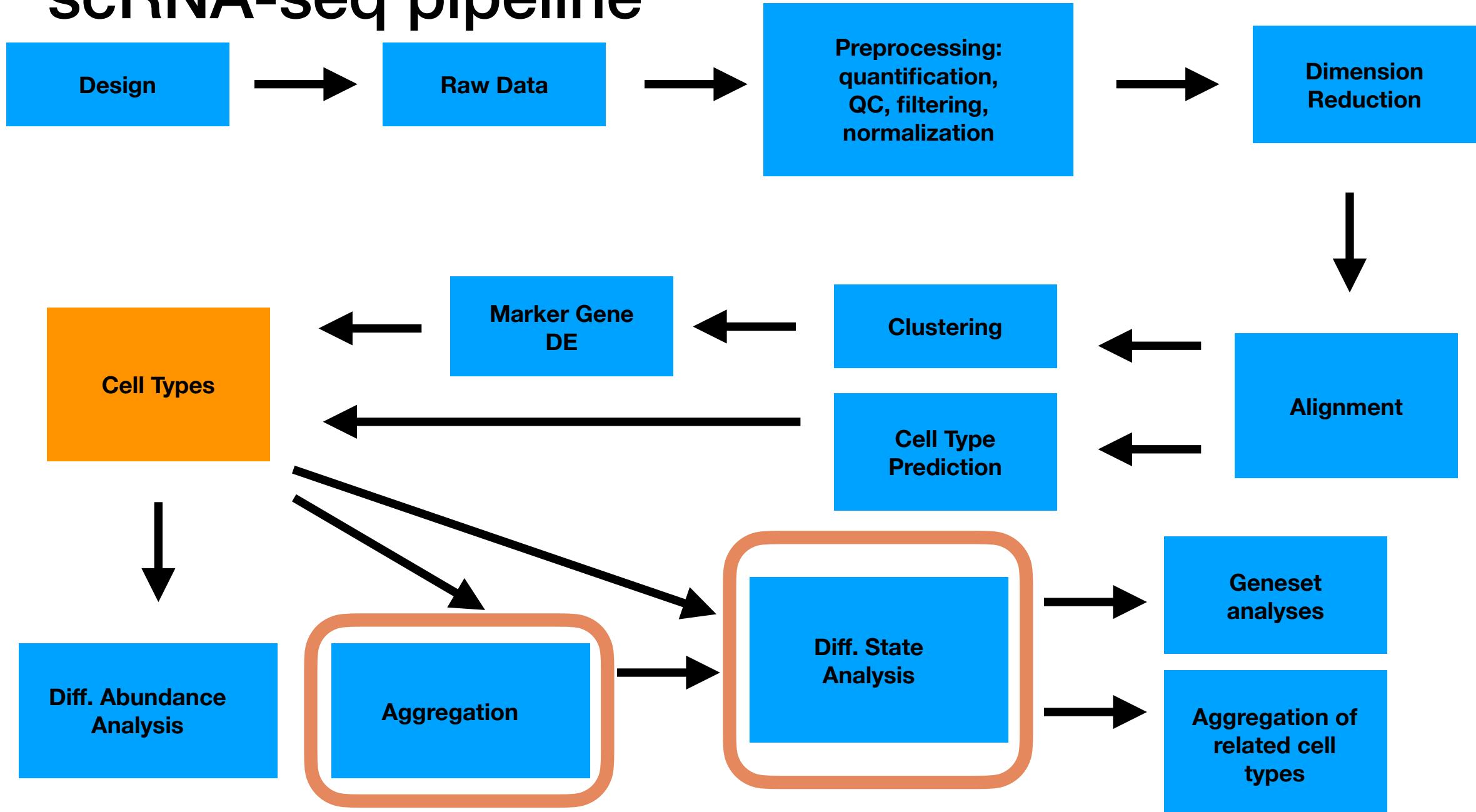




## Single cell differential state analysis (a few more details)

## Spatial omics analysis (introduction)

# scRNA-seq pipeline



# Revealing the vectors of cellular identity with single-cell genomics

Allon Wagner<sup>1</sup>, Aviv Regev<sup>2,3,5</sup> & Nir Yosef<sup>1,4,5</sup>

## Box 1 The many facets of a cell's identity

We define a cell's identity as the outcome of the instantaneous intersection of all factors that affect it. We refer to the more permanent aspects in a cell's identity as its type (e.g., a hepatocyte typically cannot turn into a neuron) and to the more transient elements as its state. Cell types are often organized in a hierarchical

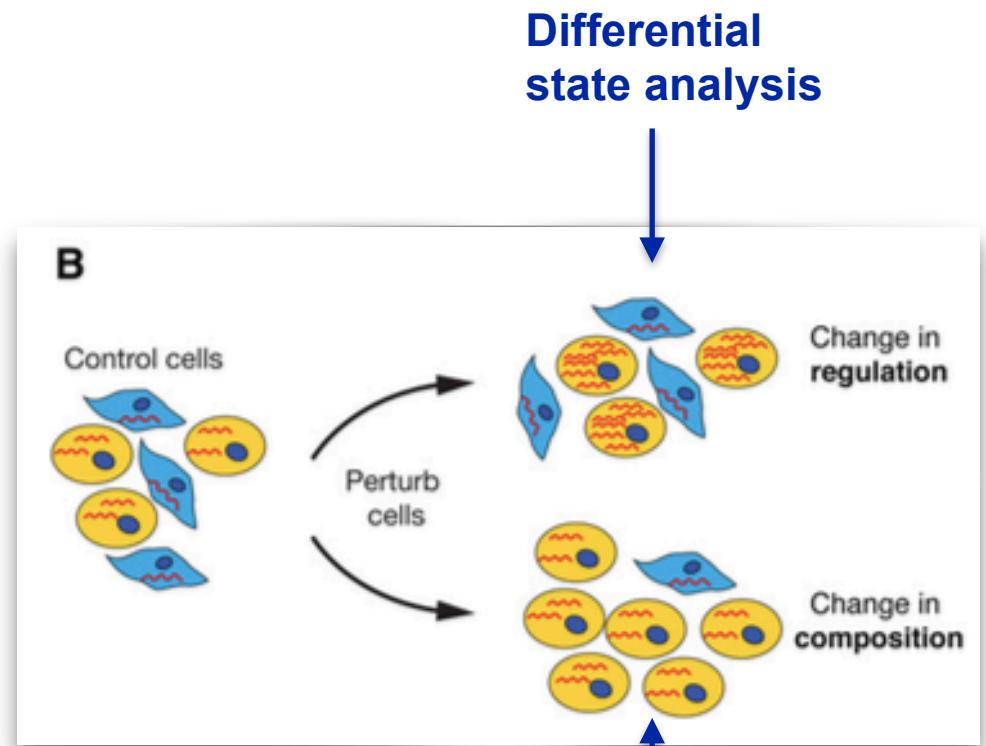
**Type:** more permanent  
**State:** more transient

Perspective

## Defining cell types and states with single-cell genomics

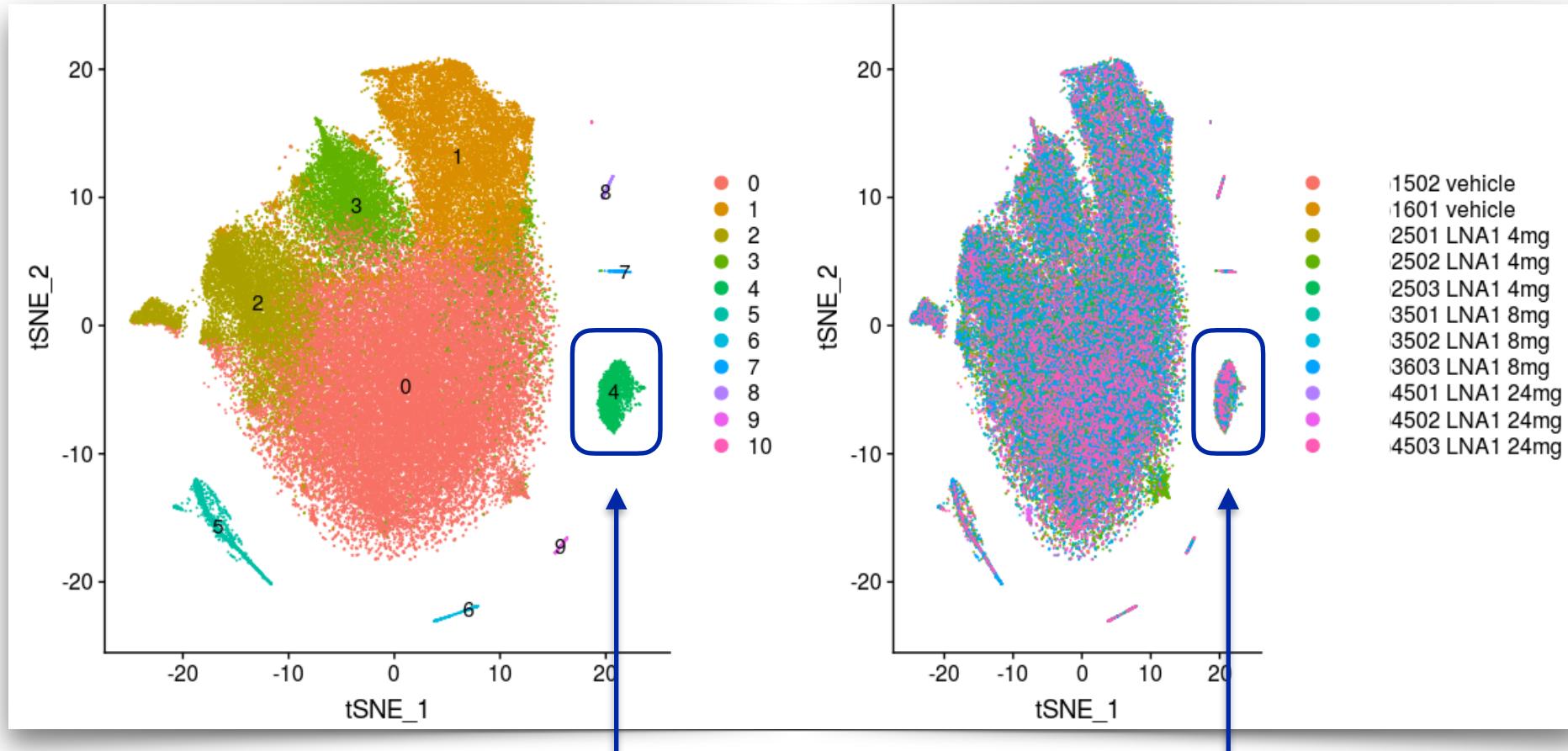
Cole Trapnell

Department of Genome Sciences, University of Washington, Seattle, Washington 98105, USA



**Differential abundance analysis**

# Two types of differential expression: marker gene DE, differential state analysis



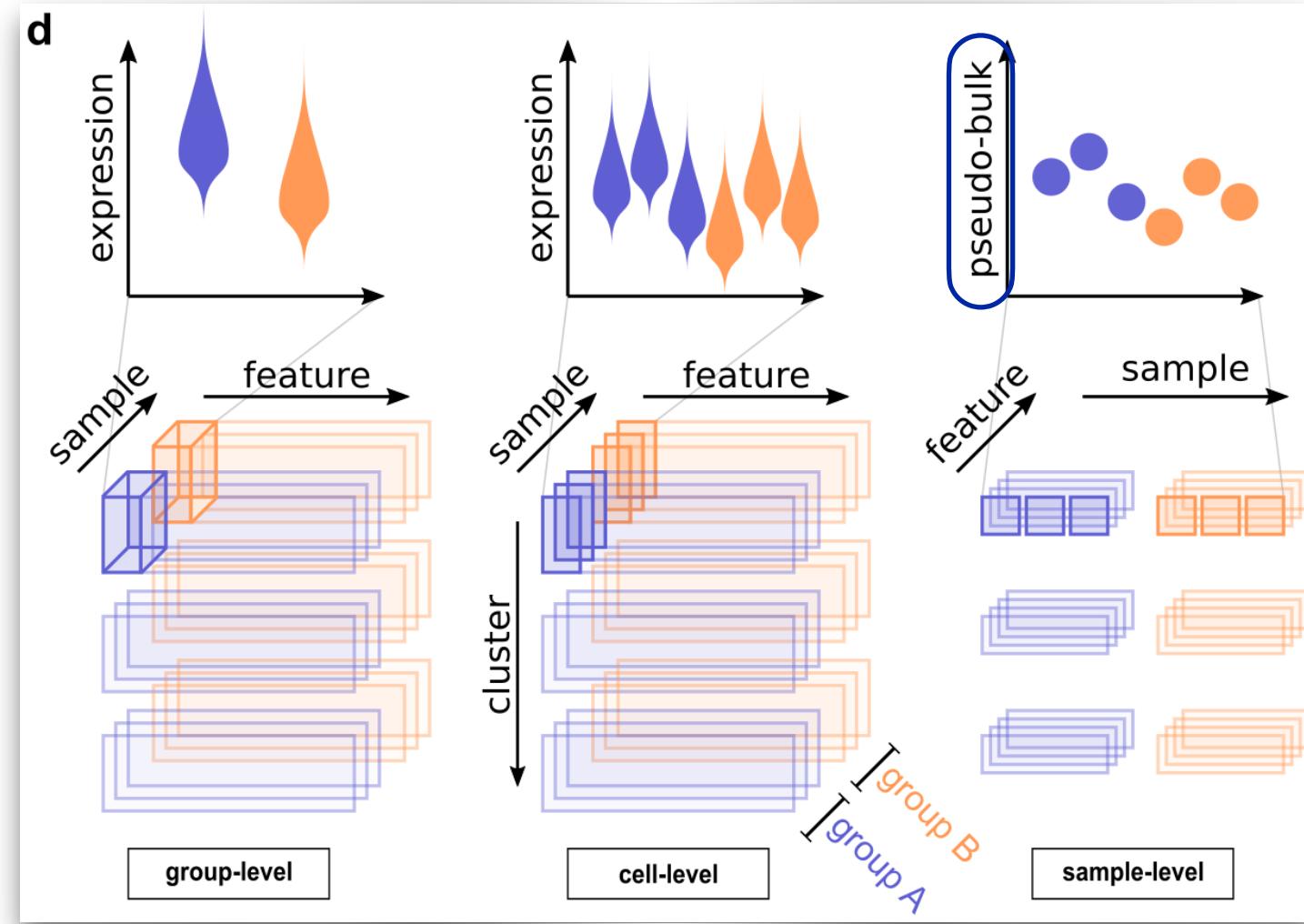
repeat for each population ..

Focus: Marker gene DE

Focus: cross-sample DE

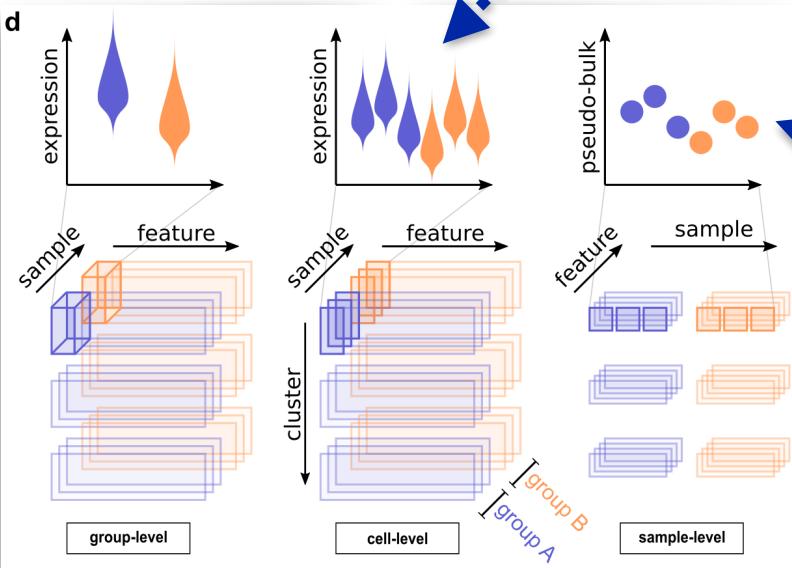
# After “Cell Type Prediction” / “Clustering”, various ways to view the inference problem

Multi-sample  
Multi-condition  
Multi-population



# Some precedent, but different contexts

Multi-sample  
Multi-condition  
Multi-population



## Batch effects and the effective design of single-cell gene expression studies

Po-Yuan Tung<sup>1,\*</sup>, John D. Blischak<sup>1,2,\*</sup>, Chiaowen Joyce Hsiao<sup>1,\*</sup>, David A. Knowles<sup>3,4</sup>, Jonathan E. Burnett<sup>1</sup>, Jonathan K. Pritchard<sup>3,5,6</sup> & Yoav Gilad<sup>1,7</sup>

mixed models

## Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data

AARON T. L. LUN\*

Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK  
aaron.lun@cruk.cam.ac.uk

JOHN C. MARONI

Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK  
EMBL European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK and Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
maroni@ebi.ac.uk

“A solution is proposed whereby counts are summed from all cells in each plate and the count sums for all plates are used in the DE analysis.”

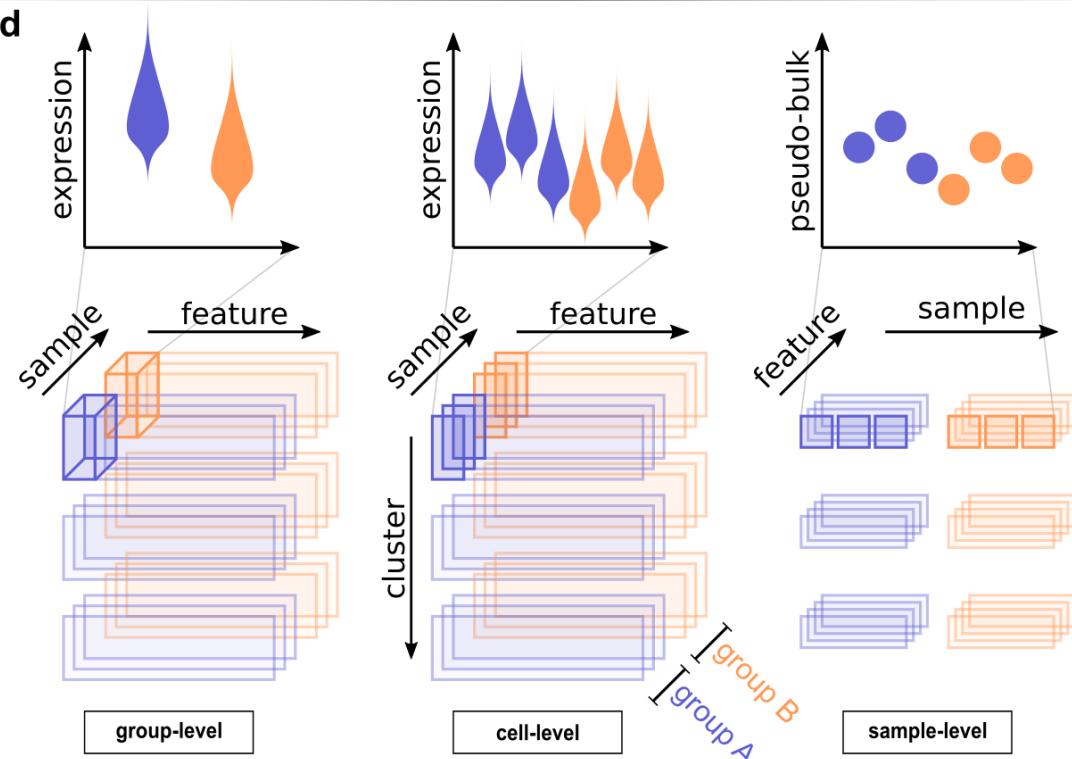
# Pseudo-bulking works well

<https://doi.org/10.1038/s41467-020-19894-4>

OPEN

*muscat* detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data

Helena L. Crowell<sup>1,2</sup>, Charlotte Soneson<sup>1,2,3,6</sup>, Pierre-Luc Germain<sup>1,4,6</sup>, Daniela Calini<sup>5</sup>, Ludovic Collin<sup>5</sup>, Catarina Raposo<sup>5</sup>, Dheeraj Malhotra<sup>5</sup> & Mark D. Robinson<sup>1,2✉</sup>



ARTICLE

<https://doi.org/10.1038/s41467-021-25960-2>

OPEN

## Confronting false discoveries in single-cell differential expression

Jordan W. Squair<sup>1,2,3</sup>, Matthieu Gautier<sup>1,2</sup>, Claudia Kathe<sup>1,2</sup>, Mark A. Anderson<sup>1,2</sup>, Nicholas D. James<sup>1,2</sup>, Thomas H. Hutson<sup>1,2</sup>, Rémi Hudelle<sup>1,2</sup>, Taha Qaiser<sup>1,3</sup>, Kaya J. E. Matson<sup>4</sup>, Quentin Barraud<sup>1,2</sup>, Ariel J. Levine<sup>1,4</sup>, Gioele La Manno<sup>1</sup>, Michael A. Skinnider<sup>1,2,5,6✉</sup> & Grégoire Courtine<sup>1,2,6✉</sup>

Matters arising

<https://doi.org/10.1038/s41467-022-35519-4>

## A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis

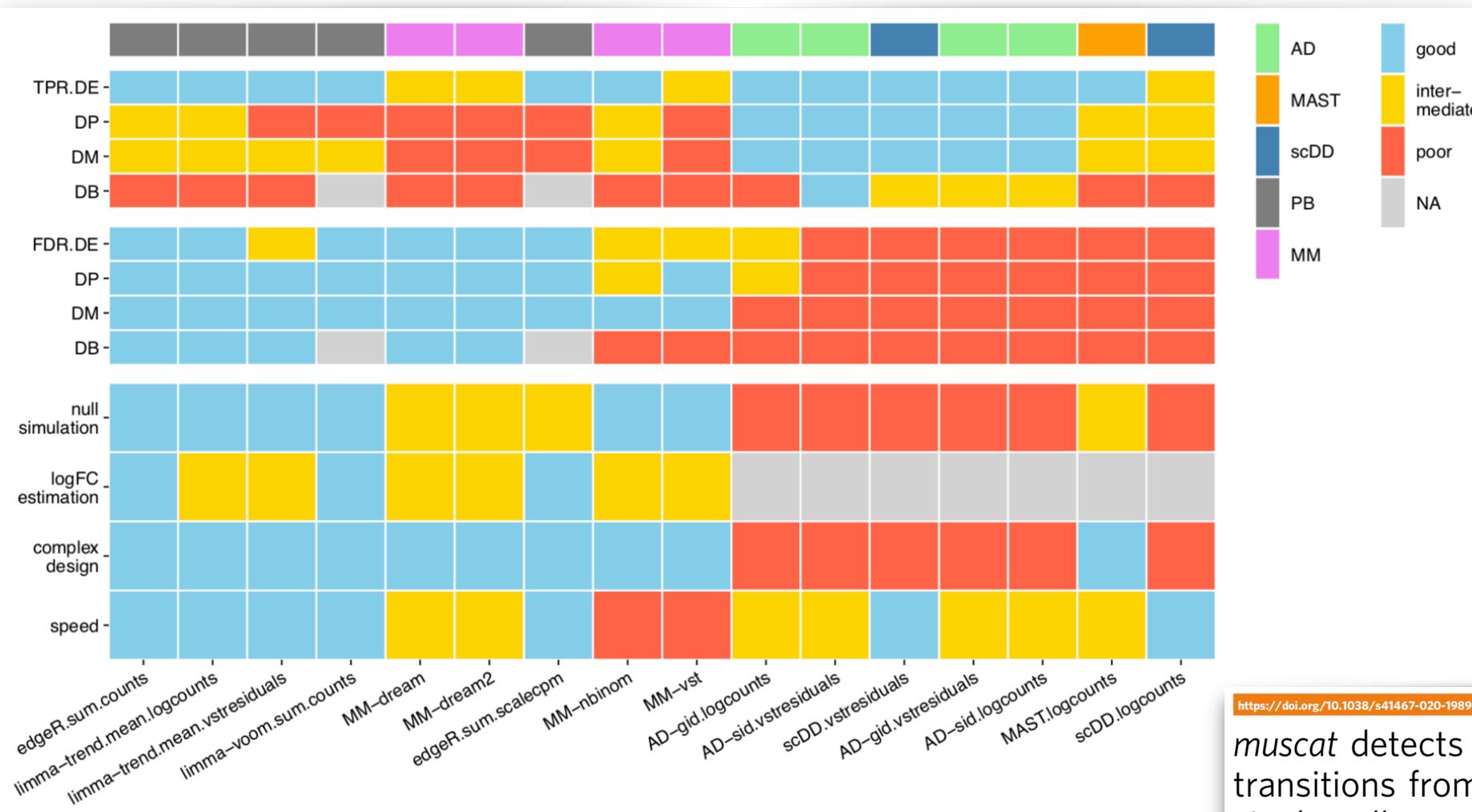
Received: 15 February 2022

Alan E. Murphy<sup>1,2✉</sup> & Nathan G. Skene<sup>1,2✉</sup>

Accepted: 8 December 2022

# Current rating: differential state analysis

PB = pseudobulk  
AD = Anderson-Darling  
MM = mixed models



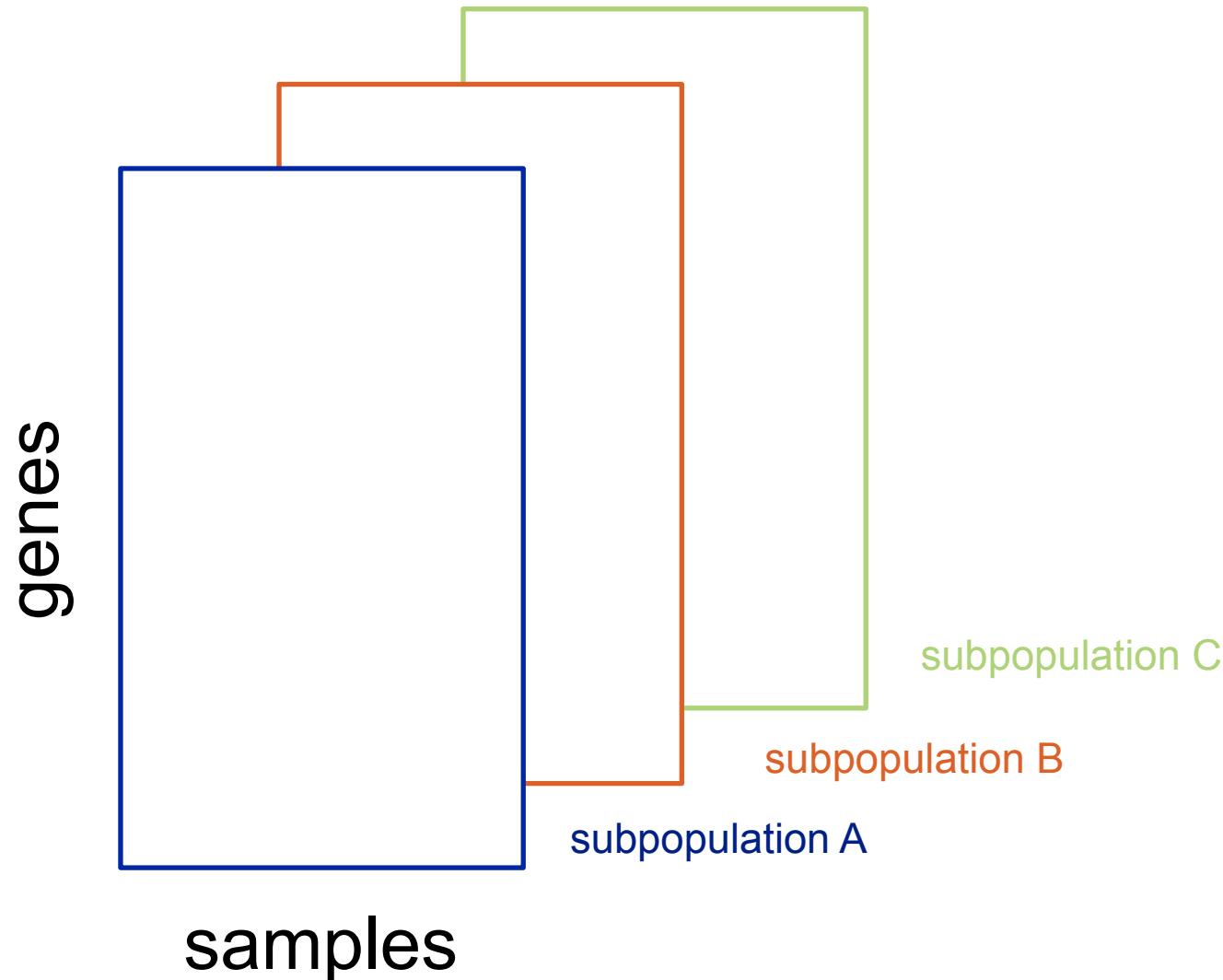
<https://doi.org/10.1038/s41467-020-19894-4>

OPEN

*muscat* detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data

Helena L. Crowell  <sup>1,2</sup>, Charlotte Soneson <sup>1,2,3,6</sup>, Pierre-Luc Germain <sup>1,4,6</sup>, Daniela Calini <sup>5</sup>, Ludovic Collin <sup>5</sup>, Catarina Raposo <sup>5</sup>, Dheeraj Malhotra <sup>5</sup> & Mark D. Robinson  <sup>1,2</sup>

# Results suggesting reducing the problem to a set of bulk RNA-seq DE analyses is fast an powerful



- At least 1 modeling question remains: should you model each population on its own, or all populations at once?
- Currently done as the former; in the latter, one could model a main effect (common response) in addition to interactions (subpopulation specific deviations)



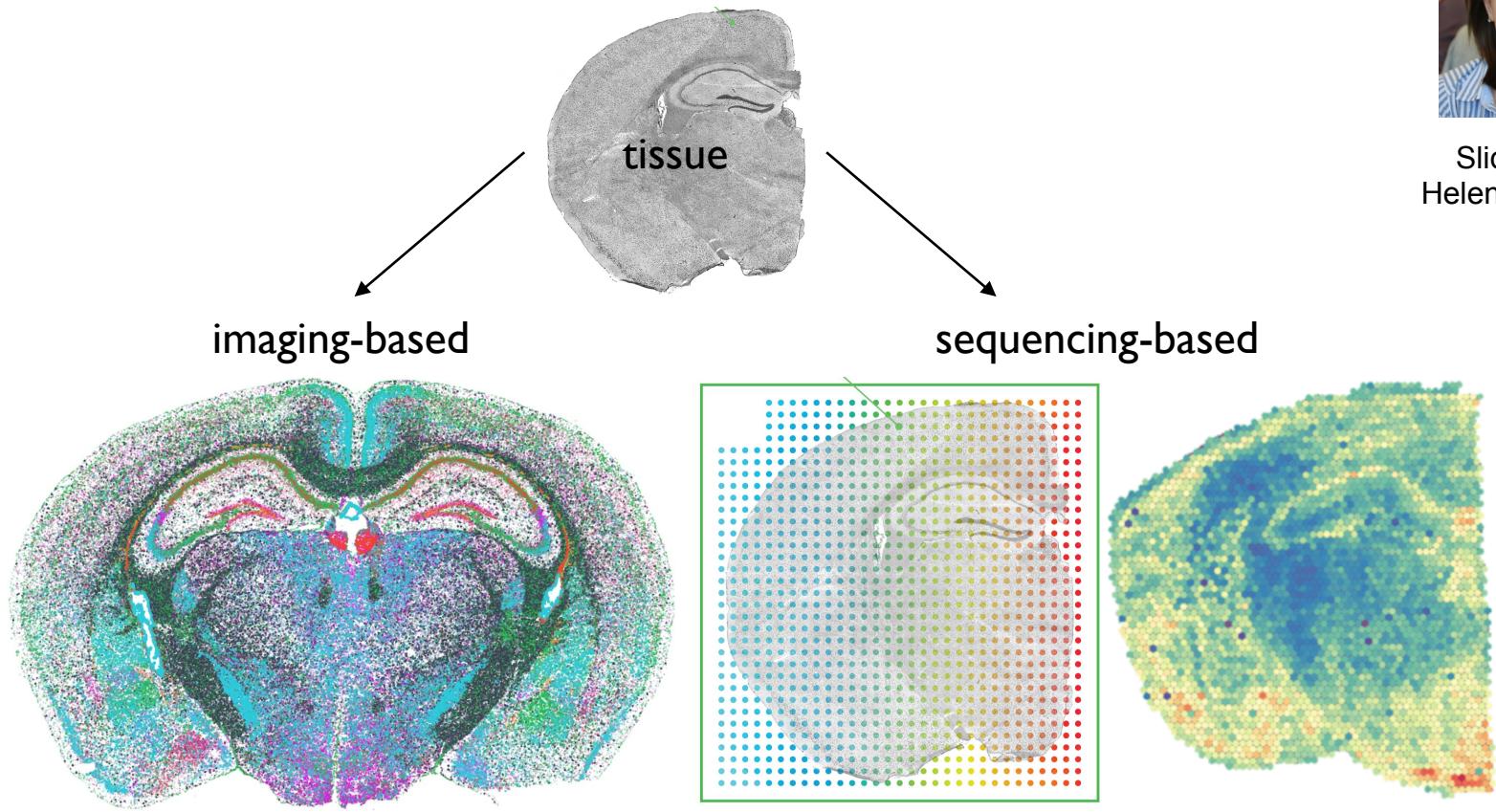
# Statistical methods for spatial omics data

- Overview on the technologies (review)
- Finding spatially-variable genes
- (Deconvoluting low-resolution spatial omics data)
- Spatially-aware dimension reduction / clustering
- Cell-cell communication —> co-localization
- Classical spatial statistics
  - Point patterns: random, clustered, intensity/correlation
  - Useful summaries / functions
  - models with spatially correlated errors
- (Integration w/ single cell RNA-seq)
- (Segmentation, preprocessing)



bulk

single-cell



- molecule-level data
- targeted panel (100s of features; 2024: 1000s)
- single-cell resolution requires segmentation

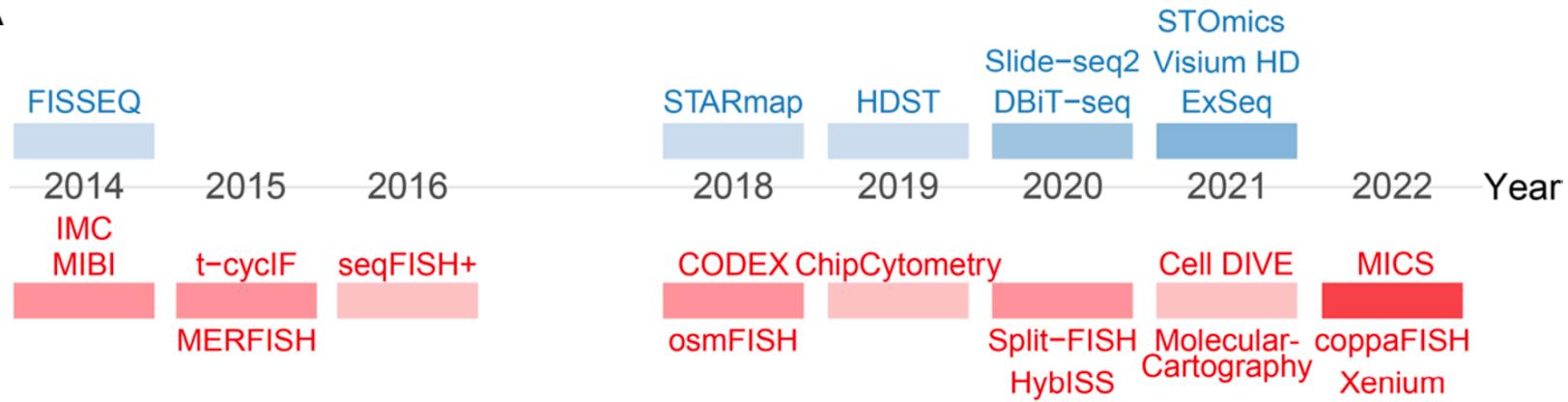
- spot-level data
- whole transcriptome (10,000s of features)
- single-cell resolutions requires aggregation or deconvolution



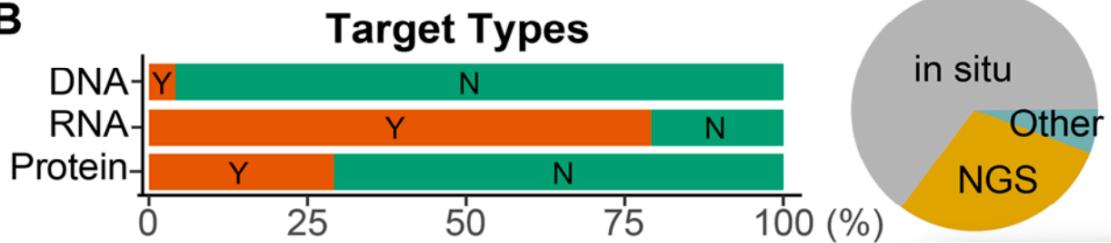
Slide from  
Helena Crowell

# (Spatial omics) Technology explosion

**A**

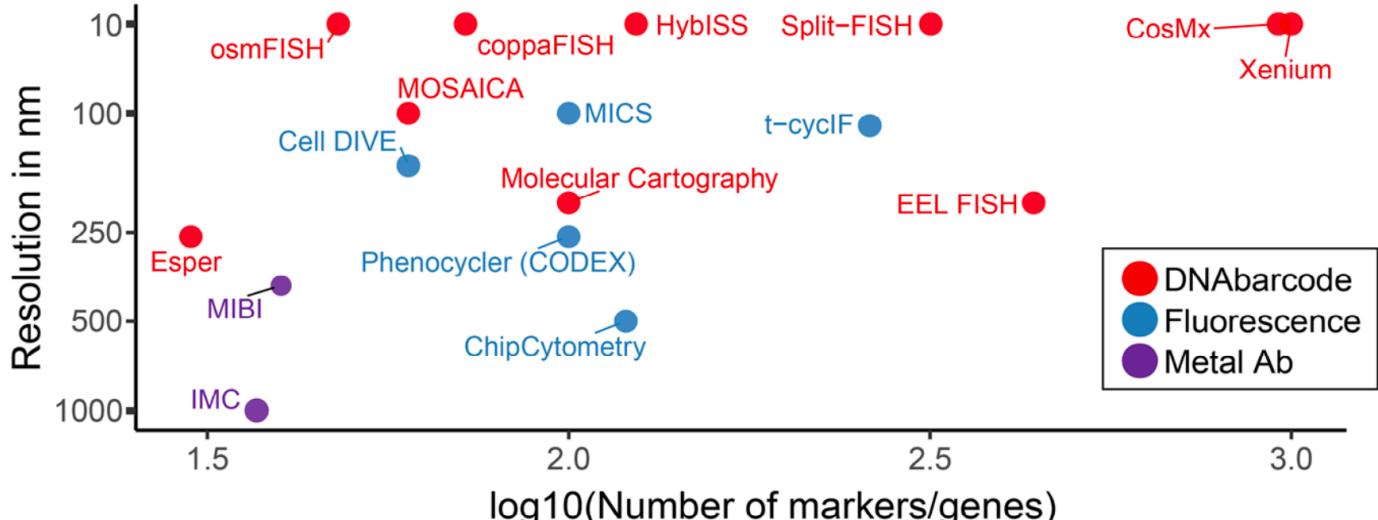


**B**

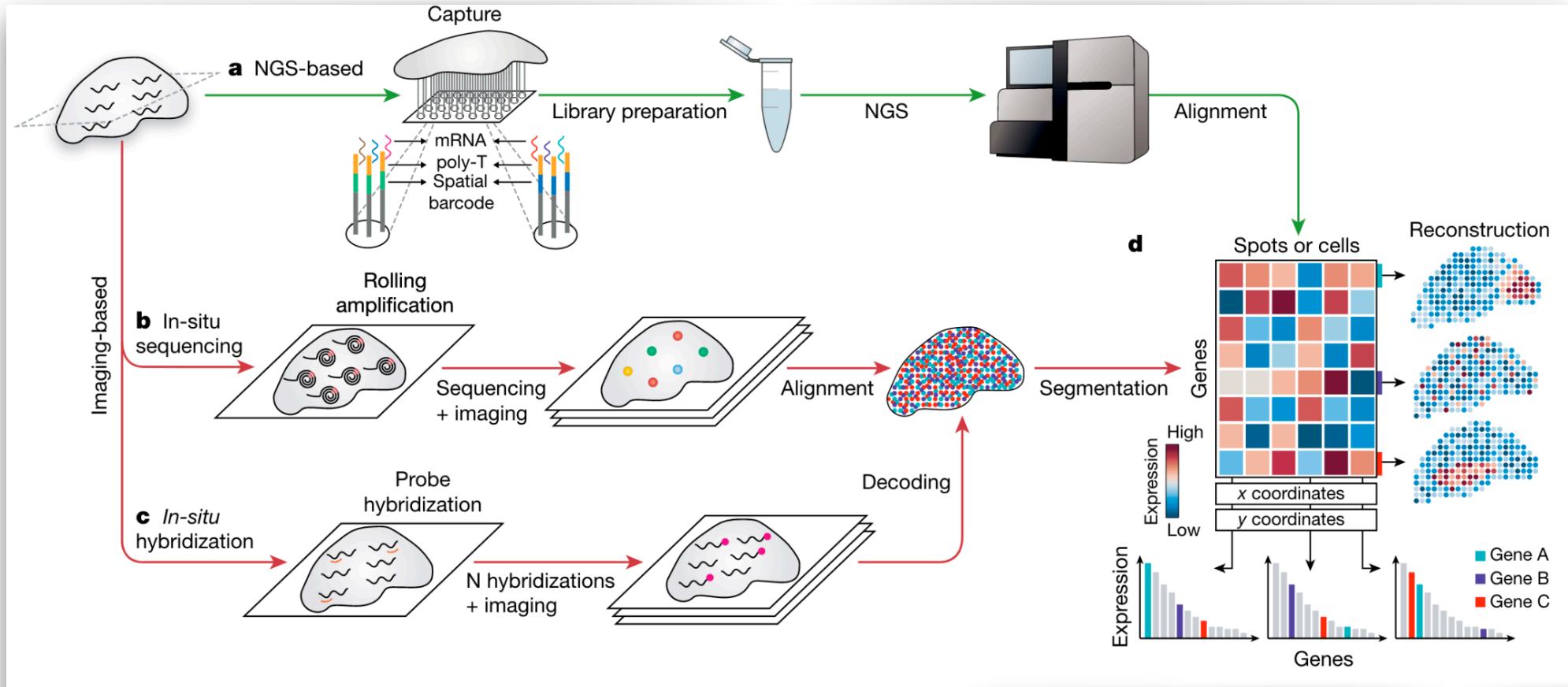


*blue colored techniques are sequencing-based while red colored techniques are multiplexed IHC/IF methodologies*

**C**



# Technology choices: expression table + coordinates



**Fig. 1 | The technologies of spatial transcriptomics provide a gene-expression matrix.** **a**, NGS-based spatial transcriptomic methods barcode transcripts according to their location in a lattice of spots. **b**, ISS approaches directly read out the transcript sequence within the tissue. **c**, ISH

methods detect target fluorescent probes in a gene-expression matrix of genes and location

## Review

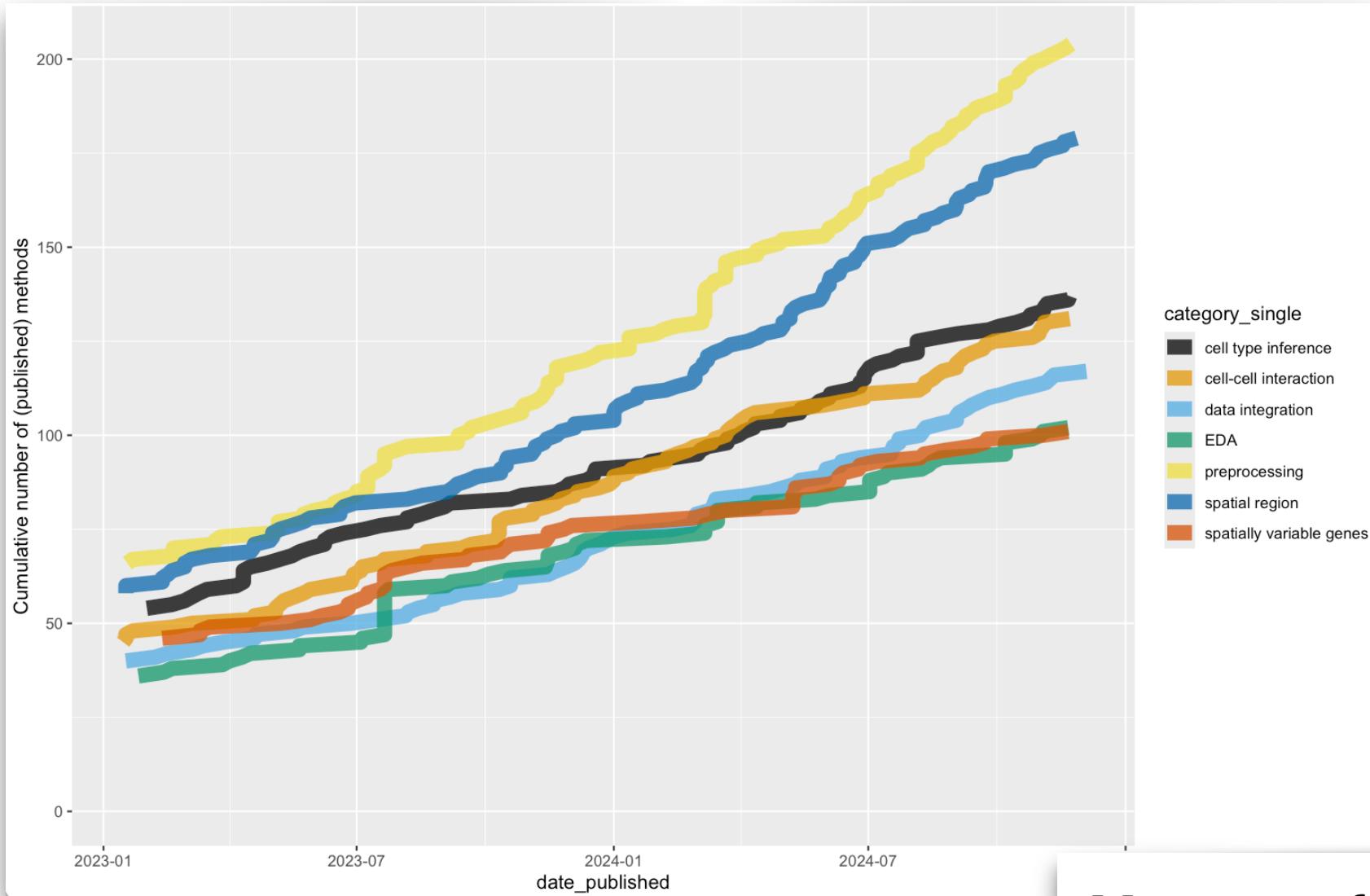
# Exploring tissue architecture using spatial transcriptomics



# Statistical methods for spatial omics data

- Overview on the technologies (review)
- Finding spatially-variable genes
- (Deconvoluting low-resolution spatial omics data)
- Spatially-aware dimension reduction / clustering
- Cell-cell communication —> co-localization
- Classical spatial statistics
  - Point patterns: random, clustered, intensity/correlation
  - Useful summaries / functions
  - models with spatially correlated errors
- (Integration w/ single cell RNA-seq)
- (Segmentation, preprocessing)

# (Spatial omics) computational method explosion



Museum of spatial transcriptomics

Lambda Moses<sup>ID</sup><sup>1</sup> and Lior Pachter<sup>ID</sup><sup>1,2</sup>✉



## Finding spatially-variable genes: SpatialDE

- SpatialDE: response = normal distribution with covariance with two components: i)  
based on distance b/w points  
- exponential decay; ii)  
constant non-spatial variance
- Null model: fit just the non-spatial variance (i.e., without sigma)
- Fit 2 models, likelihood ratio test

# SpatialDE: identification of spatially variable genes

Valentine Svensson<sup>1,2</sup> , Sarah A Teichmann<sup>1,3</sup>  
& Oliver Stegle<sup>2,4</sup>

**SpatialDE model.** SpatialDE models gene expression profiles  $y = (y_1, \dots, y_N)$  for a given gene across spatial coordinates  $X = (x_1, \dots, x_N)$ , using a multivariate normal model of the form

$$P(y | \mu, \sigma_s^2, \delta, \Sigma) = N(y | \mu \cdot 1, \sigma_s^2 \cdot (\Sigma + \delta \cdot I)) \quad (1)$$

The fixed effect  $\mu_g \cdot 1$  accounts for the mean expression level, and  $\Sigma$  denotes a spatial covariance matrix defined on the basis of the input coordinates of pairs of cells. SpatialDE uses the so-called squared exponential covariance function to define  $\Sigma$ :

$$\Sigma_{i,j} = k(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2 \cdot l^2}\right) \quad (2)$$



## Spatially variable genes

- different types (senses?) of spatially variable genes

# nnSVG for the scalable identification of spatially variable genes using nearest-neighbor Gaussian processes

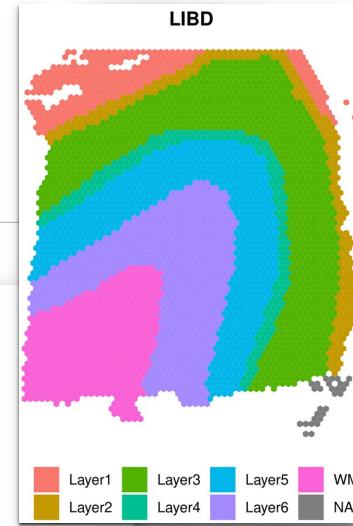
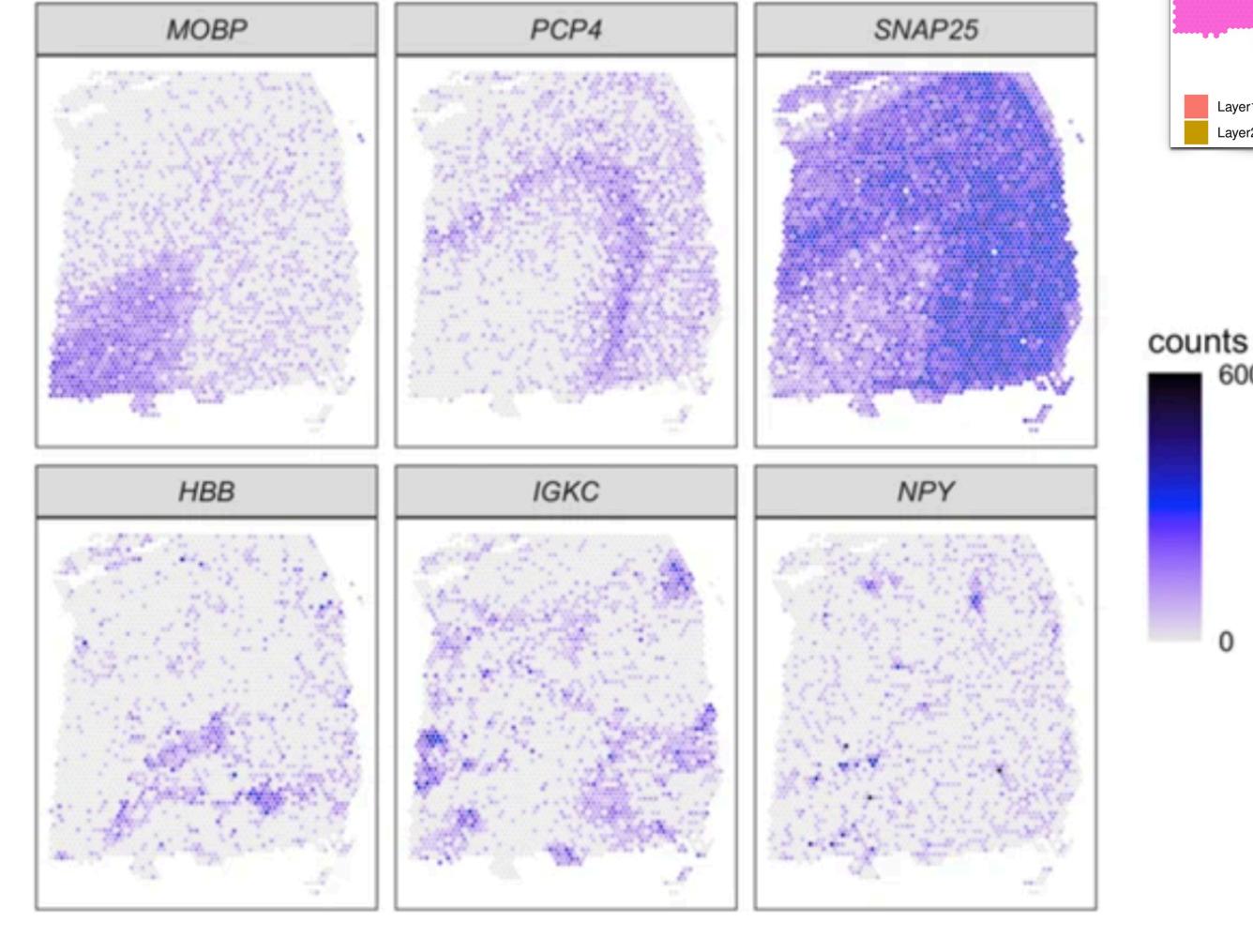
Received: 15 June 2022

Lukas M. Weber <sup>1</sup>, Arkajyoti Saha<sup>2</sup>, Abhirup Datta <sup>1</sup>, Kasper D. Hansen <sup>1</sup> &

Accepted: 23 June 2023

Stephanie C. Hicks <sup>1</sup>

Selected SVGs: human DLPFC

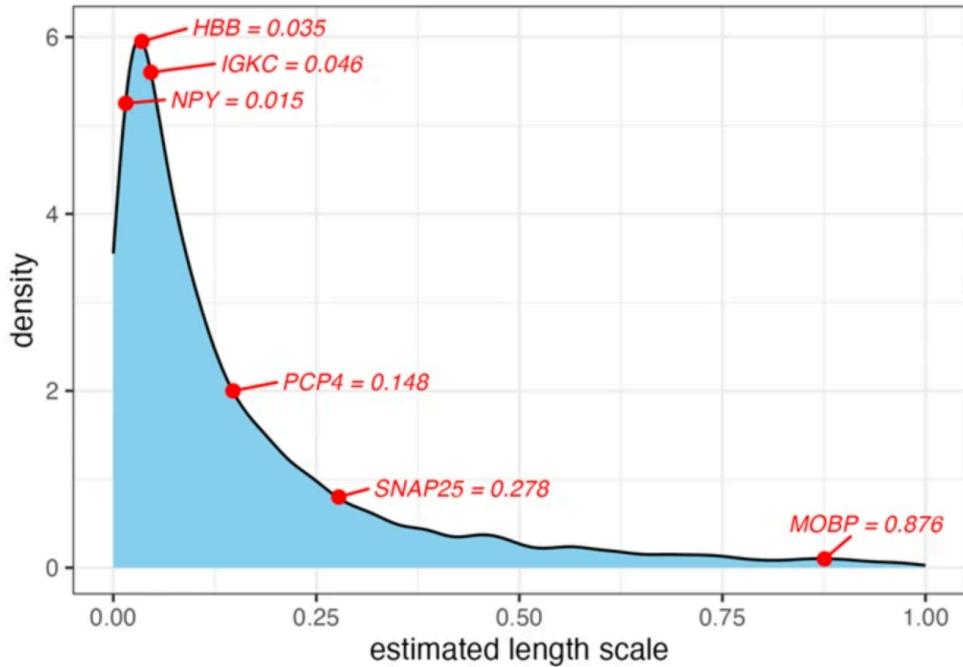




## Spatially variable genes

$$C_{ij}(\theta) = \sigma^2 \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{l}\right)$$

**b** nnSVG length scales: human DLPFC



# nnSVG for the scalable identification of spatially variable genes using nearest-neighbor Gaussian processes

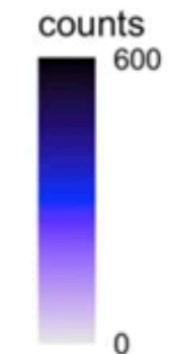
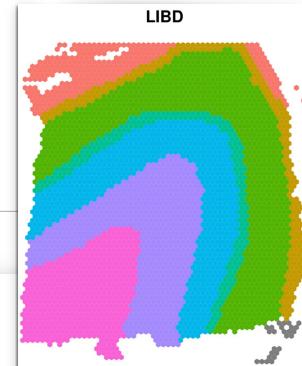
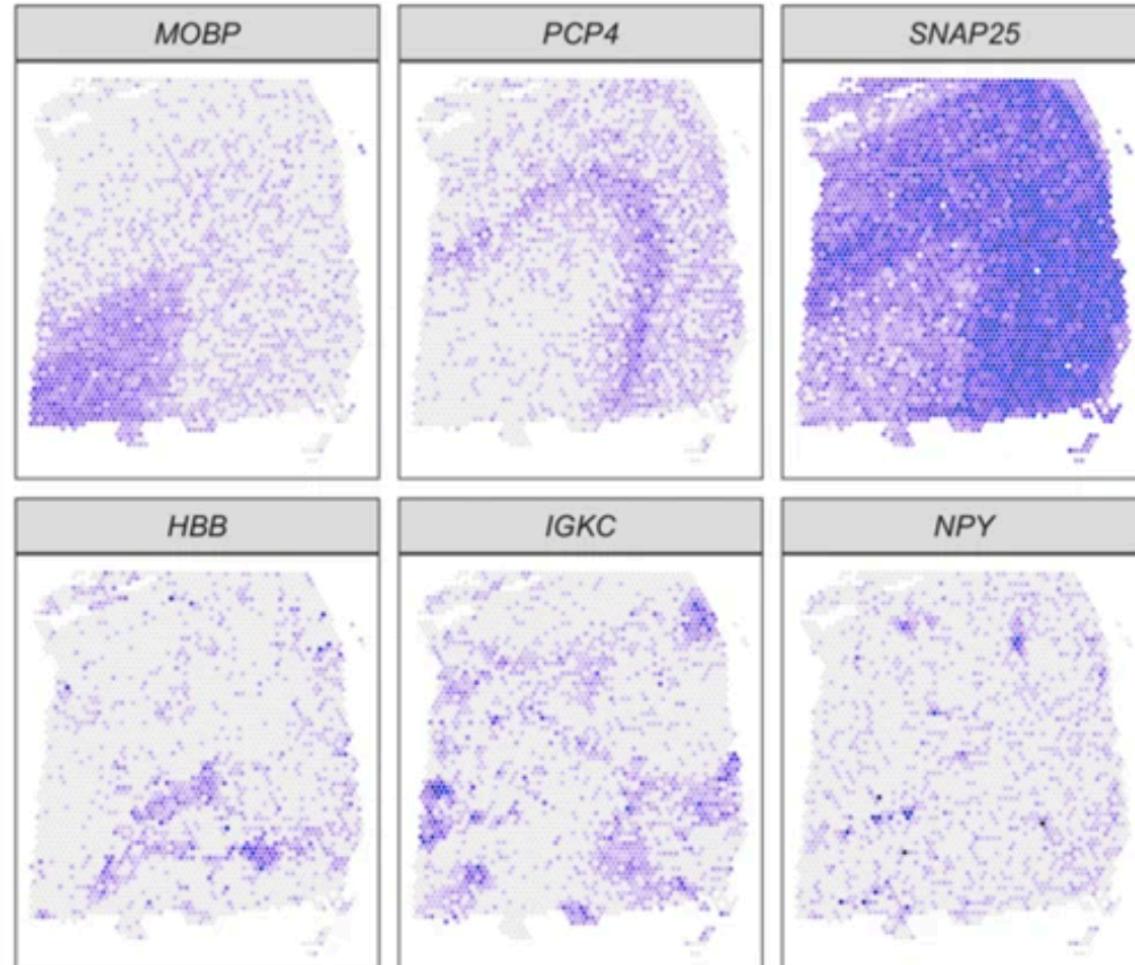
Received: 15 June 2022

Lukas M. Weber , Arkajyoti Saha<sup>2</sup>, Abhirup Datta , Kasper D. Hansen &

Accepted: 23 June 2023

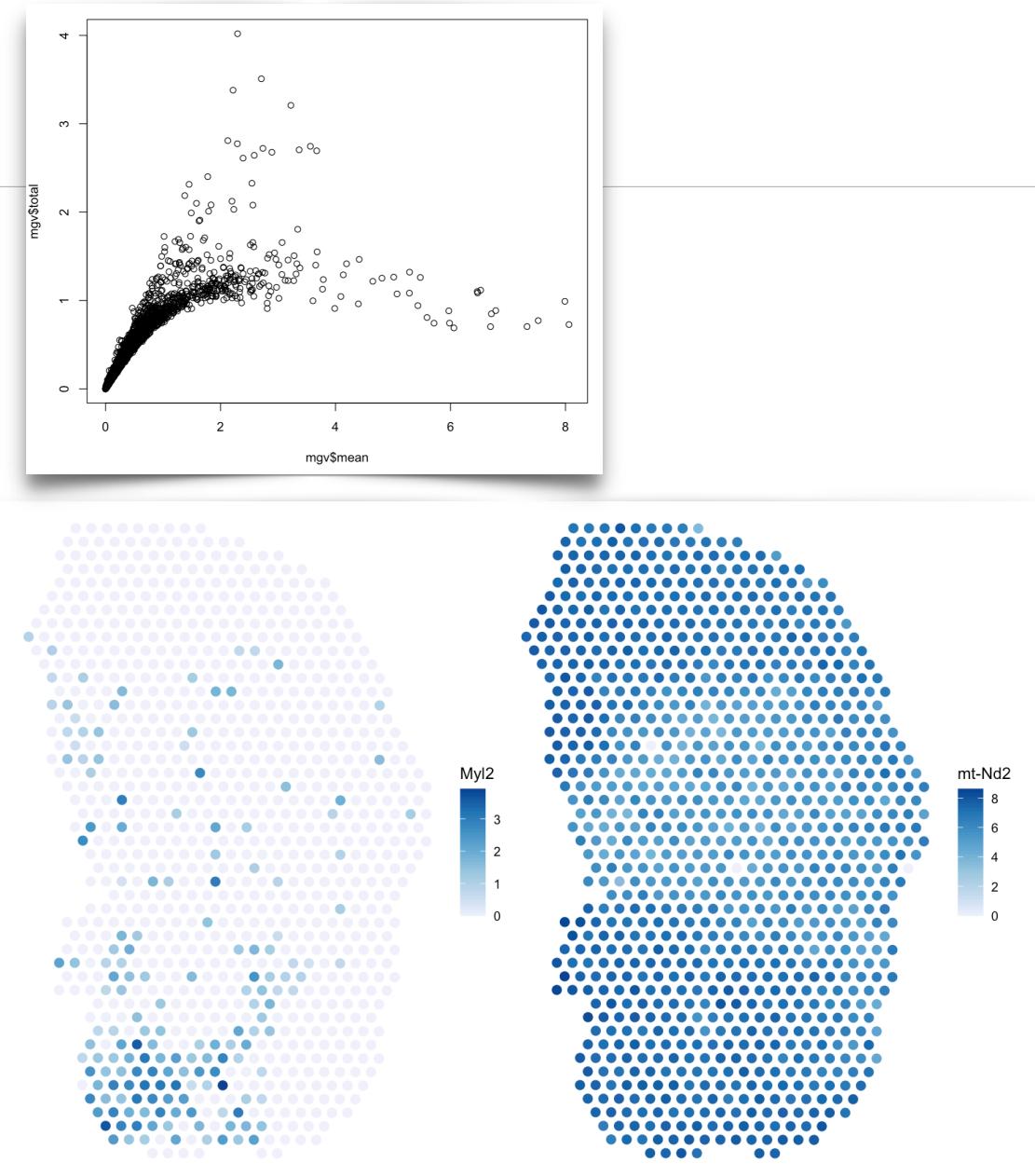
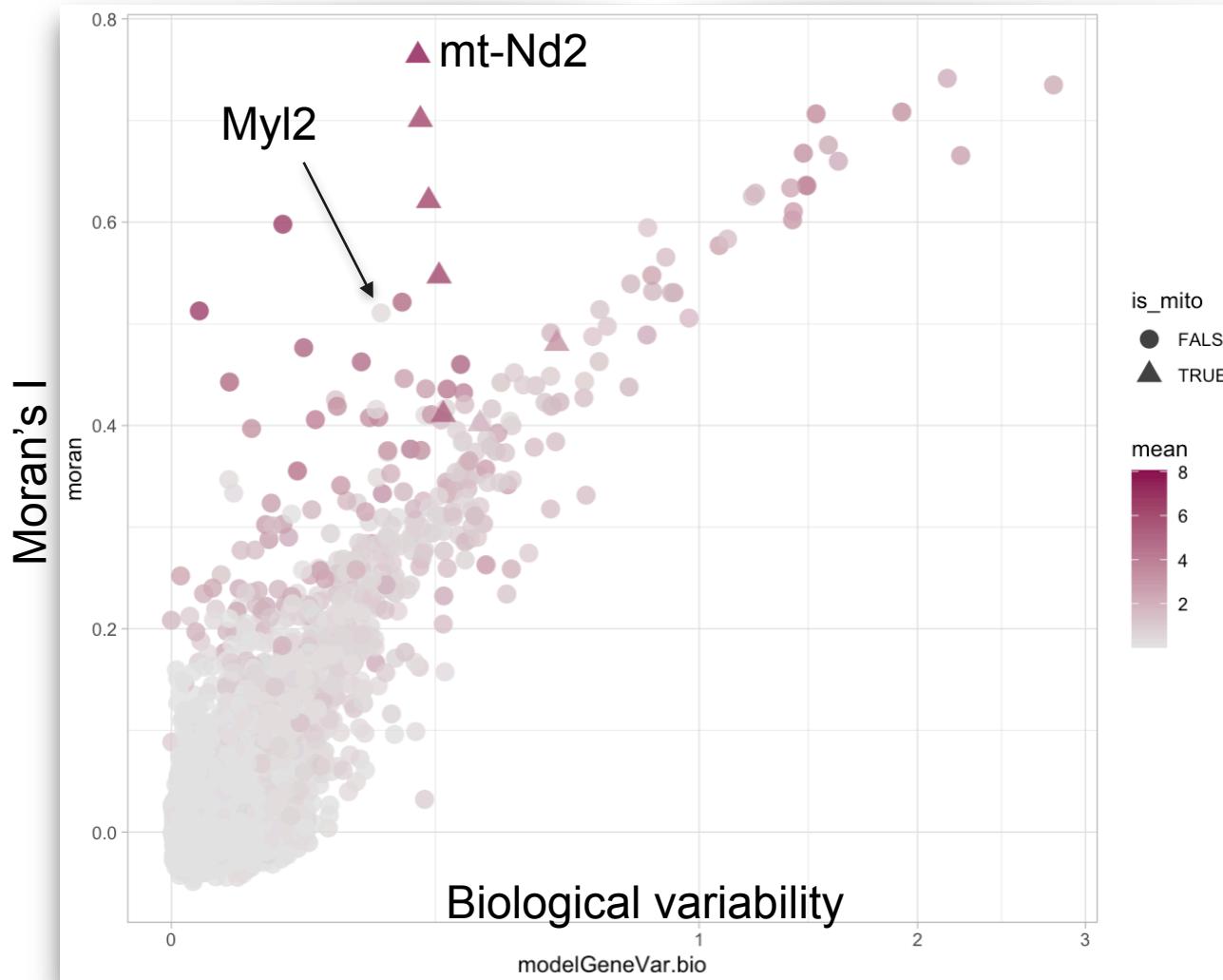
Stephanie C. Hicks

## Selected SVGs: human DLPFC



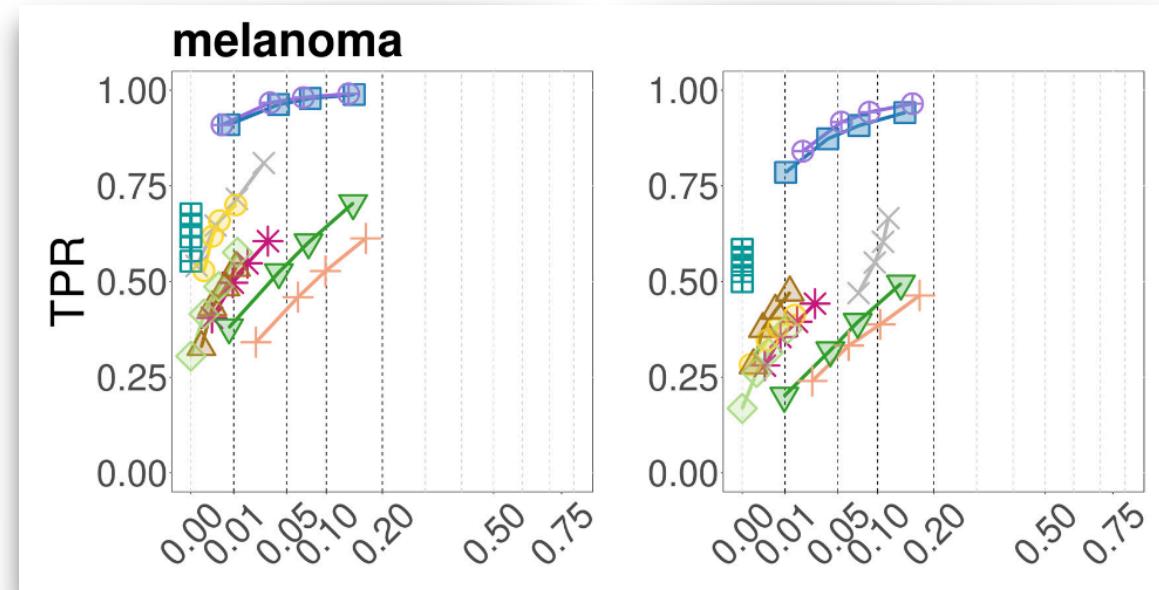
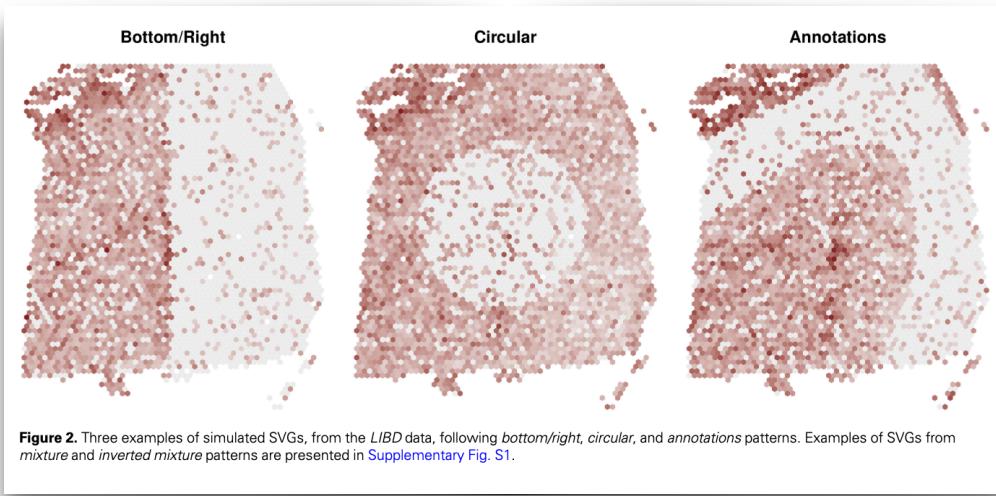


## Spatially variable versus highly variable



(More mathematical details on  
Moran's I below)

# Alternatively, spatially variable features = DE between domains



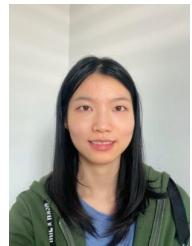
To find spatially variable genes (SVGs); spatial clustering + classical statistical method works quite well

JOURNAL ARTICLE

**DESpace: spatially variable gene detection via differential expression testing of spatial clusters** ⚡

Peiying Cai, Mark D Robinson, Simone Tiberi ✉

Simone  
Tiberi



Peiying Cai



# Statistical methods for spatial omics data

- Overview on the technologies (review)
- Finding spatially-variable genes
- (Deconvoluting low-resolution spatial omics data)
- Spatially-aware dimension reduction / clustering
- Cell-cell communication —> co-localization
- Classical spatial statistics
  - Point patterns: random, clustered, intensity/correlation
  - Useful summaries / functions
  - models with spatially correlated errors
- (Integration w/ single cell RNA-seq)
- (Segmentation, preprocessing)



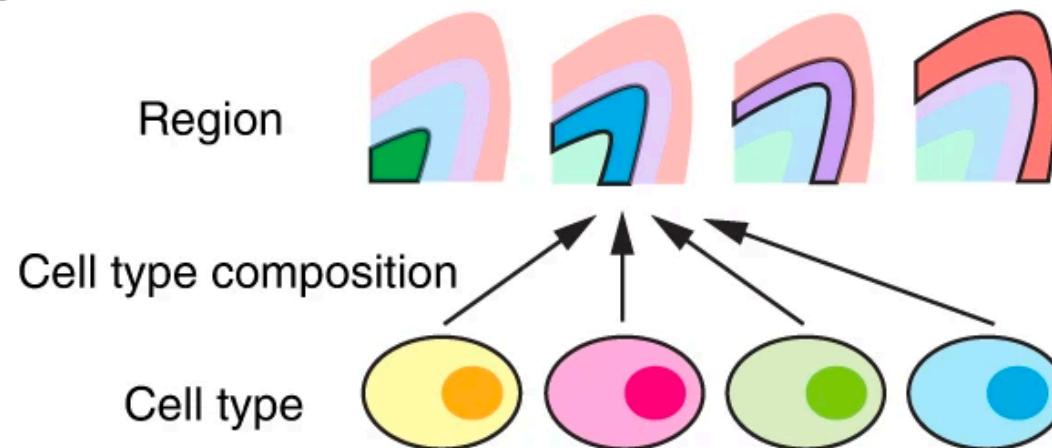
# Spatially aware dimension reduction for spatial transcriptomics

Received: 10 March 2022

Lulu Shang <sup>1,2</sup> & Xiang Zhou <sup>1,2</sup>

Spatial domain detection ~ spatially homogeneous regions ~ spatial niches

b.



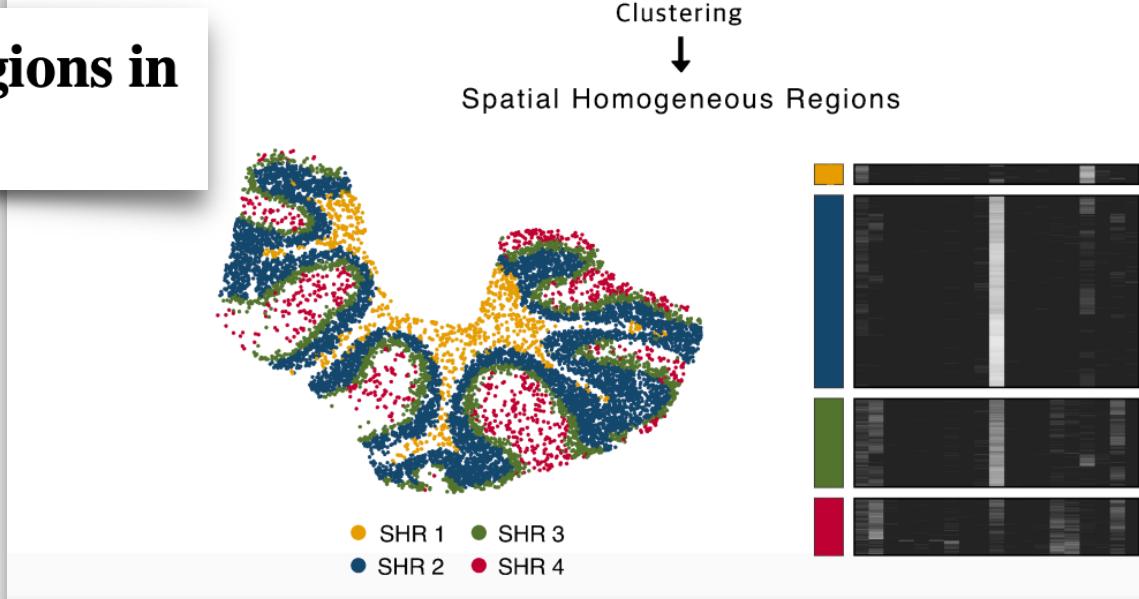
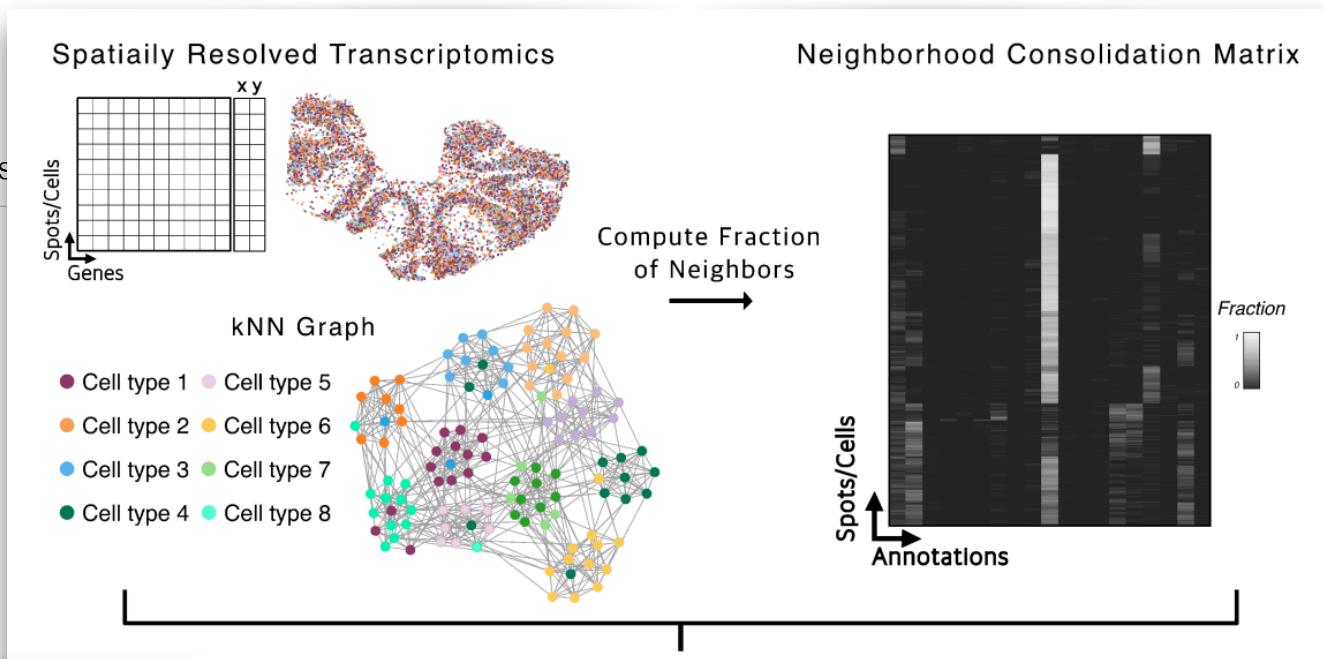
Scenario 1	70%	10%	10%	10%
Scenario 2	45%	45%	5%	5%
Scenario 3	60%	30%	5%	5%
Scenario 4	35%	30%	30%	5%

<https://www.nature.com/articles/s41467-022-34879-1>



## Spatial domain detection ~ spatially homogeneous regions

**Identification of spatial homogeneous regions in tissues with concordex**





# BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis

Spatial clustering / domain detection (BANKSY)  
→ combine transcription and spatial information

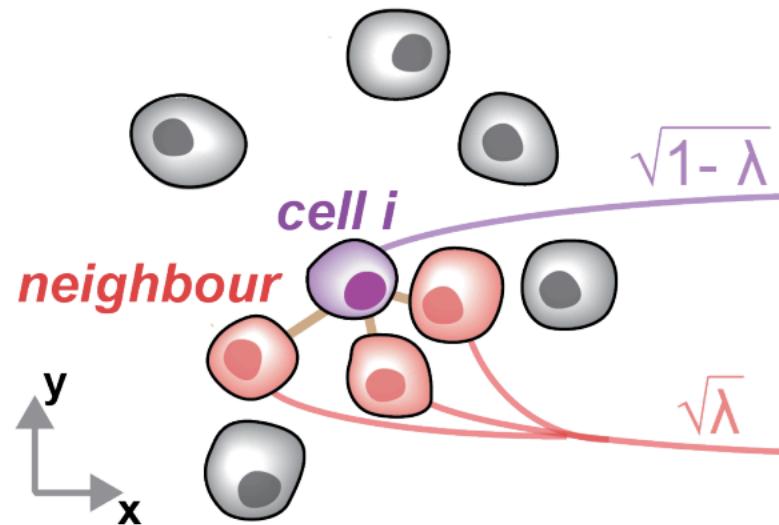
Received: 3 April 2023

Accepted: 16 January 2024

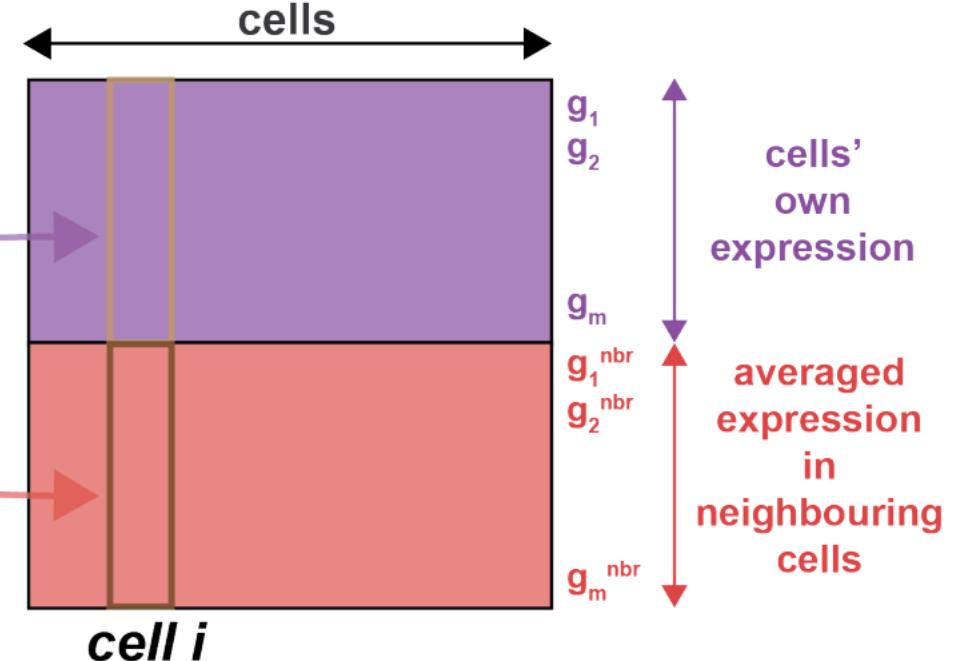
Published online: 27 February 2024

Vipul Singhal <sup>1,13</sup>, Nigel Chou <sup>1,13</sup>, Joseph Lee <sup>2</sup>, Yifei Yue <sup>3</sup>, Jinyue Liu <sup>1</sup>, Wan Kee Chock <sup>1</sup>, Li Lin <sup>4</sup>, Yun-Ching Chang <sup>5</sup>, Erica Mei Ling Teo <sup>5</sup>, Jonathan Aow <sup>1</sup>, Hwee Kuan Lee <sup>4,6,7,8,9,10</sup>, Kok Hao Chen <sup>1</sup> & Shyam Prabhakar <sup>1,11,12</sup>

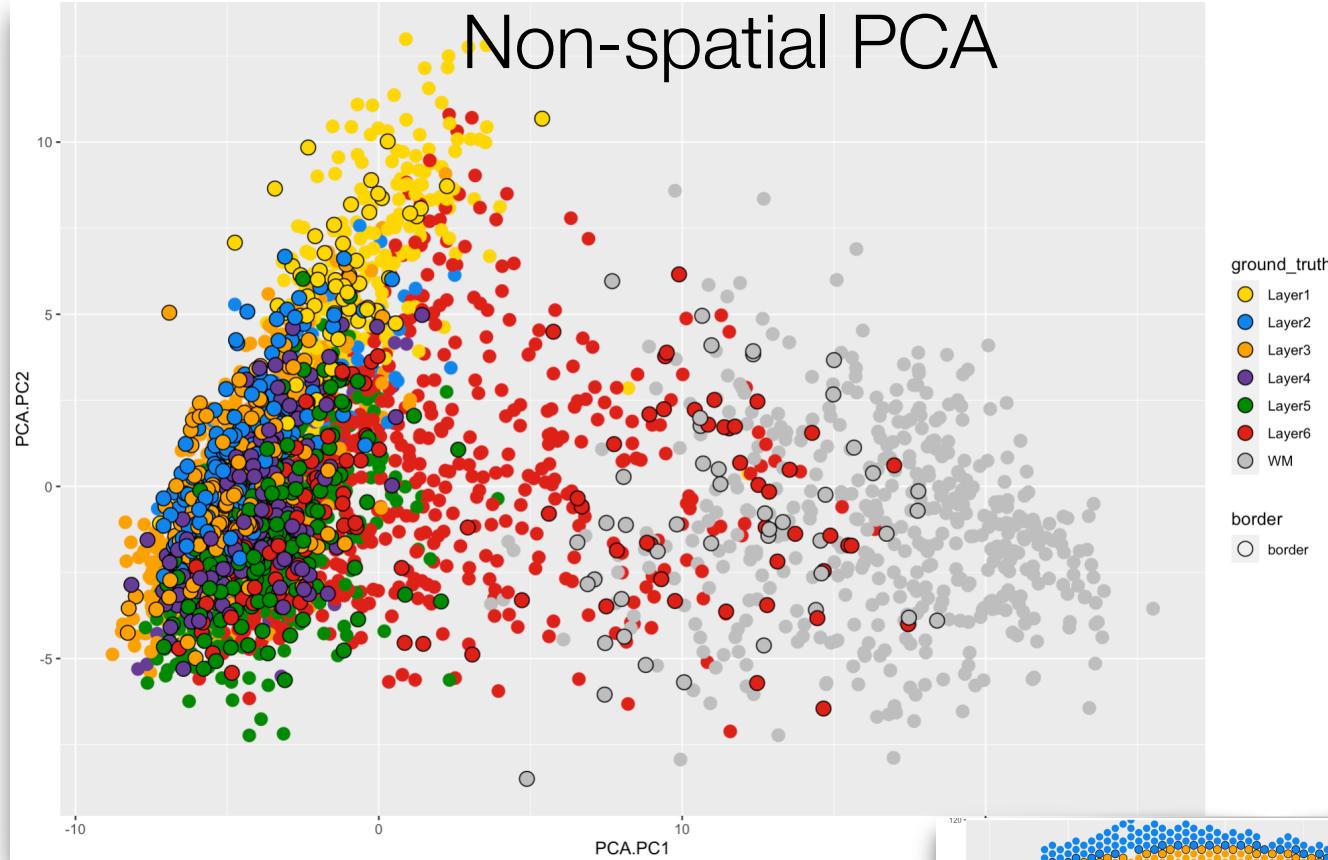
## a Cells in physical space



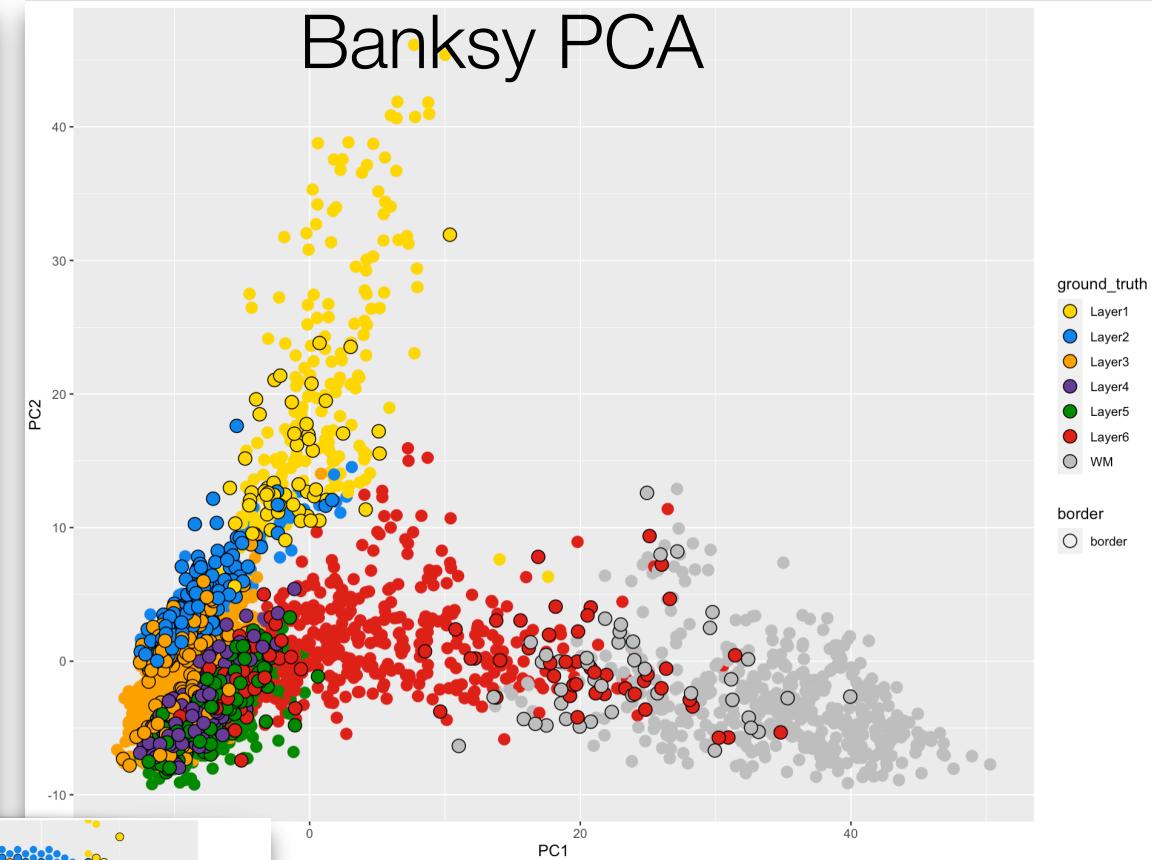
## Neighbour-augmented expression matrix



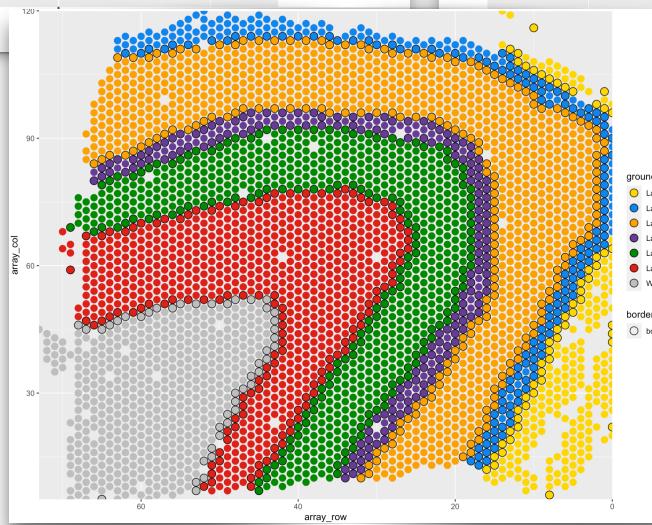
# Non-spatial PCA



# Banksy PCA



Sample 151673





# Statistical methods for spatial omics data

- Overview on the technologies (review)
- Finding spatially-variable genes
- (Deconvoluting low-resolution spatial omics data)
- Spatially-aware dimension reduction / clustering
- Cell-cell communication —> co-localization
- Classical spatial statistics
  - Point patterns: random, clustered, intensity/correlation
  - Useful summaries / functions
  - models with spatially correlated errors
- (Integration w/ single cell RNA-seq)
- (Segmentation, preprocessing)



## Spatial autocorrelation: Global Moran's I

- Global measure of auto-correlation (correlation to signal nearby in space); assume homogeneity!
- Alternative: Geary's C

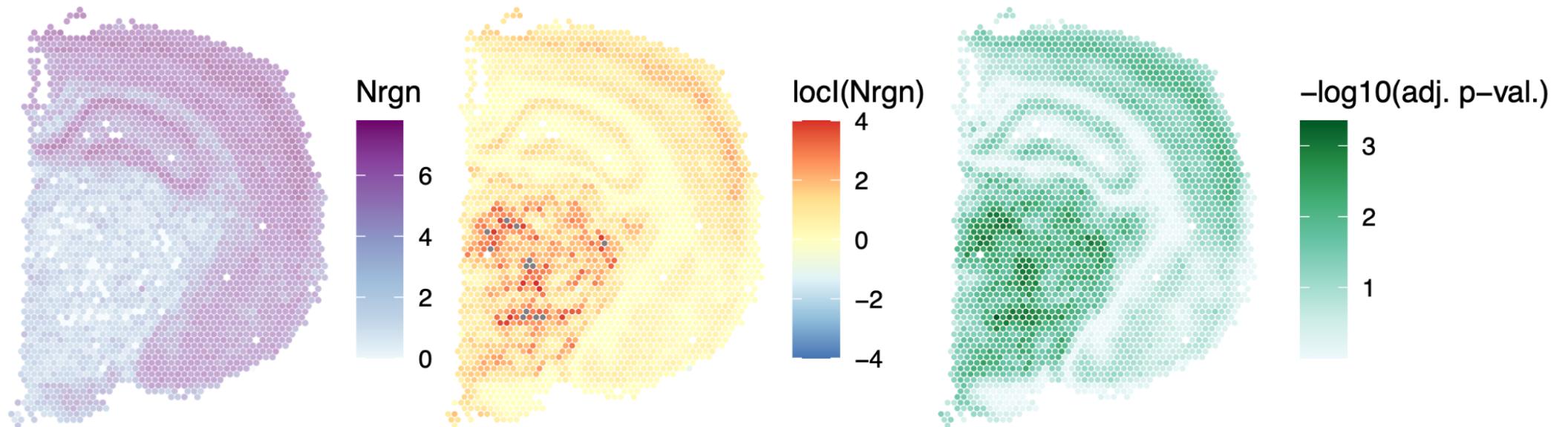
$$I = \frac{1}{\sum_{ij} w_{ij}} \frac{\sum_{ij} w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{N^{-1} \sum_i (X_i - \bar{X})^2}$$

$$C = \frac{(N - 1) \sum_i \sum_j w_{ij} (x_i - x_j)^2}{2W \sum_i (x_i - \bar{x})^2}$$

## Spatial autocorrelation: Local Moran's I

- Local measure of auto-correlation (correlation to signal nearby in space)

$$I_i = \frac{x_i - \bar{x}}{\sum_{k=1}^n (x_k - \bar{x})^2 / (n - 1)} \sum_{j=1}^n w_{ij}(x_j - \bar{x})$$





$$\text{Global Moran's } R = \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}},$$

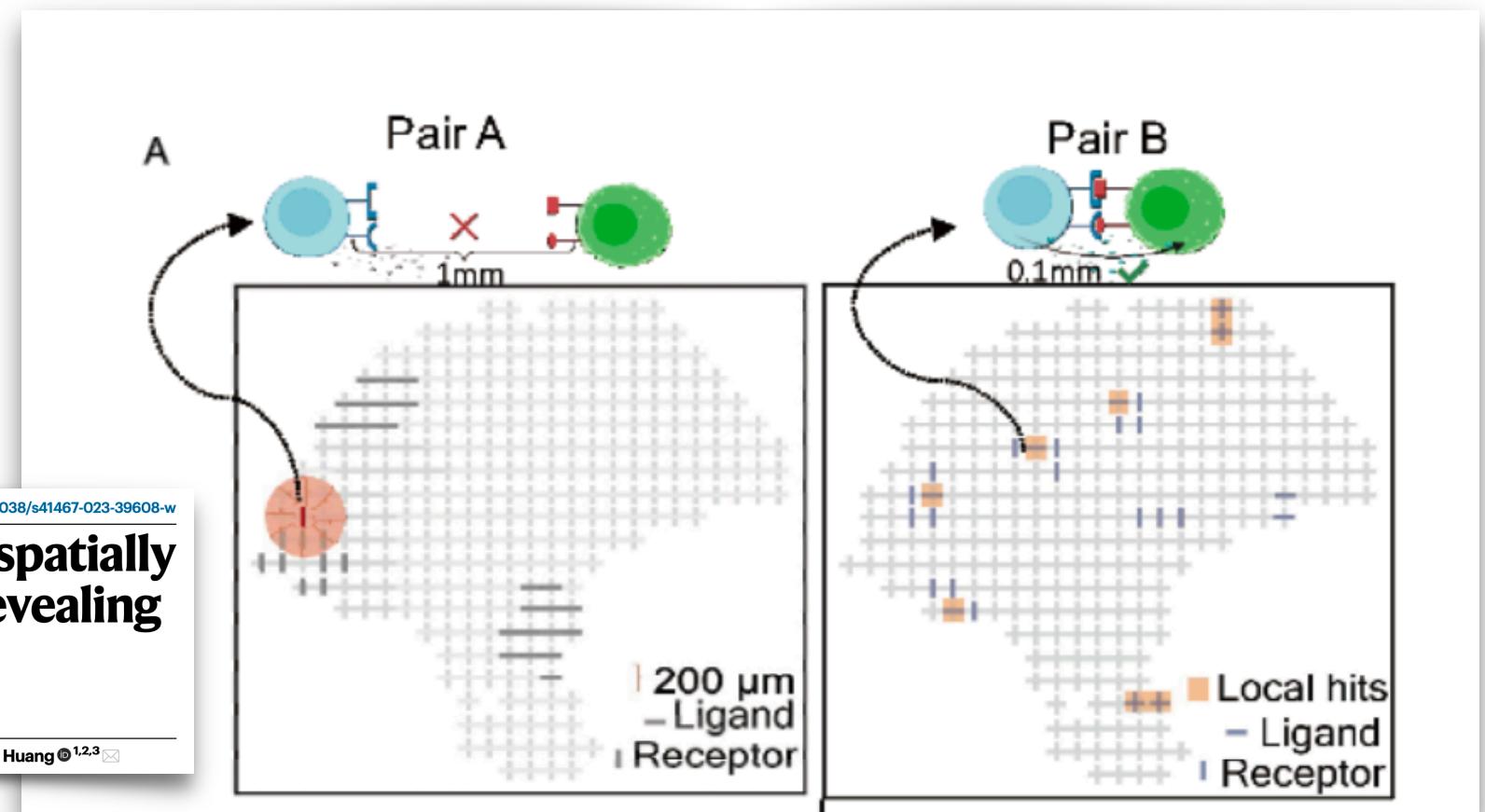
## Cell-cell communication

- SpatialDM: Global Moran's R, which is a bivariate version of Moran's I

Article <https://doi.org/10.1038/s41467-023-39608-w>

### SpatialDM for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns

Received: 28 September 2022 Zhuoxuan Li<sup>1</sup>, Tianjie Wang<sup>2</sup>, Pentao Liu<sup>1,3</sup> & Yuanhua Huang<sup>1,2,3</sup>





$$\text{Global Moran's } R = \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}},$$

## Cell-cell communication

- SpatialIDM: Global Moran's R, which is a bivariate version of Moran's I

