensive package which includes validated reagents, instrument, and data analysis software for seamless sample-to-result.



Robust *in situ* hybridization chemistry and readout



Cloud-based scalable computing and storage with interactive data viewer



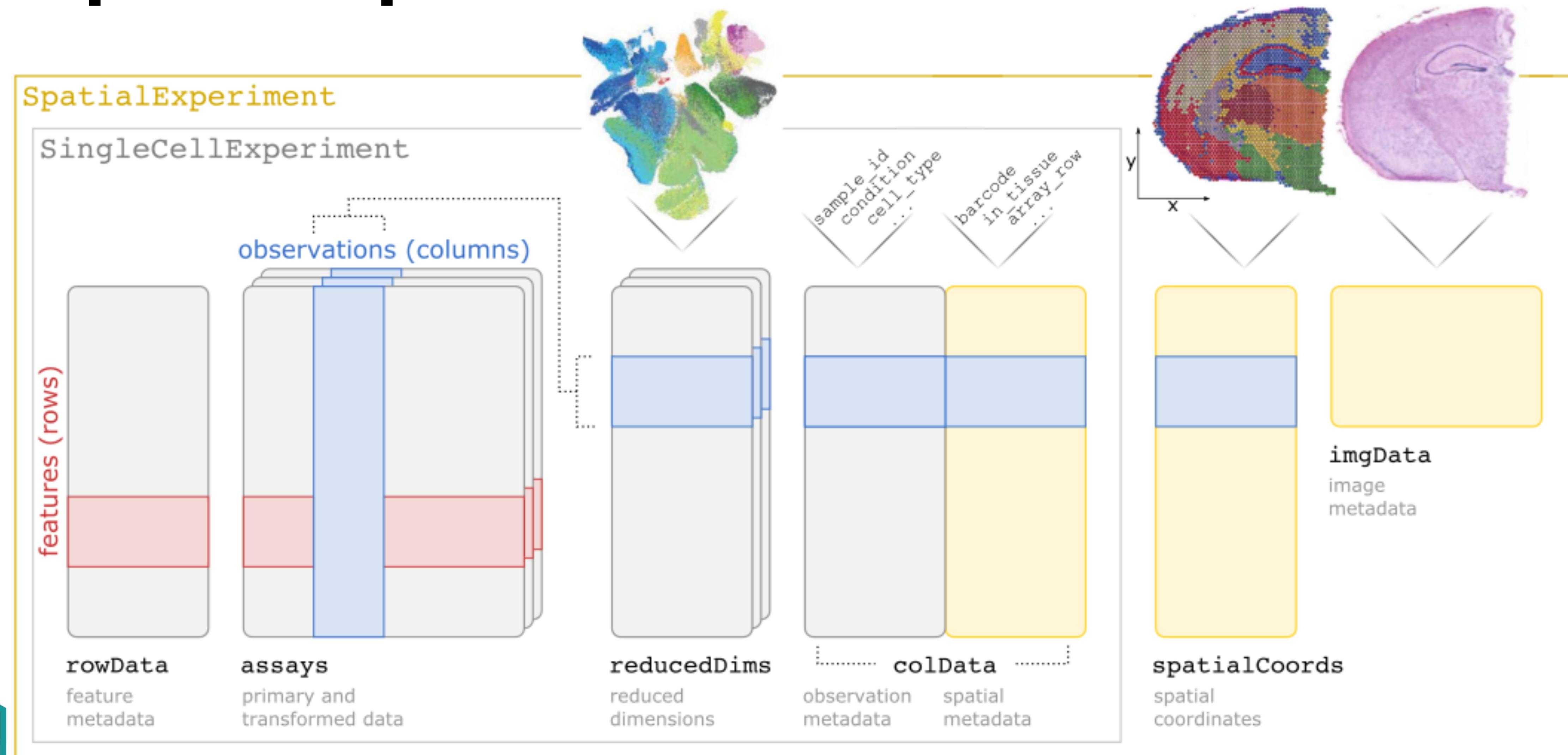
Nanostring CosMX - 10X Genomics Xenium - vizgen MERSCOPE

Not just transcriptomics

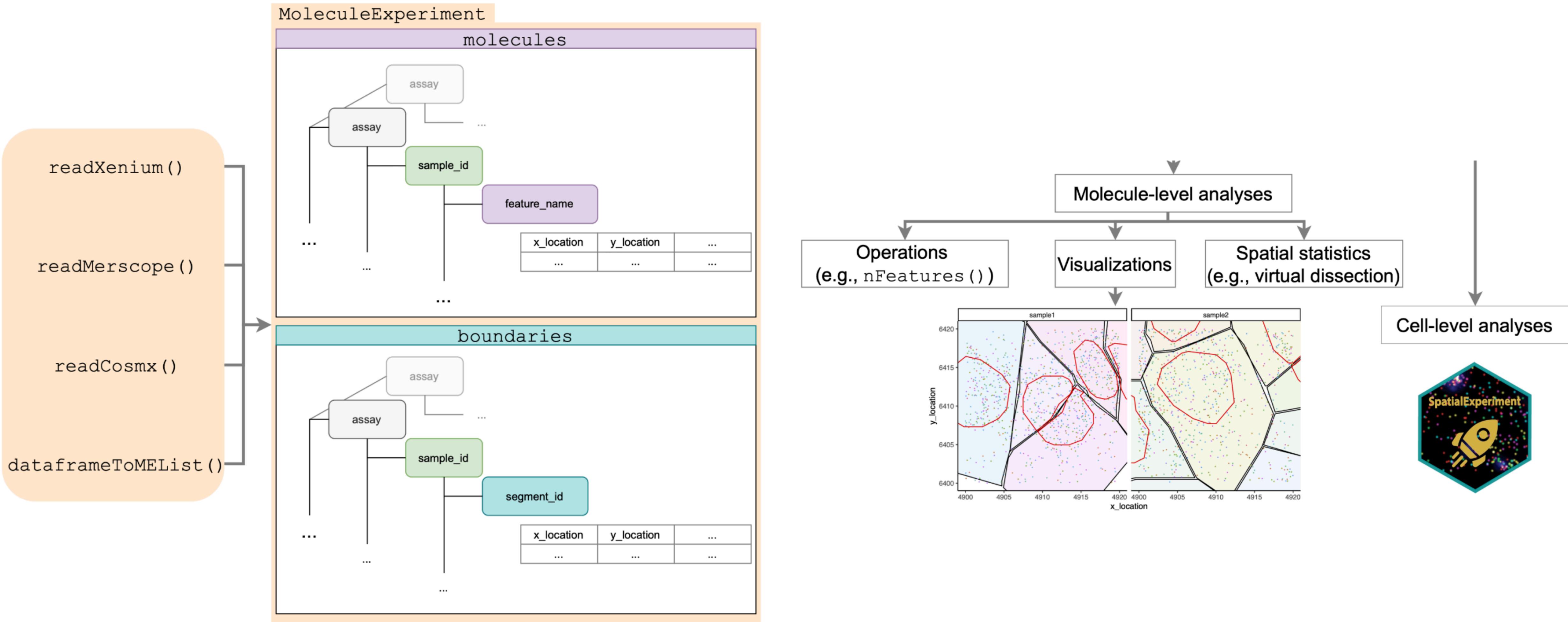
- The field of spatial proteomics is rapidly growing.
- Some methods are based on DNA-indexed antibody panels
 - E.g., CODEX (Co-Detection by Indexing) – 100+ proteins and RNA biomarkers at subcellular resolution
- Others are based on mass spectrometry (MS)
 - E.g., multiplex ion beam imaging (MIBI) and imaging mass cytometry (IMC) – up to 100 proteins at subcellular resolution

Infrastructure for spatial transcriptomics

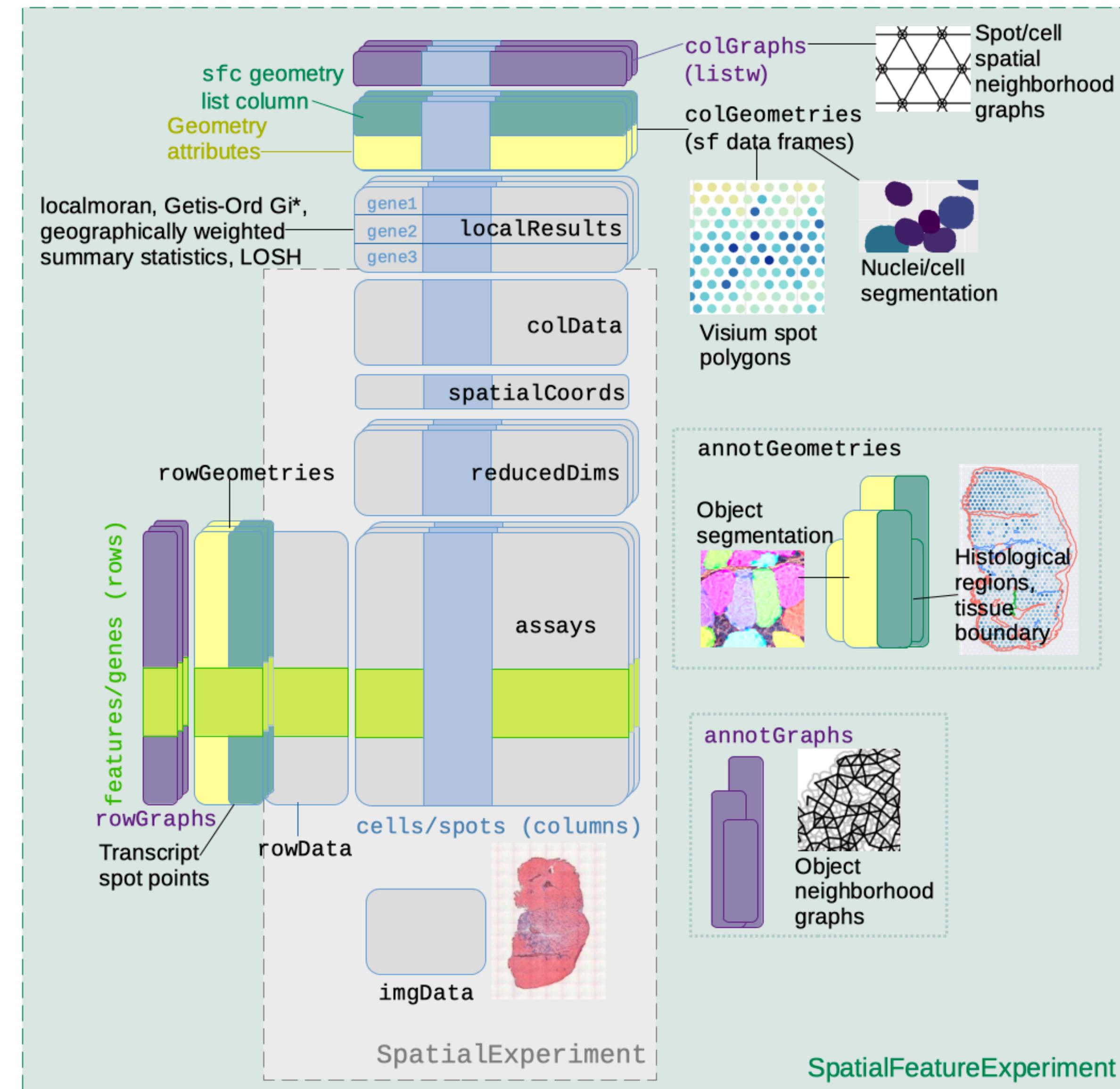
The SpatialExperiment Class



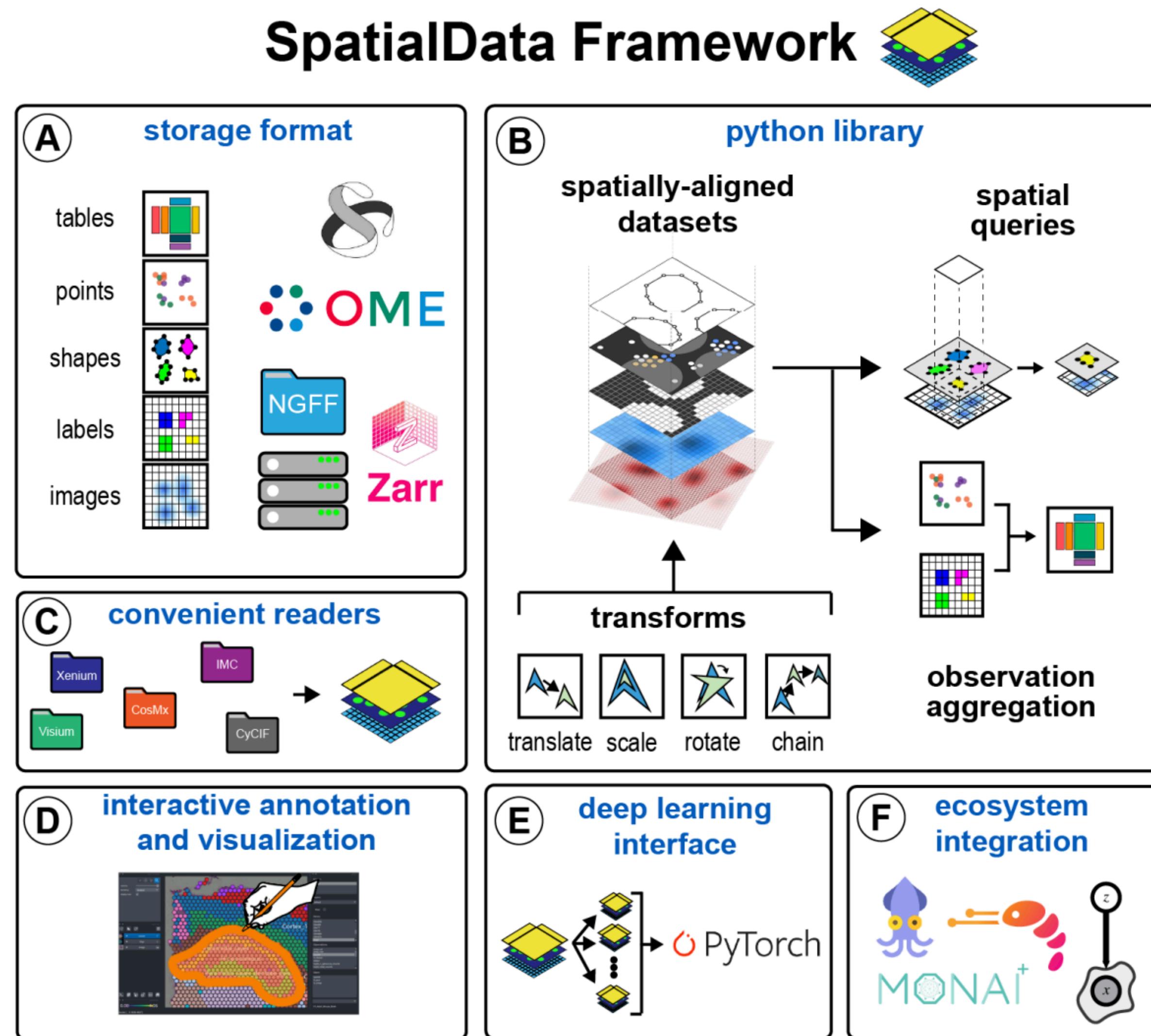
The MoleculeExperiment Class



The SpatialFeatureExperiment Class



Towards a common data structure



Spatial Data Science

Docs » Spatial Data Science with R and “terra”

Search docs

- Spatial data with *terra*
- Spatial data analysis
- Remote Sensing
- Processing MODIS data
- Case studies
- Species distribution modeling
- The *terra* package
- R companion to Geographic Information Analysis

Docs » Spatial Data Science with R and “terra”

Spatial Data Science with R and “terra”

These resources teach spatial data analysis and modeling with R. R is a widely used programming language and software environment for data science. R also provides unparalleled opportunities for analyzing spatial data and for spatial modeling.

If you have never used R, or if you need a refresher, you should start with our [Introduction to R](#).

1. Spatial data manipulation with R

Read this to learn about the basics of reading, writing, and manipulating spatial data. ([pdf](#))

2. Spatial data analysis

An introduction to methods for description, prediction and inference with spatial data. ([pdf](#))

3. Remote sensing image analysis

Introduction to remote sensing (satellite) image analysis spatial data. ([pdf](#))

4. Processing MODIS data

Introduction to remote sensing (satellite) image analysis spatial data. ([pdf](#))

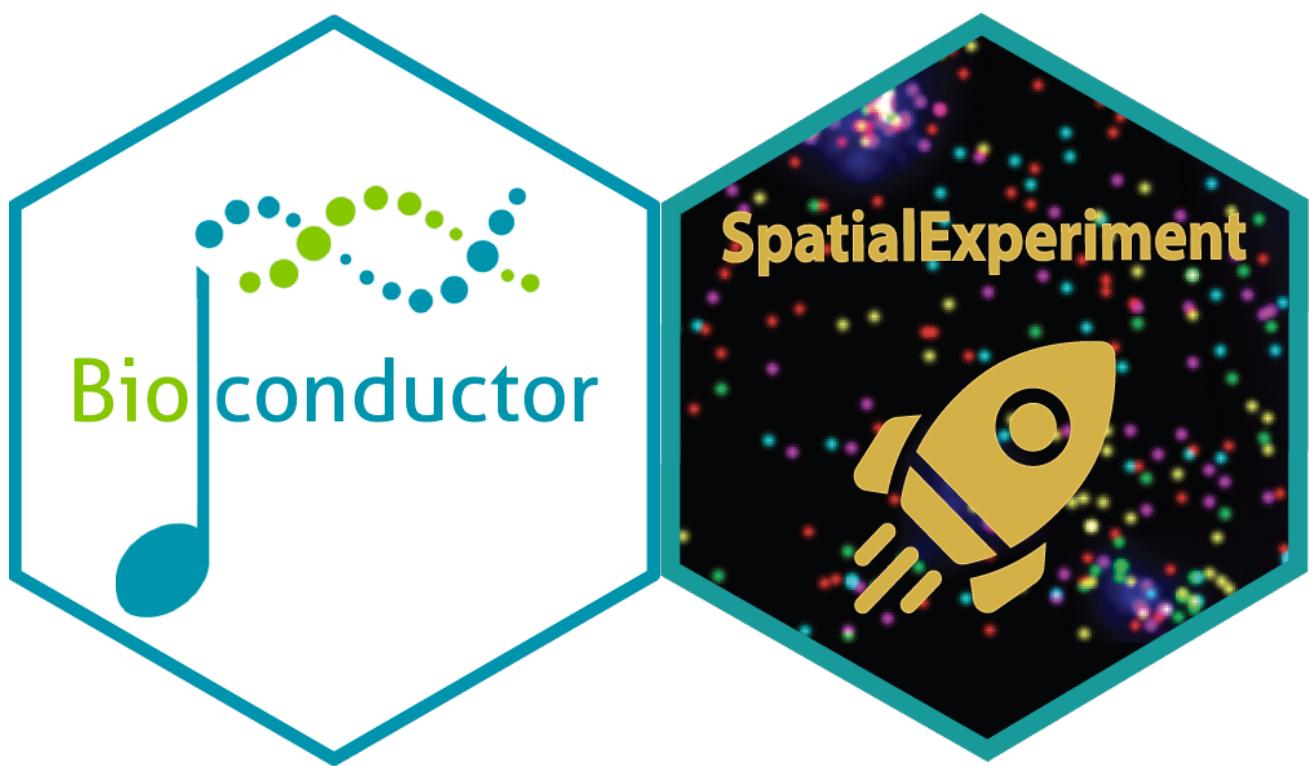
5. Case studies

A (small) collection of case studies that can help you learn more about particular topics and design your own workflows. ([pdf](#))

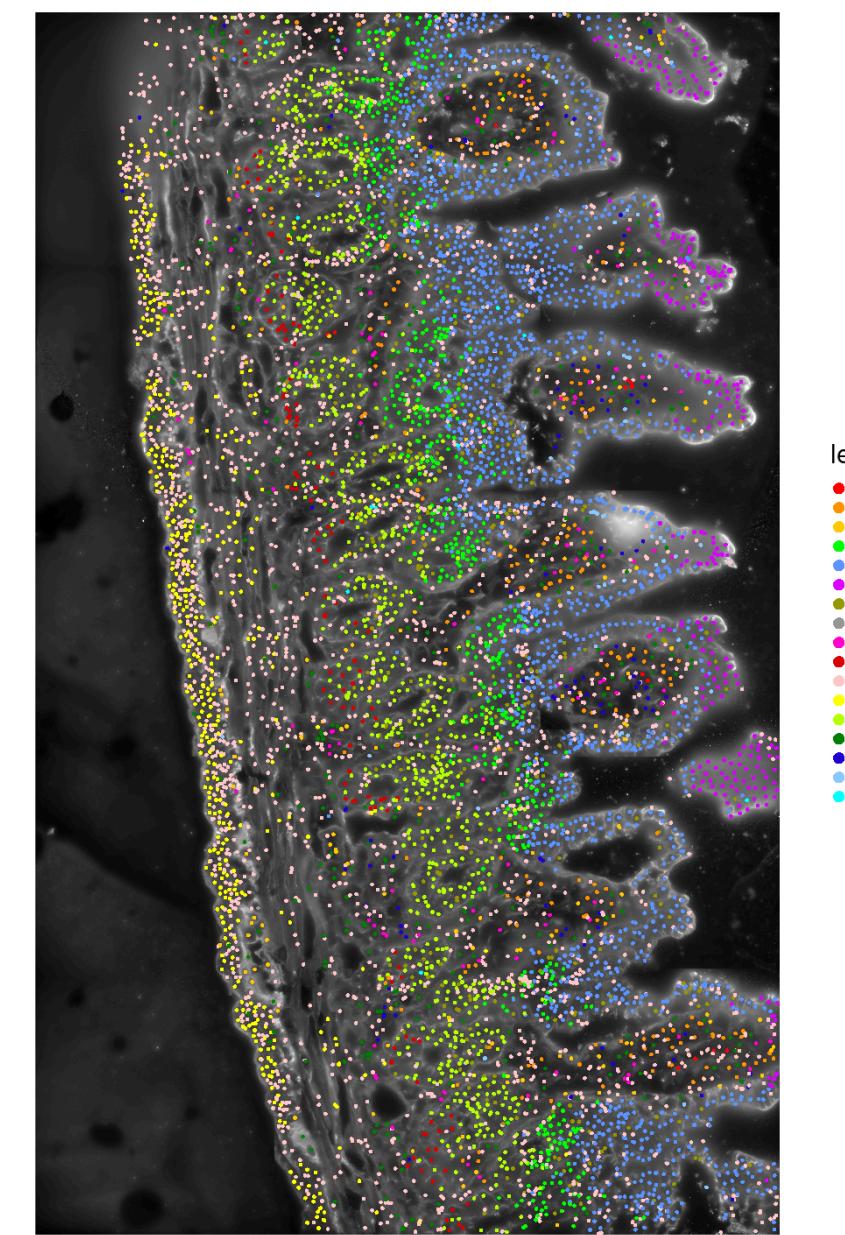
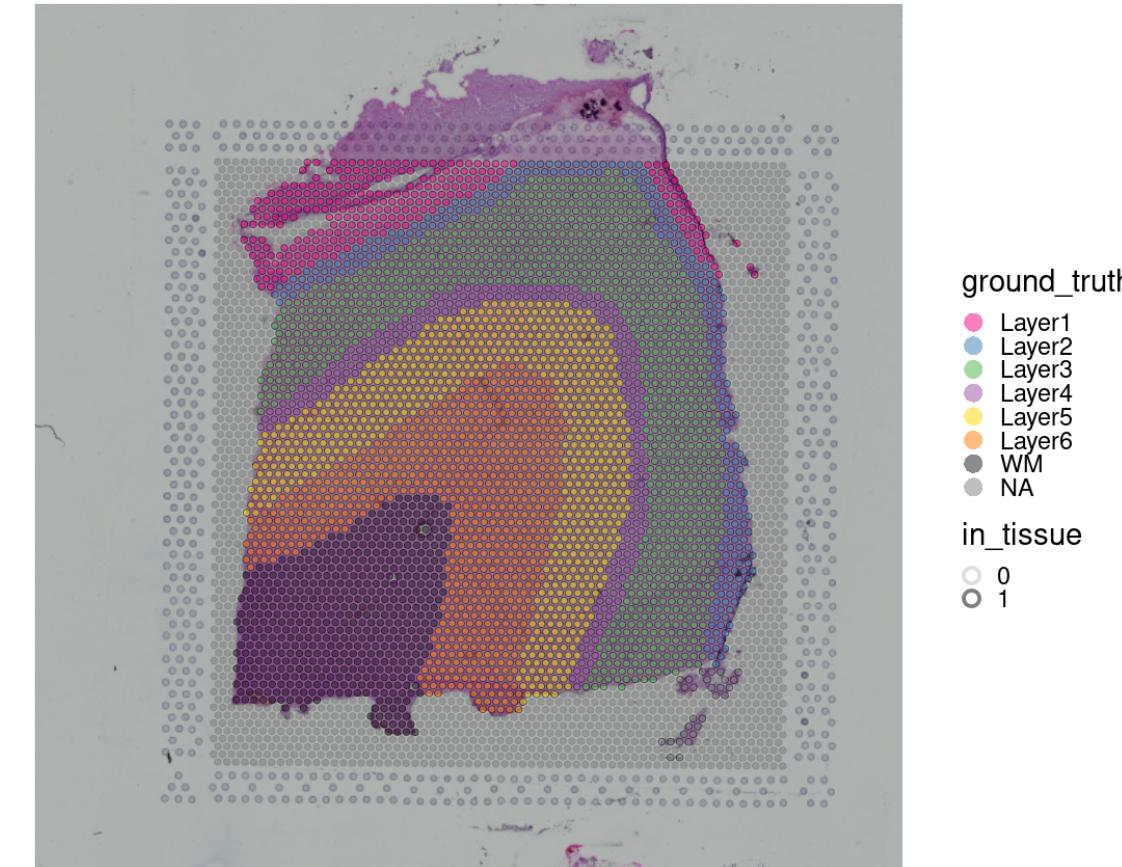
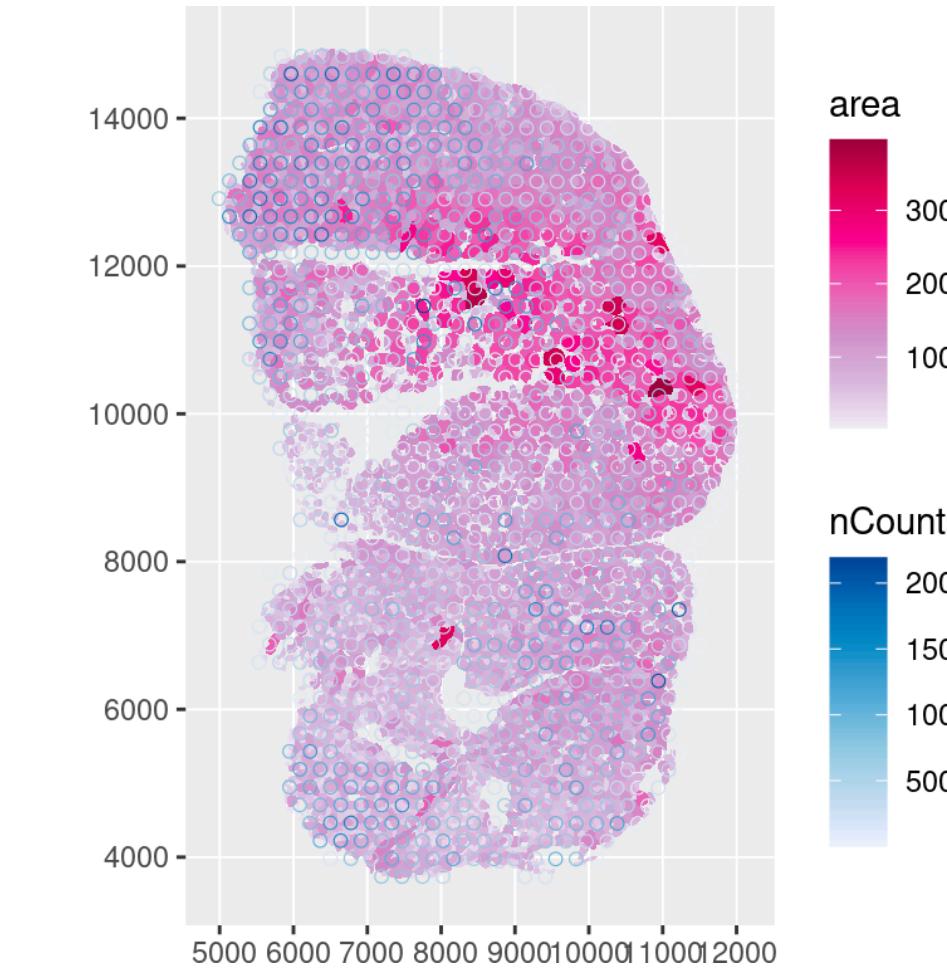
6. Species Distribution Modeling

An tutorial for predicting the geographic ranges of species (under development). ([pdf](#))

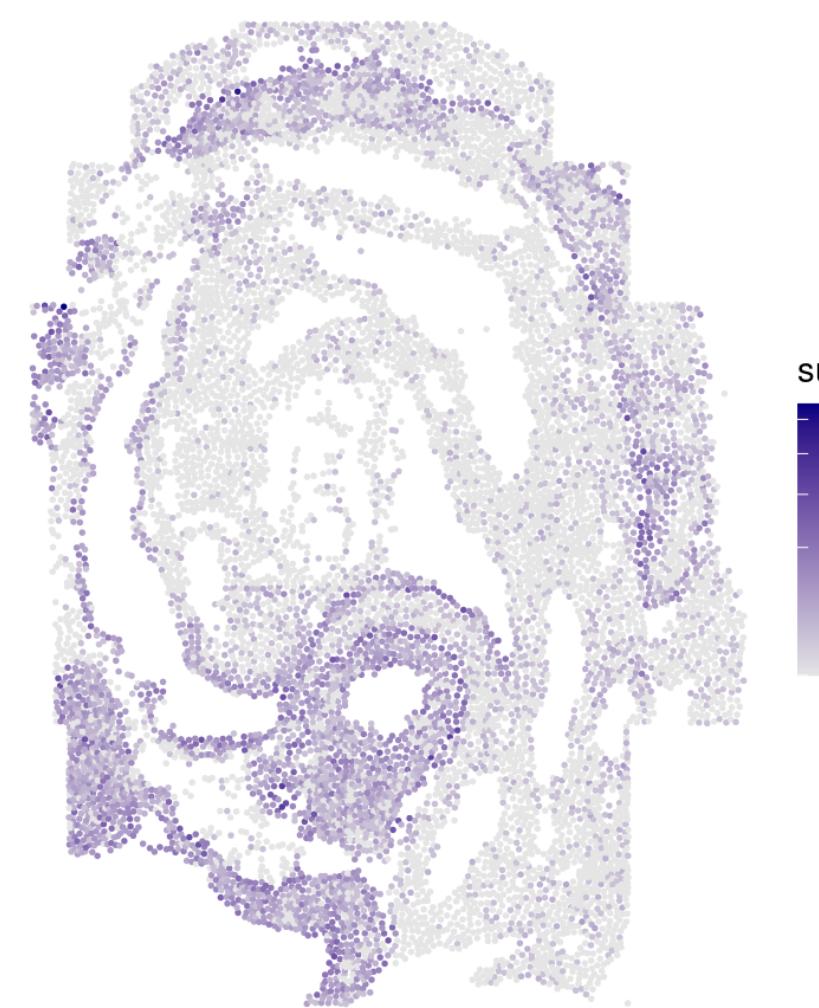
Spatial Transcriptomics Tools in Bioconductor



- 35 packages marked as “Spatial”
- 7 “SpatialData” packages (MERFISH, IMC, 10X Visium, ...)
- Deconvolution of spot expression
- Image analysis
- Visualization
- Clustering
- Spatially variable genes
- Data cleaning/normalization



Sox2



ground_truth
Layer1
Layer2
Layer3
Layer4
Layer5
Layer6
WM
NA
in_tissue
0
1

leiden_final
B (Follicular, Circulating)
B (Plasma)
Endothelial + Neuronal
Enterocyte (Bottom Villus)
Enterocyte (Mid Villus)
Enterocyte (Top Villus)
Goblet
ICC
Macrophage + DC
Paneth
Removed
Smooth Muscle
Stromal + TA
Stromal
T (CD4+)
T (CD8+)
Tuft

TENxVisiumData

Helena L. Crowell¹

¹Department of Molecular Life Sciences, University of Zurich, Zurich, Switzerland

```
library(TENxVisiumData)
spe <- HumanHeart()
```

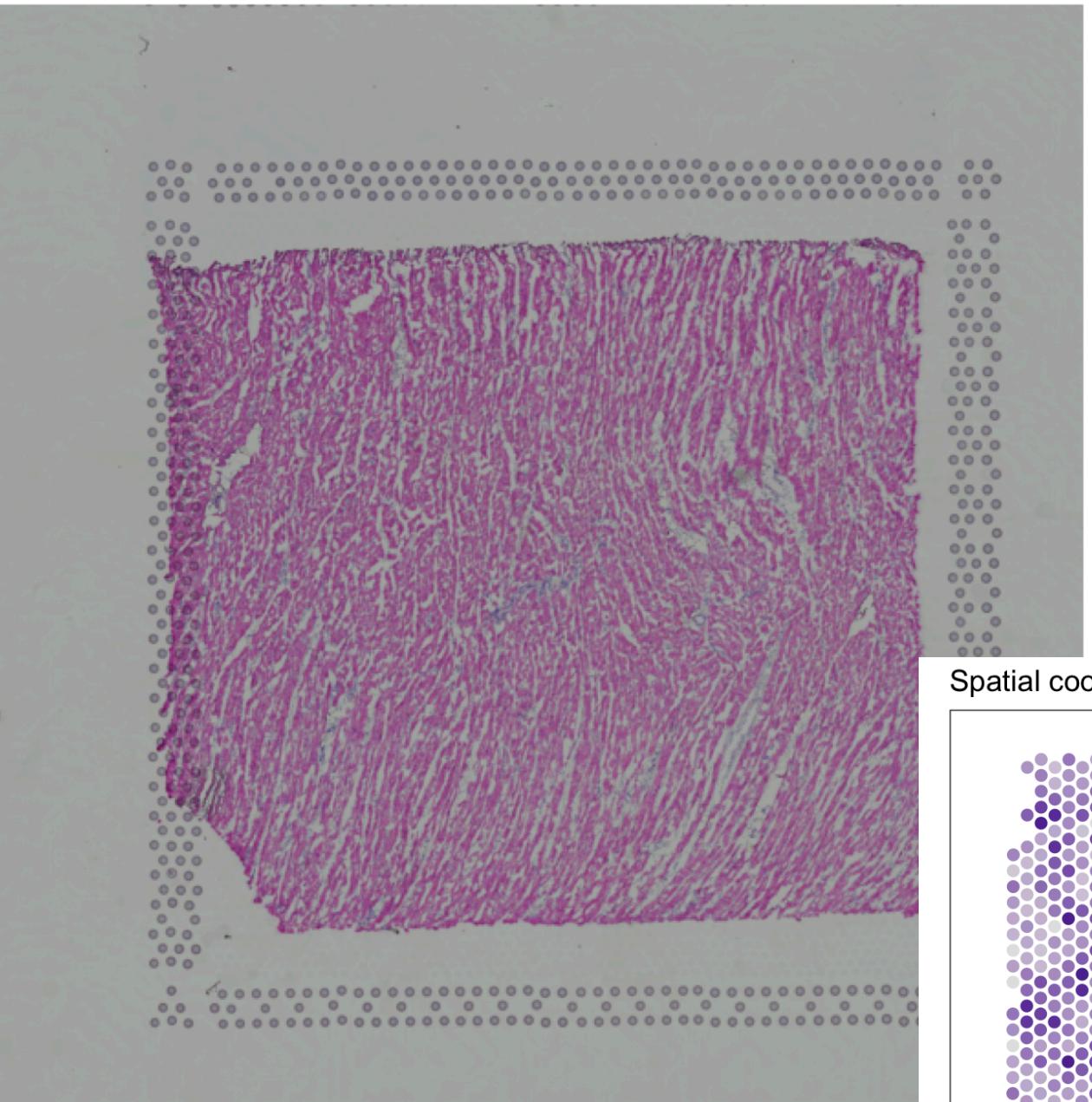
spe

```
## class: SpatialExperiment
## dim: 36601 7785
## metadata(0):
## assays(1): counts
## rownames(36601): ENSG00000243485 ENSG00000237613 ... ENSG0000027881
##   ENSG00000277196
## rowData names(1): symbol
## colnames(7785): AAACAAGTATCTCCA-1 AAACACCAATAACTGC-1 ...
##   TTGTTTGTATTACACG-1 TTGTTTGTGTAAATTG-1
## colData names(1): sample_id
## reducedDimNames(0):
## mainExpName: NULL
## altExpNames(0):
## spatialCoords names(2) : pzl_col_in_fullres pzl_row_in_fullres
## imgData names(4): sample_id image_id data scaleFactor
```

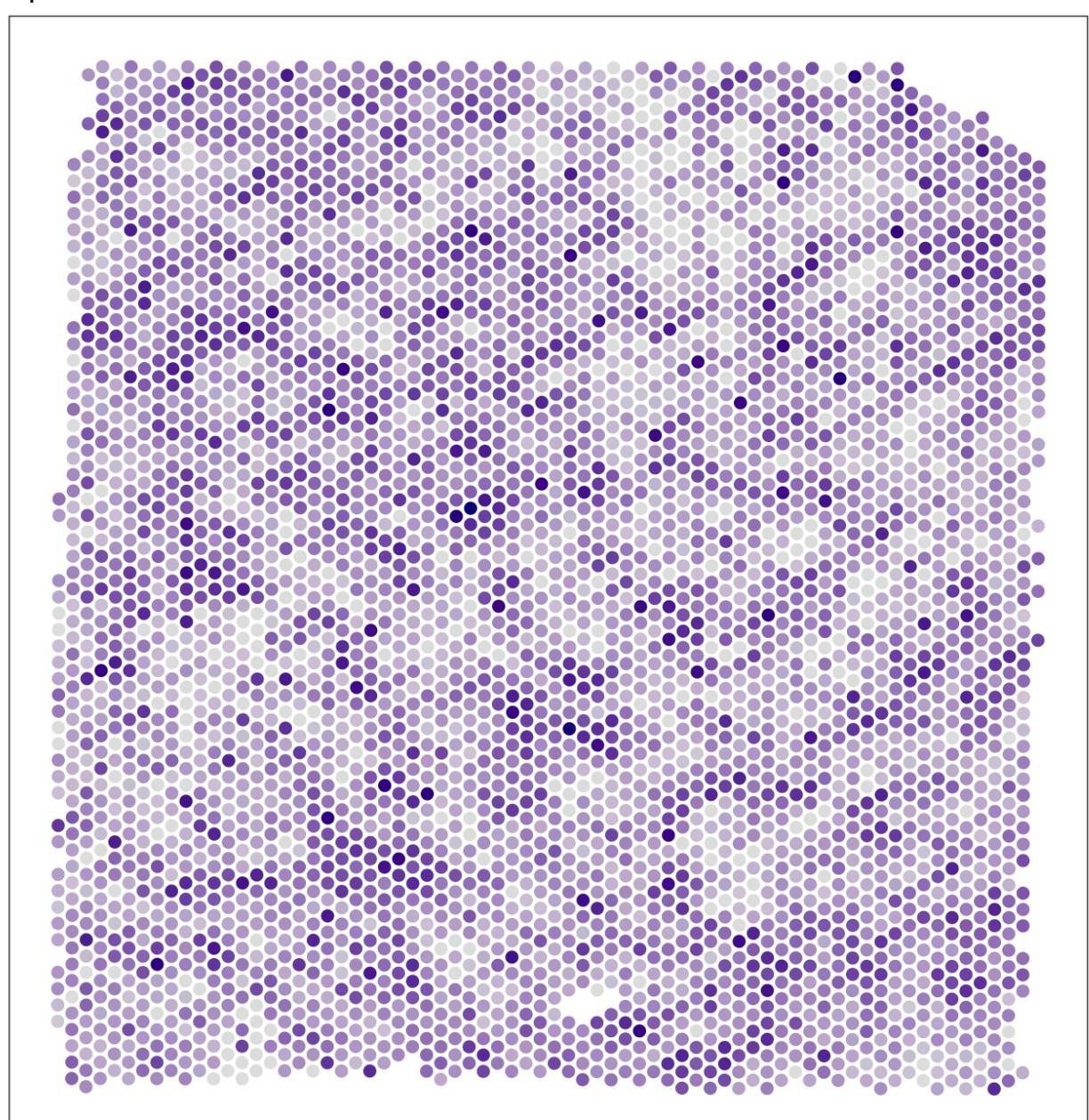
Lukas M. Weber¹ and Helena L. Crowell²

¹Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

²University of Zurich, Zurich, Switzerland



Spatial coordinates



HBB

7.5

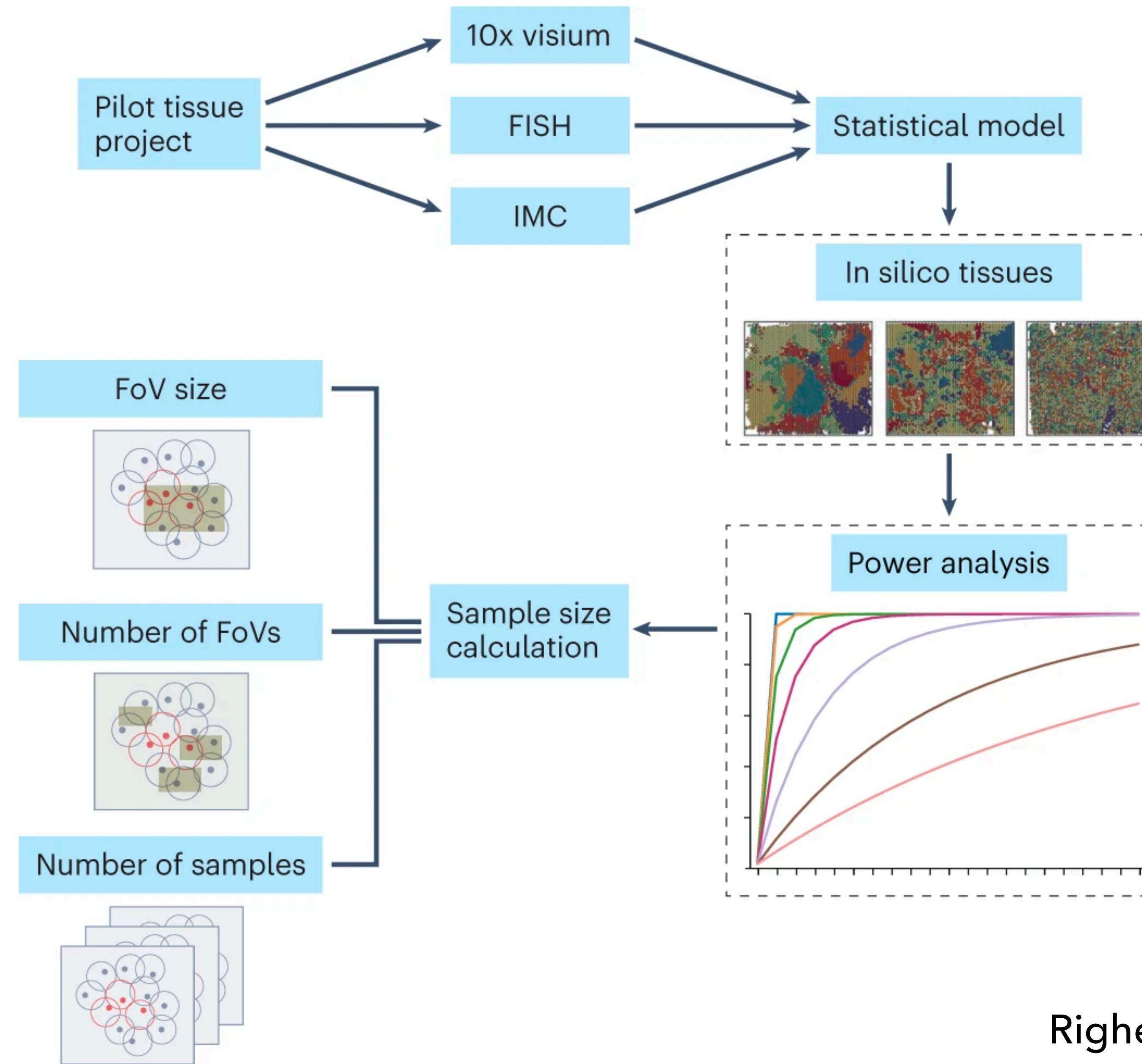
5.0

2.5

0.0

Analysis of spatial omics data

Experimental design

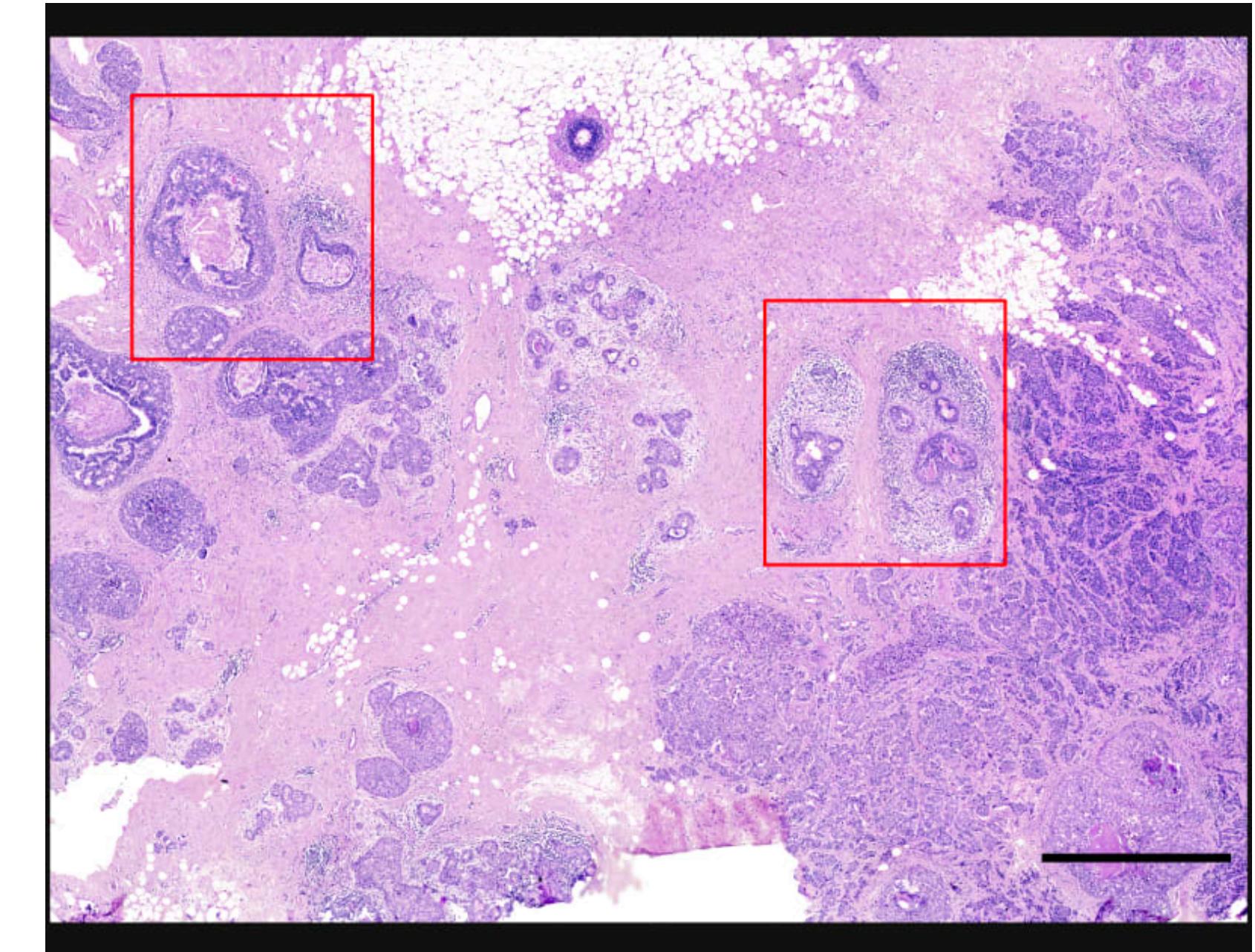


Experimental design

There is a hierarchy of replication

- ▶ Multiple donors from different cohorts,
- ▶ Multiple samples per donor
- ▶ Multiple Fields of Views (FoVs) per sample

Each level of replication involves a material cost, and it is important to find an optimal tradeoff between fully characterizing each sample versus having a large enough collection of samples, representative of the target population.



Workflow for spot-based methods

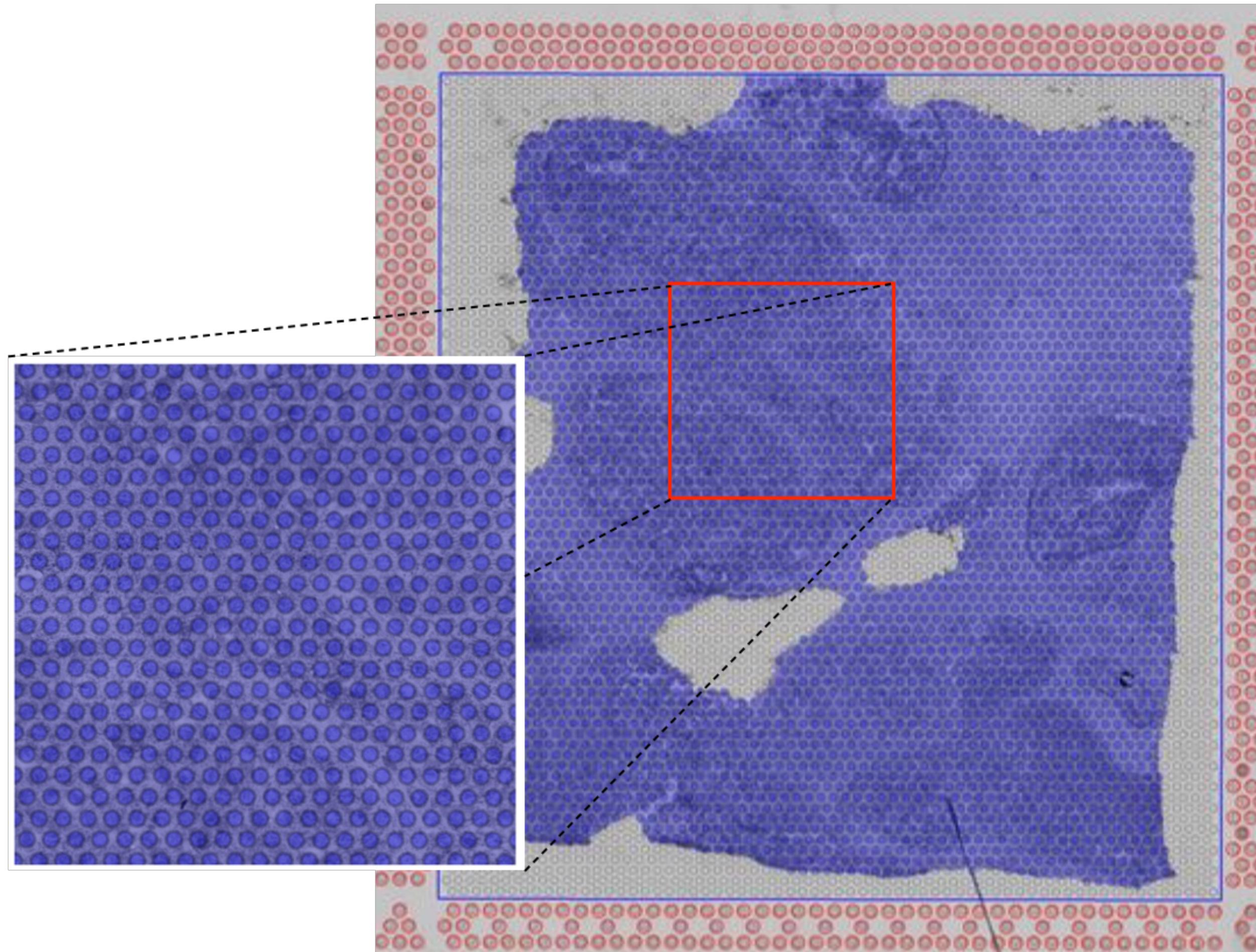
- Experimental design
- QC and preprocessing (we know a lot from RNA-seq)
- Normalization
- Cell type deconvolution
- Spatially variable genes
- Clustering (of cells and genes)

Workflow for in-situ methods

- Experimental design
- QC and preprocessing (we know a lot less!)
- Summarization at the cell level
- Normalization (technology-specific biases?)
- Cell type annotation
- Spatially variable genes
- Cell-cell communications (e.g., ligand-receptor)
- Transcript co-localization

A co-clustering model

Notation

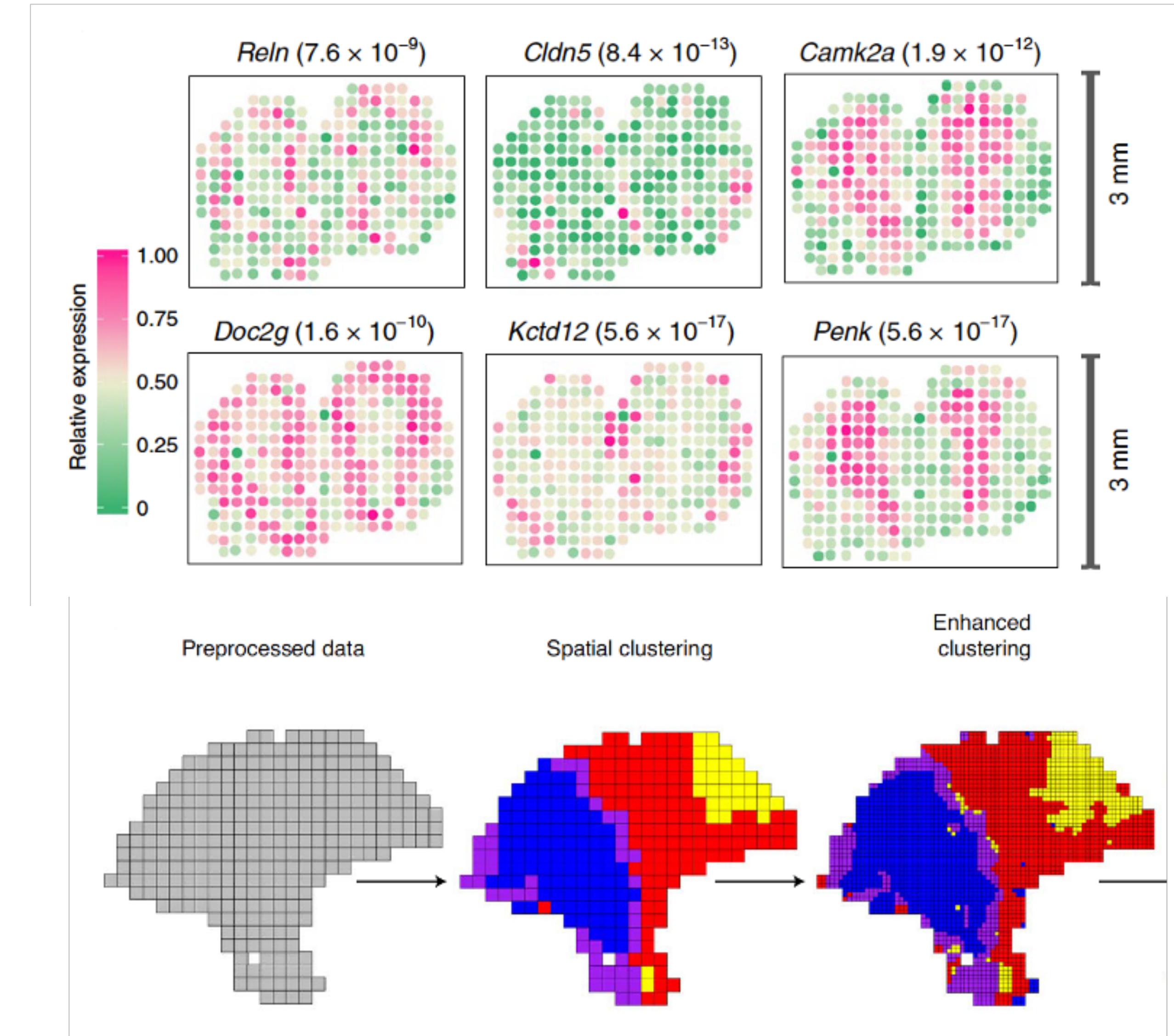


\mathbf{X} , an $n \times p$ matrix, where:

- n : the number of genes, whose expression is measured in every spot;
- p : the number of spots;
- the position of each spot on the grid is known.

Spatial transcriptomic data analysis

- **Spatially expressed (s.e.) genes:** genes whose expression in a cell is influenced by its expression in the surrounding cells ([Svensson et al., 2018](#), [Sun et al., 2020](#), ...).
- **Spatial clustering of tissue samples:** automatic recognition of cell types ([Zhao et al., 2021](#), ...).



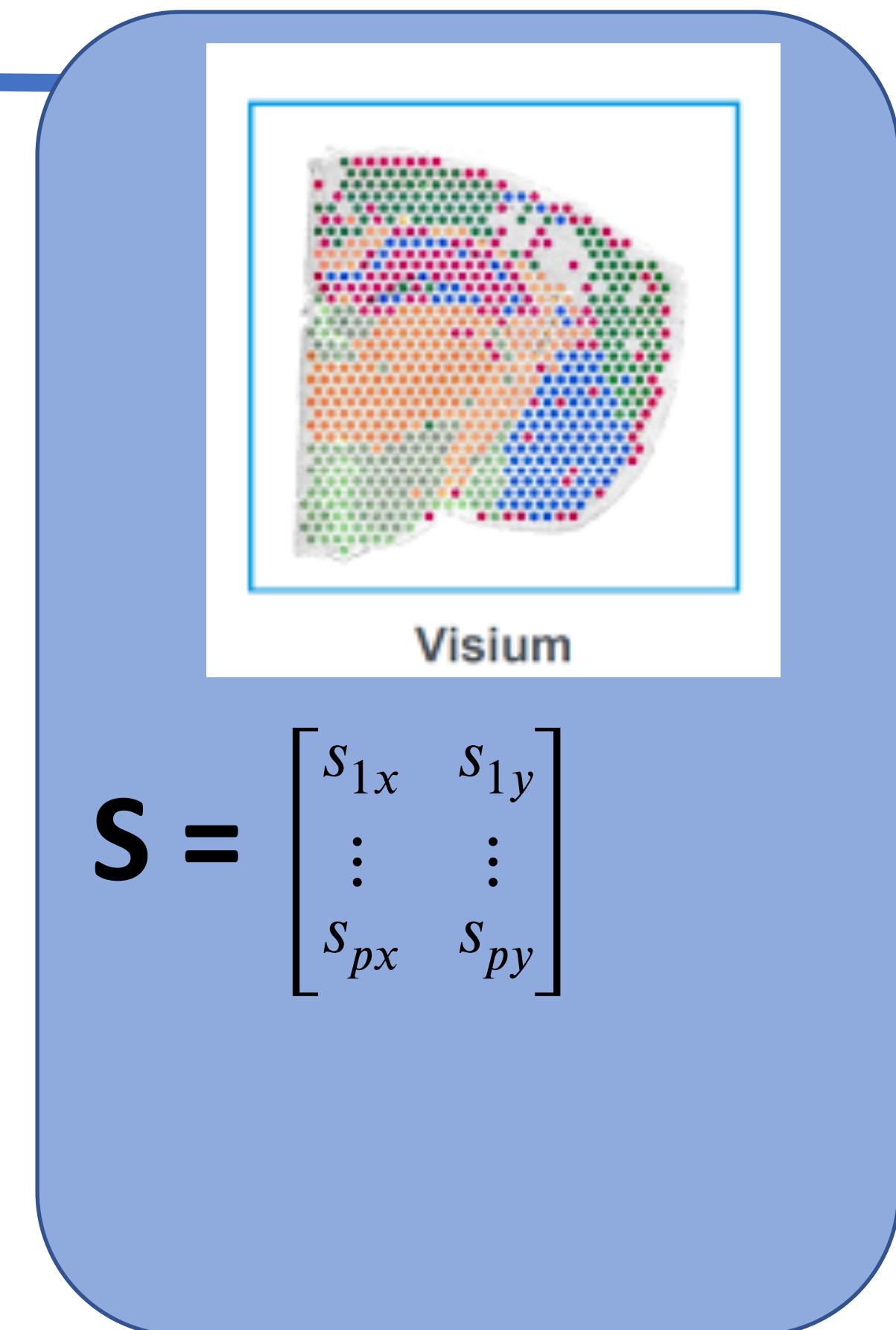
Top: different patterns of s.e. genes in H&E-stained mouse olfactory bulb sequenced with ST ([Sun et al., 2020](#)).

Bottom: melanoma ST sample ([Thrane et al., 2018](#)).

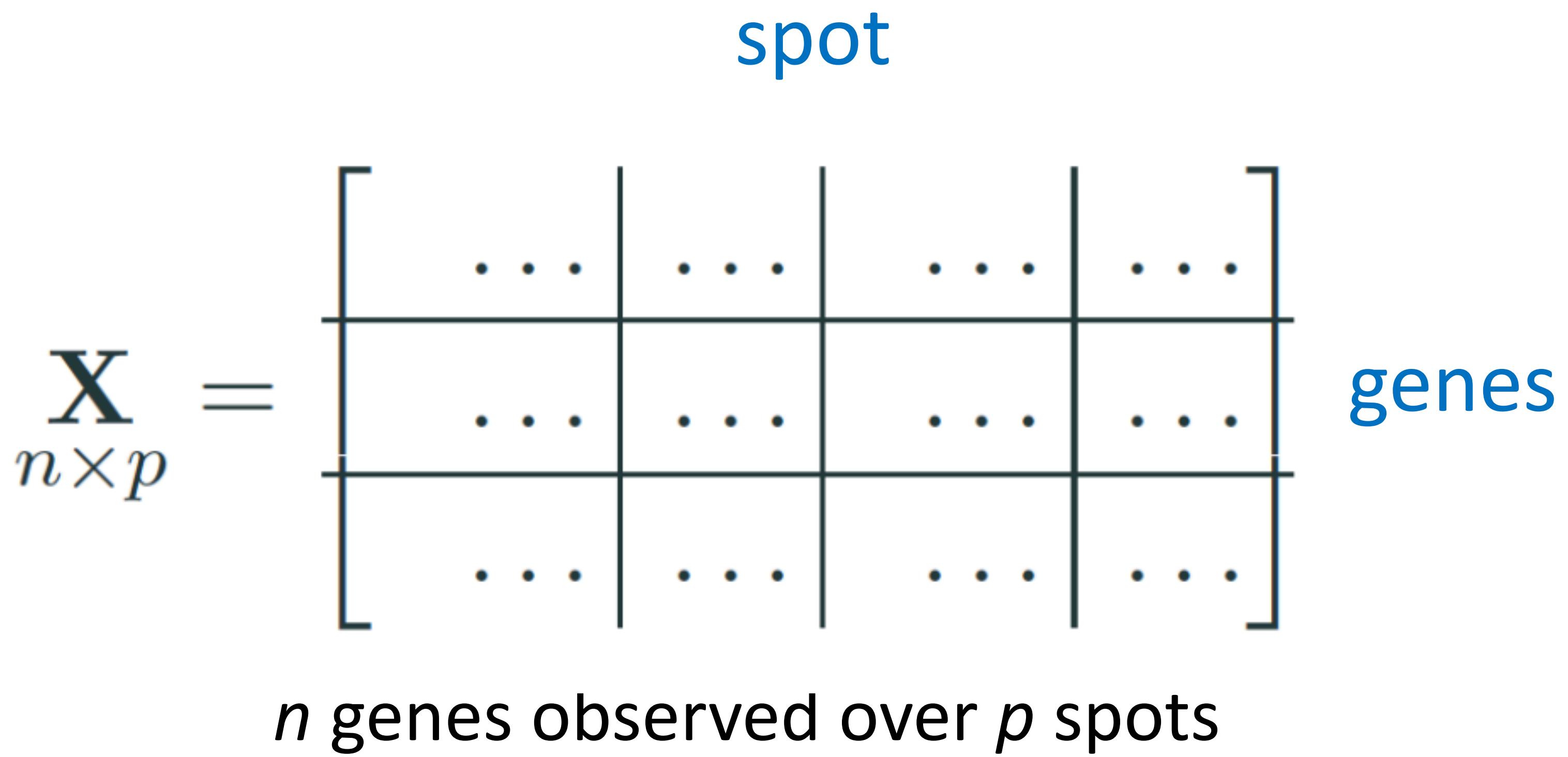
Task 1: Determine the clustering of the tissue sample according to the spatial variation of the genes

$$\mathbf{X}_{n \times p} = \begin{bmatrix} & \text{spots} \\ \cdots & | & \cdots & | & \cdots & | & \cdots \\ \cdots & | & \cdots & | & \cdots & | & \cdots \\ \cdots & | & \cdots & | & \cdots & | & \cdots \end{bmatrix} \text{genes}$$

n genes observed over p spots

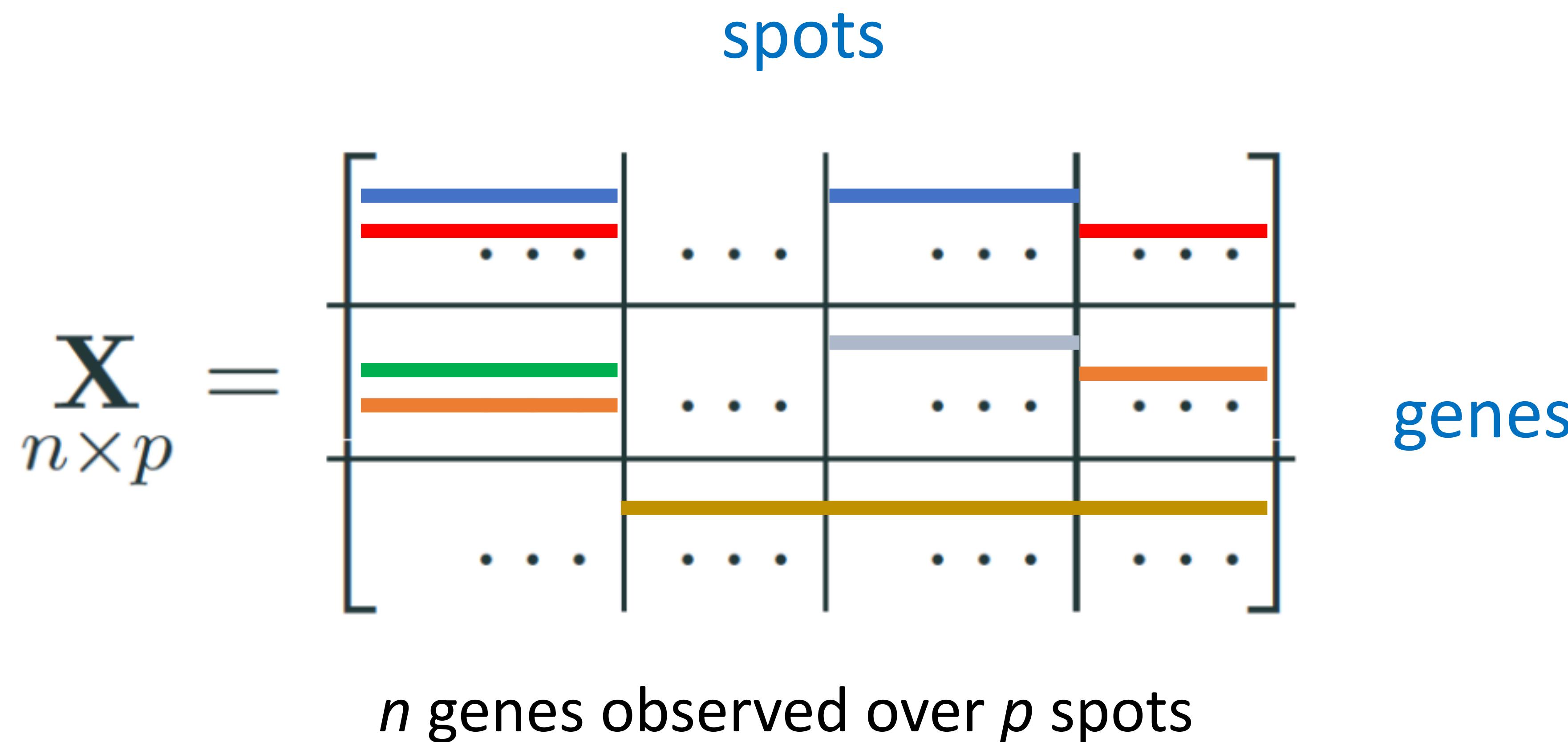


Task 2: Determine the existence of clusters of genes which are s.e. only in some of the areas discovered from Task 1



- Current methods are not designed for discovering s.e. genes in multiple clusters
- ↓
- Need of clustering methods that exploit the spatial covariance

Task 3: Determine the highly variable genes in the areas discovered from **Task 1** net of the cluster spatial effect



A co-clustering approach

$$K = R = 2$$

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{1j} & \dots & x_{np} \end{bmatrix}$$

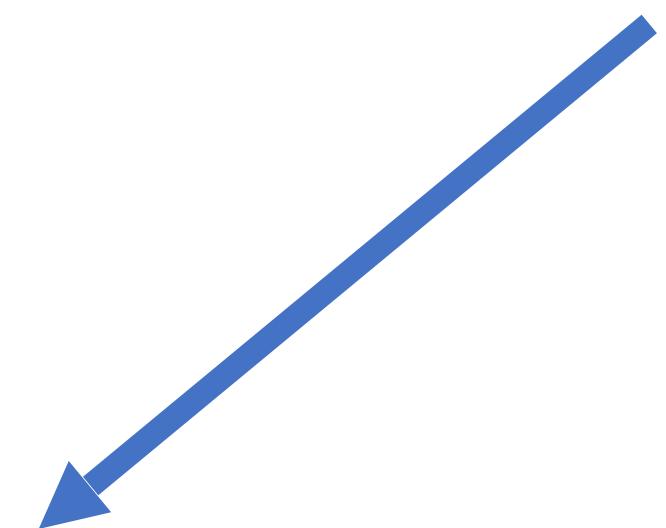
A group of genes
with a similar
expression profile in
a group of spots

(rows and columns have been reordered)

SPARTACO: Spatially Resolved Transcriptomic Co-clustering

$$\mathbf{X}_{n \times p} = \begin{bmatrix} \mathbf{X}^{11} & \dots & \mathbf{X}^{1r} & \dots & \mathbf{X}^{1R} \\ \vdots & & \vdots & & \vdots \\ \mathbf{X}^{k1} & \dots & \mathbf{X}^{kr} & \dots & \mathbf{X}^{kR} \\ \vdots & & \vdots & & \vdots \\ \mathbf{X}^{K1} & \dots & \mathbf{X}^{Kr} & \dots & \mathbf{X}^{KR} \end{bmatrix}$$

Gene-specific variance



$$\mathbf{x}_{i \cdot}^{kr} \sim \mathcal{N}(\mu_{kr} \mathbf{1}, \sigma_{kr,i}^2 \Delta_{kr})$$

$$\sigma_{kr,i}^2 \sim \mathcal{IG}(\alpha_{kr}, \beta_{kr})$$

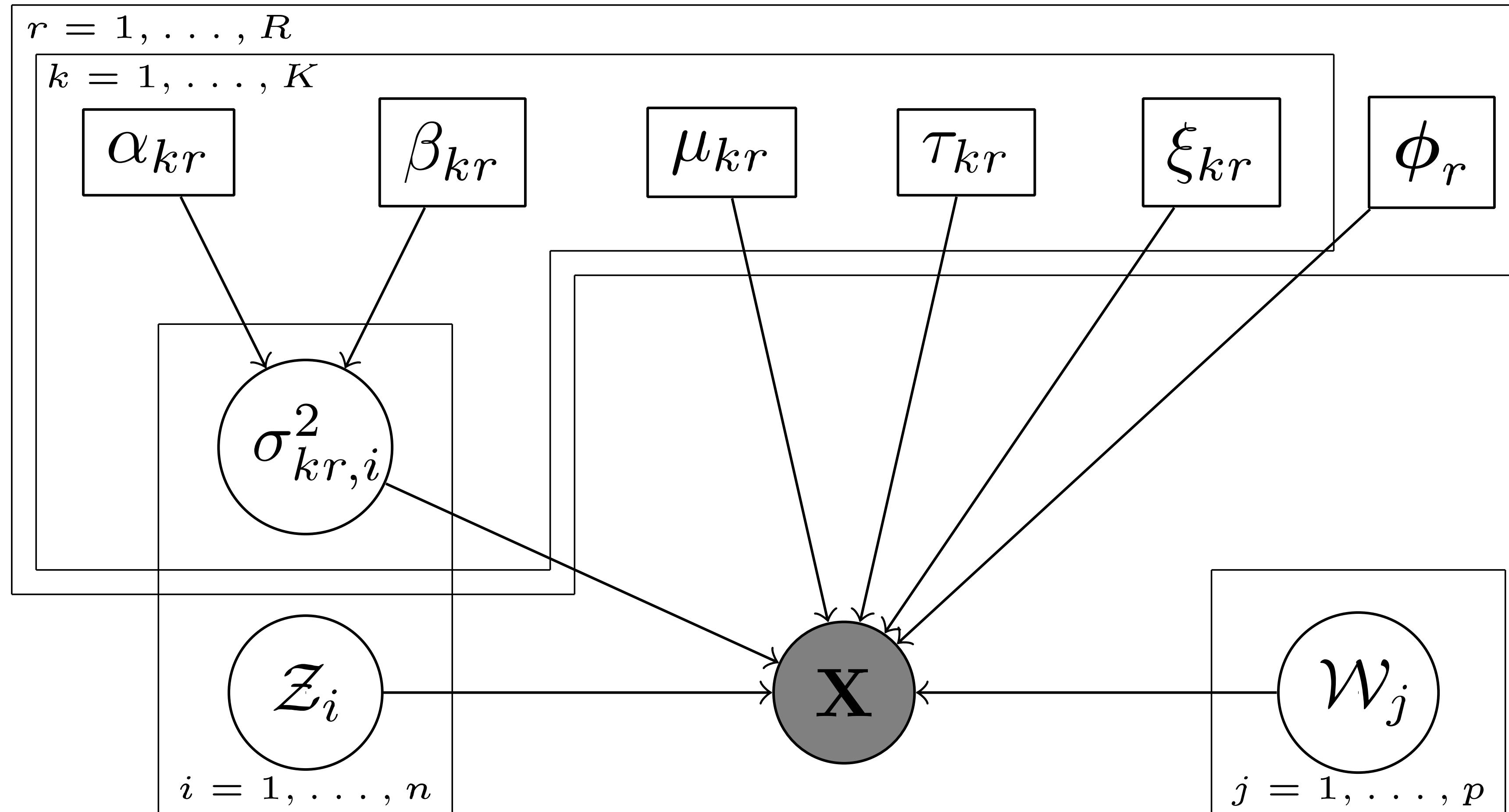
$$\Delta_{kr} = \tau_{kr} \mathbf{K}(S^r; \phi_r) + \xi_{kr} \mathbf{I}$$

Spatial variability

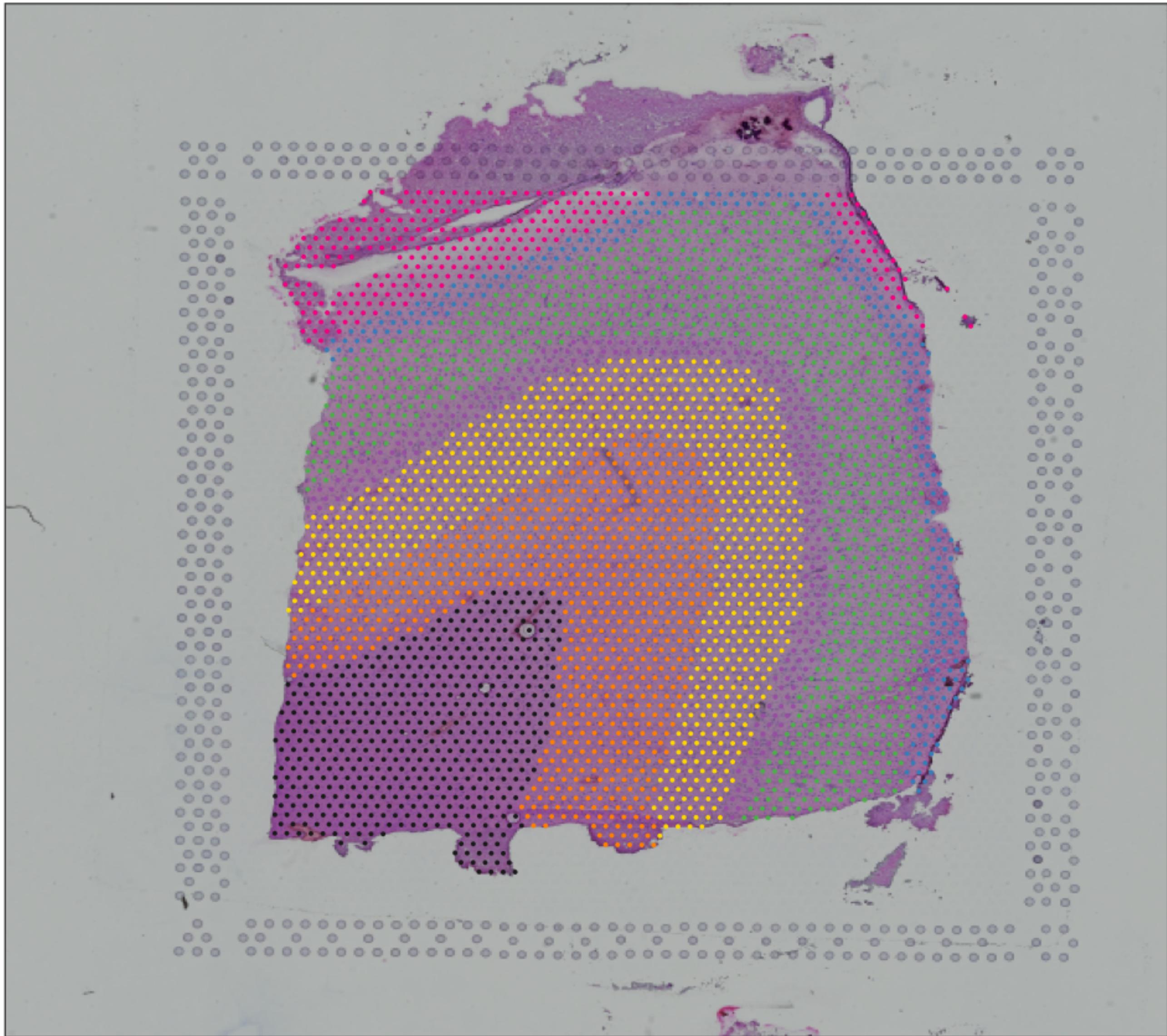


Nugget effect
(residual variance)

SPARTACO: Spatially Resolved Transcriptomic Co-clustering



Subject 151673

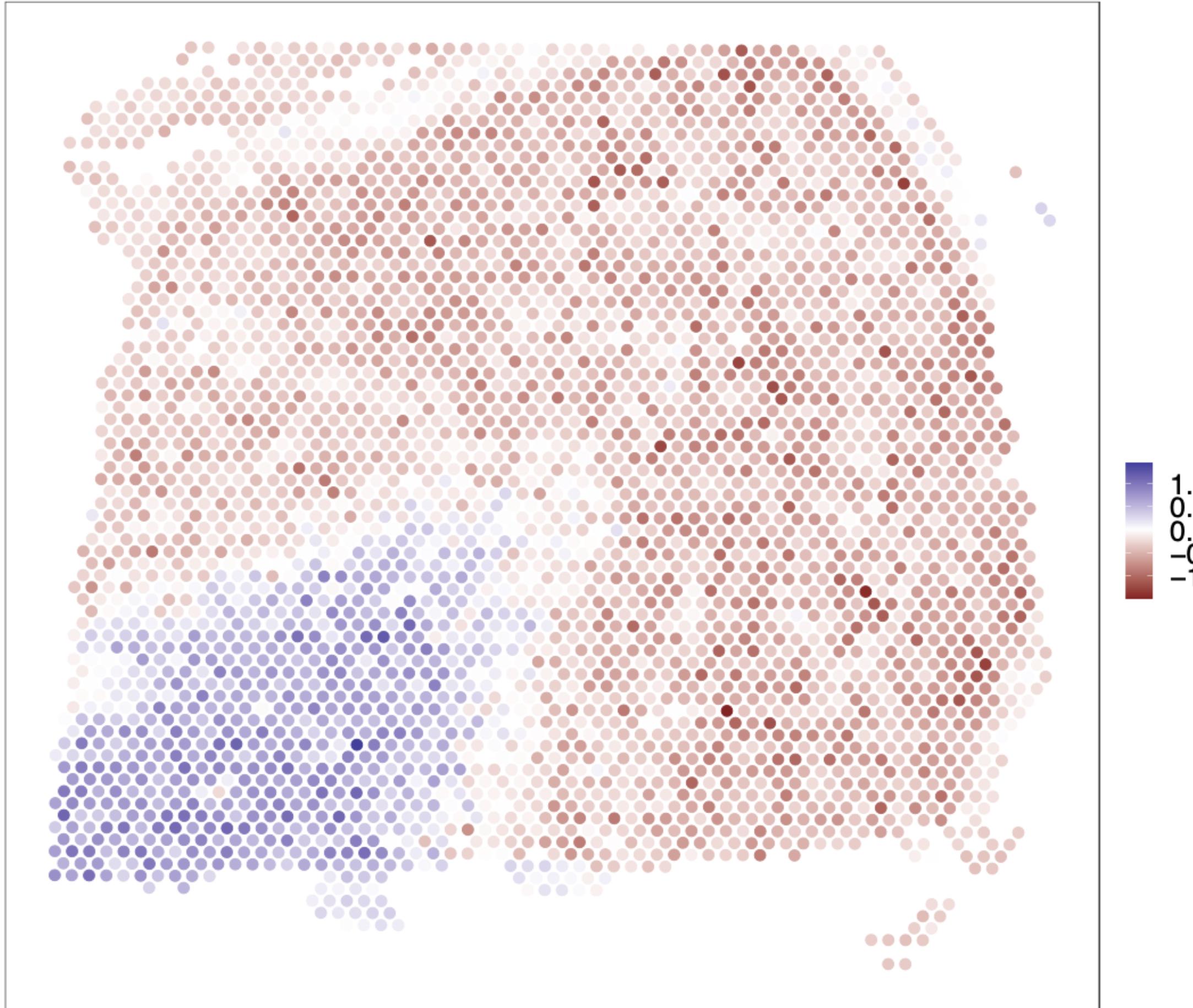


Brain sample with ID 151673 from spatialLIBD package.

- 33,538 genes measured over 3,639 spots.
- To exclude uninformative genes, we applied the gene selection procedure for UMI counts proposed by [Townes et al. \(2019\)](#) to keep the first 500 genes.
- We transform the data using the deviance residuals.
- Our information criterion based on the penalized likelihood selects **K=2 gene clusters and R=9 spot clusters**.

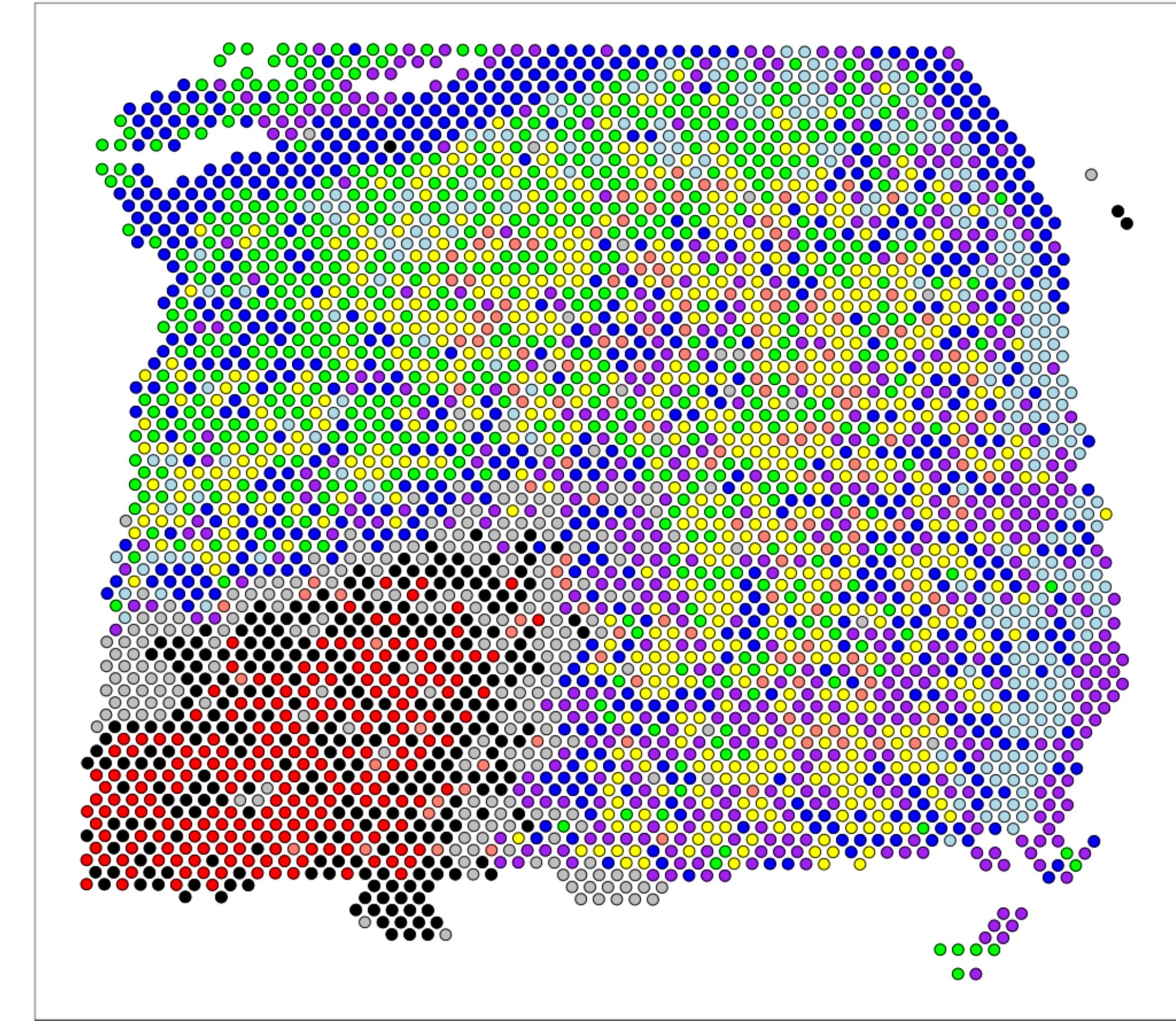
Leonardo Collado-Torres

Gene clustering



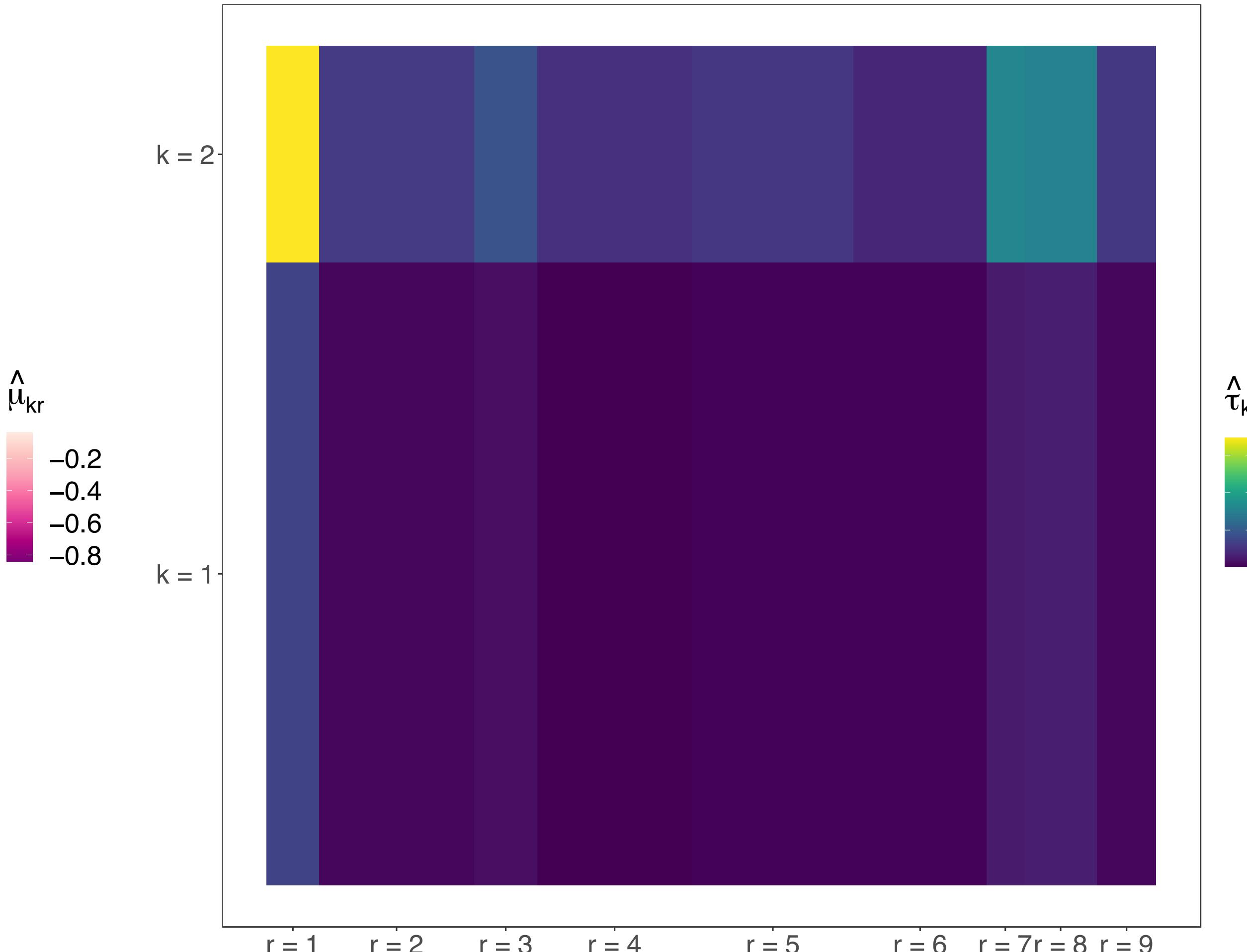
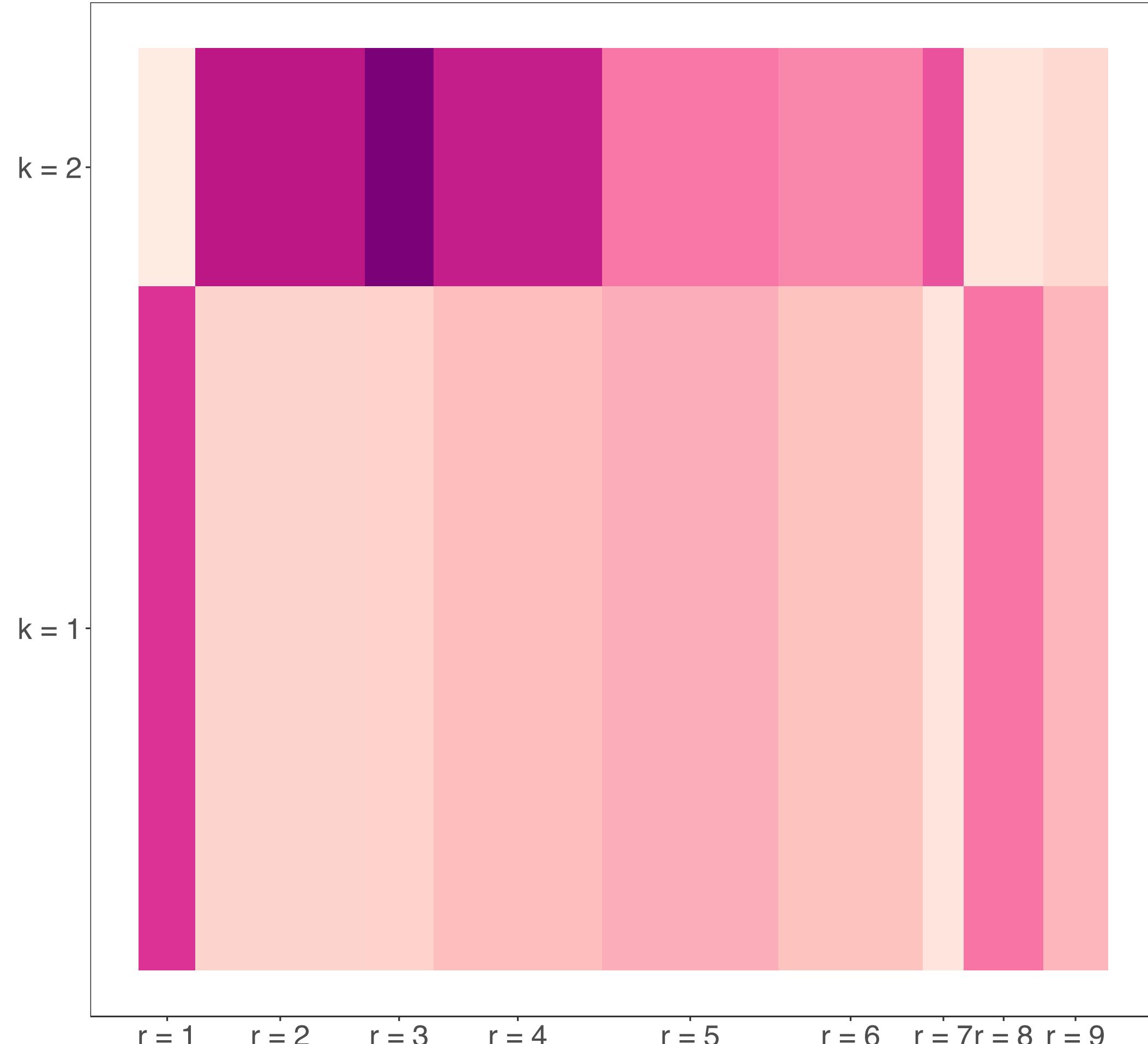
Gene clustering: average expression of gene cluster 2

Spot clustering



Spot clustering: the model discovers 9 regions across the tissue with different spatial variation profiles.

Co-clustering



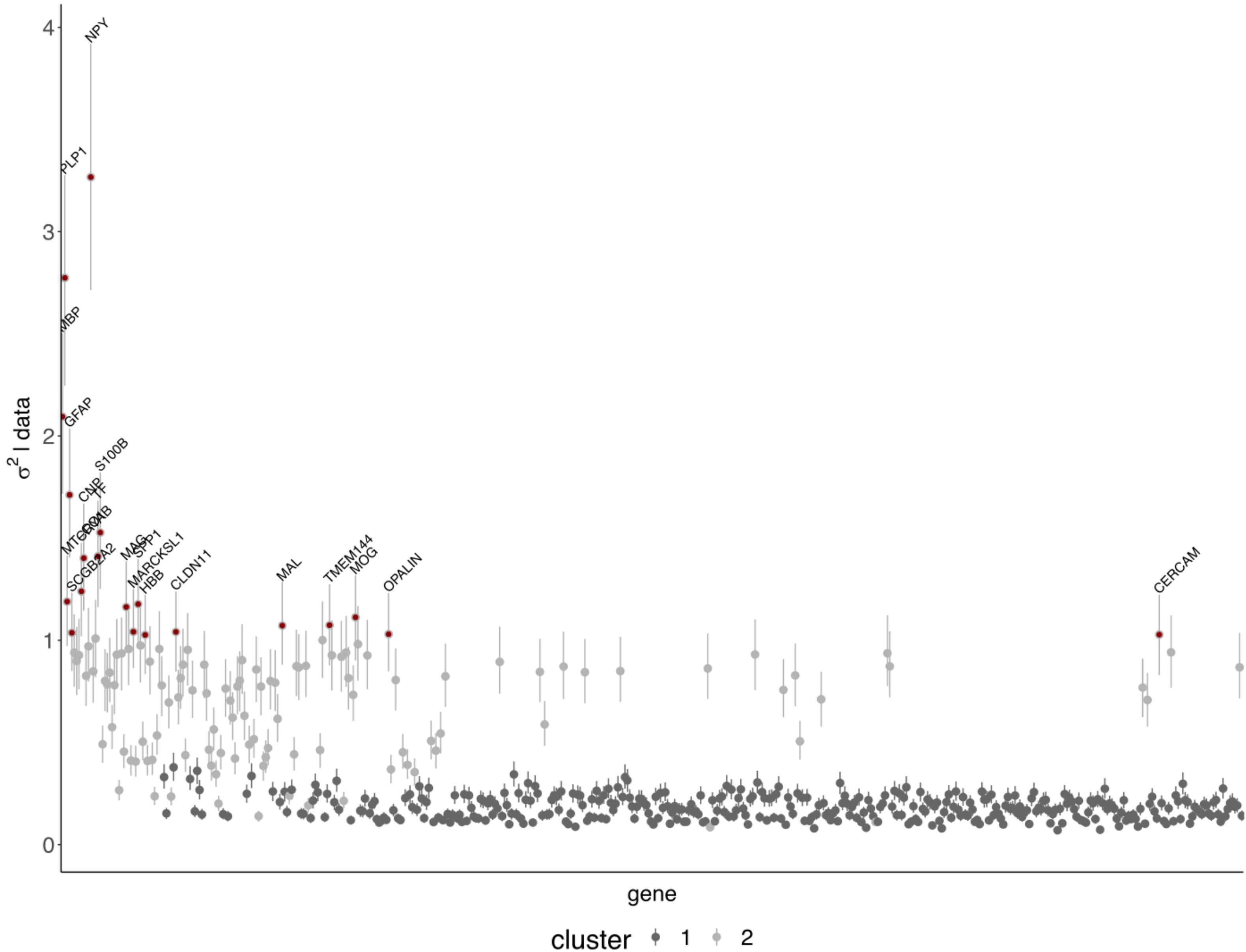
Identification of genes that exhibit high specific variation

Ranking genes by

$$\sigma_i^2 | X, \hat{\Theta}$$

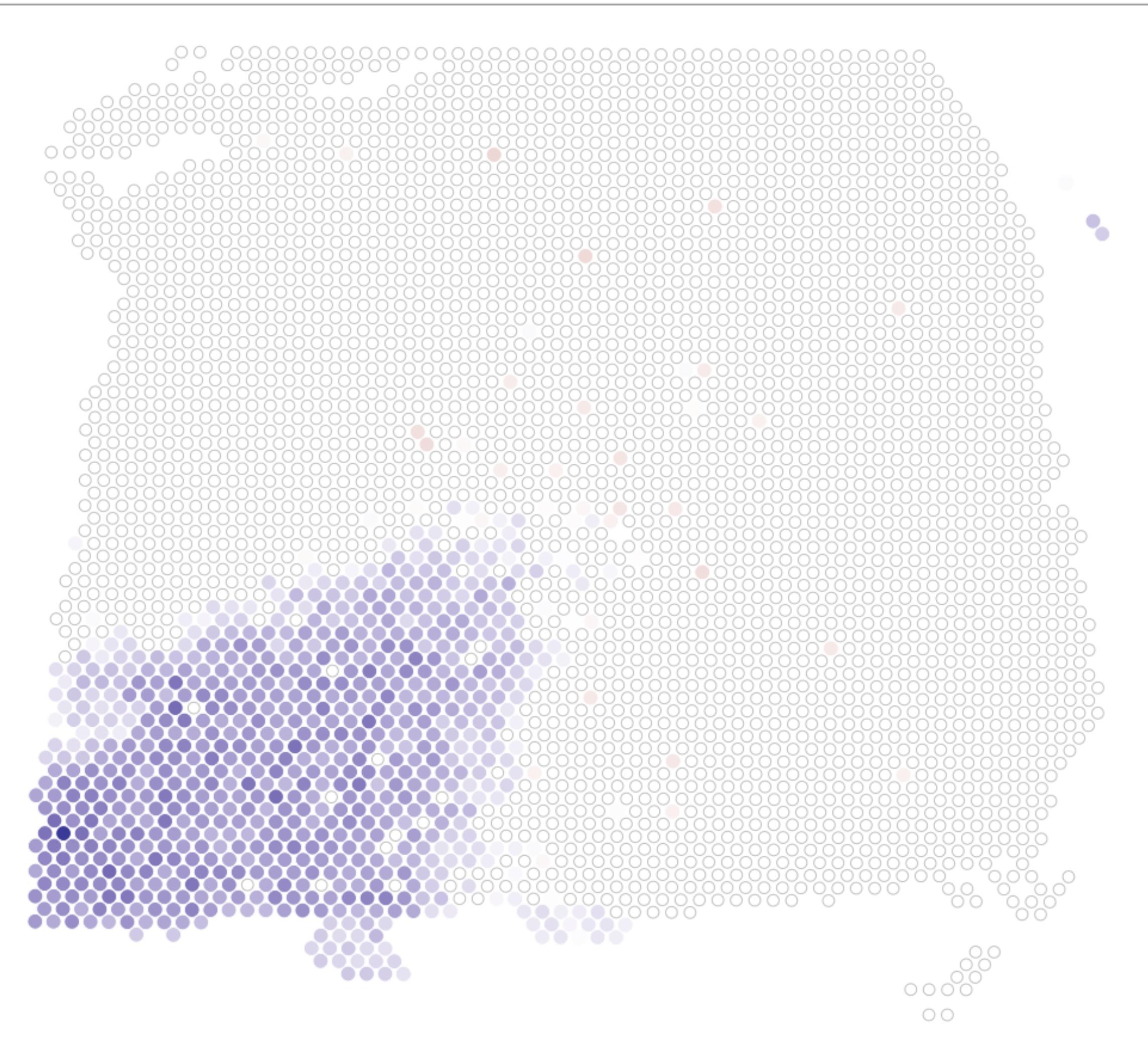
allows us to determine the **highly variable genes** net of the cluster spatial effect.

$r = 1$

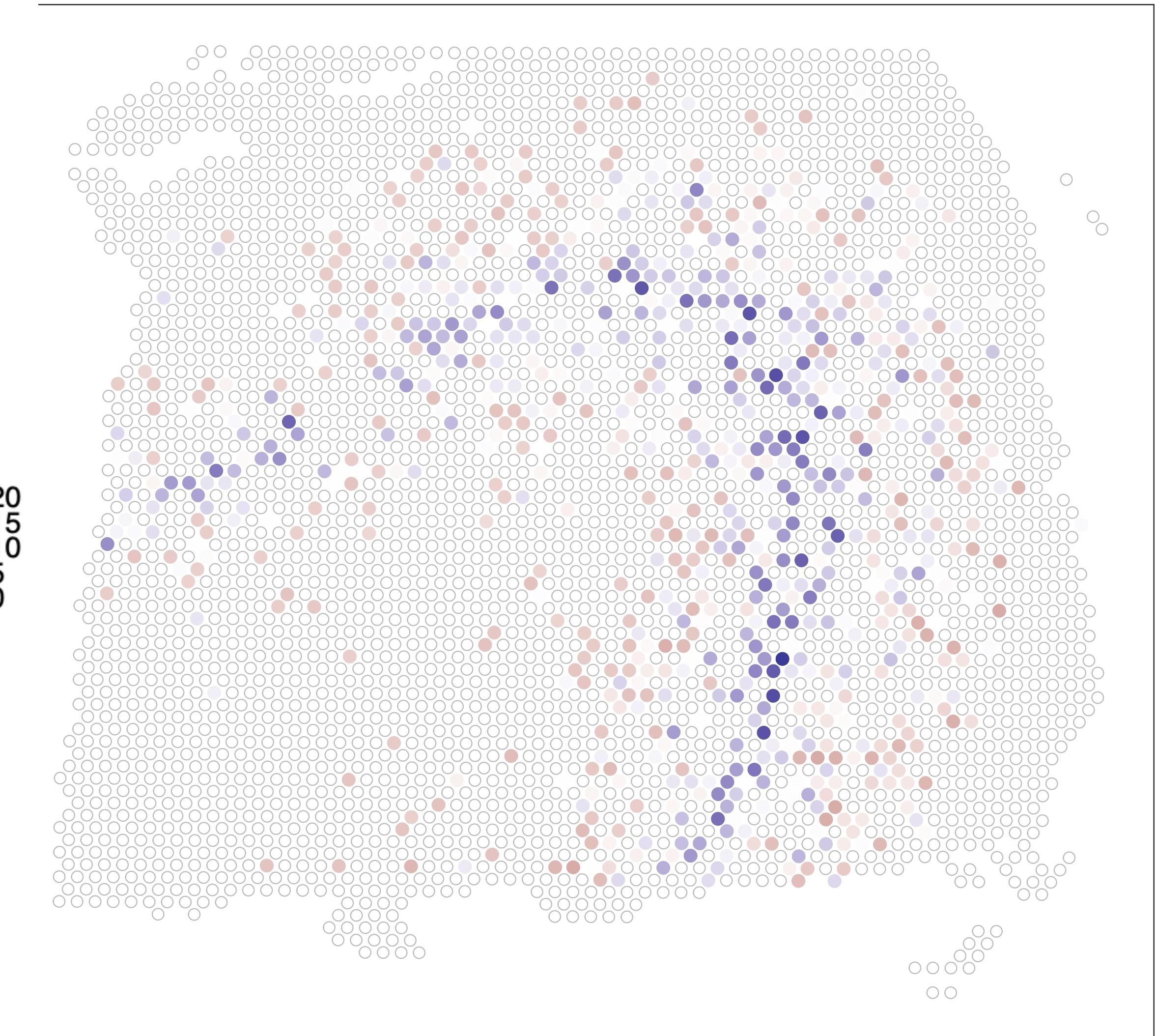


Confirm highly variable, cell-type-specific genes

MBP ($r = 1, 8, 9$)

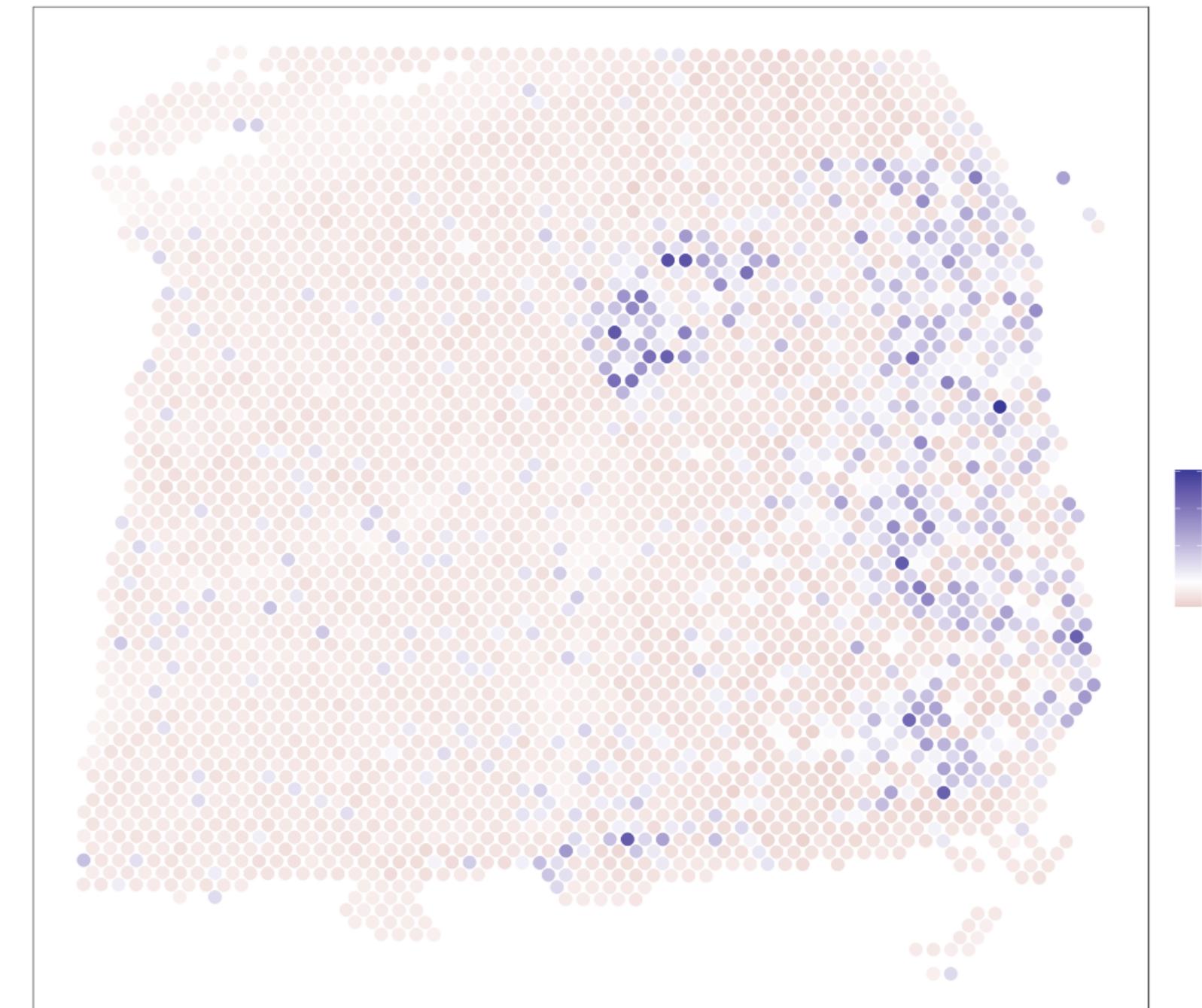
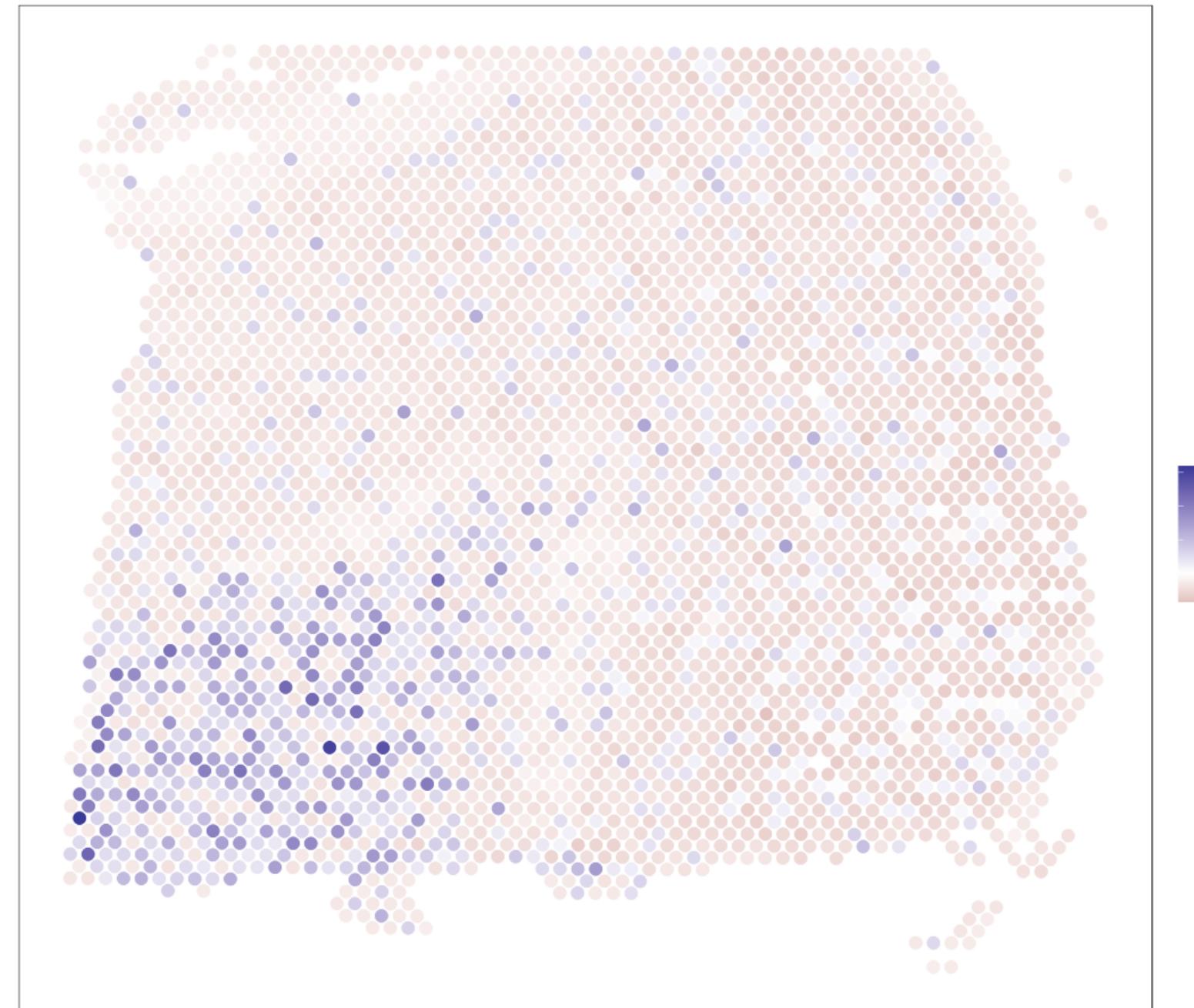
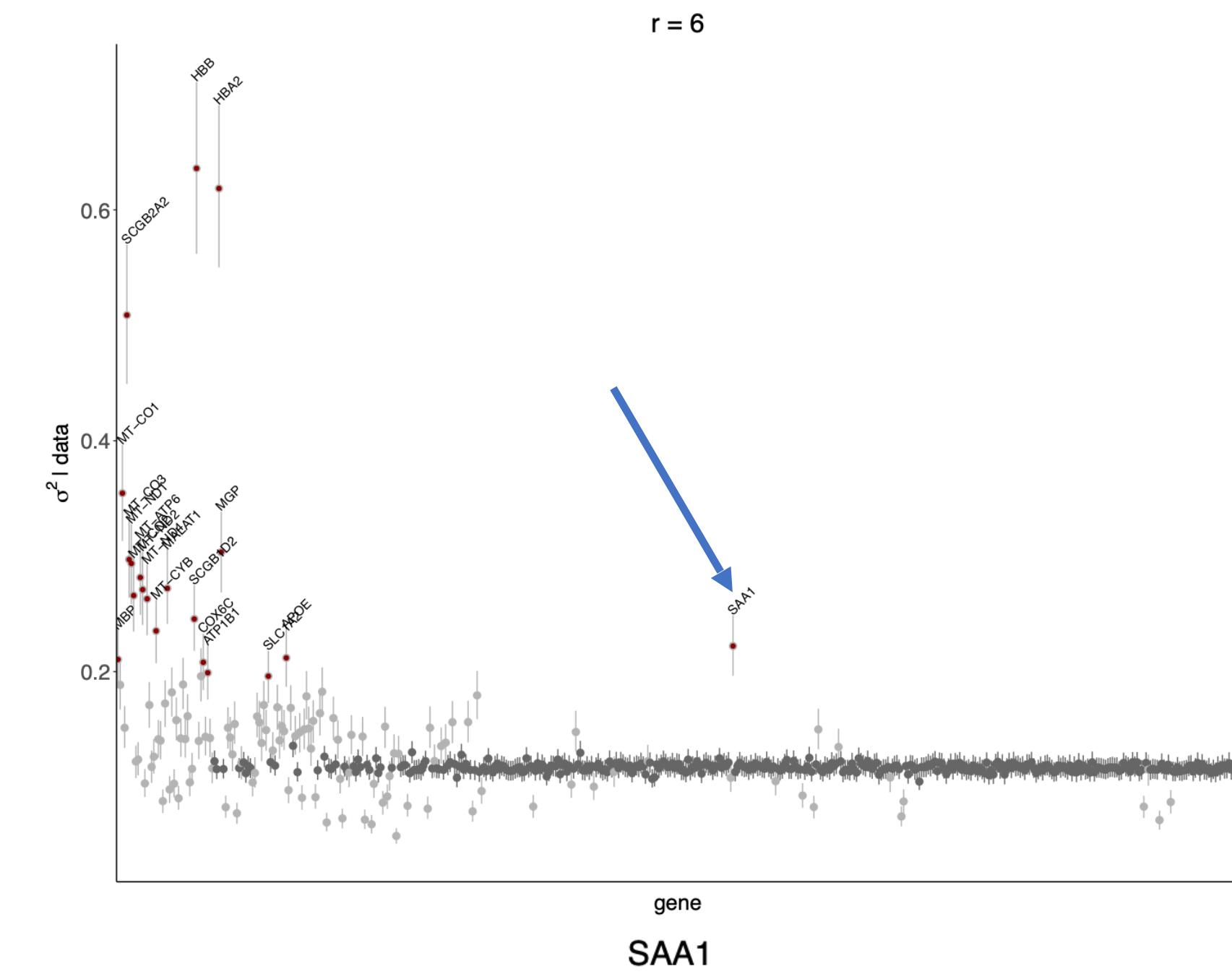
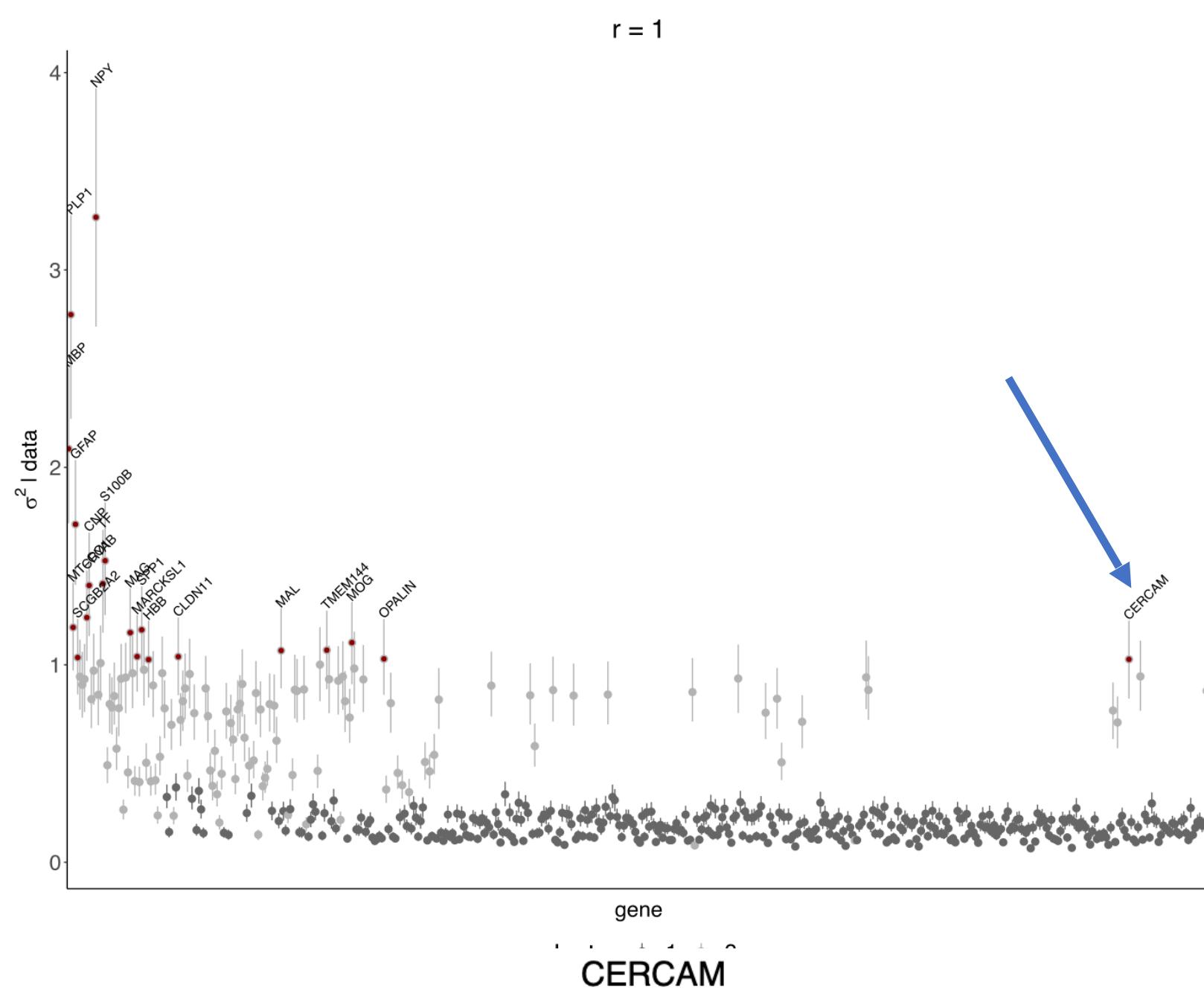


PCP4 ($r = 2, 7$)



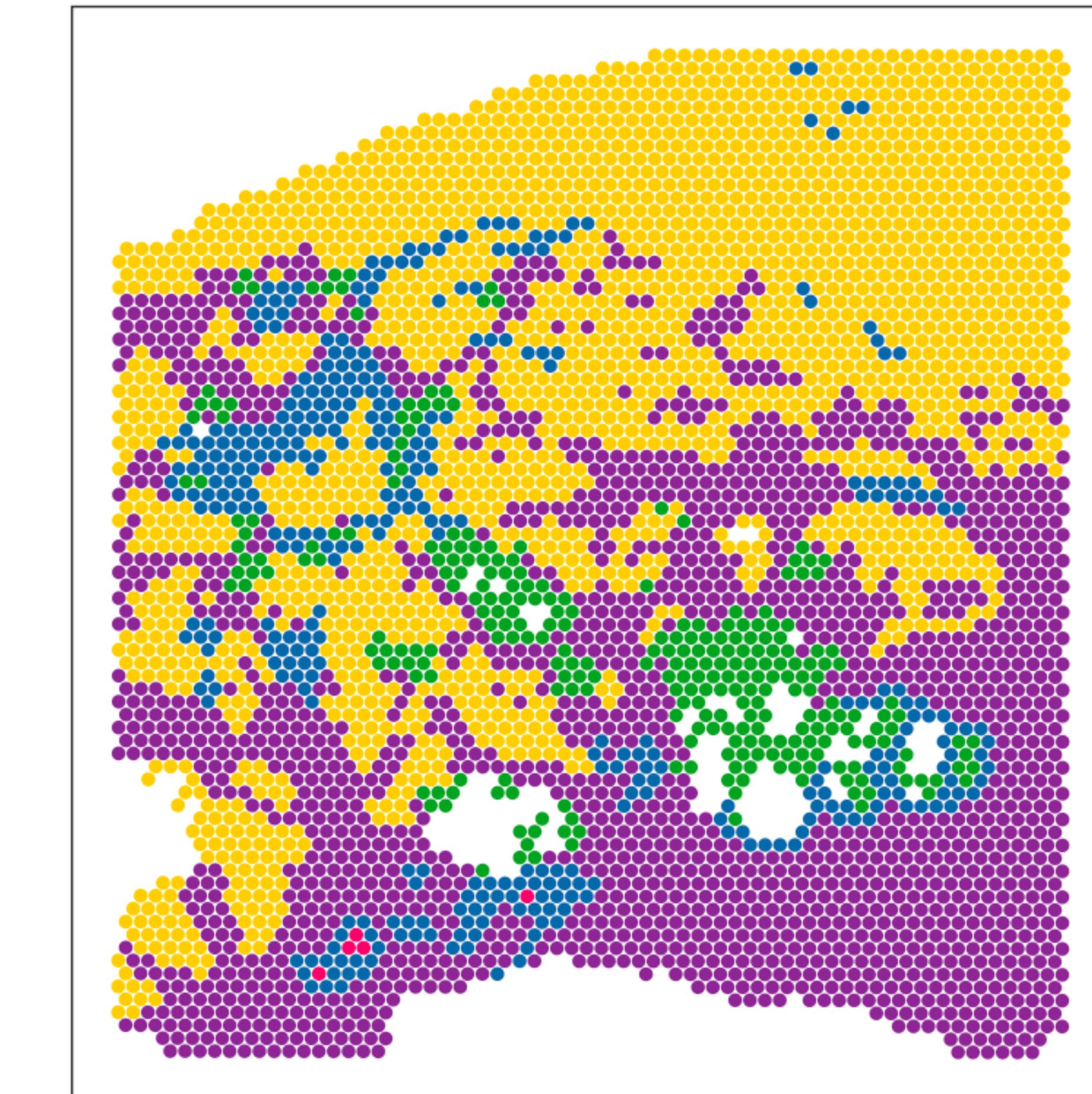
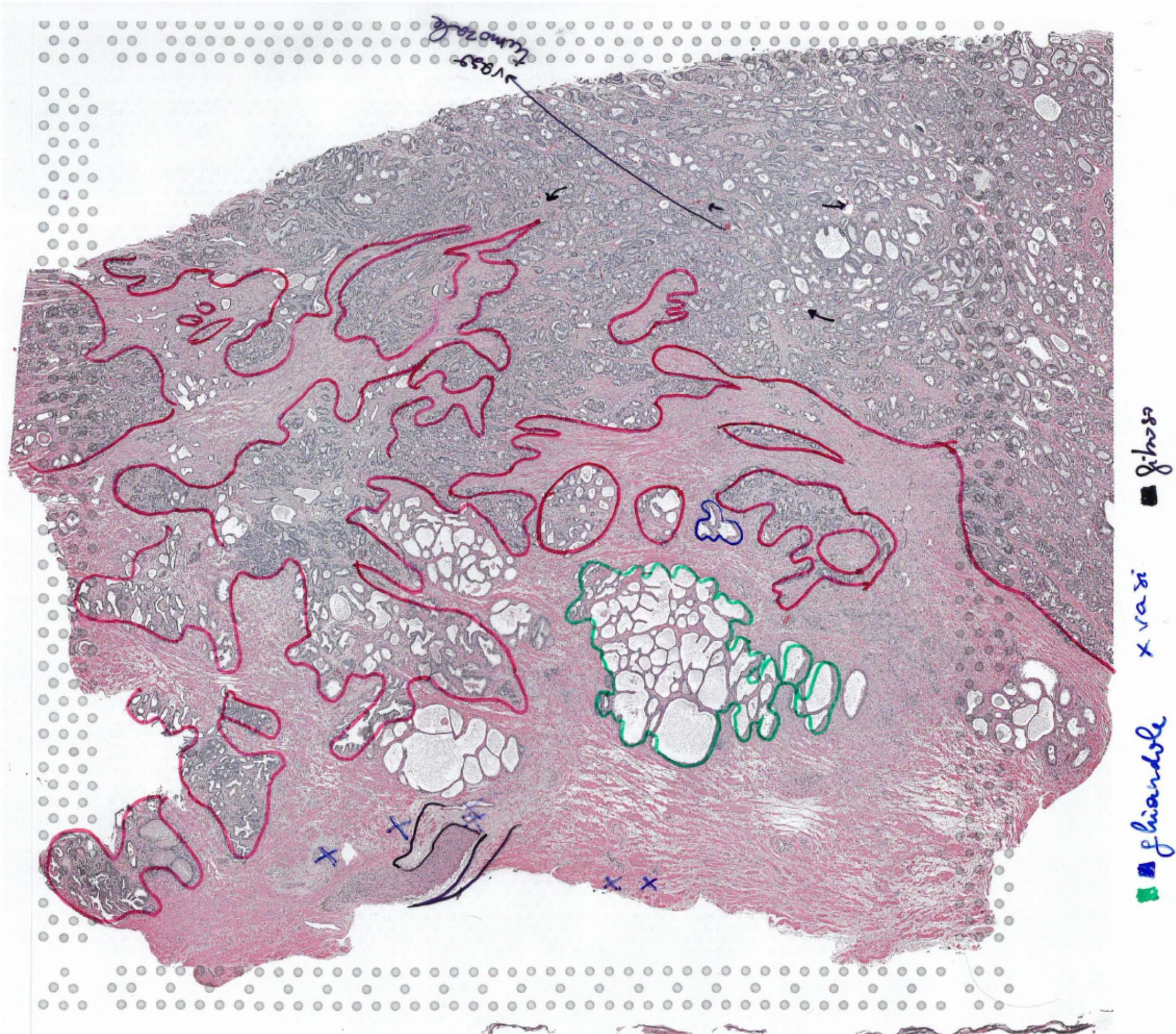
Novel insights

SAA1 “highly expressed in response to inflammation and tissue injury.”

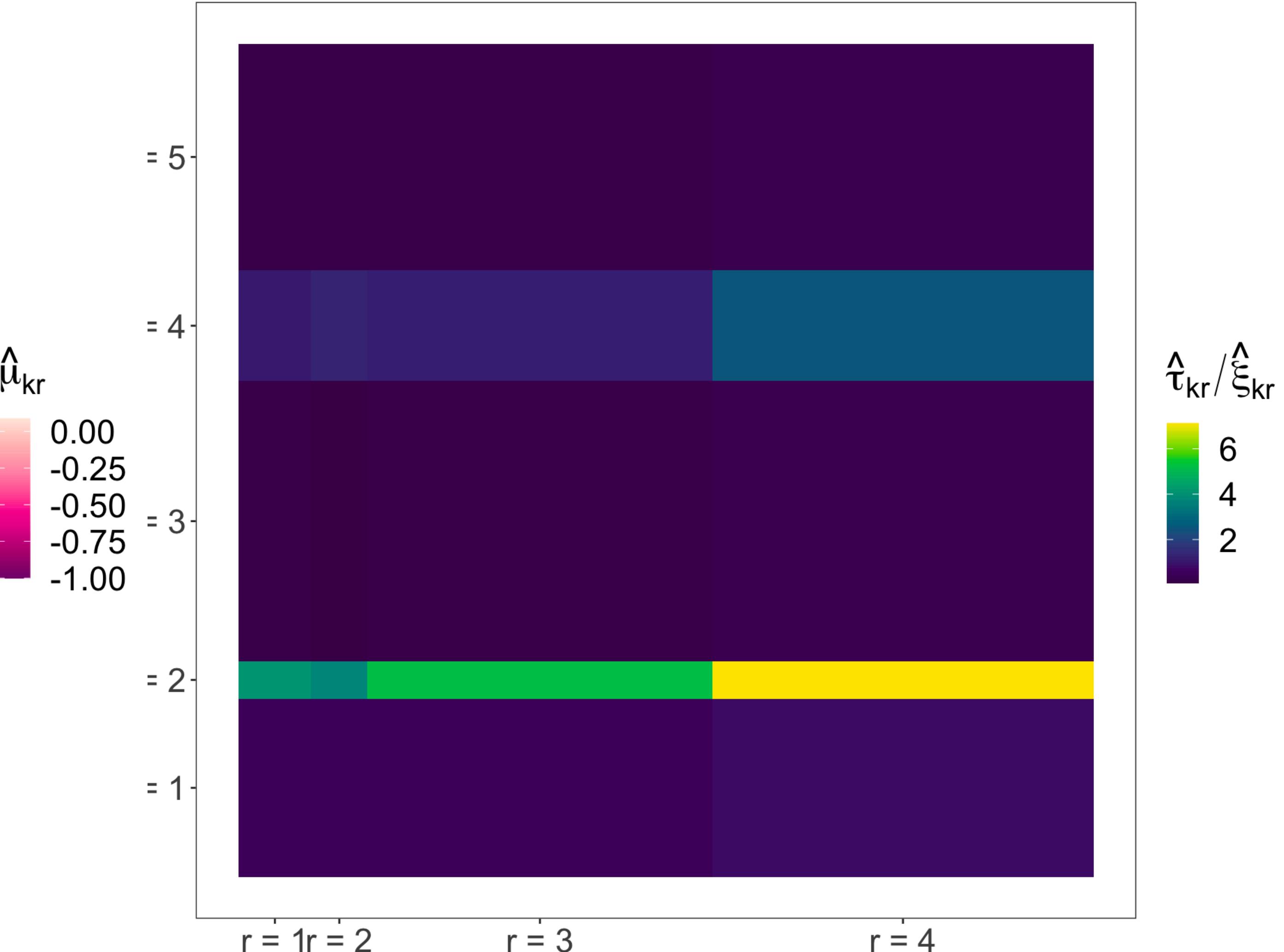
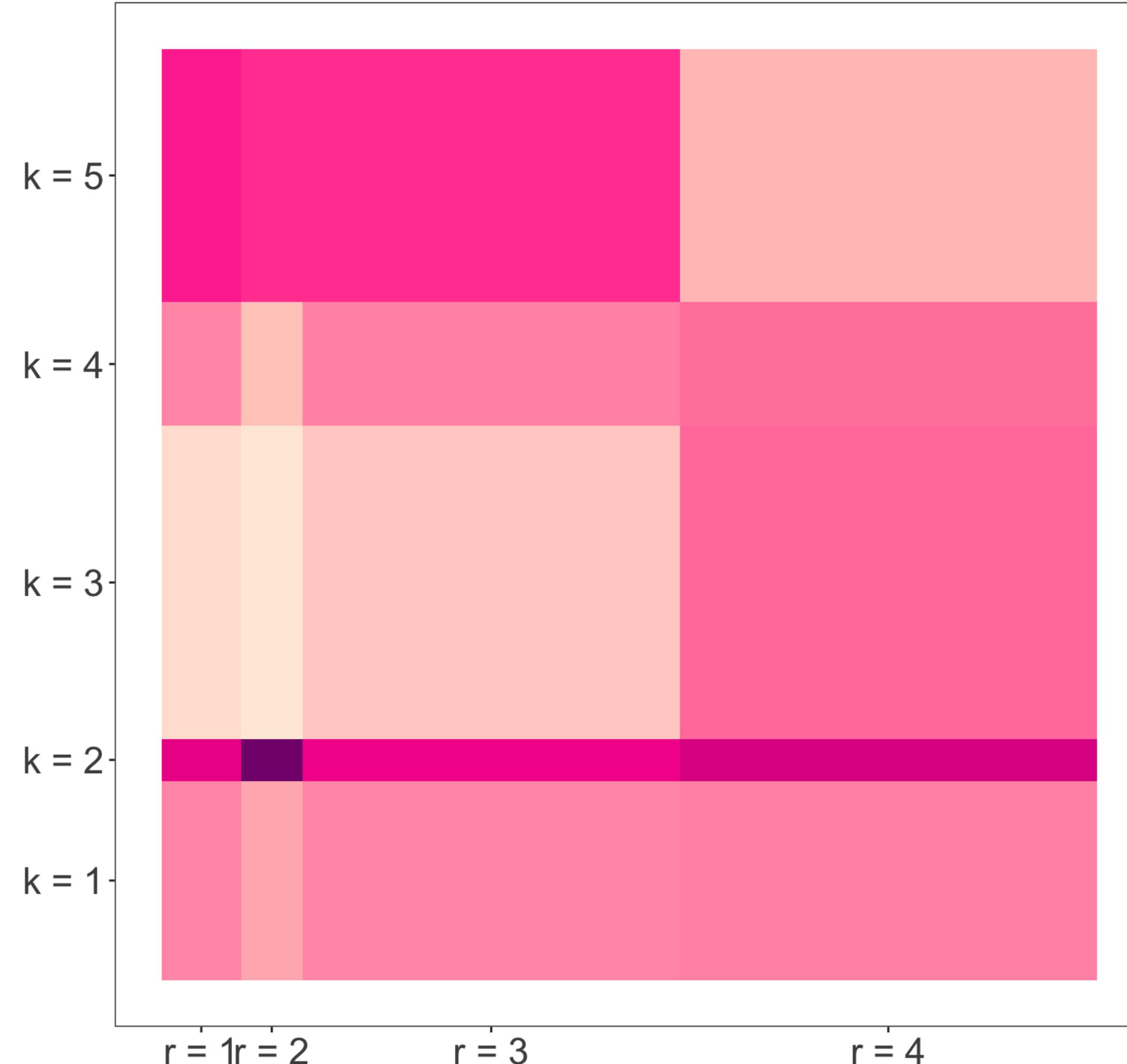


Pathologist annotation

Prostate Cancer



Supervised SPARTACO



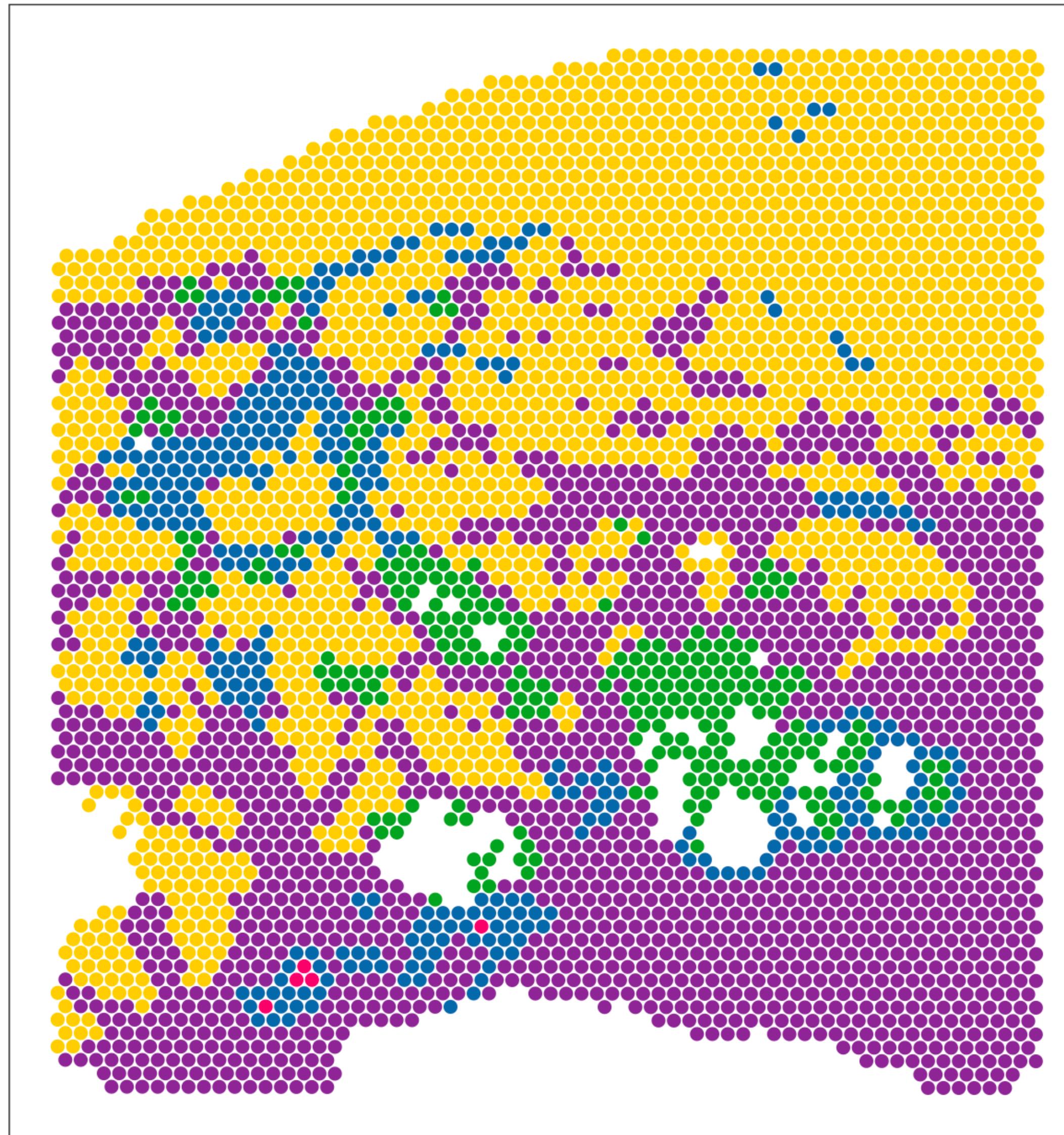
Sara A. Castiglioni

Spatially variable genes

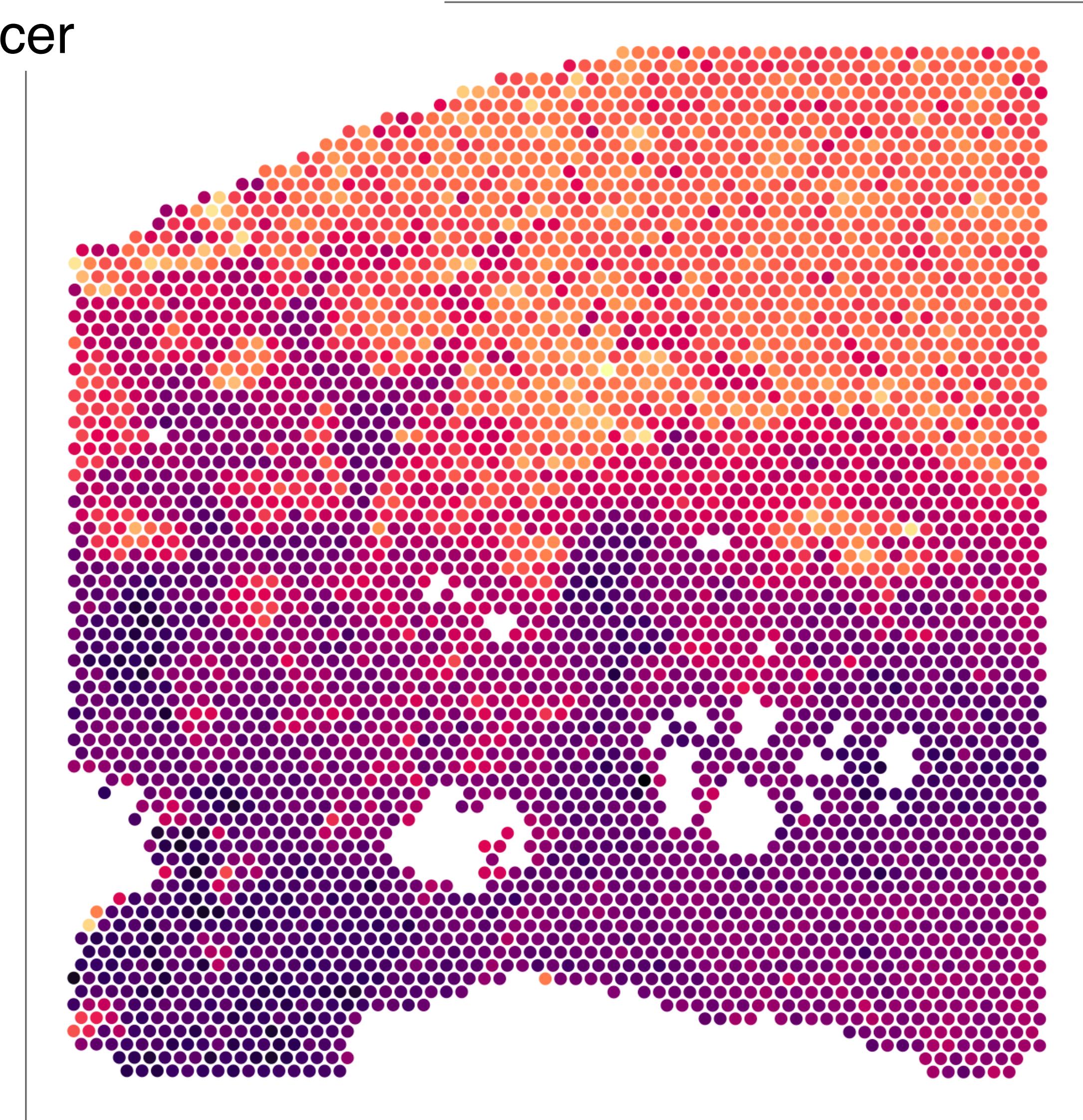
Gene	nnSVG rank	Townes rank	Gene	nnSVG rank	Townes rank
MSMB (1)(2)(3)(4)	1	1	IGHA1 (1)(2)(3)(4)	24	24
KLK3 (1)(2)(3)(4)	2	2	IGLC1 (1)(2)	28	28
MYL9 (1)(2)(3)(4)	3	3	EEF2 (4)	30	30
ACTG2 (1)(2)(3)(4)	4	4	KRT15 (1)(2)(4)	32	32
ACTA2 (1)(3)(4)	5	5	APOD (1)(3)(4)	35	35
TAGLN (1)(3)(4)	6	6	AZGP1 (2)	36	36
KLK2 (1)(2)(3)(4)	7	7	LTF (4)	42	42
TGM4 (1)(2)	8	8	FOSB (3)	43	43
PLA2G2A (4)	9	9	OLFM4 (2)(4)	50	50
IGKC (1)(2)(3)(4)	10	10	NEFH (2)	52	52
FLNA (1)(3)	11	11	JCHAIN (1)(2)	61	61
SPON2 (4)	12	12	CD74 (3)	65	65
TPM2 (1)(3)	13	13	IGHG1 (2)	70	70
TMEFF2 (4)	14	14	KRT5 (2)	78	78
MYLK (1)(3)	15	15	IGHG2 (3)	79	79
DES (1)(3)	17	17	TMSB4X (2)(4)	83	83
AMACR (4)	19	19	VIM (3)	87	87
ACP3 (1)(2)(3)(4)	20	20	MMP7 (2)	107	107
TPM1 (1)(3)	22	22	KRT7 (2)	118	118

KLK3

Spatial coordinates



KLK3 is a standard marker for prostate cancer

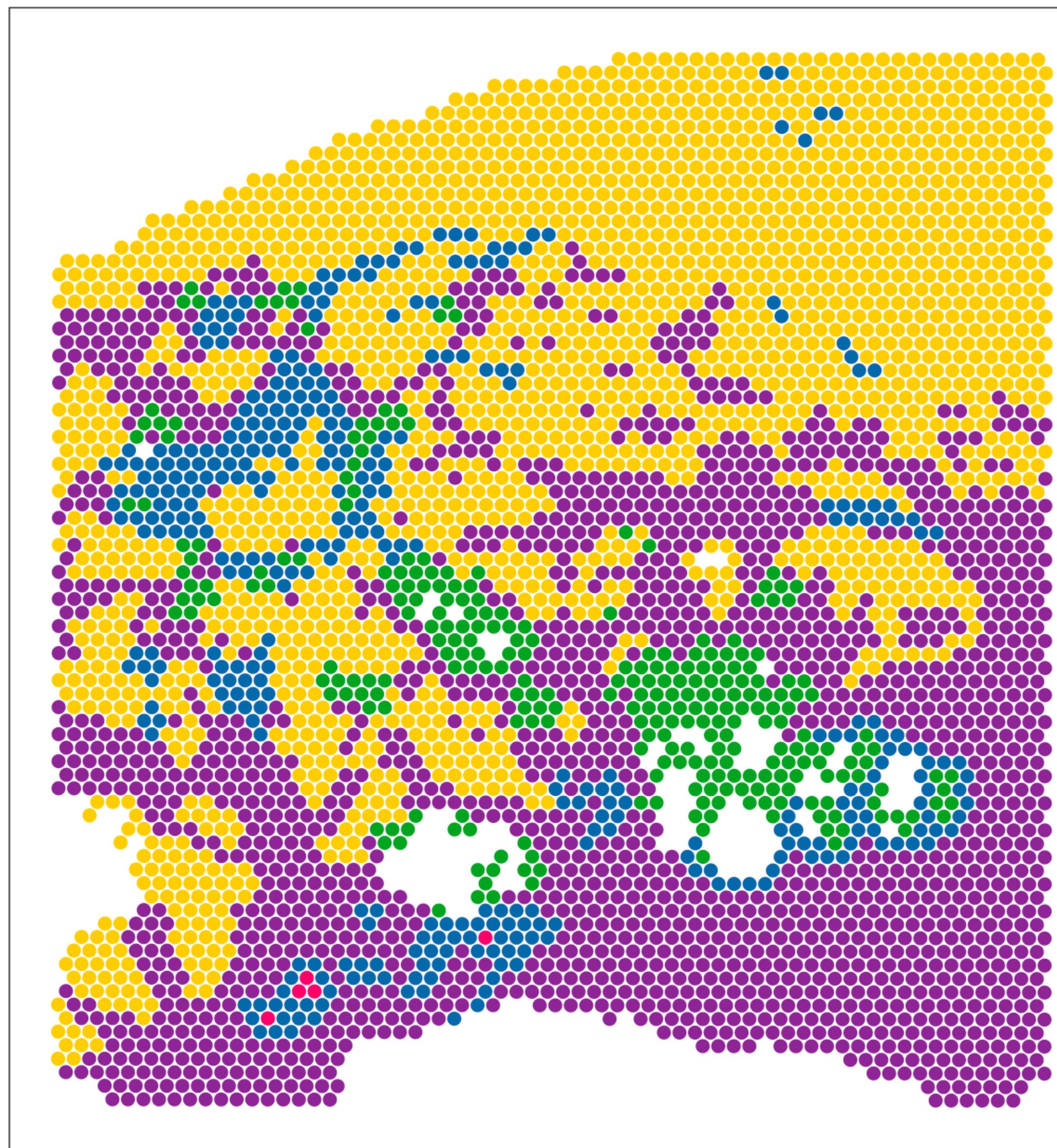


pathologist • blood vessel • fibroblast • gland • stroma • tumor

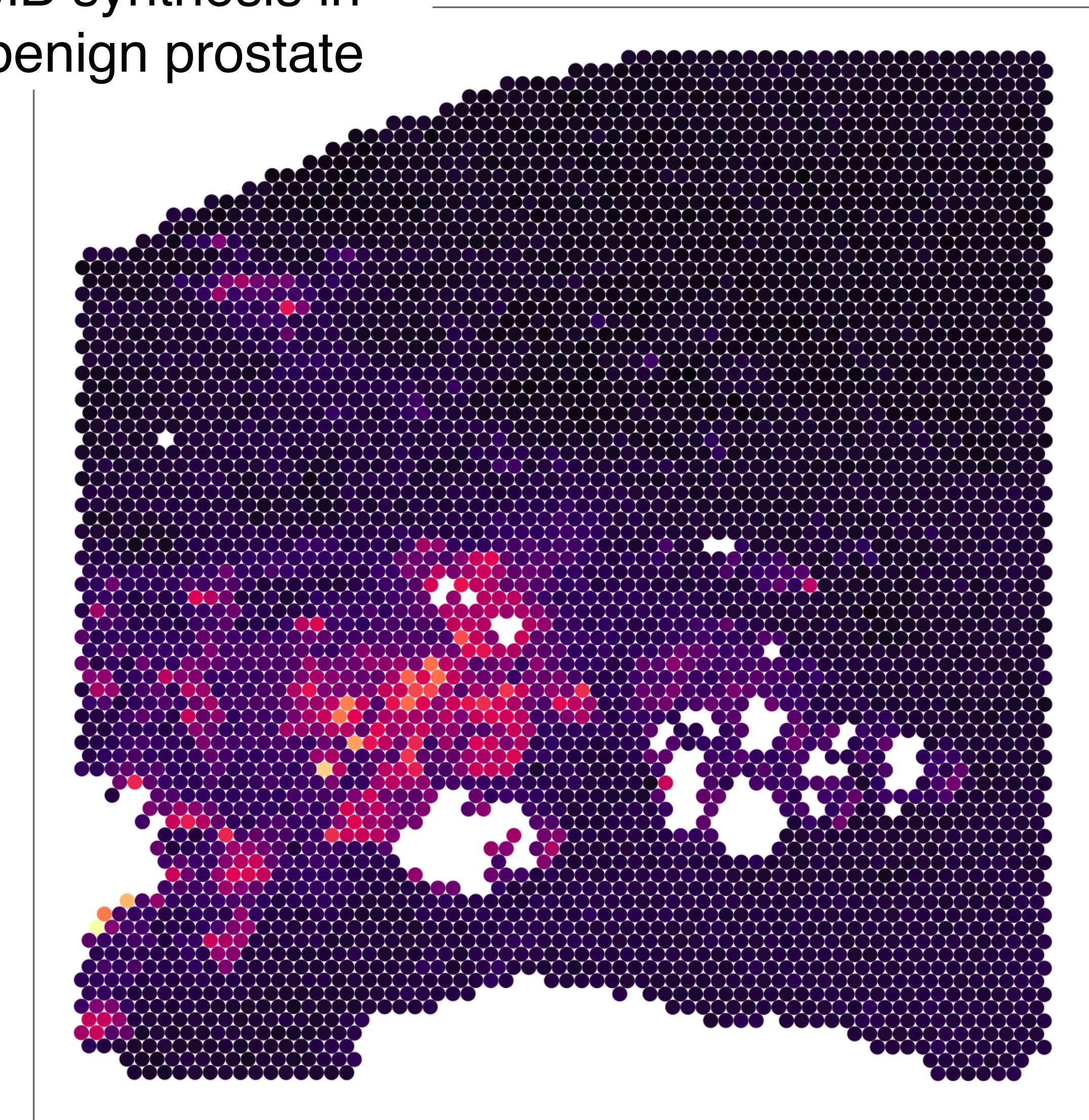
Penney et al. (2011)

MSMB

Spatial coordinates



aggressive prostate
tumors suppress
MSMB synthesis in
the benign prostate

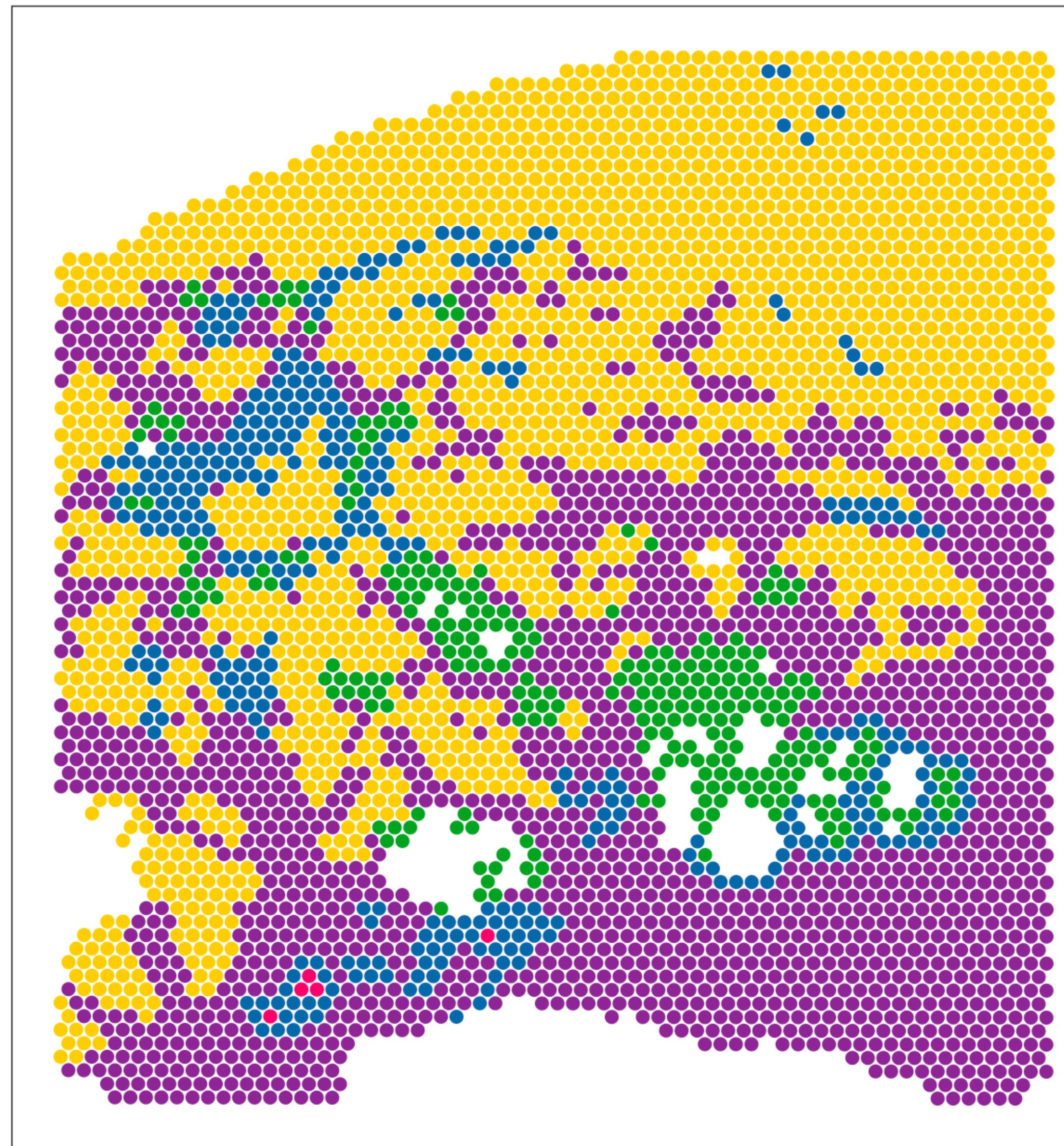


pathologist • blood vessel • fibroblast • gland • stroma • tumor

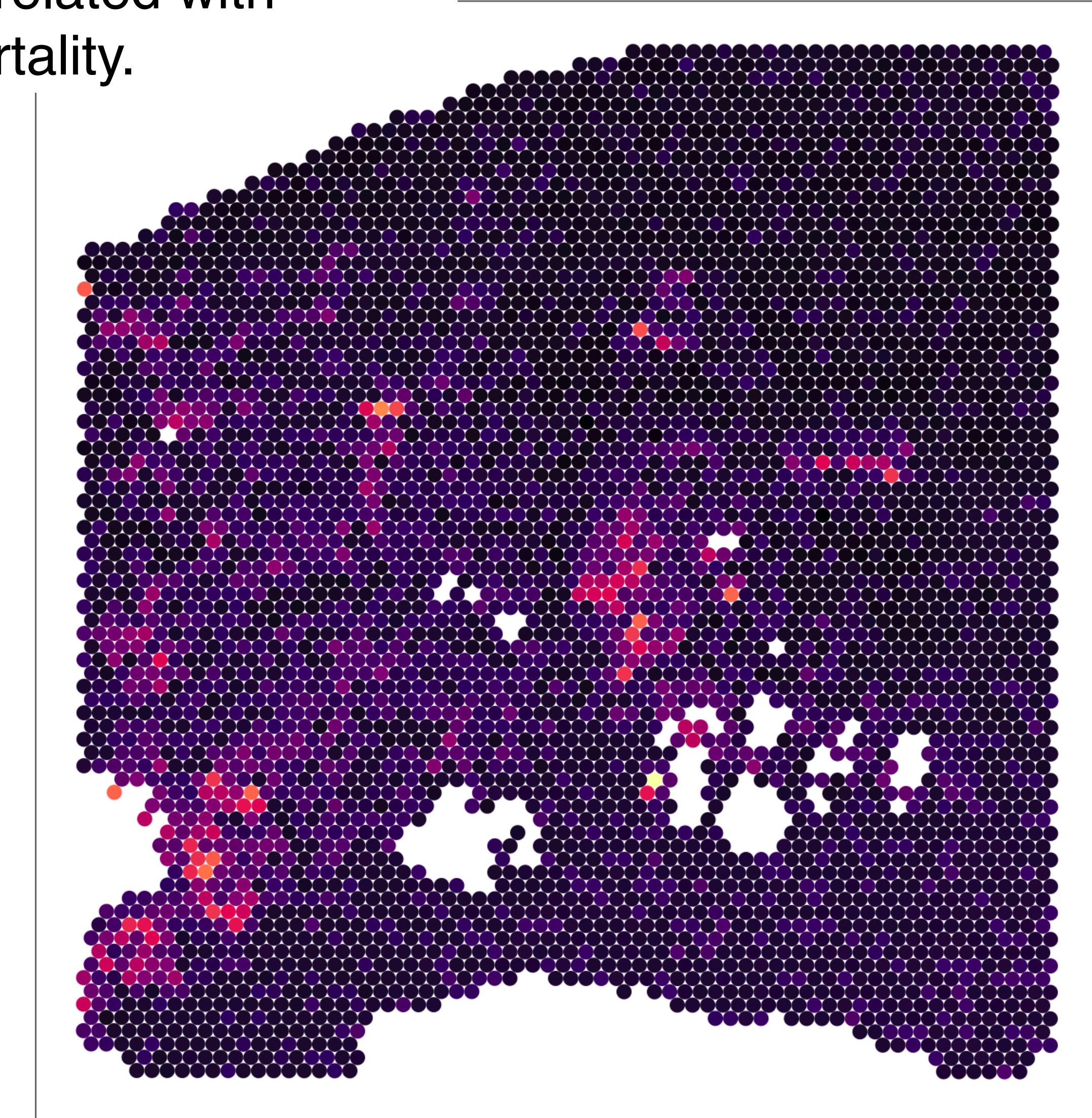
Bergstrom et al. (2018)

KRT7

Spatial coordinates



High expression in
benign glands is
correlated with
mortality.

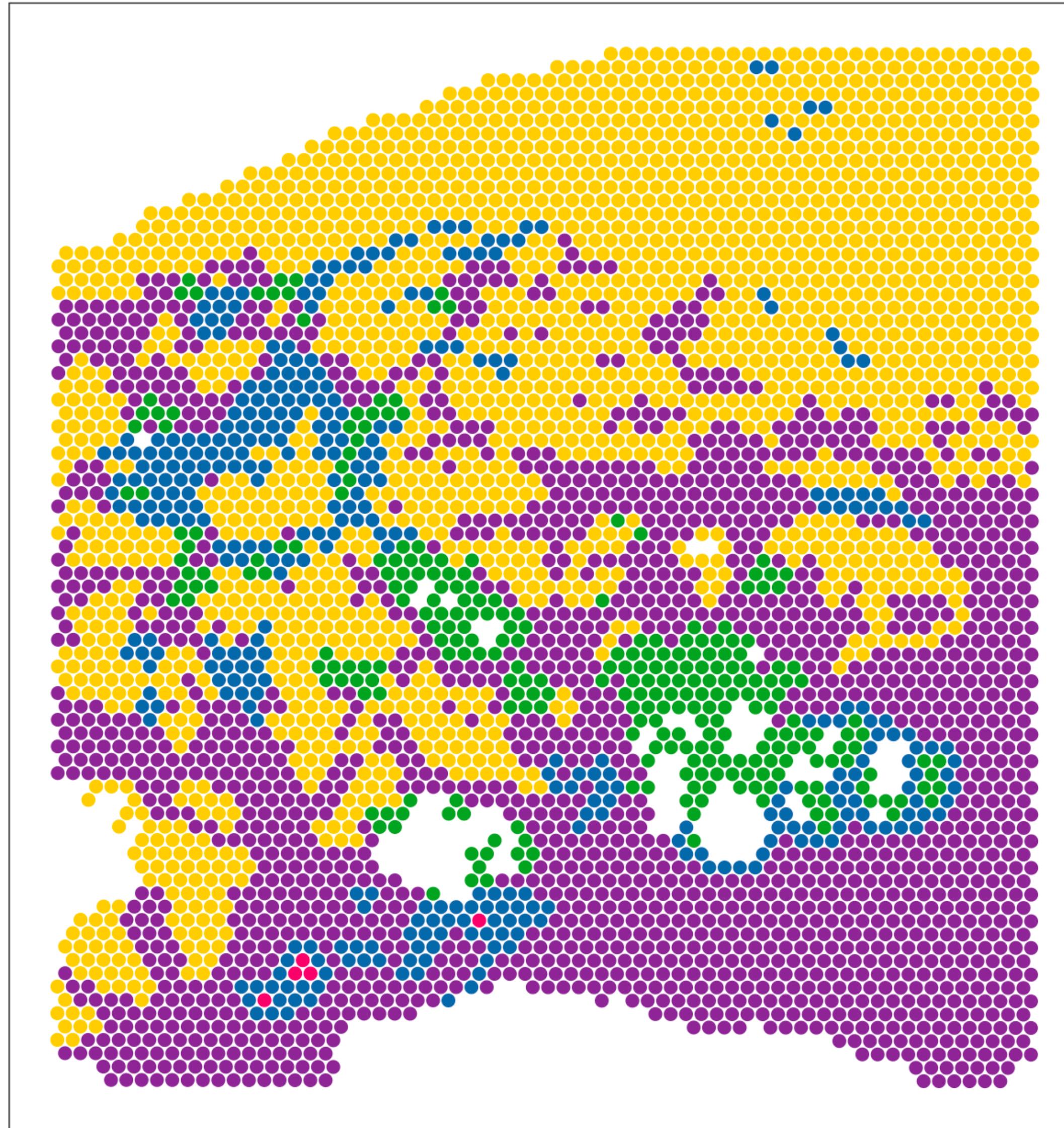


pathologist • blood vessel • fibroblast • gland • stroma • tumor

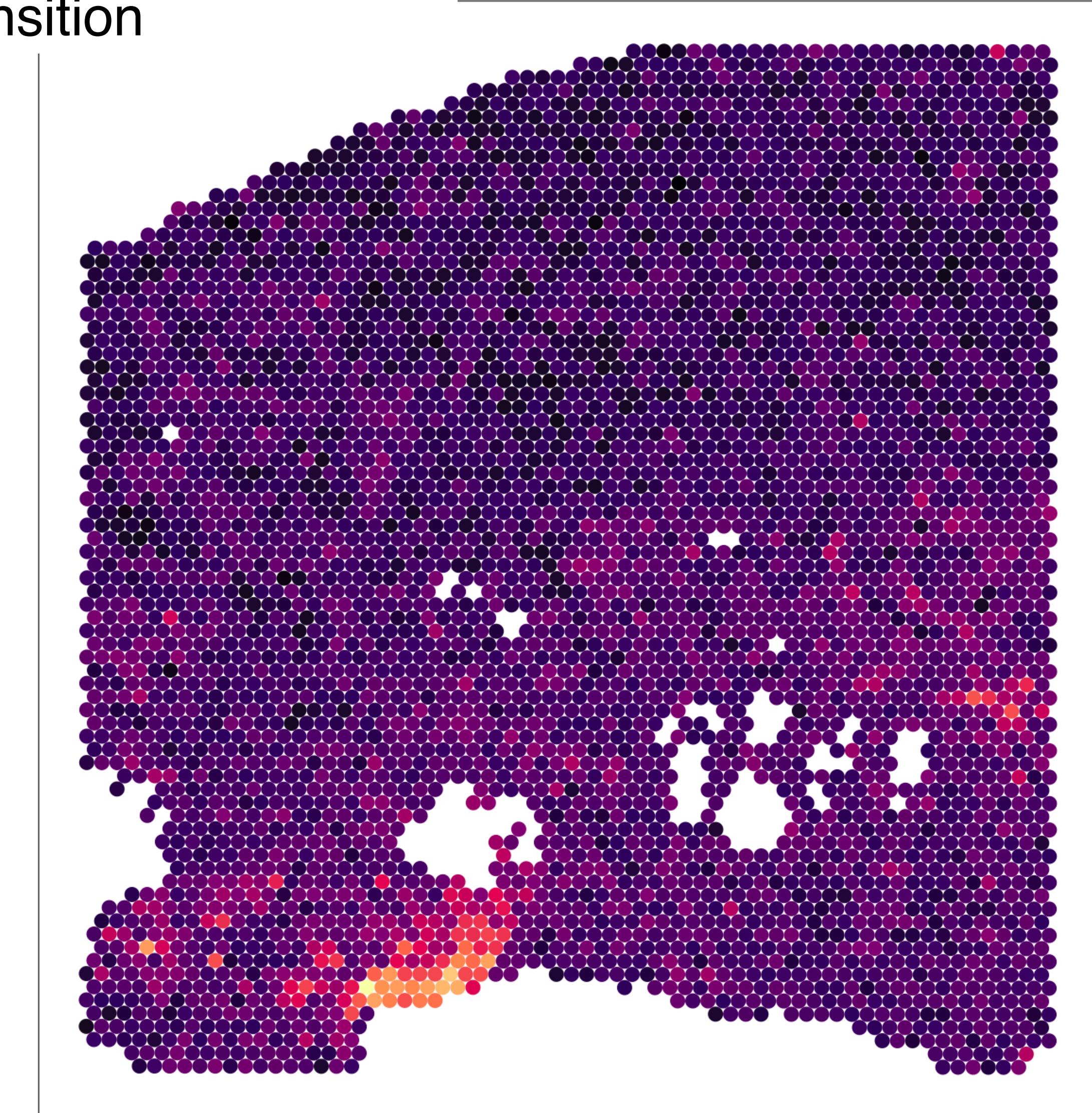
Dariane et al. (2022)

VIM

Spatial coordinates



Involved in epithelial-mesenchymal transition

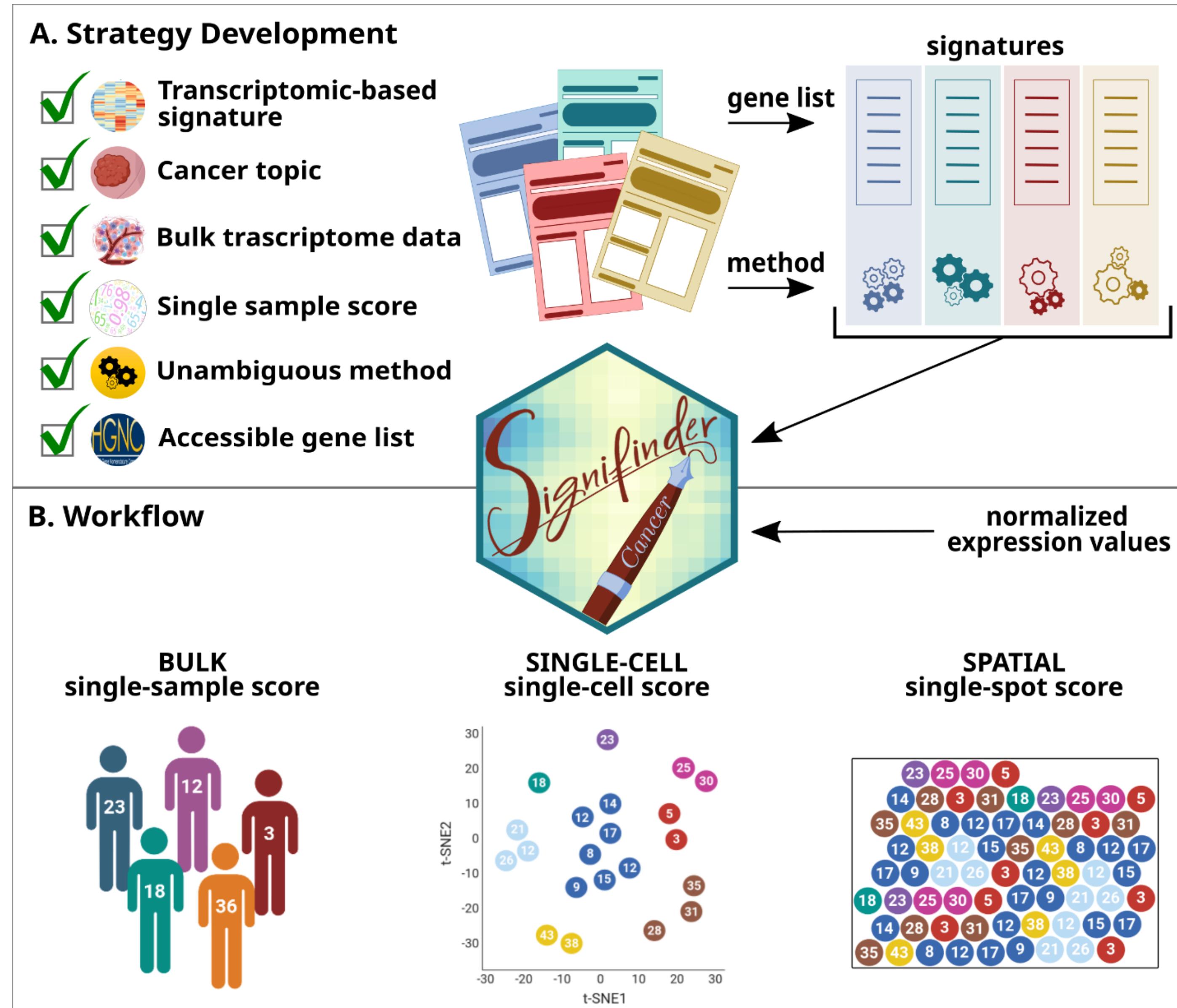


pathologist • blood vessel • fibroblast • gland • stroma • tumor

Wei et al. (2018)

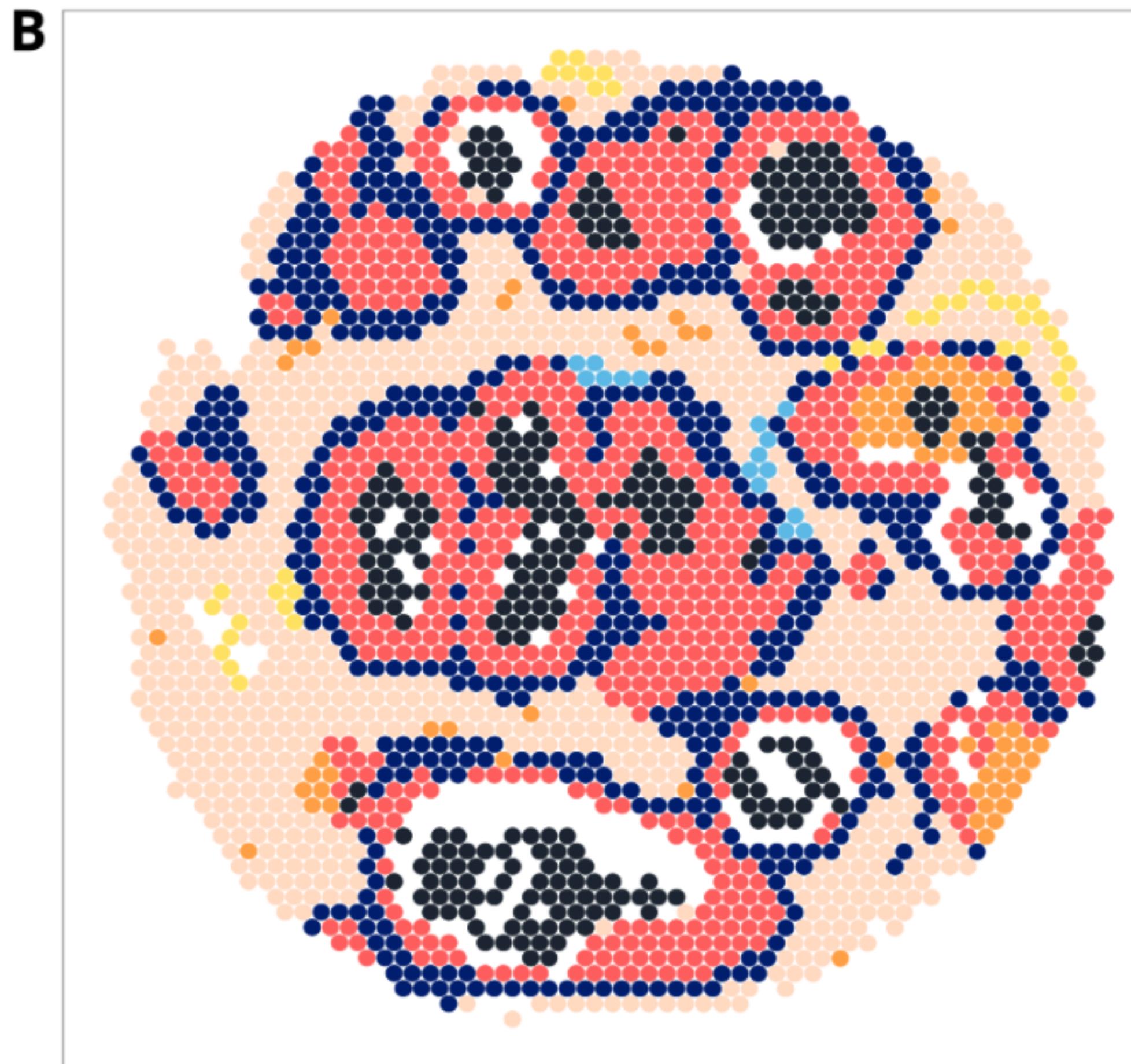
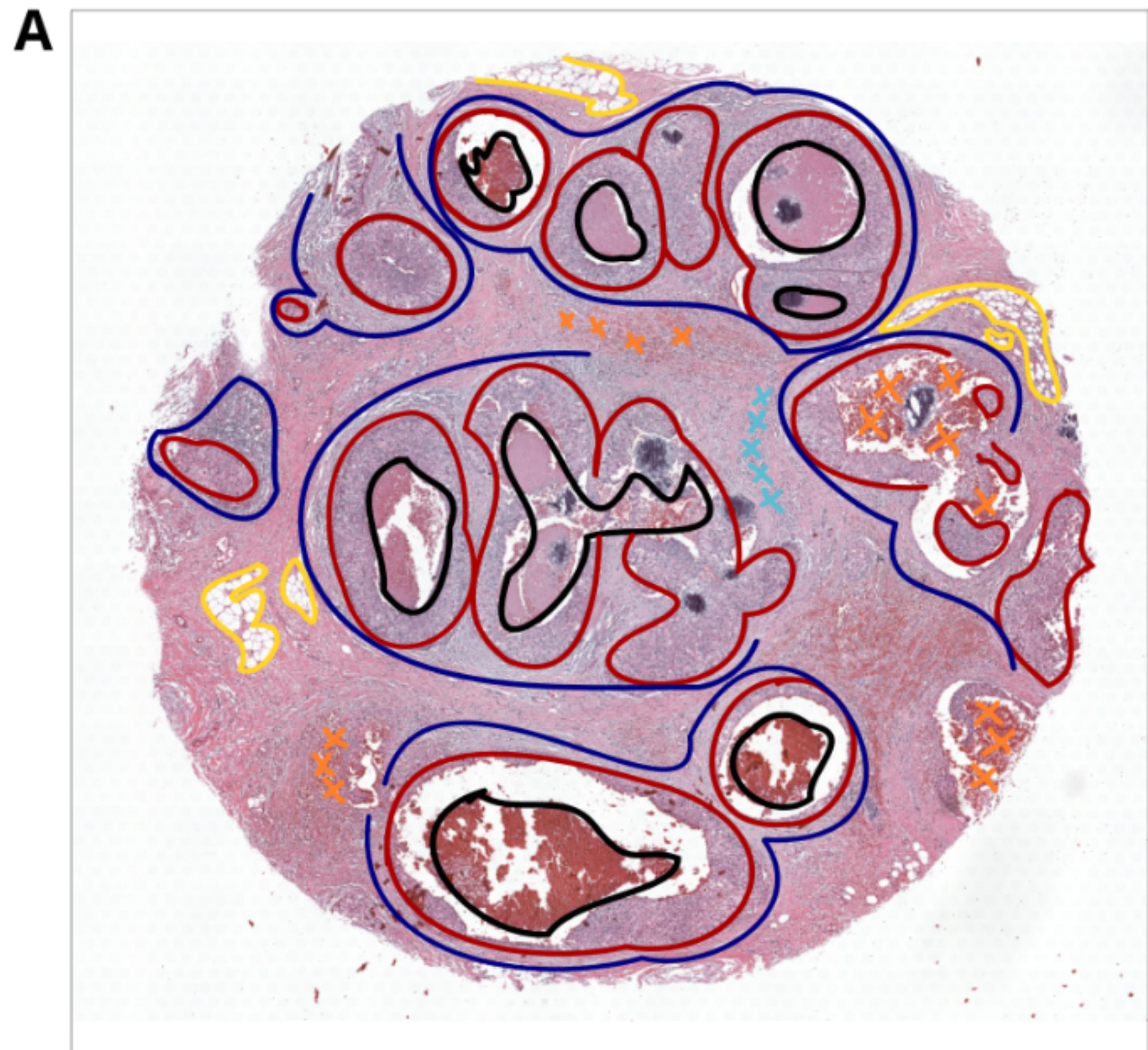
Gene signatures from bulk studies

bioconductor.org/package/signifinder

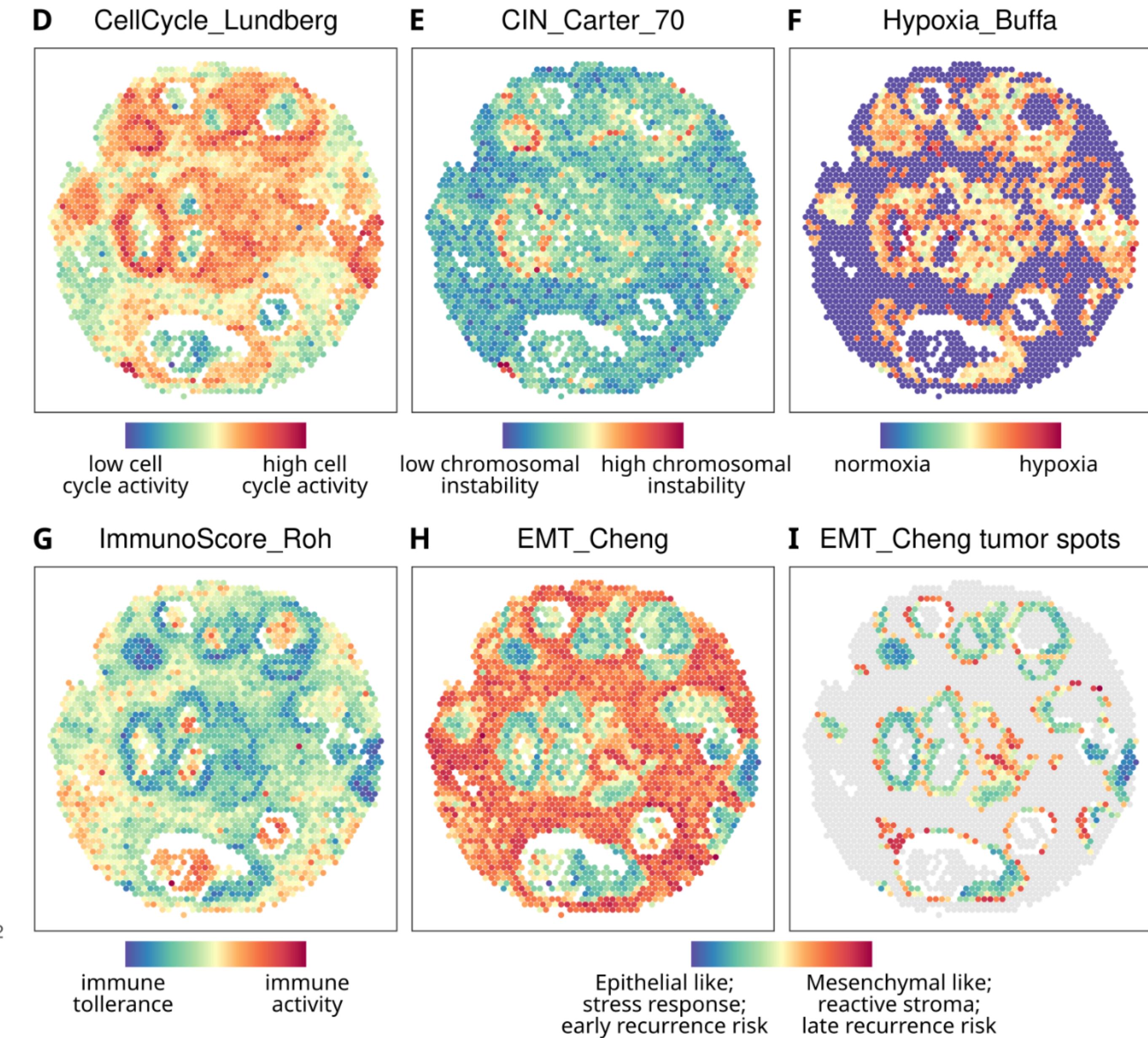
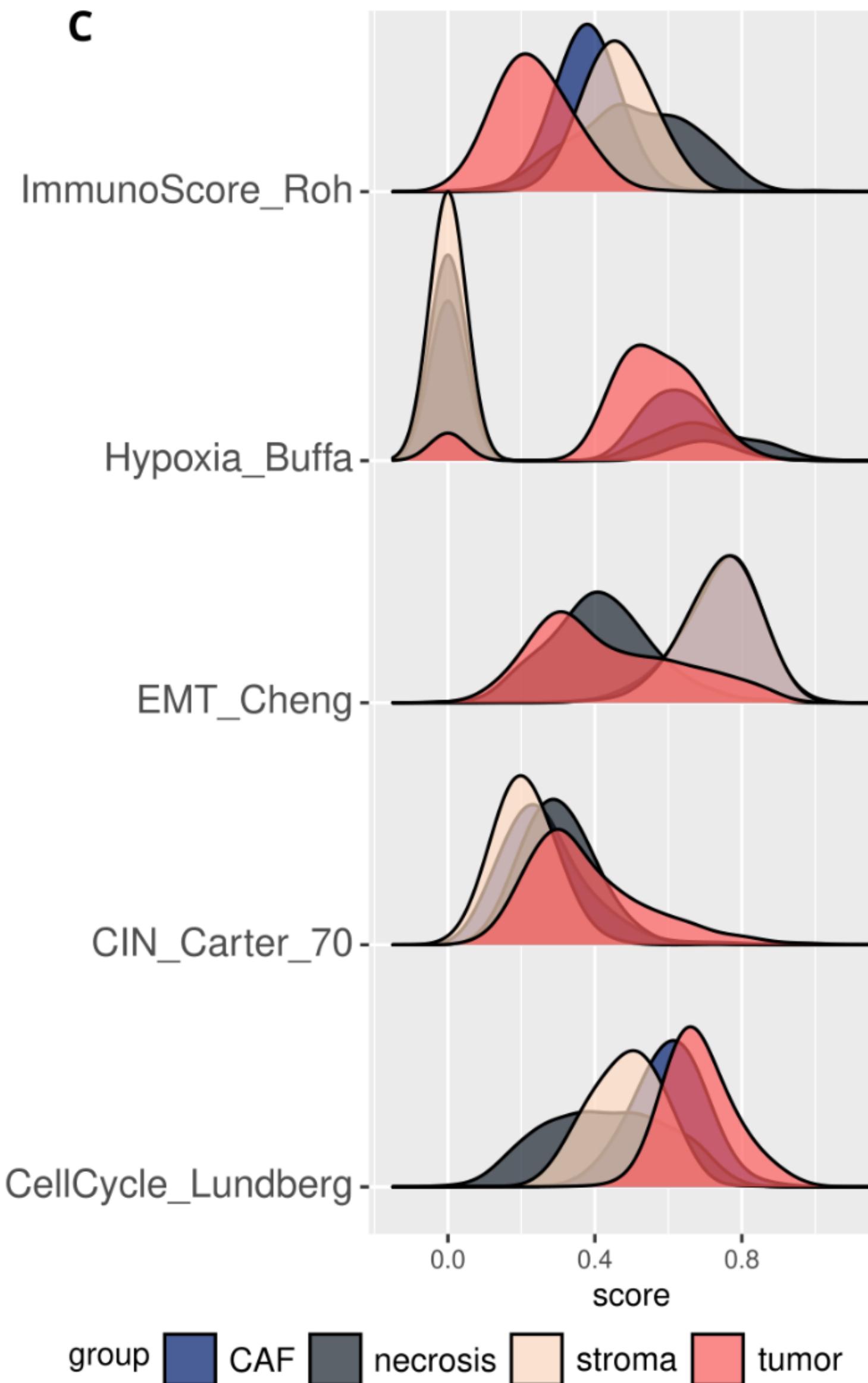


Pathologist annotation

Ductal Breast Cancer



Signature-based annotation



**THANKS FOR YOUR
ATTENTION!**

