Swiss Institute of Bioinformatics

INTRODUCTION TO SEQUENCING-BASED TRANSCRIPTOMICS DATA ANALYSIS

Intermediate processing

Joana Carlevaro Fita

June 3-5, 2026

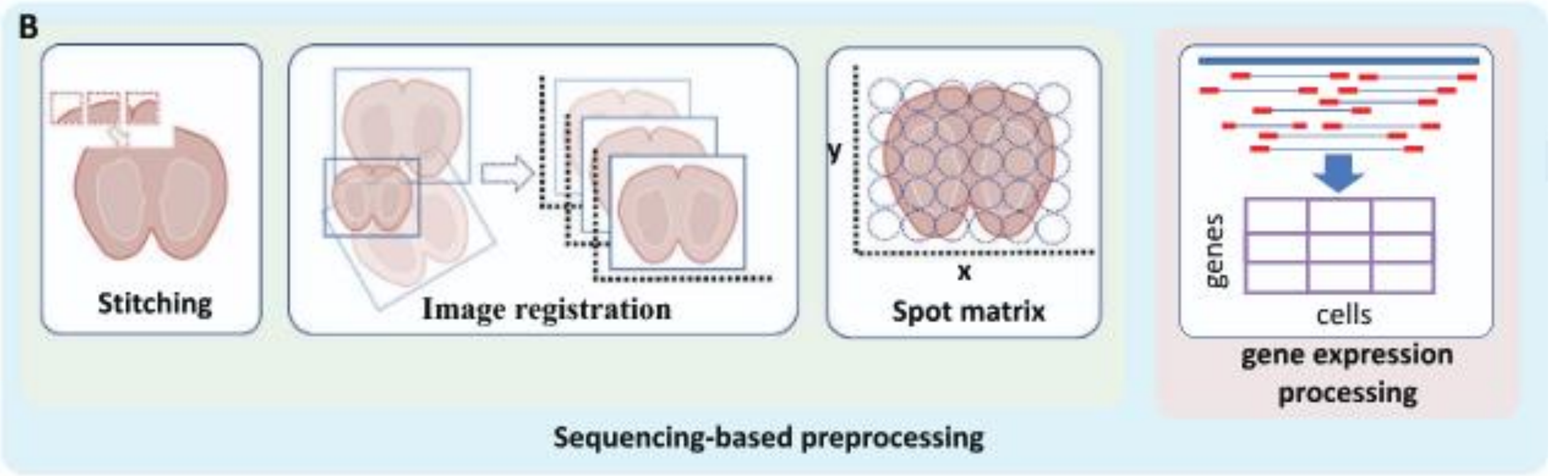
SWITZERLAND

Learning objectives

- Describe intermediate processing steps applied to spatial transcriptomics data analysis
- List spatially-aware methods used at each step, and explain the importance of testing them over standard scRNAseq methods
- Exaplain what a spatial domain is, and key aspects in their identification

Normalisation, feature selection, dimensionality reduction

Several scRNAseq methods are used for intermediate processing of spot-based ST



Cells ~ Spots

Adapted from: Wang et al. Small methods, 2025

Standard scRNAseq methods applied

1. Select Features:

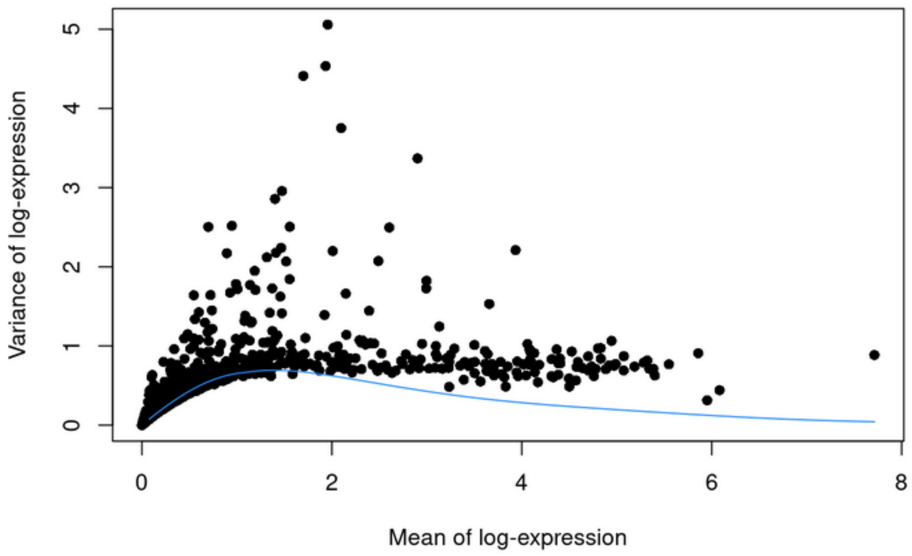
Identify highly variable genes (HVG) to screen a proportion of genes

- Reduce noise from random variation from biologically uninformative genes
- Improve computatuinal efficiency

Two-step procedure:

1) Model the mean-variance relationship, which decomposes variance into a technical component (smooth fit) and biological component (deviation thereof).

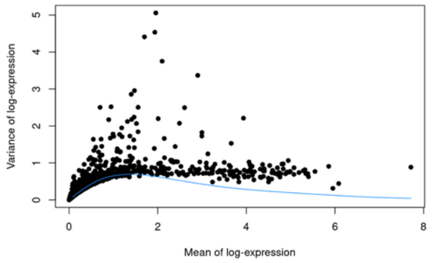
2) Select top HVG (fixed number or proportion of genes)



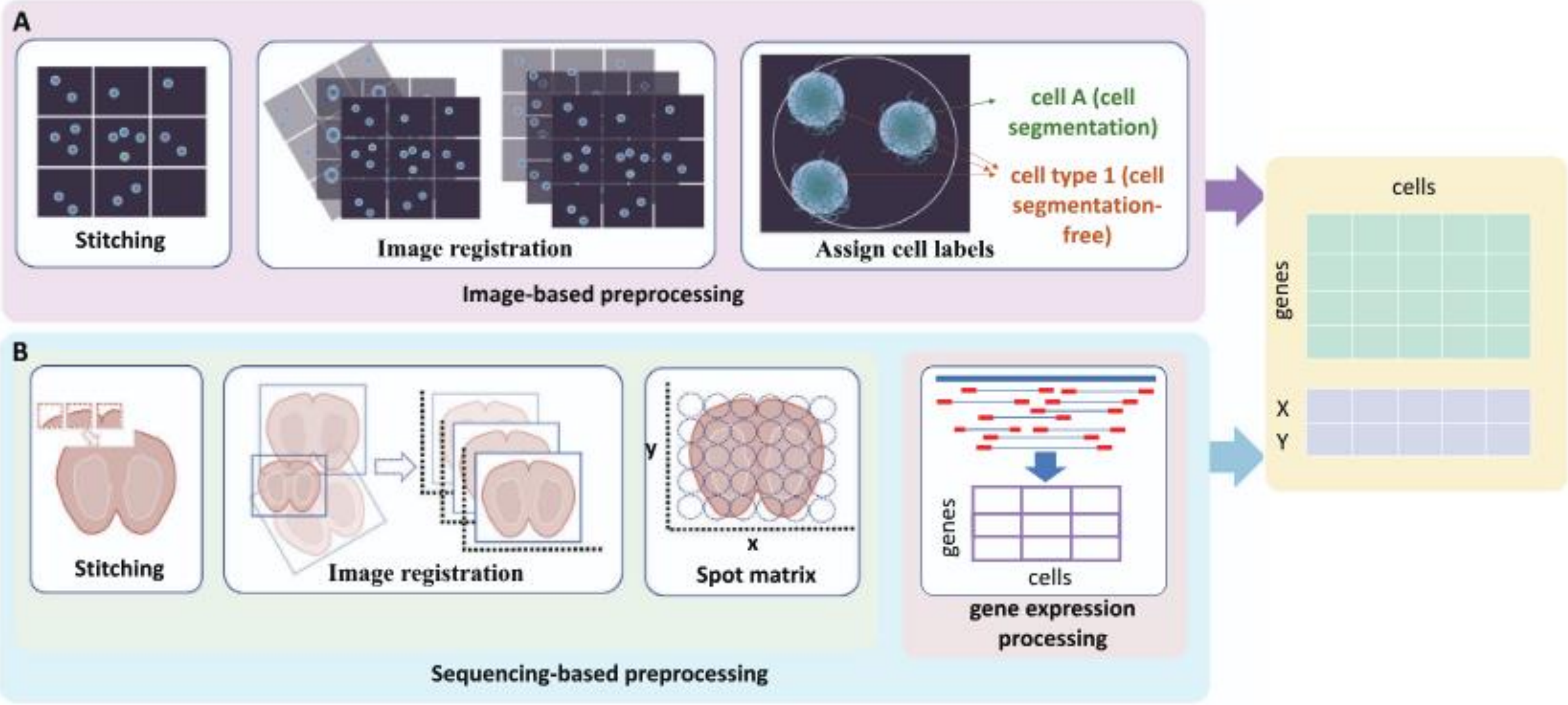
Standard scRNAseq methods applied

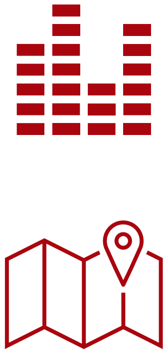
1. HVG

Identify highly variable genes to screen a proportion of genes



Spatial information





Wang et al. Small methods, 2025

Feature selection

HVG: Highly variable genes

Defined only based on molecular features (gene expression).

Do not incorporate spatial information

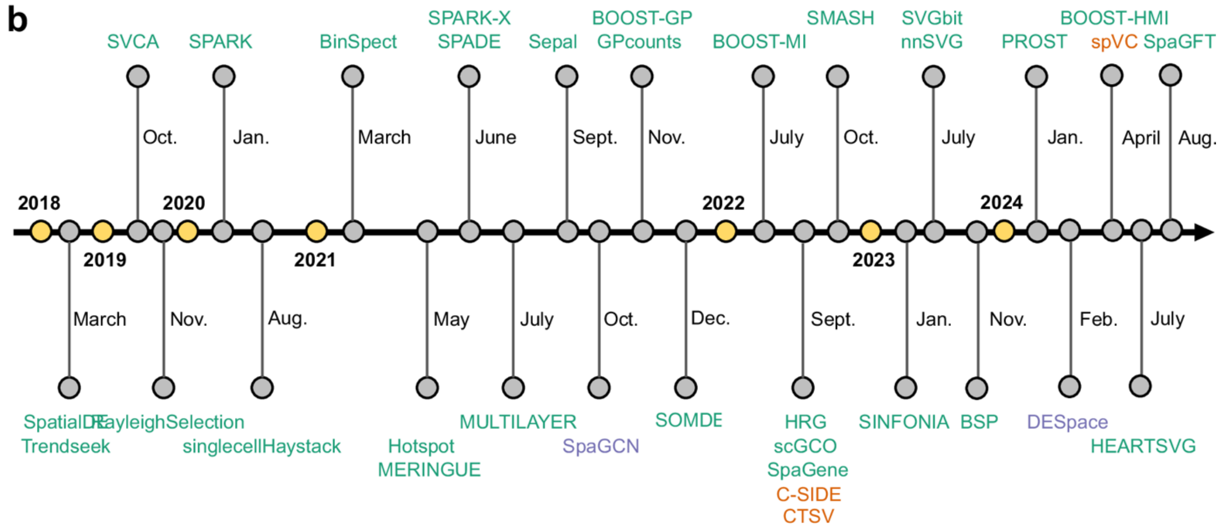
SVG: Spatially variable genes

Take spatial coordinates of the measurements also into account

Select genes with non-random, informative spatial patterns –> more biologically informative ranking of genes

Used either instead of or complementary to HVGs in subsequent steps.

Several SVG methods:
- de novo (nnSVG)
- pre-computed spatial clusters (by morphology or clustering methods (DESpace))

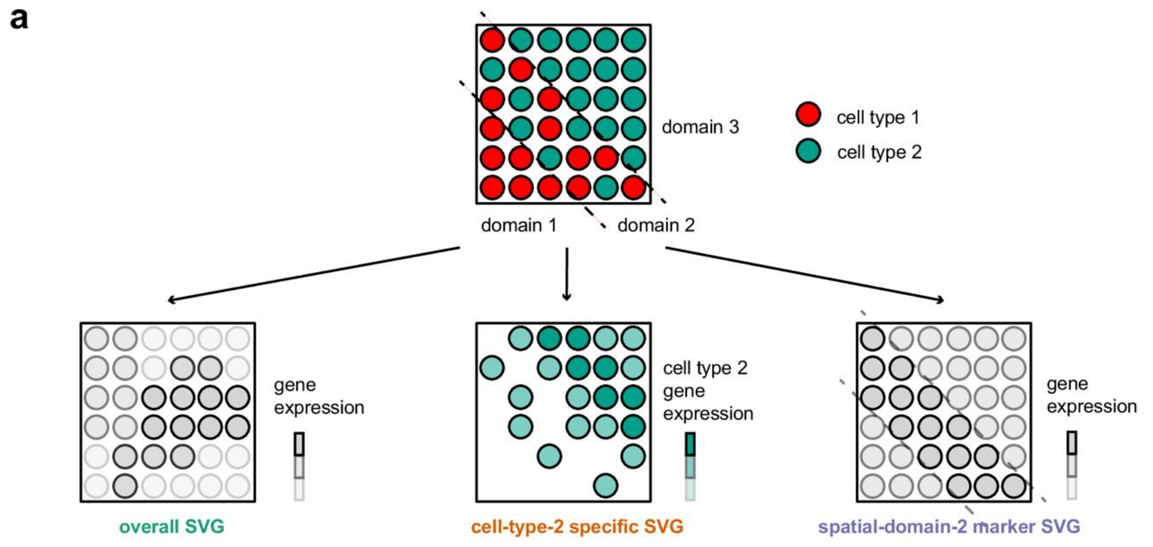


Yan, Hua, Li. Nat Comm,, 2025

SVG methods

Classification of SVG methods

[Yan, Hua, and Li. Nat Comm, 2025](#) → Categorise 34 SVG methods in 3 categories:



1. Overall SVG:

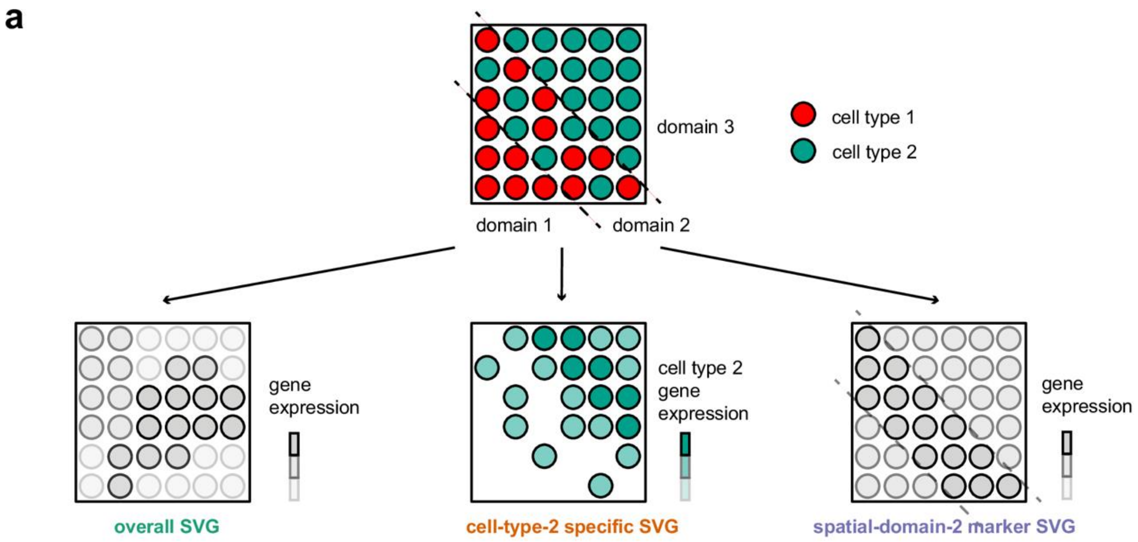
- Based on Gaussian process model: spatialDE,nnSVG
- Rank genes based on spatial autorcorrelation: Moran's I, Geary's C
- Non-parametric test of covariance matrices: SPARK

- Used as feature selection step (screen for informative genes)
- To identify spatial domains . Cluster spots (graph-based clustering) using SVGs expression and location information.
- Not necessary for all spatial domain detection methods (BayesSpace based on HVG)
- Identification of spatial-gene modules

SVG methods

Classification of SVG methods

[Yan, Hua, and Li. Nat Comm, 2025](#) → Categorise 34 SVG methods in 3 categories:



2. Spatial domain-specific:

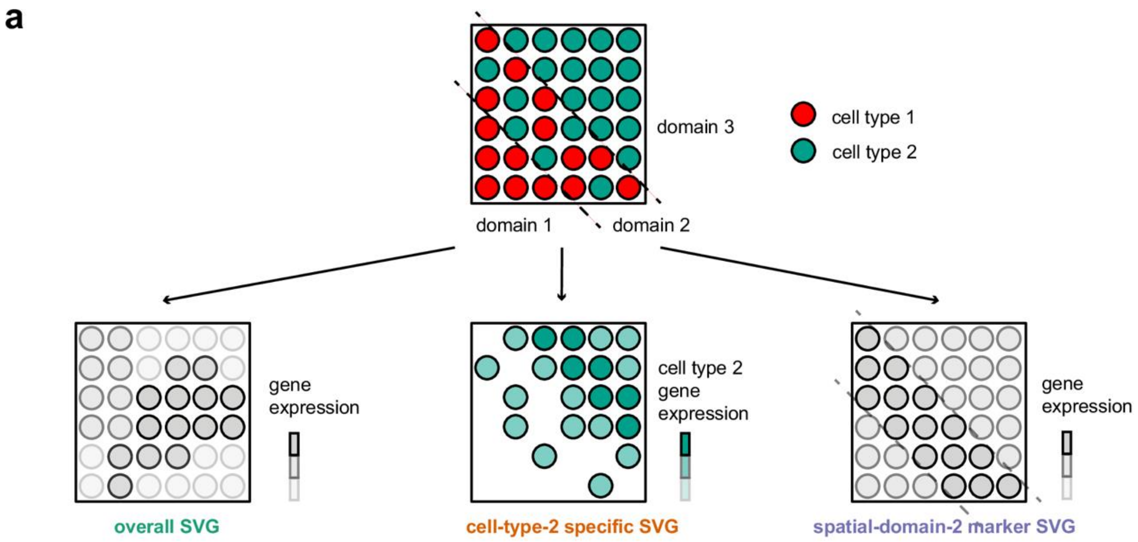
DESpace, SpaGCN

- Genes that change significantly between domains and summarise spatial information
- Identified using spatial domains
- Insights to molecular mechanisms of spatial domains/tissue layers
- Spatial domain marker SVGs can help with other domains annotation.

SVG methods

Classification of SVG methods

[Yan, Hua, and Li. Nat Comm, 2025](#) → Categorise 34 SVG methods in 3 categories:

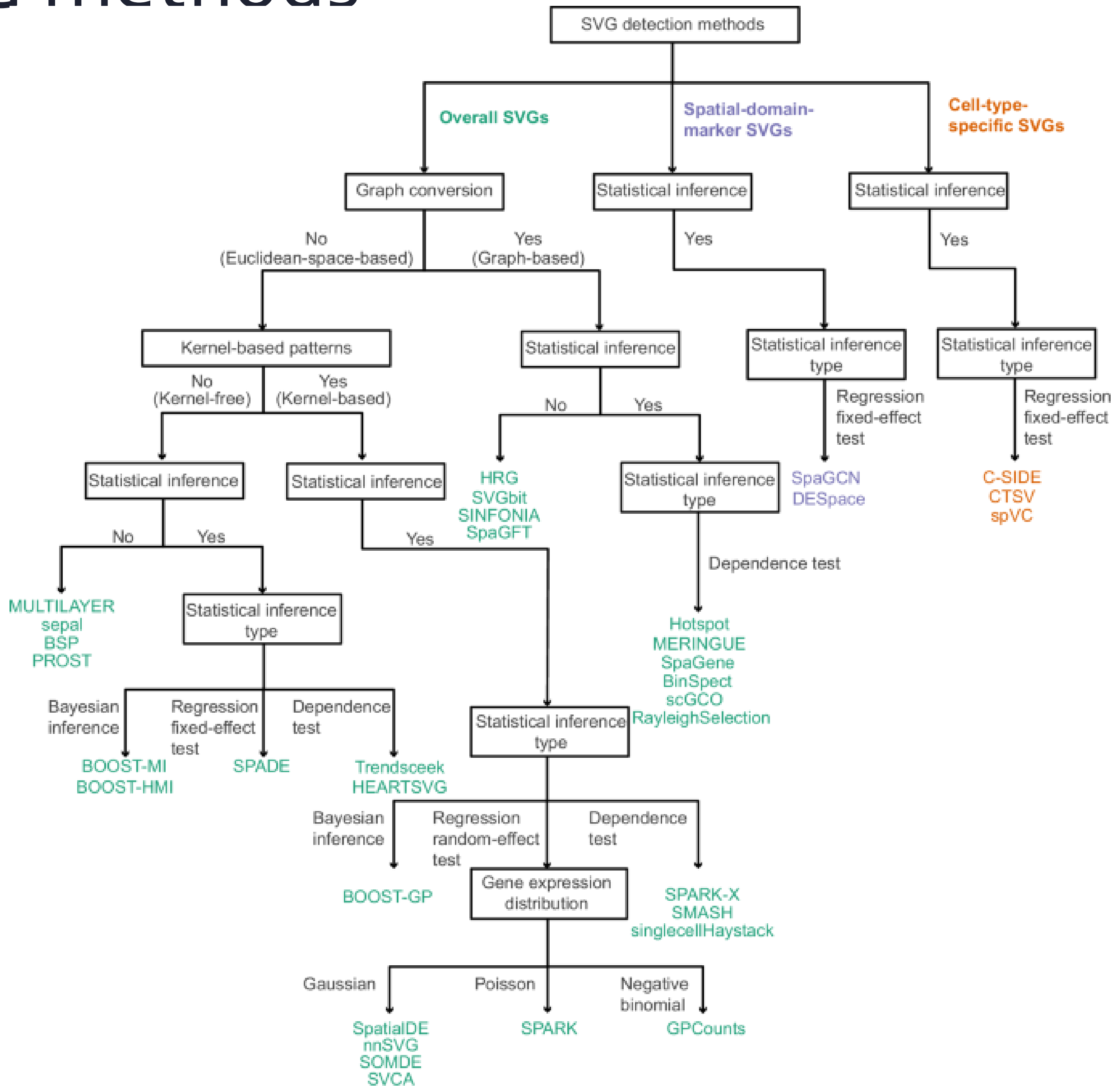


3. **Cell type-specific:**

   CTSV, C-SIDE, spVC

- Genes that exhibit non-random spatial expression patterns within a cell type.
- Indentified using external cell type annotations
- Identify cell subpopulations or cell states across the tissue section

SVG methods



Supplementary Table 1: Summary of three existing benchmark studies for SVG detection methods

	Charitakis *et al.*, 2023 [1]	Chen *et al.*, 2024 [2]	Li *et al.*, 2023 [3]
SVG Detection Methods	1. SpatialDE 2. SPARK-X 3. Squidpy 4. Seurat ("markvariogram") 5. SpaGCN 6. scGCO	1. SpatialDE 2. SPARK-X 3. SOMDE 4. Giotto 5. nnSVG 6. MERINGUE 7. Seurat ("moransi")	1. Moran's I 2. Spanve 3. scGCO 4. SpaGCN 5. SpaGFT 6. Sepal 7. SpatialDE 8. SpatialDE2 9. SPARK 10. SPARK-X 11. BOOST-GP 12. GPcounts 13. nnSVG 14. SOMDE

[…] Datasets, Evaluation Metrics, Conclusions

Yan, Hua, Li. Nat Comm,, 2025

Dimensionality reduction (DR)

- Non spatially-aware

DR based on the cell's molecular profile only: PCA, NMF, LDA..
Often combined with spatially-aware clustering methods

- Spatially-aware

Takes spatial information also into account: BANKSY, SpatialPCA, STAMP
Often combined with standard clustering approaches from scRNAseq (SNN graph), scRNAseq trajectory inference models

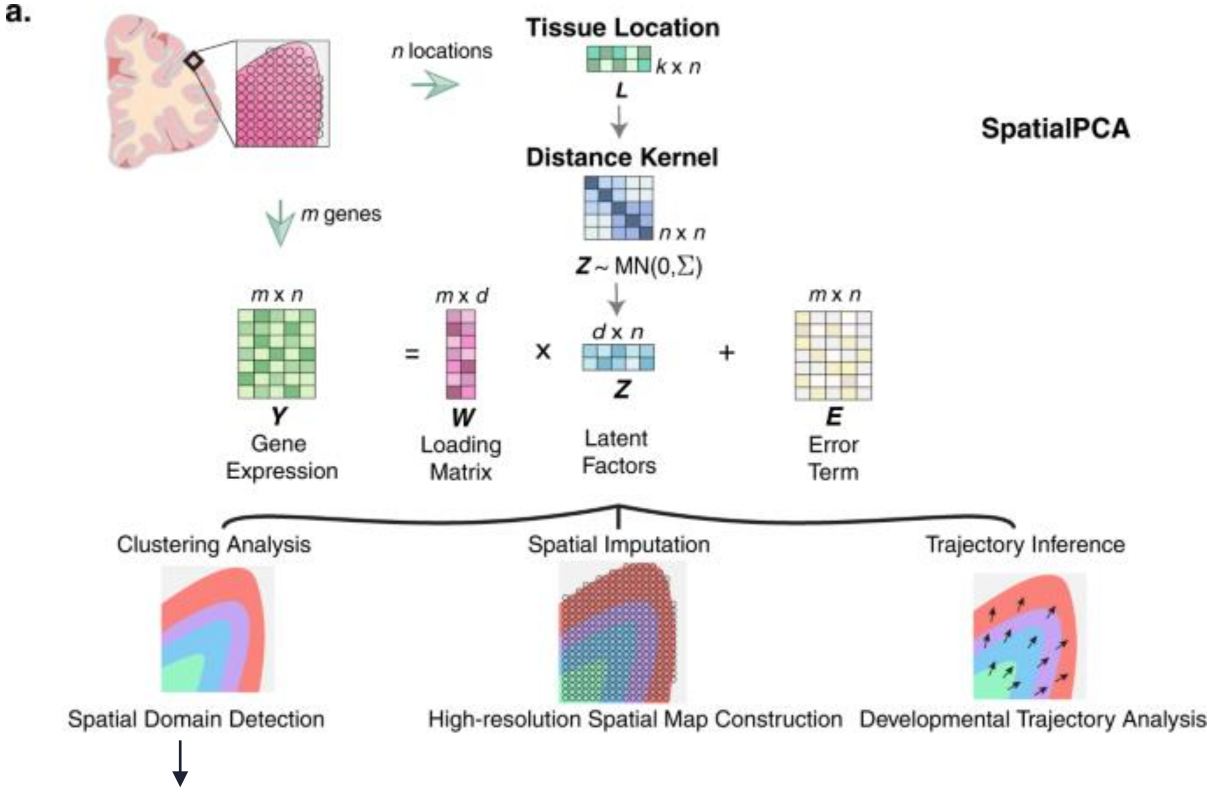
Dimensionality reduction (DR)

SpatialPCA (Shang and Zhou, Nat Comm. 2022)

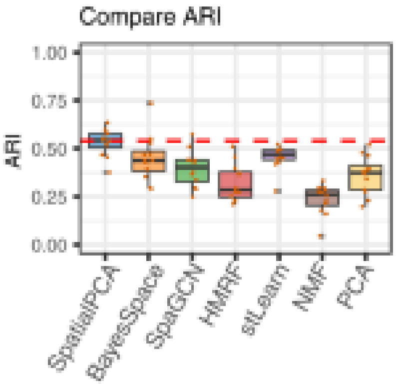
Neighbouring regions share similar cell type compositions

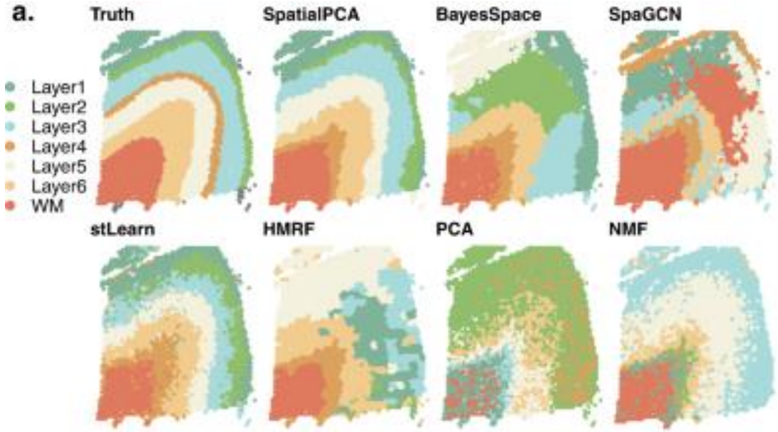
“SpatialPCA builds upon the probabilistic version of PCA, incorporates localization information as additional input, and uses a kernel matrix to explicitly model the spatial correlation structure across tissue locations.”

Spatial PCs can be paired with clustering analysis on the low-dimensional components that contain spatial correlation information.



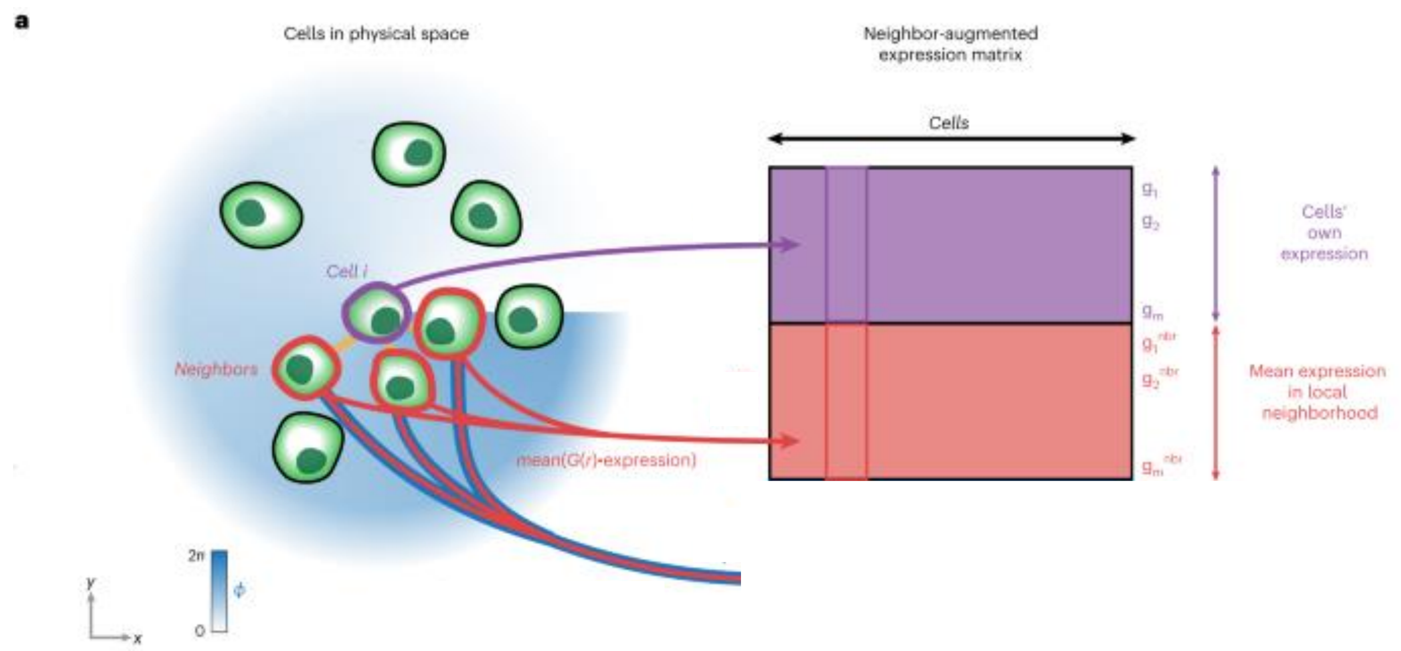
Adjusted Rand Index (ARI)
Statistical measure to quantify similarity beween clusterings





Dimensionality reduction (DR)

Building Aggregates with a Neighborhood Kernel and Spatial Yardstick (BANKSY)

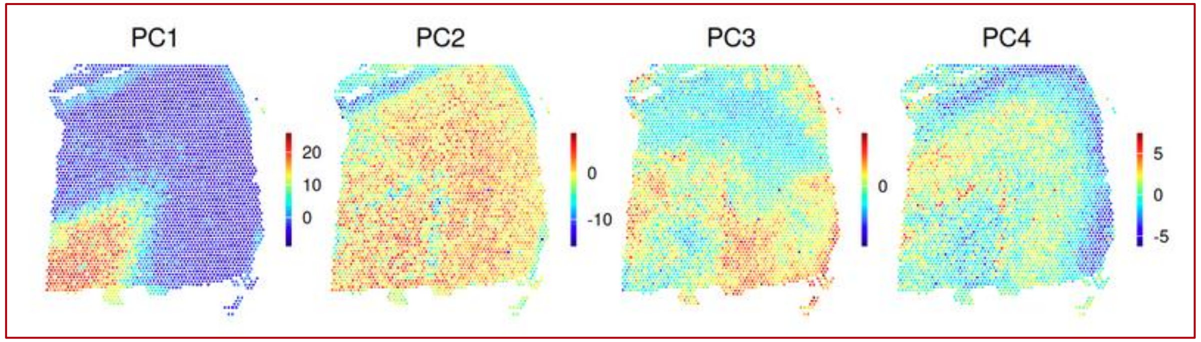


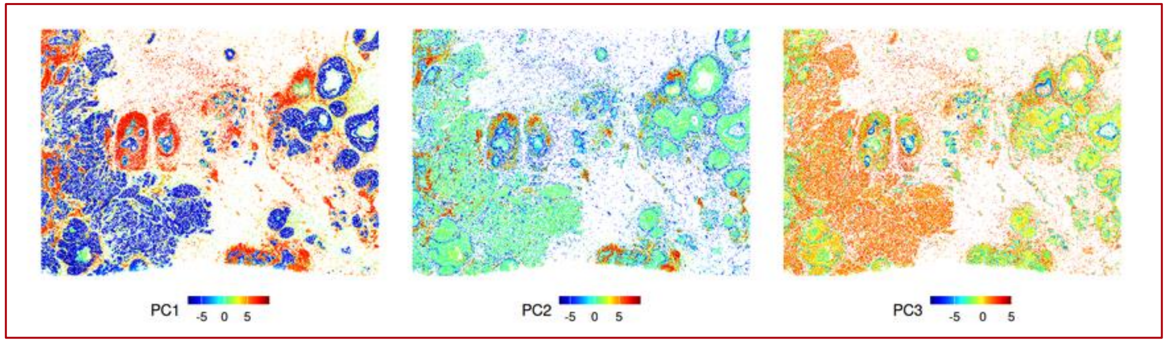
Adapted from Singhal et al. Nat Gen, 2024

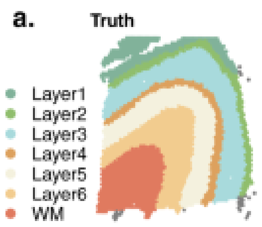
Dimensionality reduction (DR)

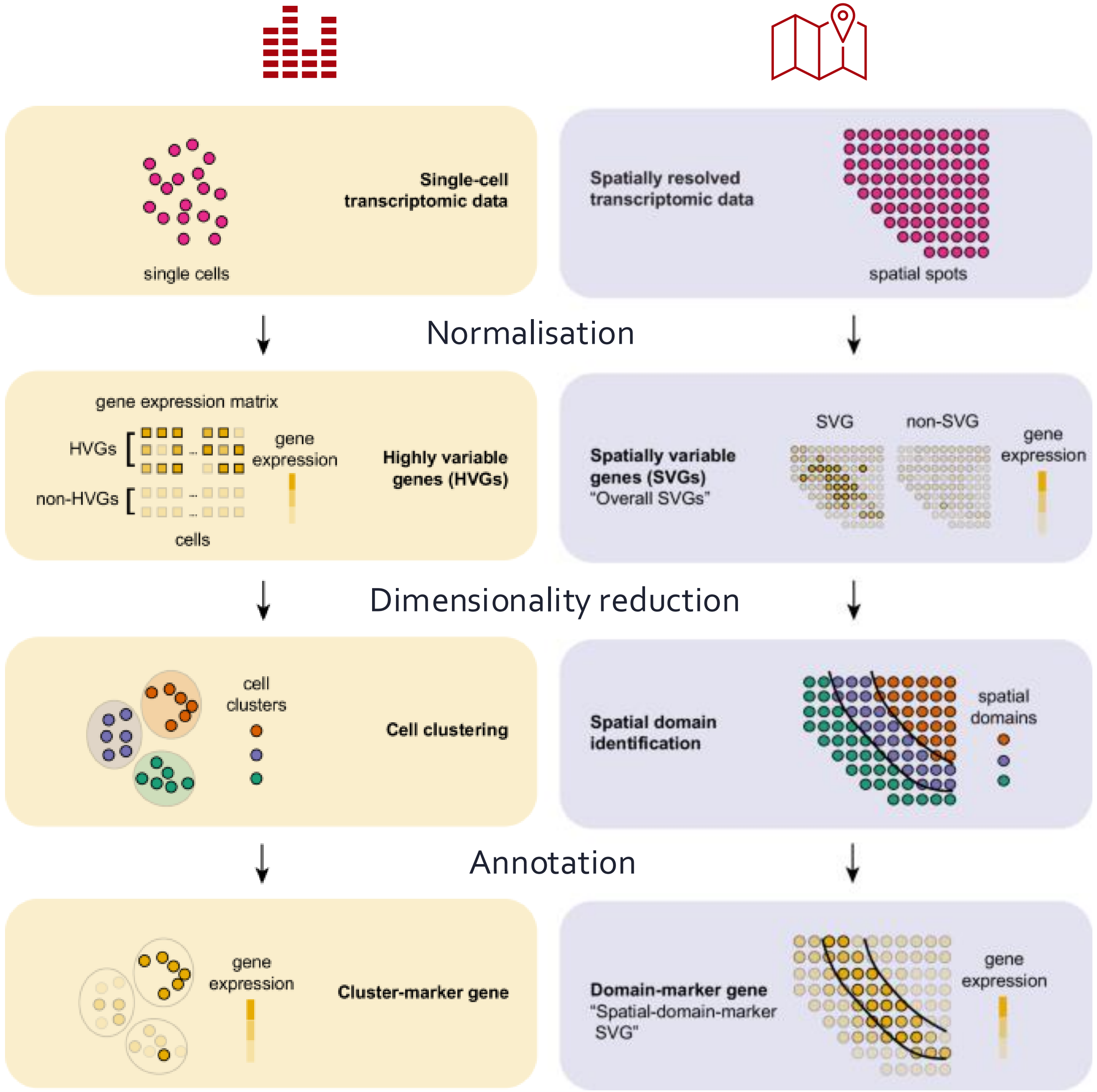
Visualise PCs in spatial context

Visualise PCs in the tissue slice in order to observe if the main sources of variation explain distinguishable tissue structres/spatial locations









Adapter from: Yan, Hua, Li. Nat Comm,, 2025

Thank you

DATA SCIENTISTS FOR LIFE

sib.swiss