Spatial Transcriptomics

Lu Zhang
Assistant Professor
Division of Biostatistics
Keck School of Medicine

USC University of Southern California

What is spatial transcriptomics?

Spatial transcriptomics, or spatially resolved gene expression, is a quantitative readout of either whole transcriptome or targeted gene expression mapped to specific locations in a tissue section, and a proven powerful method to understand cellular composition and activity in the native tissue context. —10X Genomics

Editorial | Published: 06 January 2021

Method of the Year 2020: spa transcriptomics

Nature Methods **18**, 1 (2021) | Cite this article

50k Accesses | **152** Citations | **246** Altmetric | Me

Spatially resolved transcriptomics methods ar complex tissues.

nature > nature methods > focus

Focus | **10 December 2024**

Method of the Year 2024: spatial proteomics

Spatial proteomics is our pick for Method of the Year 2024, for the impact that these technologies have had on the understanding of the organization, structure and function of complex tissues, including in global tissue atlas projects.

Spatial Biology Technologies

REVIEW ARTICLE
https://doi.org/10.1038/s41592-022-01409-2

nature methods

Museum of spatial transcriptomics

Lambda Moses[1] and Lior Pachter[1,2] ✉

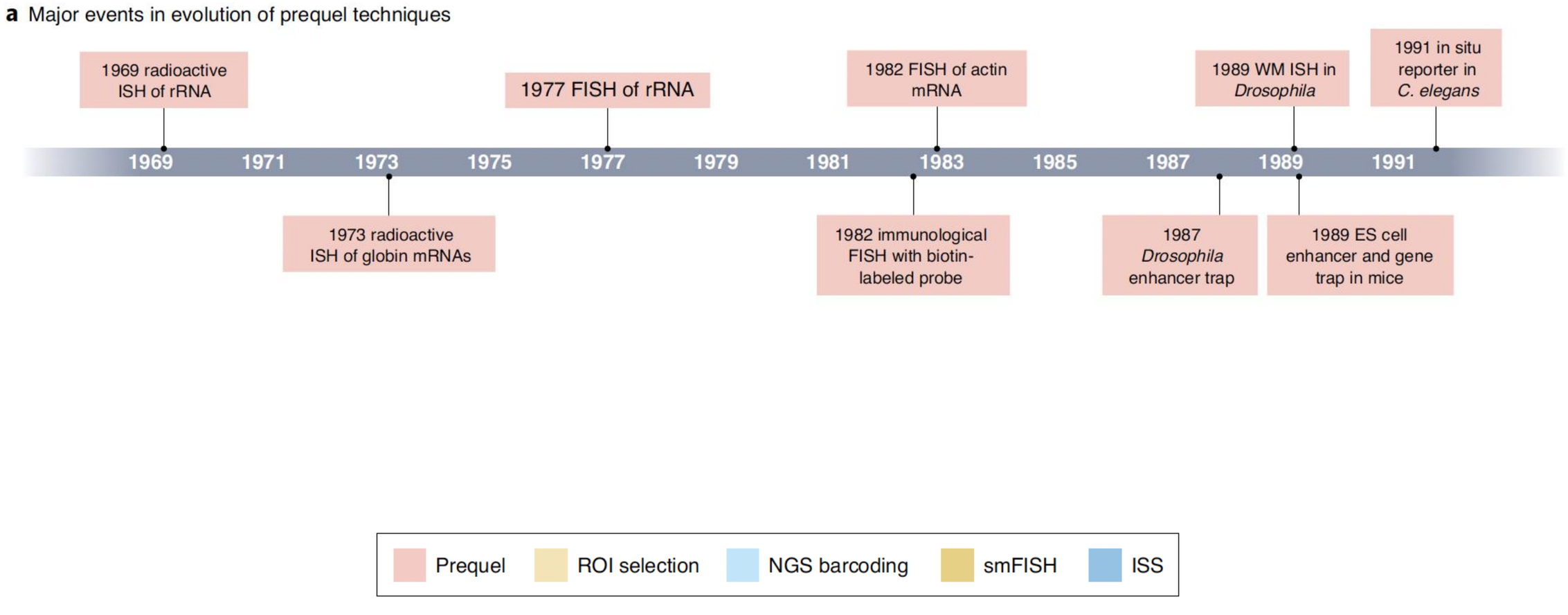
The function of many biological systems, such as embryos, liver lobules, intestinal villi, and tumors, depends on the spatial organization of their cells. In the past decade, high-throughput technologies have been developed to quantify gene expression in space, and computational methods have been developed that leverage spatial gene expression data to identify genes with spatial patterns and to delineate neighborhoods within tissues. To comprehensively document spatial gene expression technologies and data-analysis methods, we present a curated review of literature on spatial transcriptomics dating back to 1987, along with a thorough analysis of trends in the field, such as usage of experimental techniques, species, tissues studied, and computational approaches used. Our Review places current methods in a historical context, and we derive insights about the field that can guide current research strategies. A companion supplement offers a more detailed look at the technologies and methods analyzed: https://pachterlab.github.io/LP_2021/.

It has long been recognized that in biological systems ranging from the *Drosophila* embryo to the hepatic lobule, many genes need to be properly regulated in space for the system to function. To study the spatial patterns of gene expression, many different spatial transcrip-

Prequel era

By "spatial transcriptomics", we mean attempts to quantify mRNA expression of large numbers of genes within the spatial context of tissues and cells. Some important technologies enabling spatial

Spatial Biology Technologies



Spatial Biology Technologies

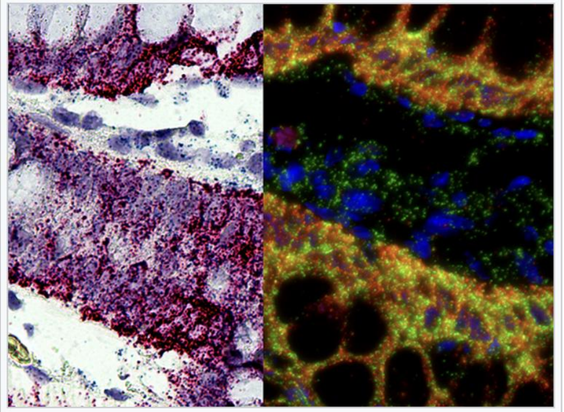
In situ hybridization

14 languages

Article Talk | Read Edit View history Tools

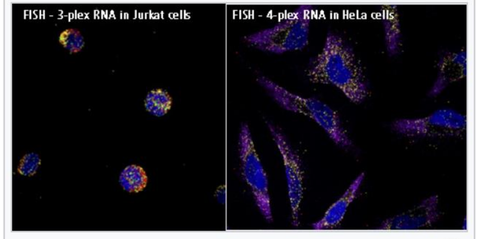
From Wikipedia, the free encyclopedia

***In situ* hybridization (ISH)** is a type of hybridization that uses a labeled complementary DNA, RNA or modified nucleic acid strand (i.e., a probe) to localize a specific DNA or RNA sequence in a portion or section of tissue (*in situ*) or if the tissue is small enough (e.g., plant seeds, *Drosophila* embryos), in the entire tissue (whole mount ISH), in cells, and in circulating tumor cells (CTCs). This is distinct from immunohistochemistry, which usually localizes proteins in tissue sections.



RNA *in situ* hybridization - KRT5 and housekeeping gene in human melanoma FFPE tissue section - visualized under brightfield and fluorescence microscope

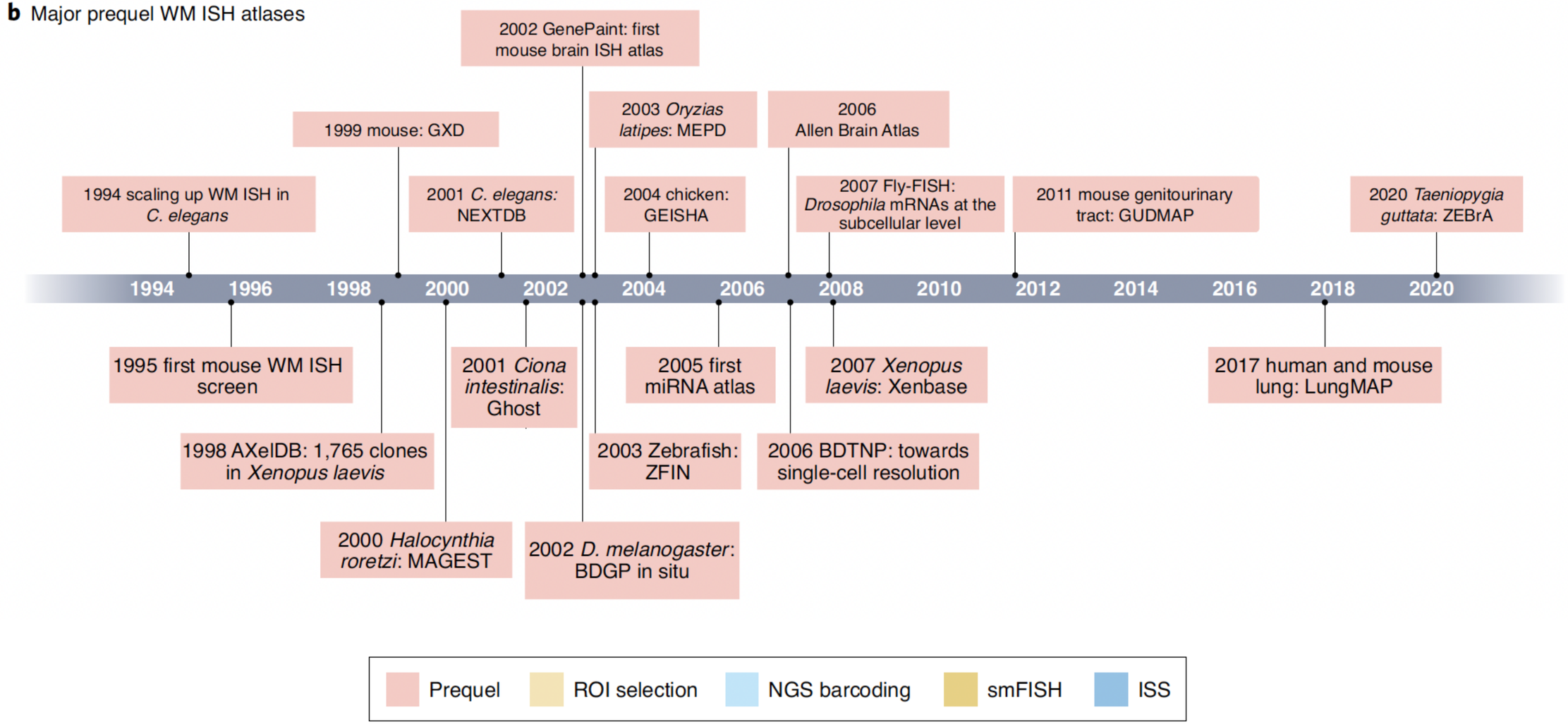
In situ hybridization is used to reveal the location of specific nucleic acid sequences on chromosomes or in tissues, a crucial step for understanding the organization, regulation, and function of genes. The key techniques currently in use include *in situ* hybridization to mRNA with oligonucleotide and RNA probes (both radio-labeled and hapten-labeled), analysis with light and electron microscopes, whole mount *in situ* hybridization, double detection of RNAs and RNA plus protein, and fluorescent *in situ* hybridization to detect chromosomal sequences. DNA ISH can be used to determine the structure of chromosomes. Fluorescent DNA ISH (FISH) can, for example, be used in medical diagnostics to assess chromosomal integrity. RNA ISH (RNA *in situ* hybridization) is used to measure and localize RNAs (mRNAs, lncRNAs, and miRNAs) within tissue sections, cells, whole mounts, and circulating tumor cells (CTCs). *In situ* hybridization was invented by American biologists Mary-Lou Pardue and Joseph G. Gall.[1][2][3]



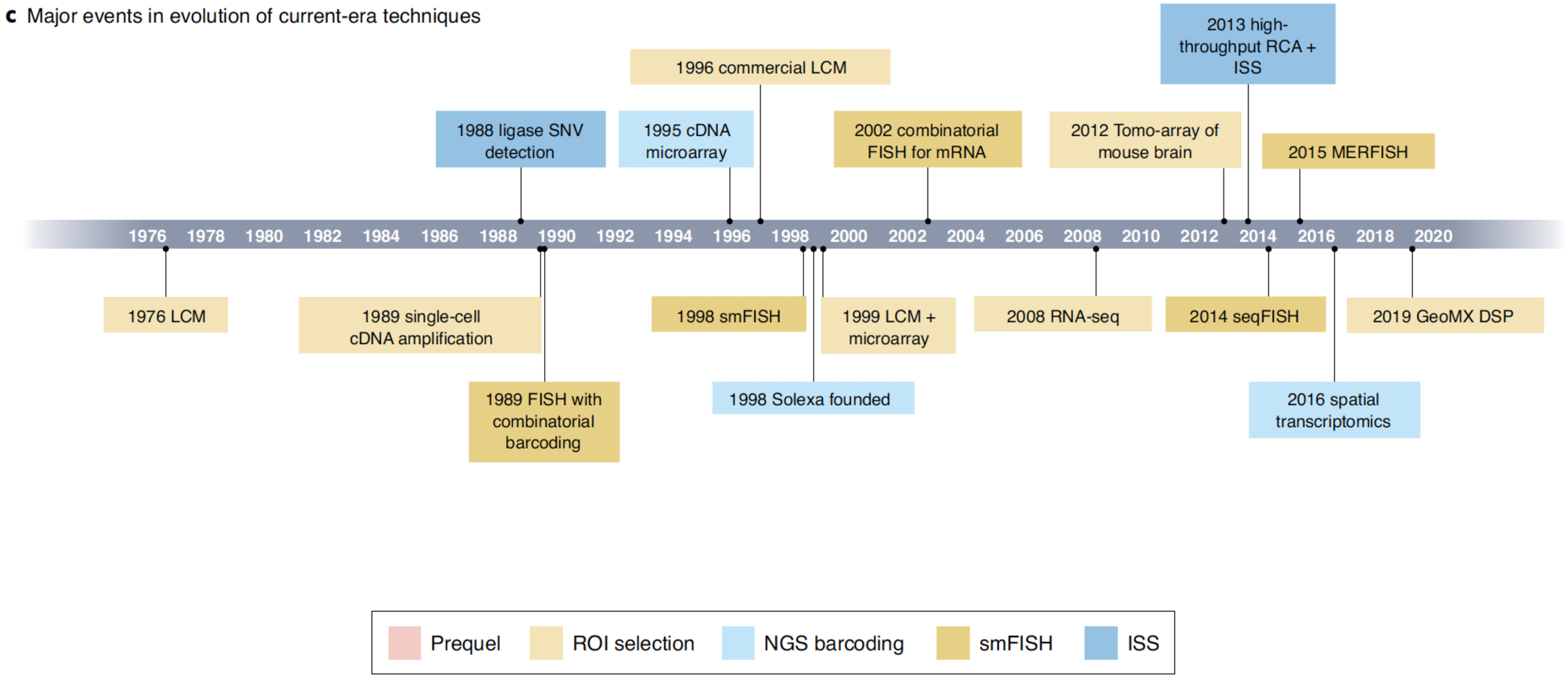
Multiplex RNA visualization in cells using ViewRNA FISH Assays

Challenges of in-situ hybridization [edit]

Spatial Biology Technologies



Spatial Biology Technologies



Spatial Biology Technologies

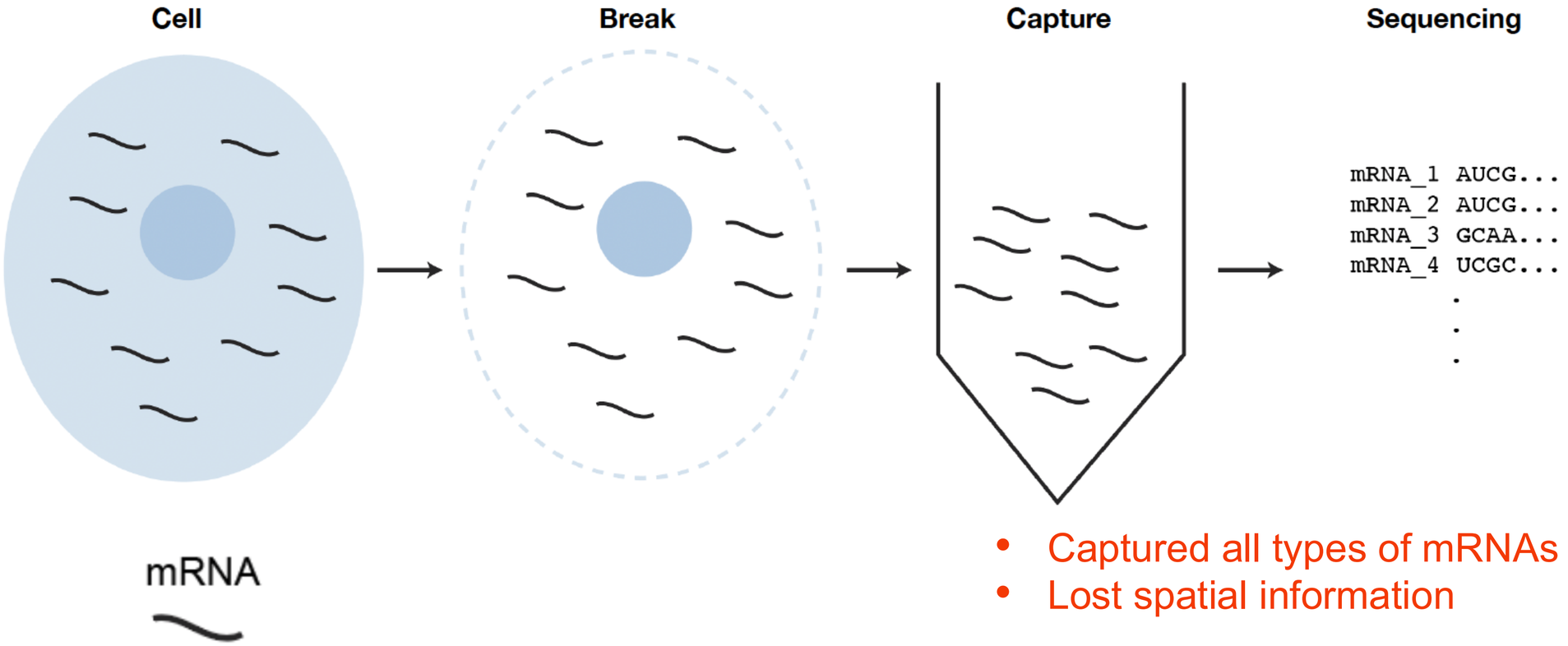
Method	First published	Category	Max # genes	Min spot diameter (μm)
voxelation	2002-01-31	ROI selection	Tx wide	NA
PA-GFP	2010-11-11	ROI selection	Tx wide	NA
SRM seqFISH	2012-06-02	smFISH	32	single cell
Tomo-array	2012-09-18	ROI selection	Tx wide	NA
iceFISH	2013-02-16	smFISH	20	single cell
ISS	2013-07-14	ISS	256	single cell

203 entries!

Slide from Kelly Street

How can mRNA be *seen*?

Method 1: Sequencing



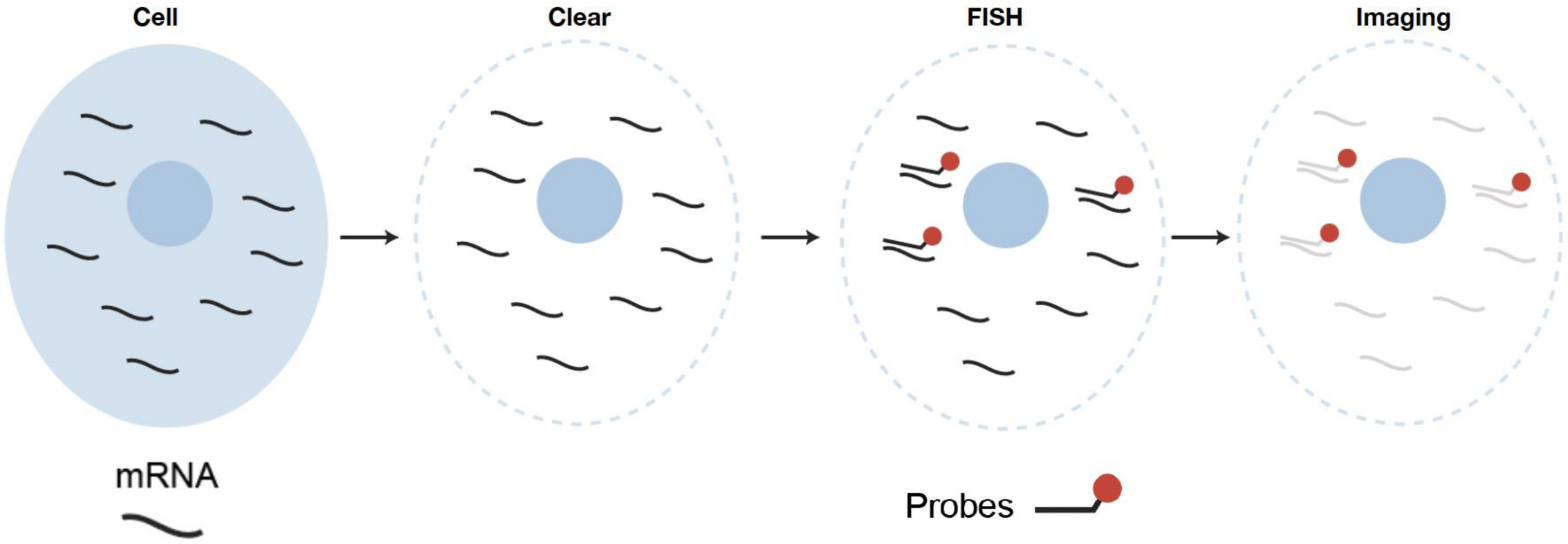
- Captured all types of mRNAs (all genes)
- Lost spatial information

Slide From Fangming Xie

How can mRNA be *seen*?

Method 2: Staining and Imaging

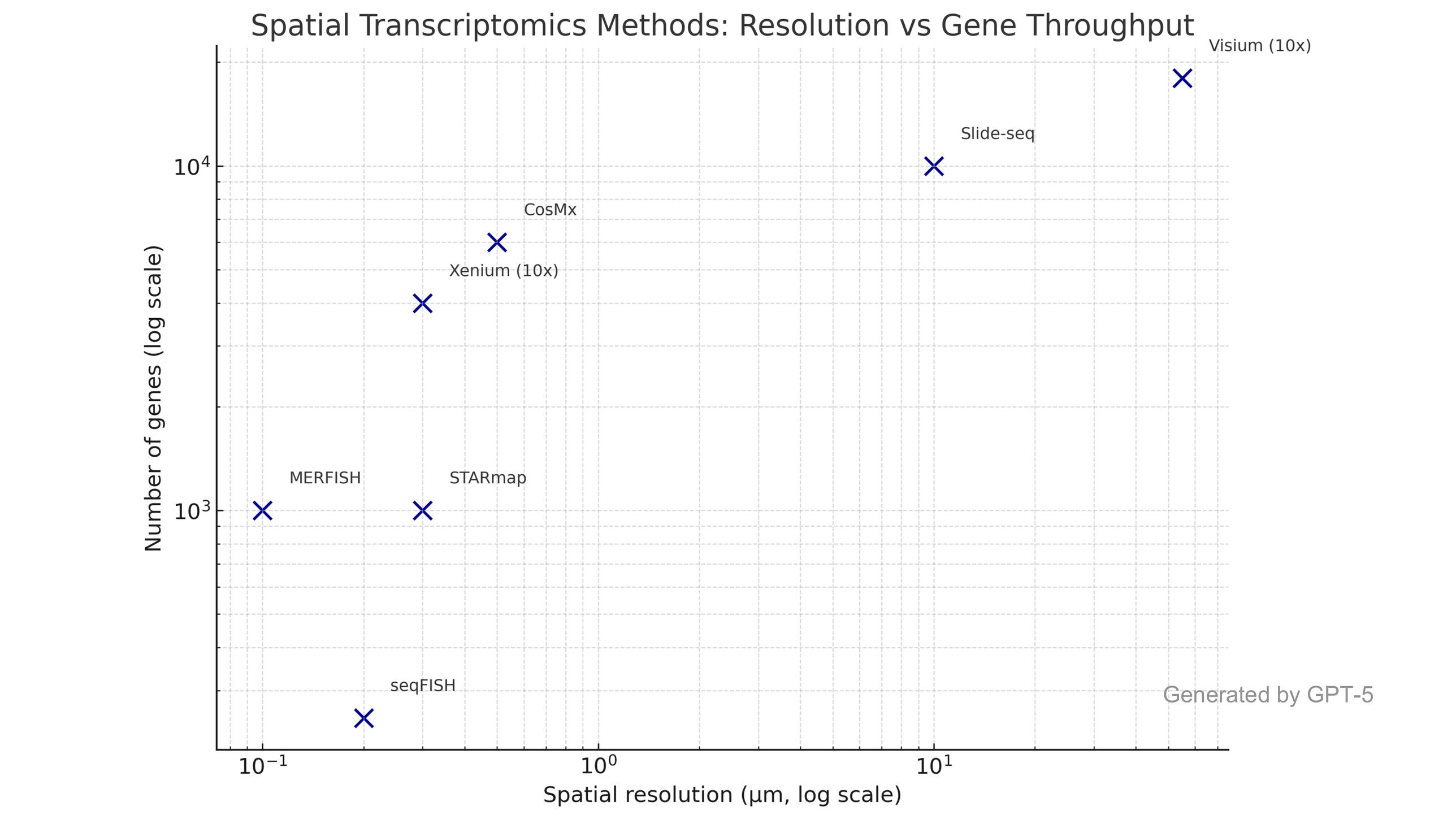
- Captured one types of mRNA (1 gene)
- Preserved spatial information



Slide From Fangming Xie

Two major branches of technologies: sequencing-based vs imaging-based assays

- **Imaging-based:**
  - seqFISH (sequential Fluorescence In Situ Hybridization)
  - MERFISH (Multiplexed Error-Robust FISH)
  - STARmap (Spatially Resolved Transcript Amplicon Readout Mapping)
  - CosMx (NanoString CosMx Spatial Molecular Imager)
  - **10x Genomics Xenium** → *despite being from 10x, this is actually **imaging-based**, not sequencing-based. It uses in situ hybridization and imaging (similar to CosMx)*
  - …
- **Sequencing-based:**
  - Slide-seq → *uses barcoded beads on a surface to capture RNAs*
  - **10X Visium**→ *Sequencing-based, array of barcoded capture spots on slides; whole-transcriptome coverage at spot resolution.*
  - …


Spatial Transcriptomics Methods: Resolution vs Gene Throughput
Visium (10x)
Slide-seq
CosMx
Xenium (10x)
MERFISH
STARmap
seqFISH
Number of genes (log scale)
Spatial resolution (µm, log scale)
10^4
10^3
10^{-1}
10^0
10^1
Generated by GPT-5

Spatial Transcriptomics Methods Summary

Method	Resolution (µm)	Genes (typical)	Approx. Cellular Resolution
MERFISH	0.1	~1,000	Subcellular / single molecule
seqFISH	0.2	~250	Subcellular / single molecule
STARmap	0.3	~1,000	Subcellular / single cell
Xenium (10x)	0.3	~4,000	Subcellular / single cell
CosMx	0.5	~6,000	Single cell / subcellular
Slide-seq	10	~10,000	Near single-cell (~1 cell/spot)
Visium (10x)	55	~18,000	~1–10 cells/spot (often dozens)

Summarized by GPT-5

Slide-seq

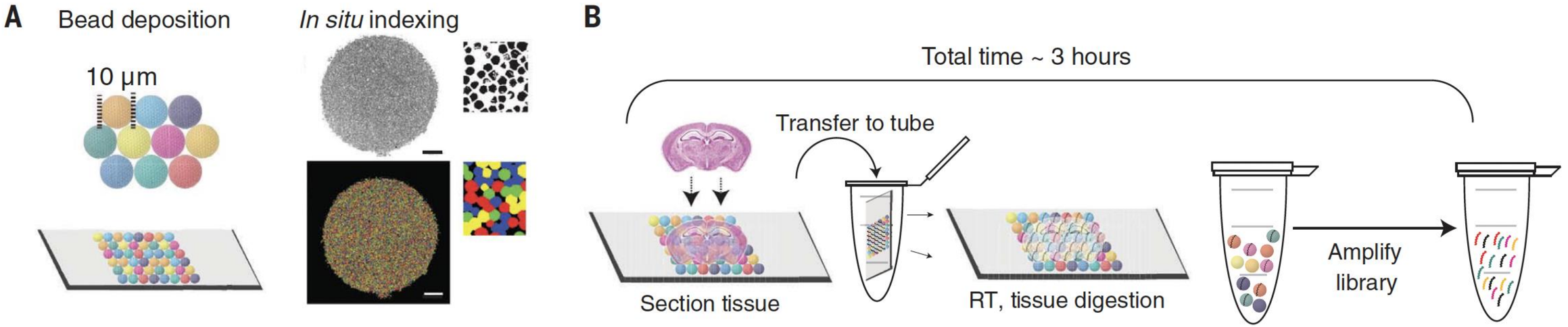
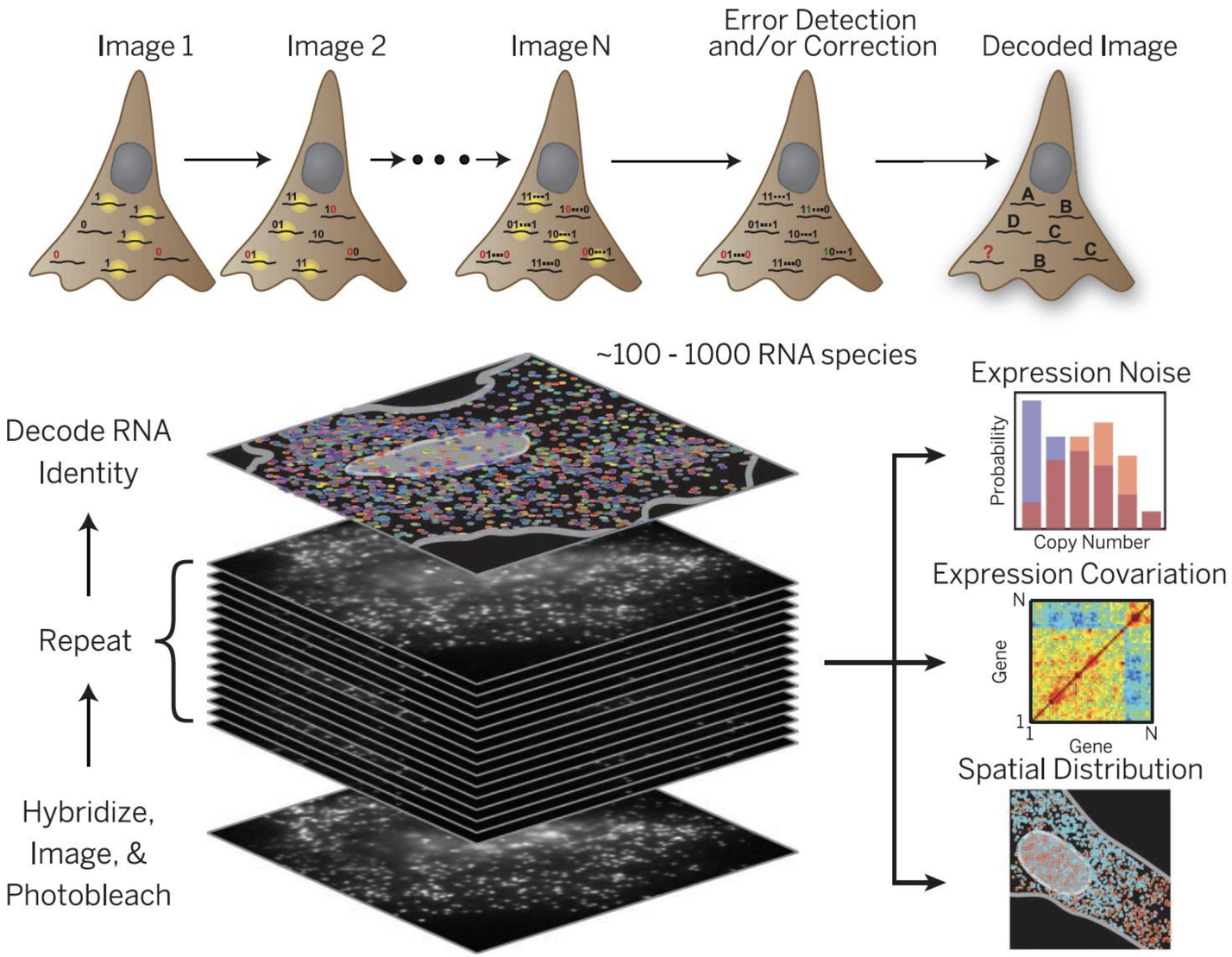


Fig. 1. High-resolution RNA capture from tissue by Slide-seq. (**A**) (Left) Schematic of array generation. A monolayer of randomly deposited, DNA barcoded beads (a "puck") is spatially indexed by SOLiD sequencing. (Top right) Representative puck with sequenced barcodes shown in black. (Bottom right) Composite image of the same puck colored by the base calls for a single base of SOLiD sequencing. (**B**) Schematic of the sample preparation procedure. RT, reverse transcription. (**C**) (Top left) t-distributed

Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution (Rodriques et al. 2019 Science)

MERFISH



Spatially resolved, highly multiplexed RNA profiling in single cells (Chen et al. 2015 Science)

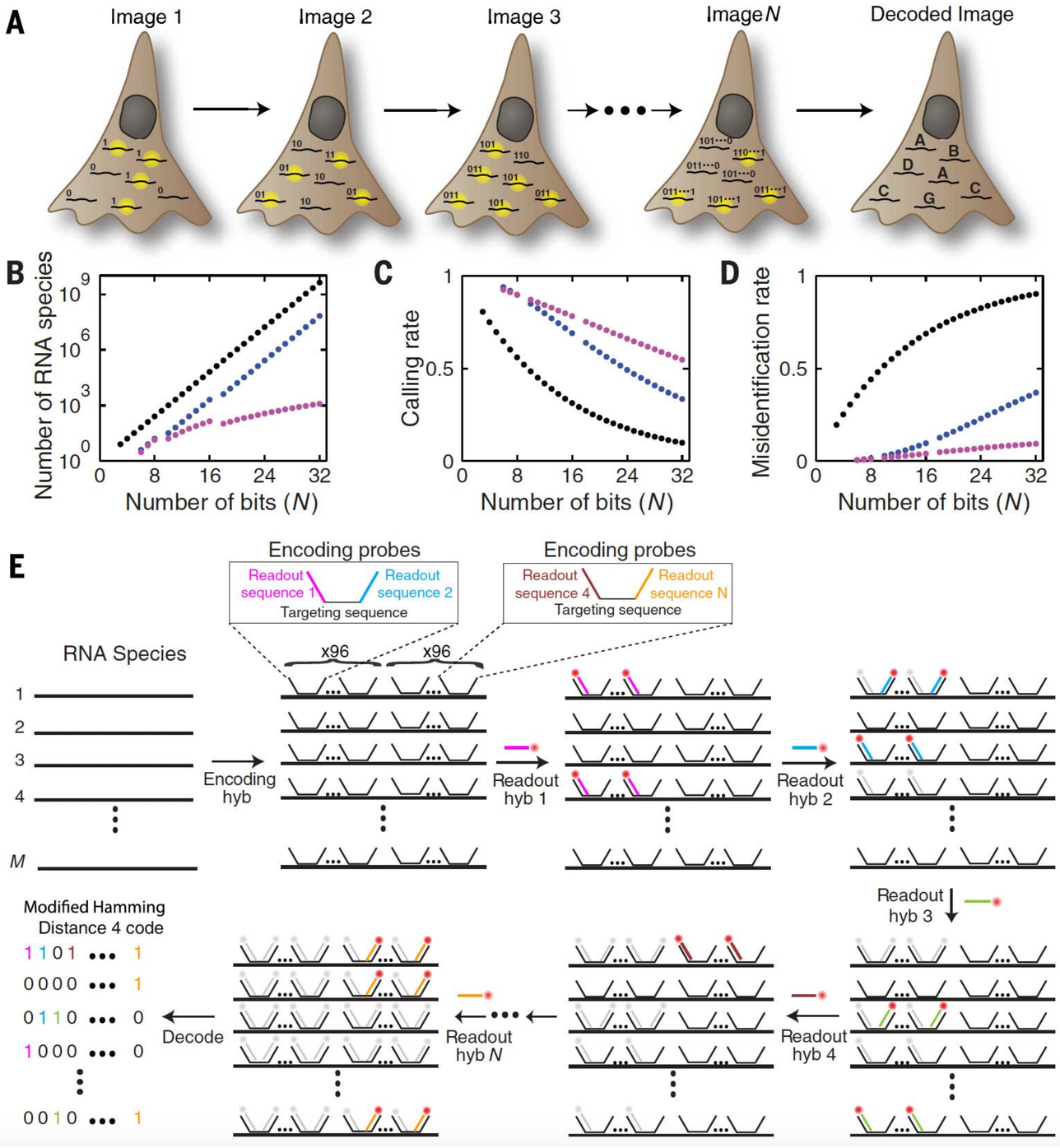


Fig. 1. MERFISH: A highly multiplexed smFISH approach enabled by combinatorial labeling and error-robust encoding. (**A**) Schematic depiction of the identification of multiple RNA species in *N* rounds of imaging. Each RNA species is encoded with a *N*-bit binary word, and during each round of imaging, only the subset of RNAs that should read 1 in the corresponding bit emit signal. (**B** to **D**) The number of addressable RNA species (B); the rate at which these RNAs are properly identified—the "calling rate" (C); and the rate at which RNAs are incorrectly identified as a different RNA species—the "misidentification rate" (D); plotted as a function of the number of bits (*N*) in the binary words encoding RNA. Black indicates a simple binary code that includes all 2^N-1 possible binary words. Blue indicates the HD4 code in which the Hamming distance separating words is 4. Purple indicates a modified HD4 (MHD4) code where the number of 1 bits are kept at four. The calling and misidentification rates are calculated with per-bit error rates of 10% for the 1→0 error and 4% for the 0→1 error. (**E**) Schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a specific combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA-targeting region flanked by two readout sequences, drawn from a pool of *N* different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a particular combination of four of the *N* readout sequences, which correspond to the four hybridization rounds in which this RNA should read 1. *N* subsequent rounds of hybridization with the fluorescent readout probes are used to probe the readout sequences (hyb 1, hyb 2, ..., hyb *N*). The bound probes are inactivated by photobleaching between successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

10x Genomics and 10x in Spatial Transcriptomics

- Founded in **2012**, headquartered in Pleasanton, California.
- A leading biotechnology company specializing in single-cell and spatial technologies.
- Known for creating the Chromium platform (single-cell RNA-seq, ATAC-seq, immune profiling).
- Leveraged its success in single-cell genomics to expand into spatial biology.

- Visium (launched ~2019): sequencing-based platform for spatial transcriptomics at mesoscale resolution (~55 μm spots). *10x Genomics Visium Page*
- How it works | Visium Spatial Gene Expression Solution (2mins video)

10x Genomics and 10x in Spatial Transcriptomics

- Founded in **2012**, headquartered in Pleasanton, California.
- A leading biotechnology company specializing in single-cell and spatial technologies.
- Known for creating the Chromium platform (single-cell RNA-seq, ATAC-seq, immune profiling).
- Leveraged its success in single-cell genomics to expand into spatial biology.

- Xenium (released ~2022): imaging-based platform for in situ detection of RNA at subcellular resolution with targeted panels. *10x Genomics Xenium Page*

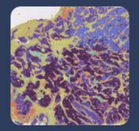
Datasets

Explore and download datasets created by 10x Genomics.

Chromium Single Cell - Featured

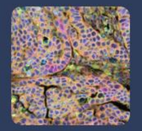
320k scFFPE From 8 Human Tissues 320k, 16-Plex

Visium Spatial - Featured



Visium HD 3' Gene Expression Library, Ovarian Cancer (Fresh Frozen)

Xenium In Situ - Featured



Xenium In Situ Gene and Protein Expression data for FFPE Human Renal Cell Carcinoma

Search datasets

Top searches PBMC Xenium HD GEM-X Flex Cell Segmentation Breast Cancer Mouse Brain Brain Lung FFPE

Filter datasets

10x Genomics product

- Platform
- Product
- Chemistry version
- Additional application
- Software

Datasets (Showing 755 datasets)	**Product**	**Species**	**Sample type**	**Cells or nuclei**	**Preservation**
Visium HD 3' Gene Expression Library, Mouse Brain (Fresh Frozen)	HD 3' Spatial Gene Expression v1.0	Mouse	Brain	N/A	Fresh Frozen
Visium HD Spatial Gene Expression Library, Human Pancreas (FFPE)	HD Spatial Gene Expression v1.0	Human	Pancreas	N/A	FFPE
Visium HD Spatial Gene Expression Library, Human Breast Cancer (Fresh Frozen), Ultima	HD Spatial Gene Expression v1.0	Human	Breast	N/A	Fresh Frozen

Part 2 Spatial Transcriptomics Analysis

Orchestrating Spatial Transcriptomics Analysis with Bioconductor

Published: August 6, 2025

https://lmweber.org/OSTA/

Welcome

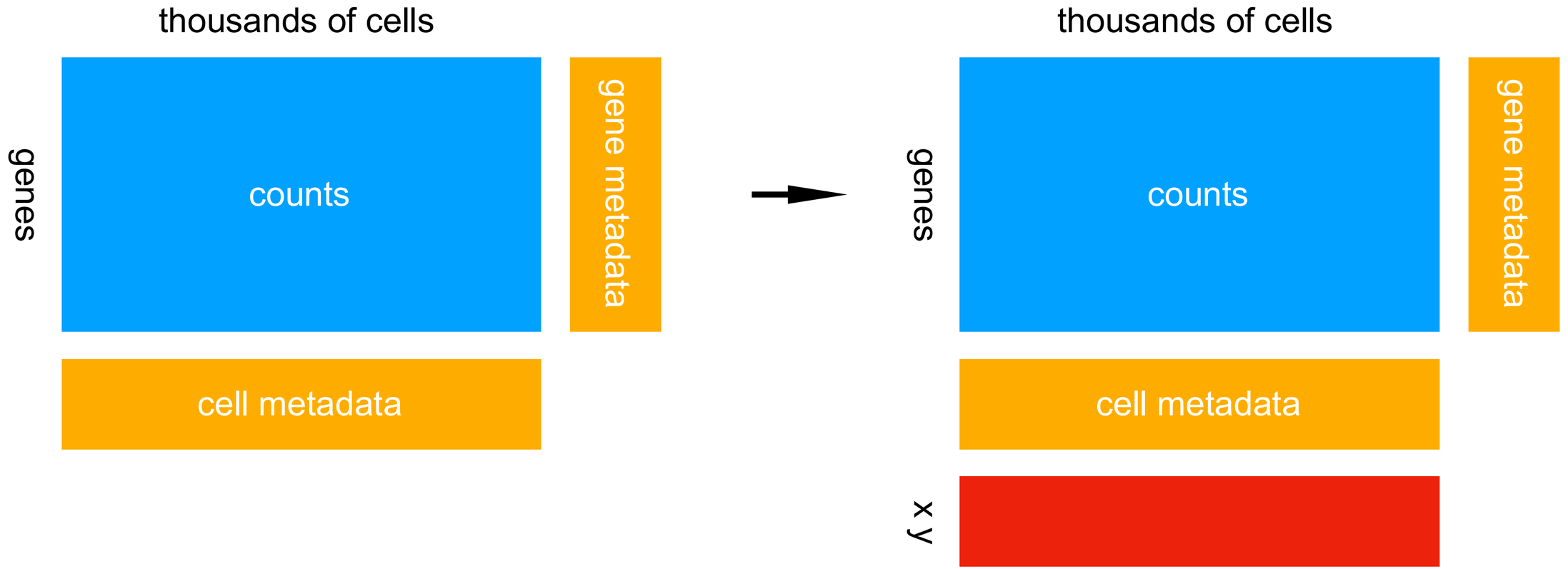
This is the website for the online book **Orchestrating Spatial Transcriptomics Analysis with Bioconductor**.

This book provides reproducible examples and discussion on computational analysis workflows for spatial omics data using Bioconductor in R. The book contains chapters describing individual analysis steps as well as extended workflows, each with examples including R code and datasets. In some examples, R code is also integrated with Python tools.

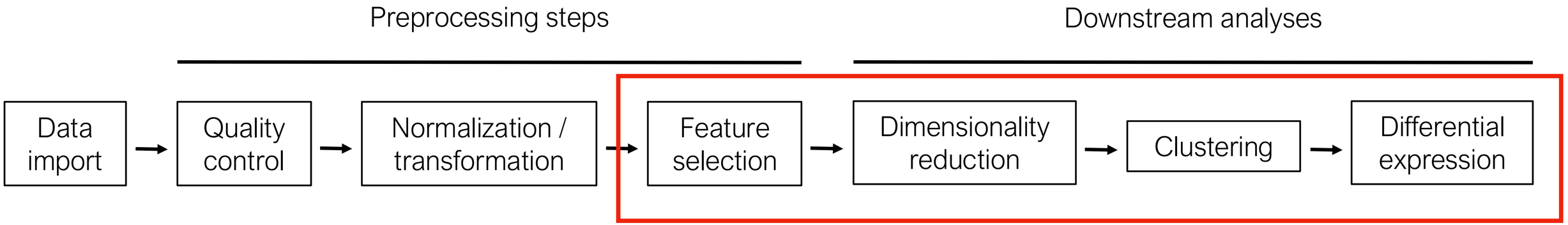
The book is organized into several parts, consisting of introductory materials, and analysis steps and workflows for the two main streams of spatial omics data: sequencing-based and imaging-based.

Additional materials on analysis workflows for single-cell (non-spatial) data, as well as further introductory materials on R and Bioconductor, can be found in the related book Orchestrating Single-Cell Analysis with Bioconductor (OSCA).

Data Structure



Typical Spatial transcriptomics analysis workflows



Seurat tutorial: https://satijalab.org/seurat/articles/seurat5_spatial_vignette_2

Feature selection: Highly variable genes (HVGs)

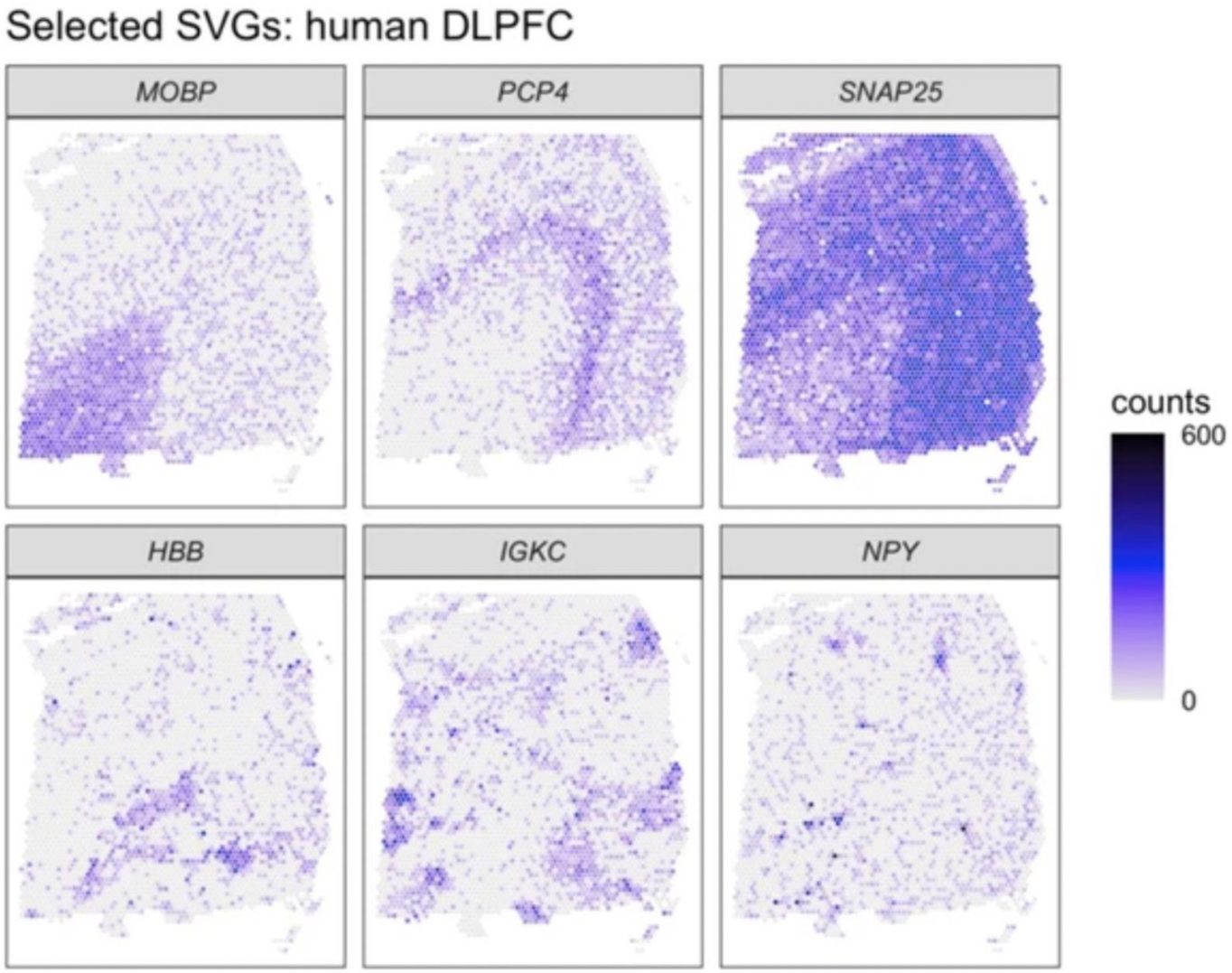
```
FindVariableFeatures()
```

"We next calculate a subset of features that exhibit high cell-to-cell variation in the dataset (i.e, they are highly expressed in some cells, and lowly expressed in others). We and others have found that focusing on these genes in downstream analysis helps to highlight biological signal in single-cell datasets."

It also just makes the computation a lot easier
Cuts down on multiple testing

Adapted from Kelly Street's slide

Feature selection: Spatially variable genes (SVGs)



{nnSVG}

- Rank genes by (spatial) variability
- Select the top N genes

\+ takes advantage of spatial information

\- could struggle with very disperse cell types

Slide from Kelly Street

Spatially Variable Genes

Selected SVGs: human DLPFC

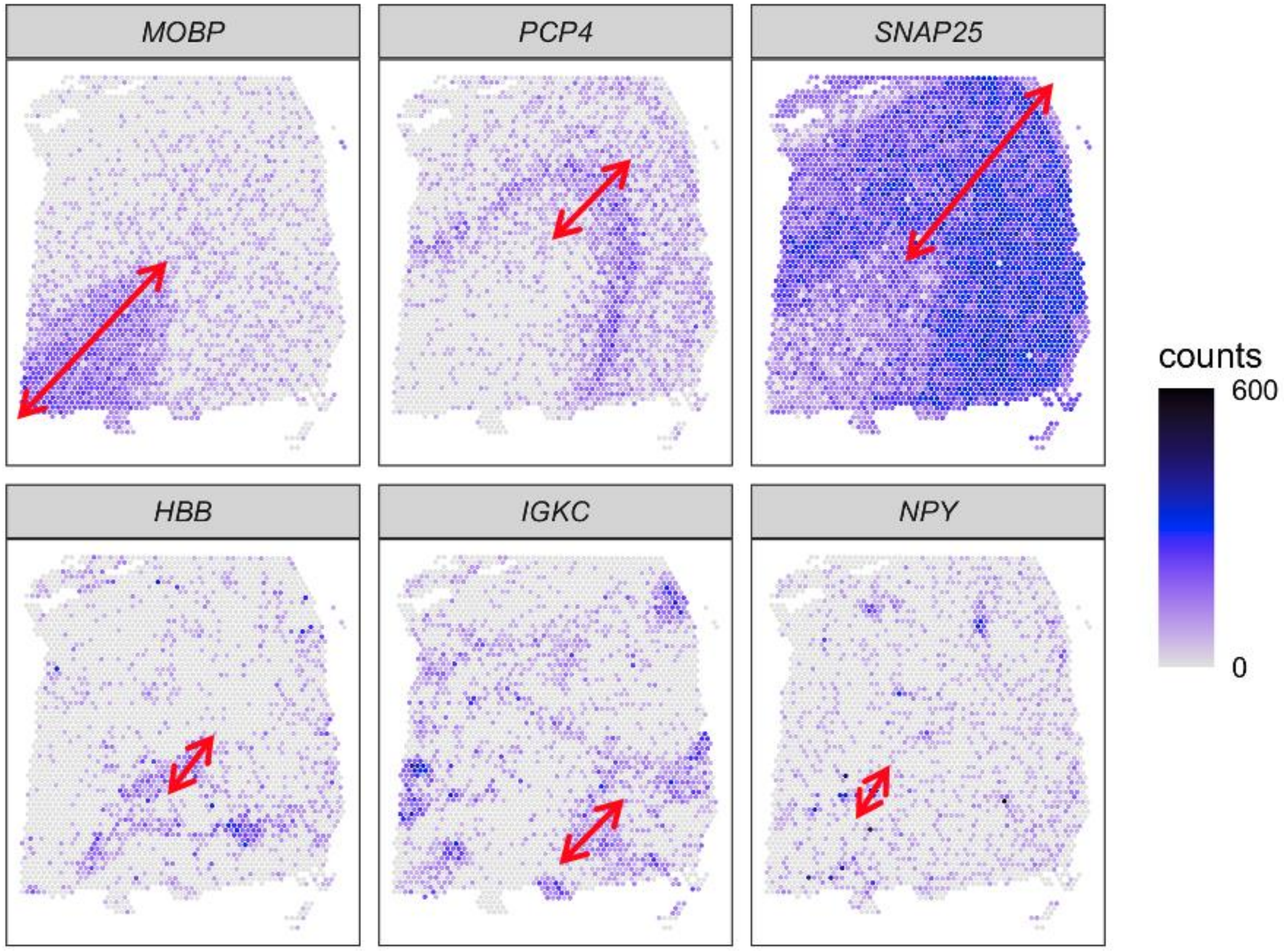


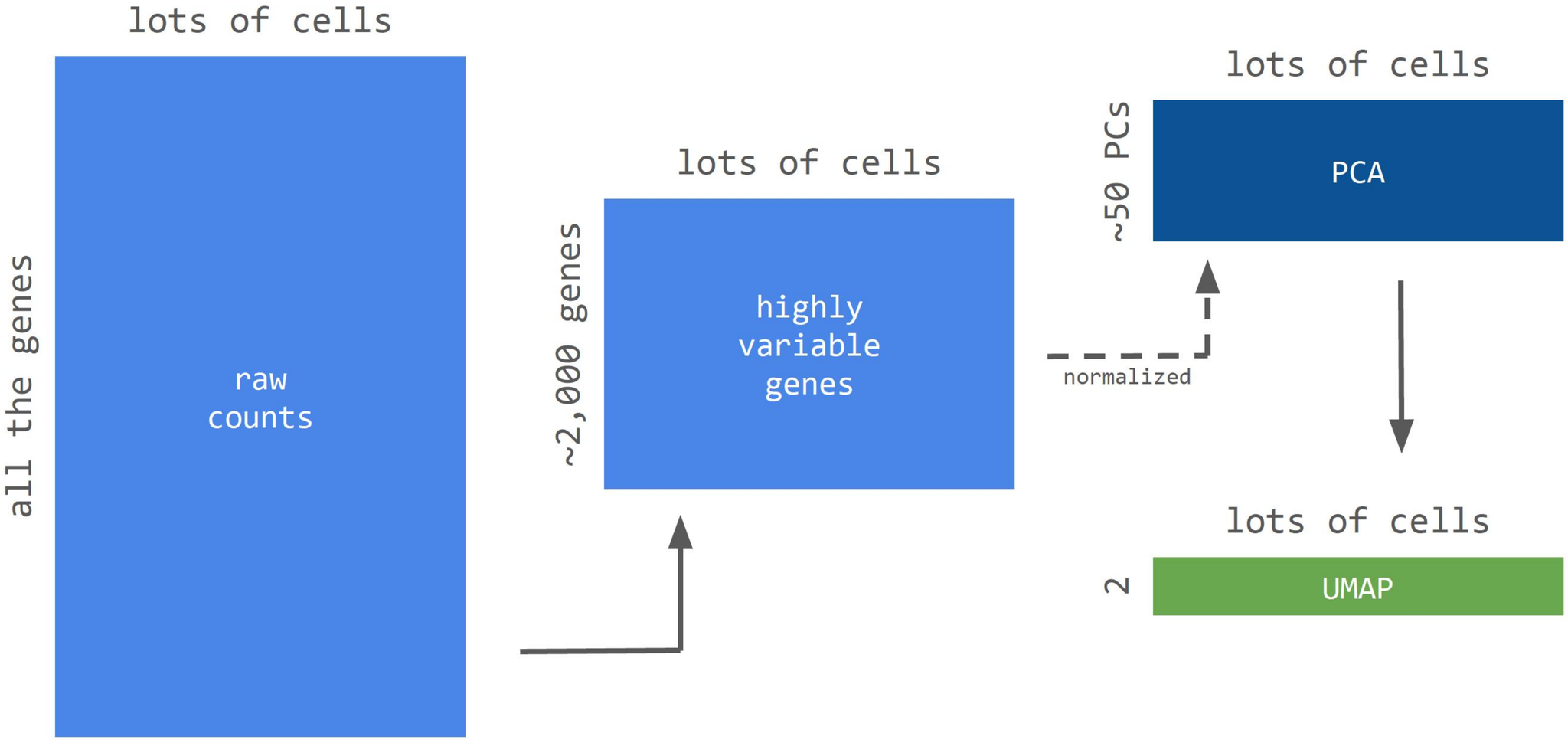
Table 1 | Summary of characteristics of methods included in the performance evaluations and runtime comparisons

Method	Spatial information	Flexible length scale parameters	Covariates for spatial domains	Runtime
nnSVG	●	●	●	◐
SPARK-X	●	○	●	●
HVGs	○	○	○	●
Moran's I	●	○	○	◐
SpatialDE	●	●	●	○
SPARK	●	●	●	○

For each method, the columns indicate whether (●) or not (○) the method: (i) takes spatial information into account, (ii) fits models with flexible gene-specific length scale parameters, (iii) provides an option to include covariates for spatial domains in the models, and (iv) provides fast runtimes. Half-filled circles (◐) indicate intermediate scores. The scalable methods are shown in the first 4 rows, and the earlier cubically scaling methods (SpatialDE and SPARK) are shown in the last 2 rows.

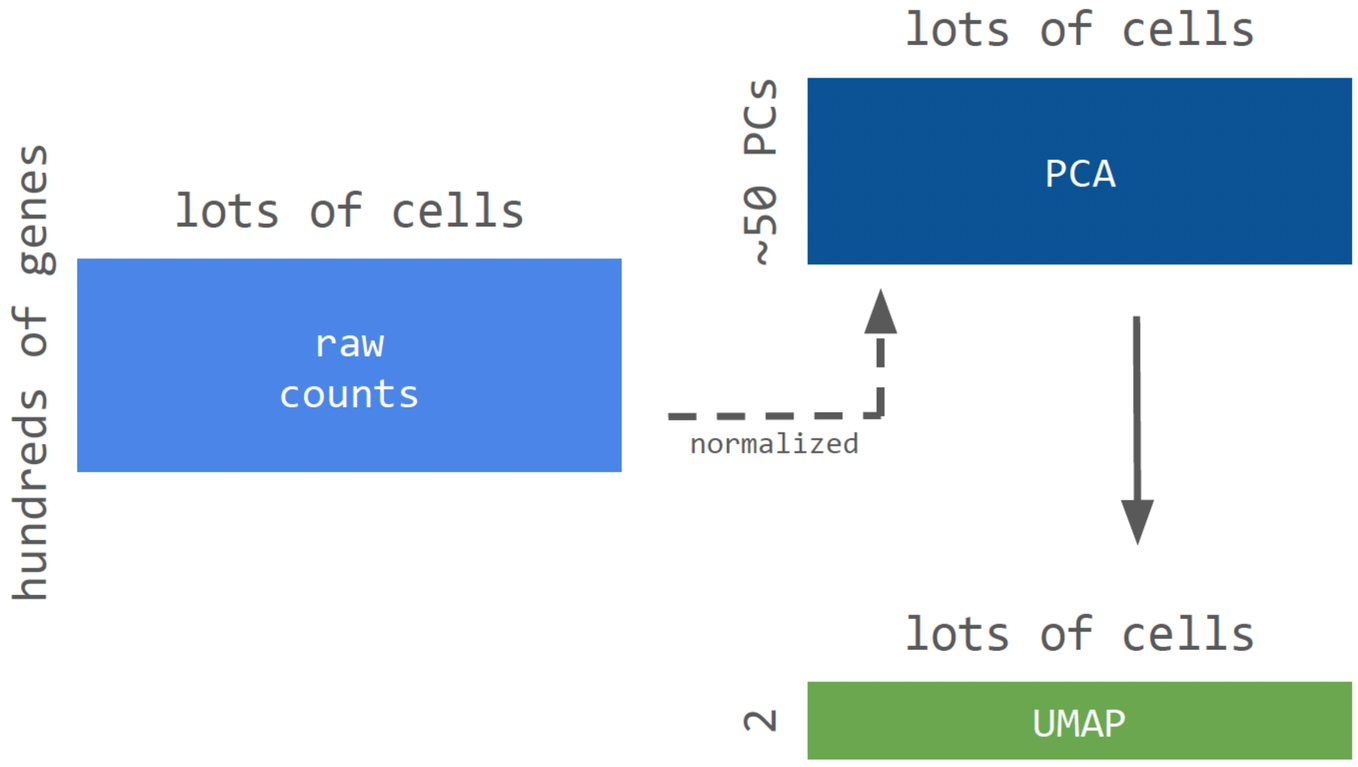
nnSVG: Lukas, etc., 2023, Nat Comm

Dimensionality Reduction



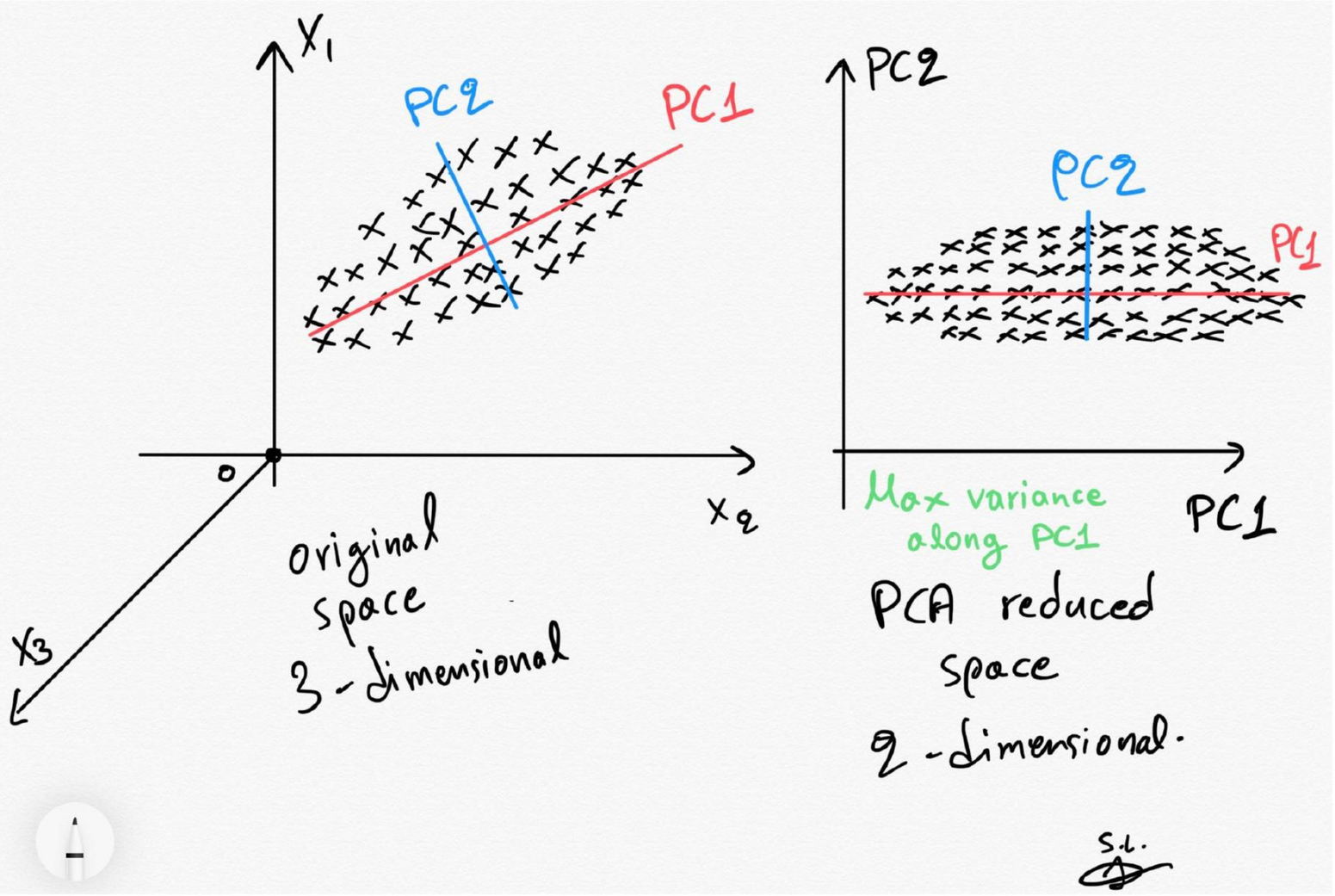
Slide from Kelly Street

Dimensionality Reduction



Slide from Kelly Street

PCA



Distills the signal(s) in the data

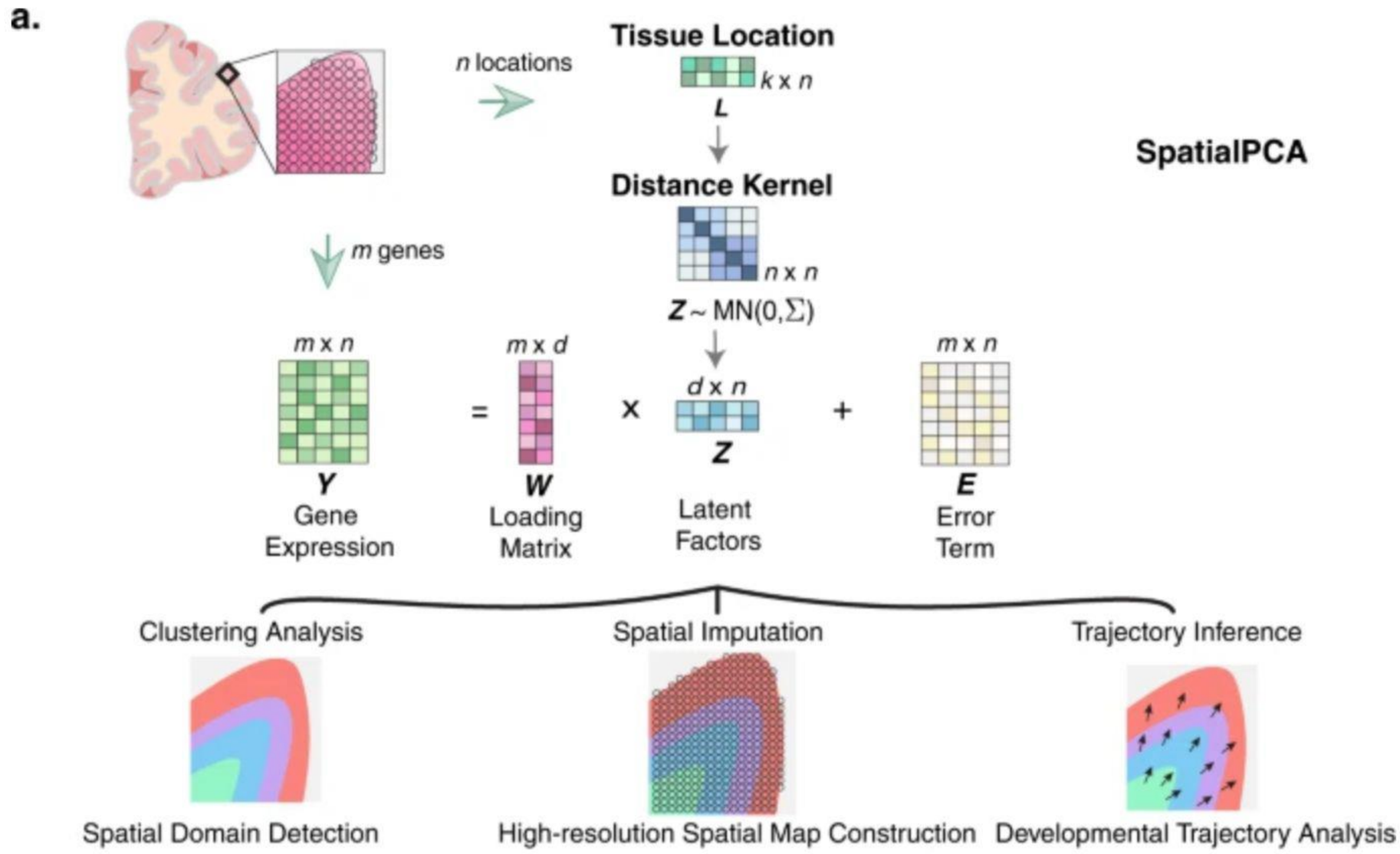
Can identify multiple axes of variation

Further reduces computational complexity

Slide from Kelly Street

Image: Serafeim Loukas
https://towardsdatascience.com/pca-clearly-explained-how-when-why-to-use-it-and-feature-importance-a-guide-in-python-7c274582c37e/

Fig. 1: Method schematic of SpatialPCA and simulation results.



Shang, L., Zhou, X. Spatially aware dimension reduction for spatial transcriptomics. Nature Communications, 2022.

Projected Spatial Factor Model

Statistics > Methodology

[Submitted on 1 Jun 2025]

ProjMC^2: Scalable and Stable Posterior Inference for Bayesian Spatial Factor Models with Application to Spatial Transcriptomics

Lu Zhang

Factor models exhibit a fundamental tradeoff among flexibility, identifiability, and computational efficiency. Bayesian spatial factor models, in particular, face pronounced identifiability concerns and scaling difficulties. To mitigate these issues and enhance posterior inference reliability, this work proposes Projected Markov Chain Monte Carlo (ProjMC^2), a novel Markov Chain Monte Carlo (MCMC) sampling algorithm employing projection techniques and conditional conjugacy. ProjMC^2 is showcased within the context of spatial factor analysis, significantly improving posterior stability and MCMC mixing efficiency by projecting posterior sampling of latent factors onto a subspace of a scaled Stiefel manifold. Theoretical results establish convergence to the stationary distribution irrespective of initial values. Integrating this approach with scalable univariate spatial modeling strategies yields a stable, efficient, and flexible modeling and sampling methodology for large-scale spatial factor models. Simulation studies demonstrate the effectiveness and practical advantages of the proposed methods. The practical utility of the methodology is further illustrated through an analysis of spatial transcriptomic data obtained from human kidney tissues, showcasing its potential for enhancing the interpretability and robustness of spatial transcriptomics analyses.

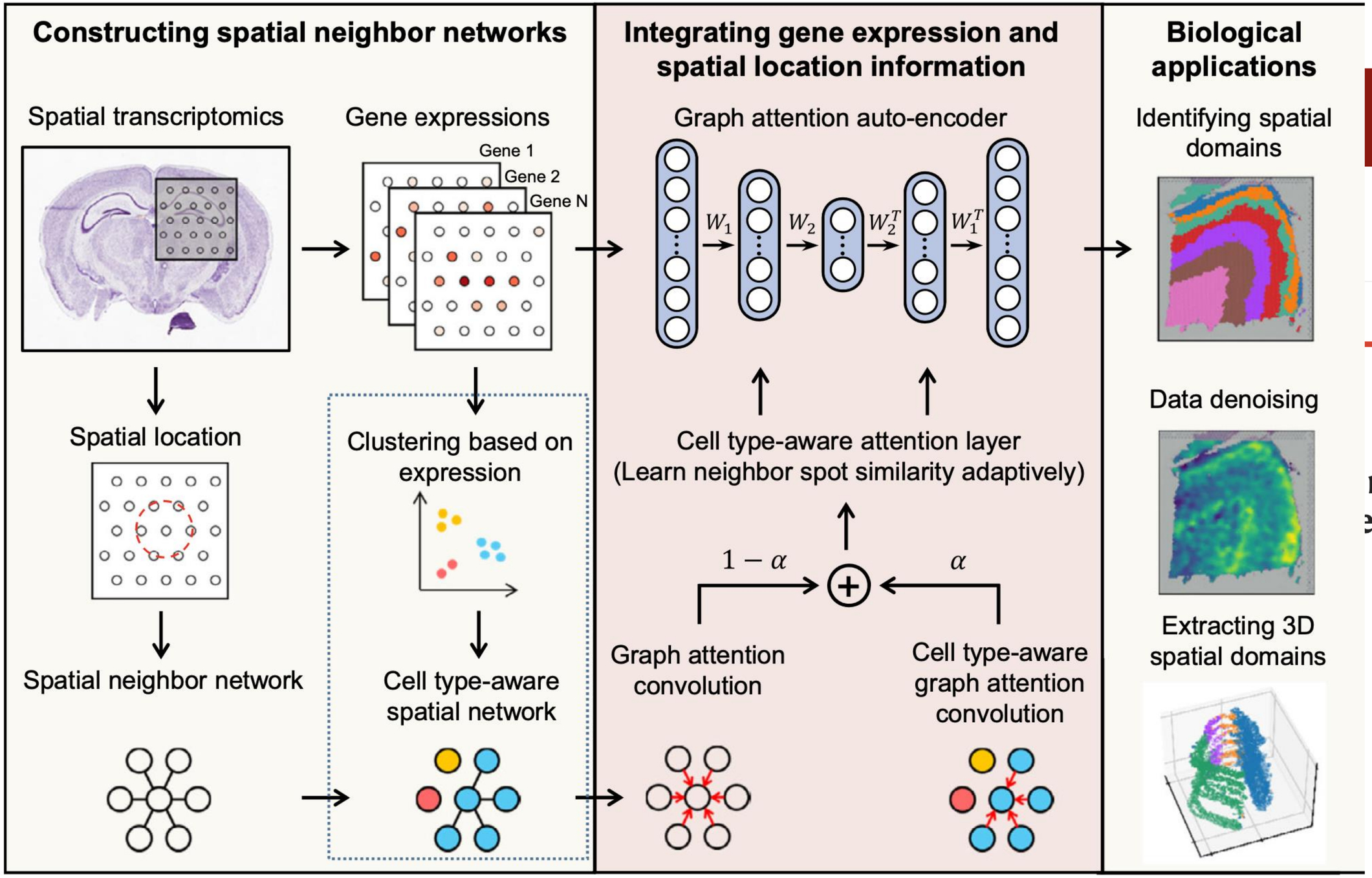
Comments: 32 pages, 5 figures
Subjects: **Methodology (stat.ME)**
Cite as: arXiv:2506.01098 **[stat.ME]**
(or arXiv:2506.01098v1 **[stat.ME]** for this version)
https://doi.org/10.48550/arXiv.2506.01098

Submission history

From: Lu Zhang [view email]
[v1] Sun, 1 Jun 2025 17:46:03 UTC (27,039 KB)

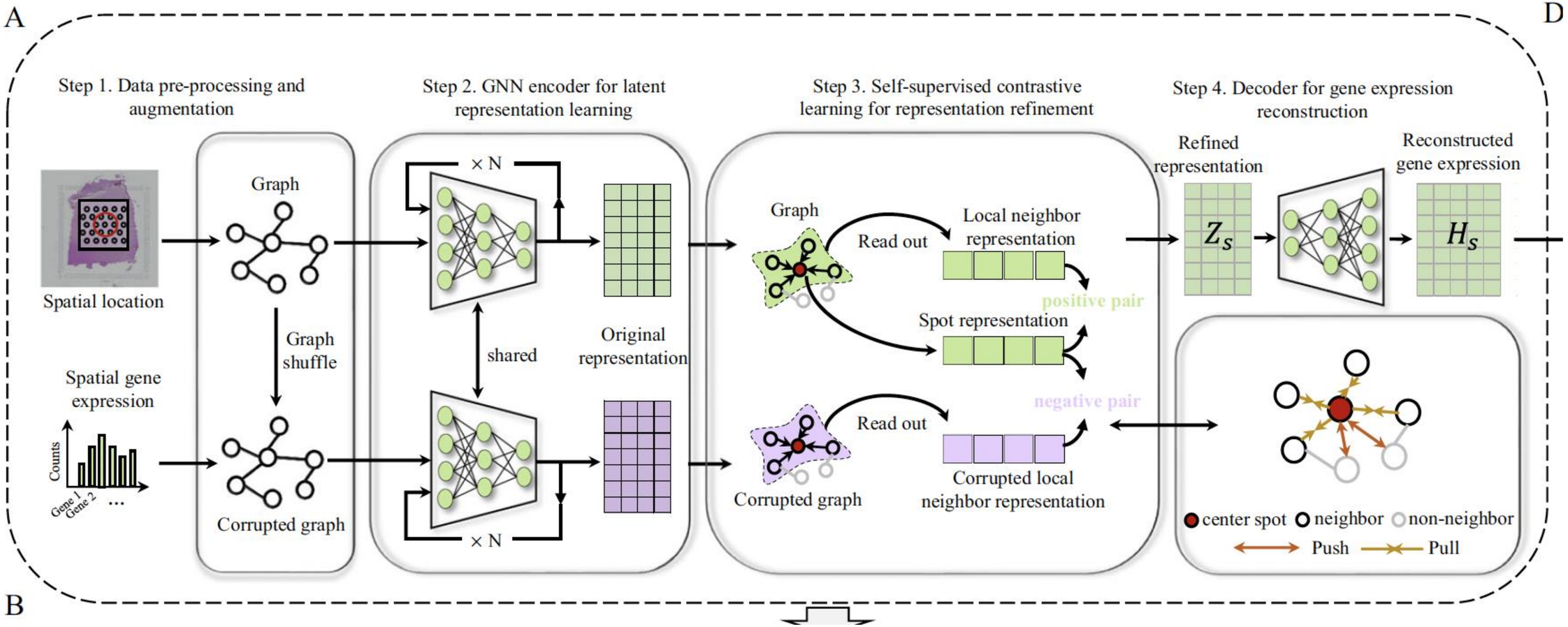
$$\mathbf{y}(\mathbf{s}) = \boldsymbol{\beta}^\top \mathbf{x}(\mathbf{s}) + \boldsymbol{\Lambda}^\top \mathbf{f}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \ , \ \mathbf{s} \in \mathcal{D} \ ,$$

analytically tractable, explainable, scalable, UQ


Constructing spatial neighbor networks
Spatial transcriptomics
Gene expressions
Gene 1
Gene 2
Gene N
Spatial location
Clustering based on expression
Spatial neighbor network
Cell type-aware spatial network
Integrating gene expression and spatial location information
Graph attention auto-encoder
W_1
W_2
W_2^T
W_1^T
Cell type-aware attention layer
(Learn neighbor spot similarity adaptively)
$1-\alpha$
α
Graph attention convolution
Cell type-aware graph attention convolution
Biological applications
Identifying spatial domains
Data denoising
Extracting 3D spatial domains

Caption

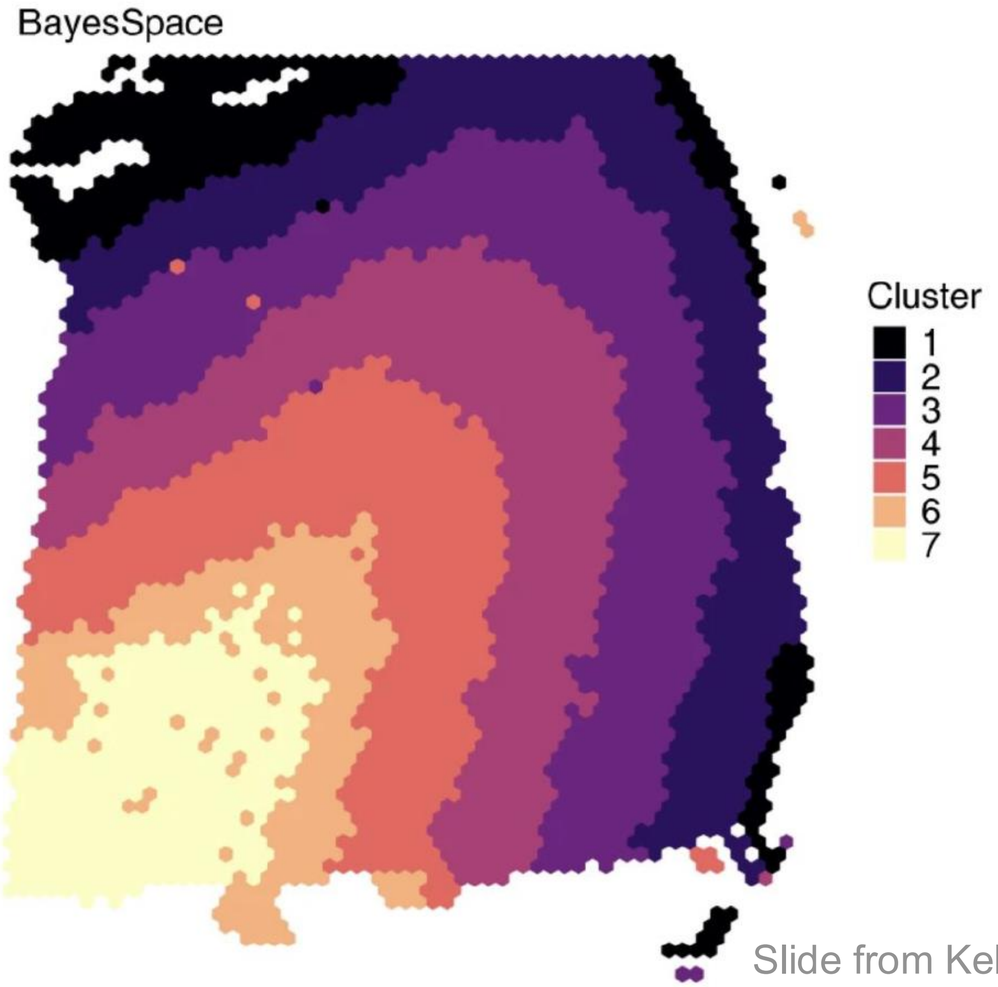
GraphST


A
D
B
Step 1. Data pre-processing and augmentation
Step 2. GNN encoder for latent representation learning
Step 3. Self-supervised contrastive learning for representation refinement
Step 4. Decoder for gene expression reconstruction
Spatial location
Spatial gene expression
Counts
Gene 1
Gene 2
...
Graph
Graph shuffle
Corrupted graph
× N
shared
× N
Original representation
Graph
Read out
Local neighbor representation
positive pair
Spot representation
negative pair
Read out
Corrupted graph
Corrupted local neighbor representation
Refined representation
Z_s
Reconstructed gene expression
H_s
center spot
neighbor
non-neighbor
Push
Pull

Clustering / Domain Detection

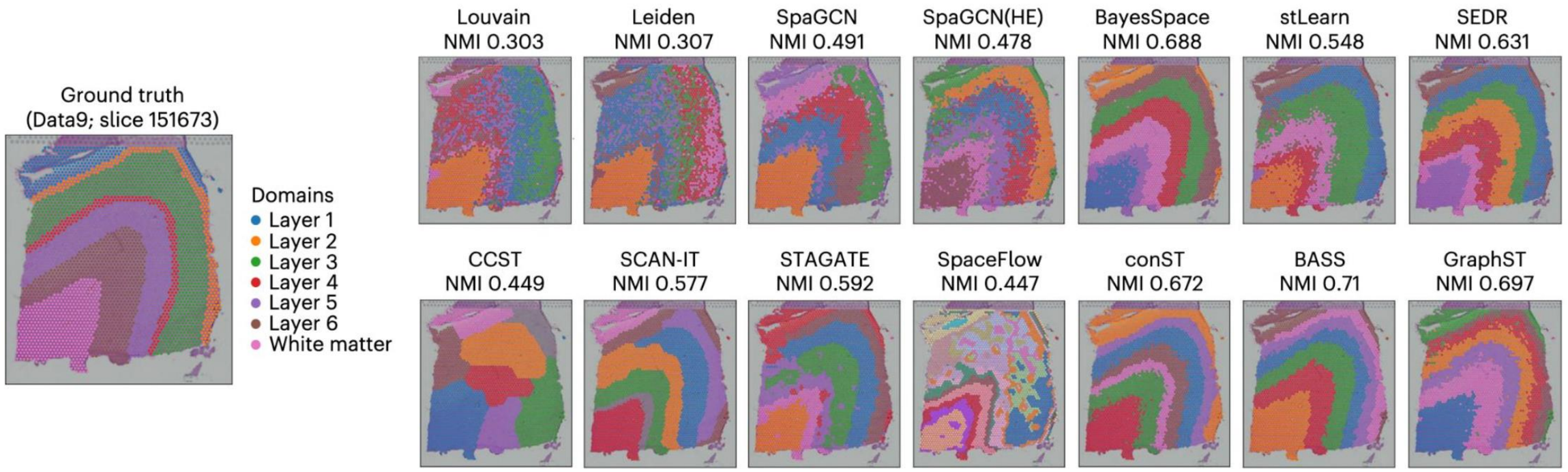
Clustering finds groups of cells with similar gene expression patterns.

Domain detection finds spatial regions with similar gene expression patterns.

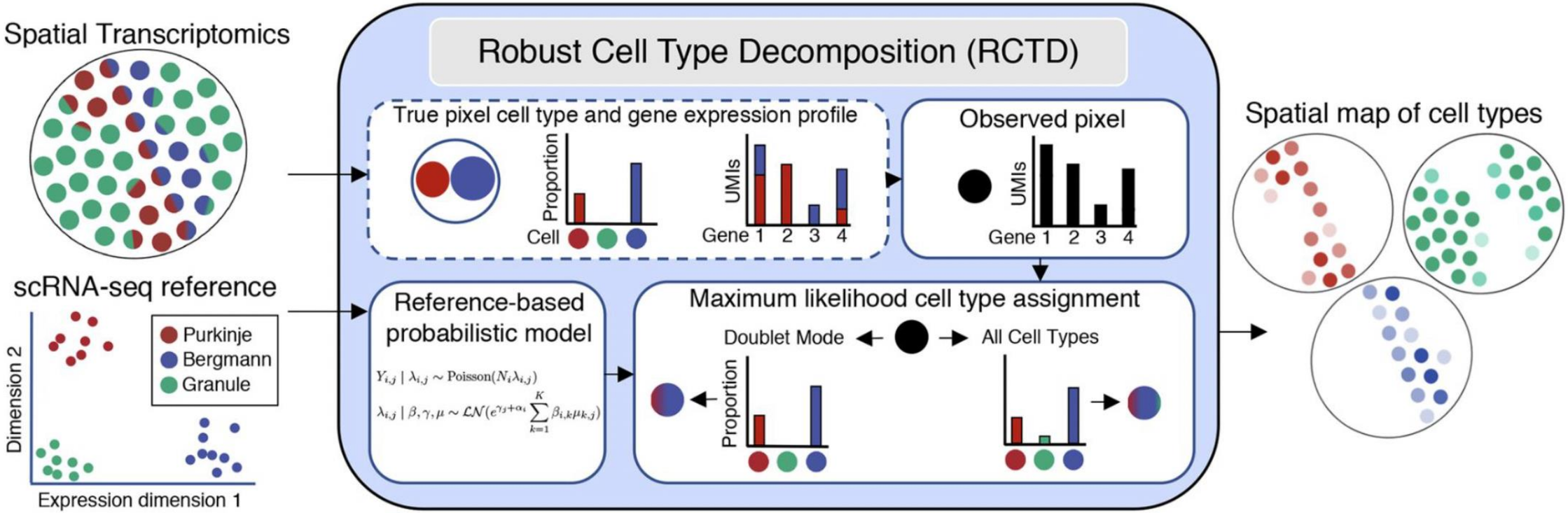


Slide from Kelly Street

Clustering / Domain Detection

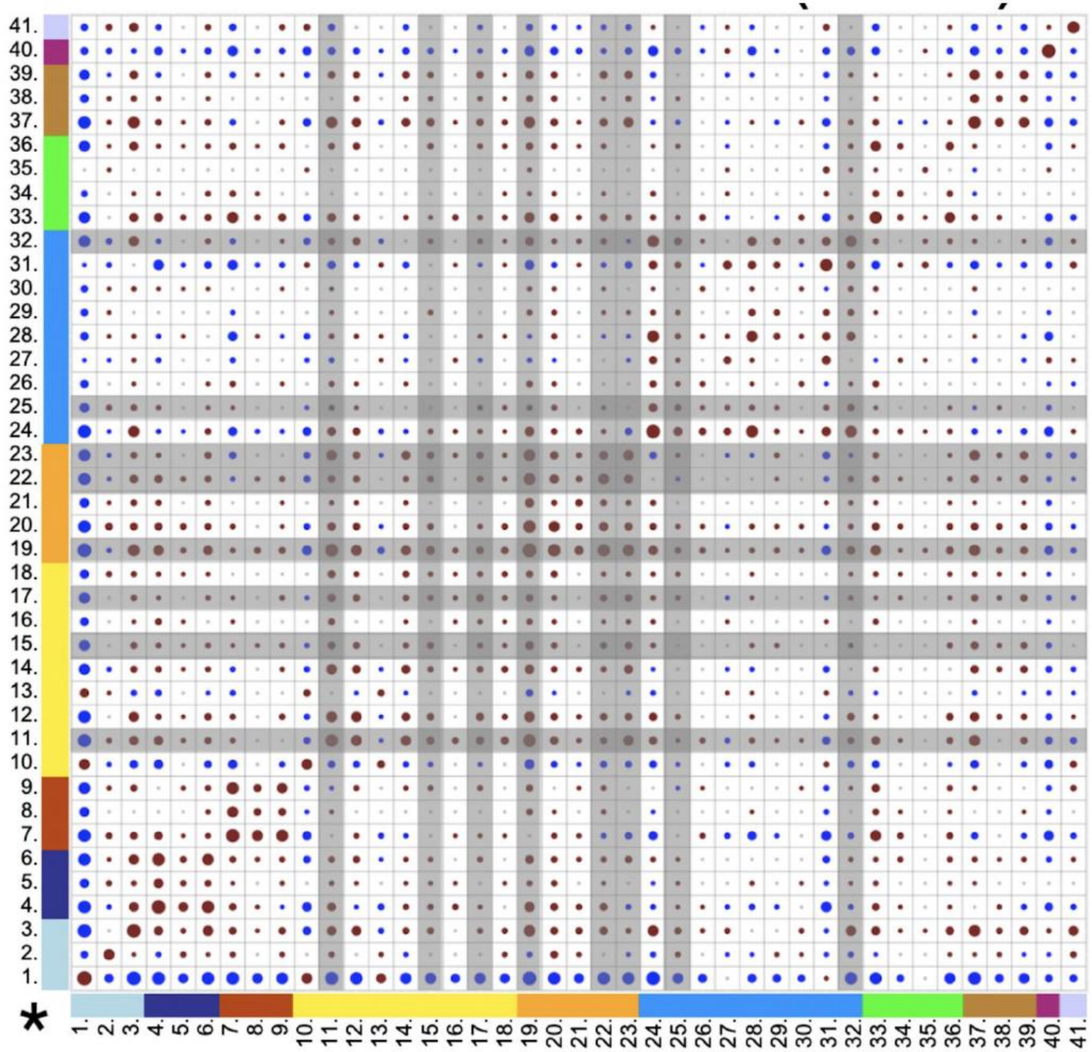


Deconvolution



Slide from Kelly Street

Attraction/Avoidance



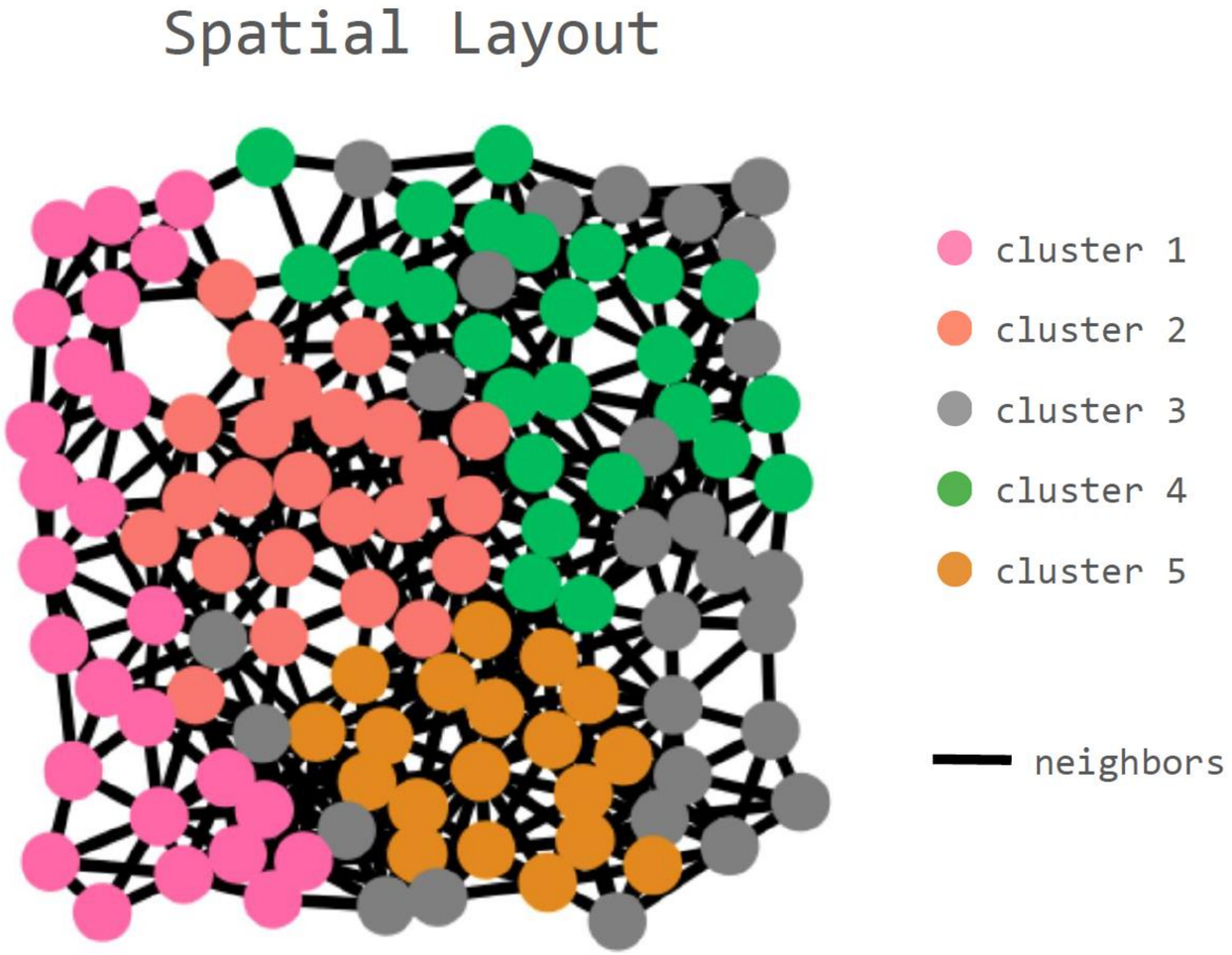
`{imcRtools}`

`testInteractions()`

For each pair of cell types, do the tend to co-localize or not?

Slide from Kelly Street

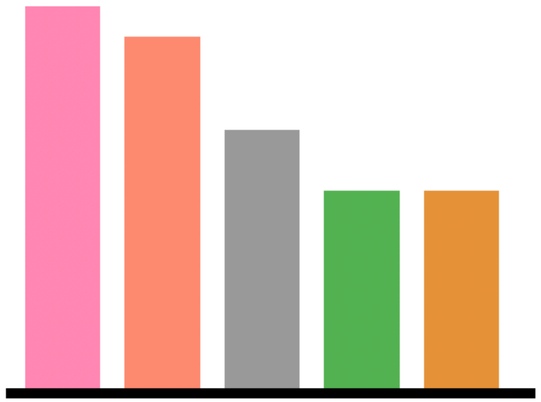
Neighborhood Analysis



Slide from Kelly Street

Neighborhood Analysis

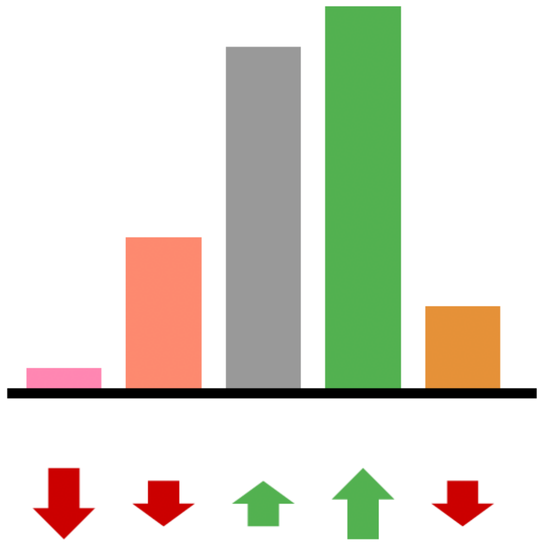
All Cells



Null distribution

Neighbors of

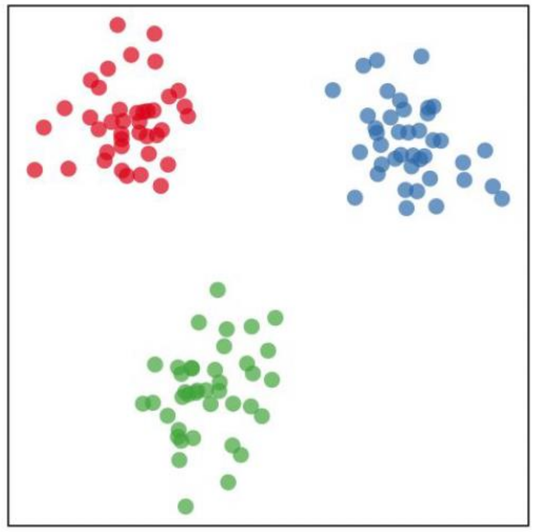
cluster 4



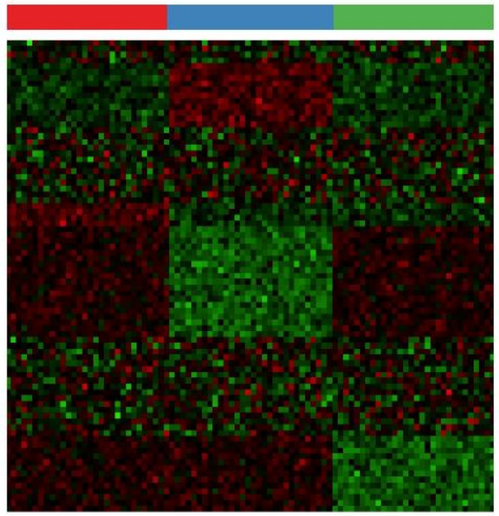
Slide from Kelly Street

Downstream Analysis

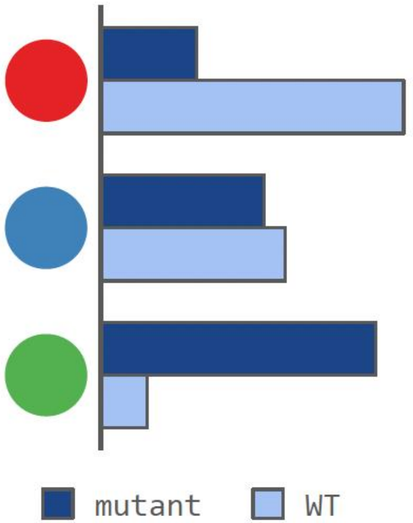
Clustering



Differential Expression



Differential Abundance



https://lmweber.org/OSTA/

Slide from Kelly Street