Introduction to Sequencing-based Spatial Transcriptomics Data Analysis

Spatially variable genes and differential spatial patterns

Outline

- Spatially variable gene (SVG) methods

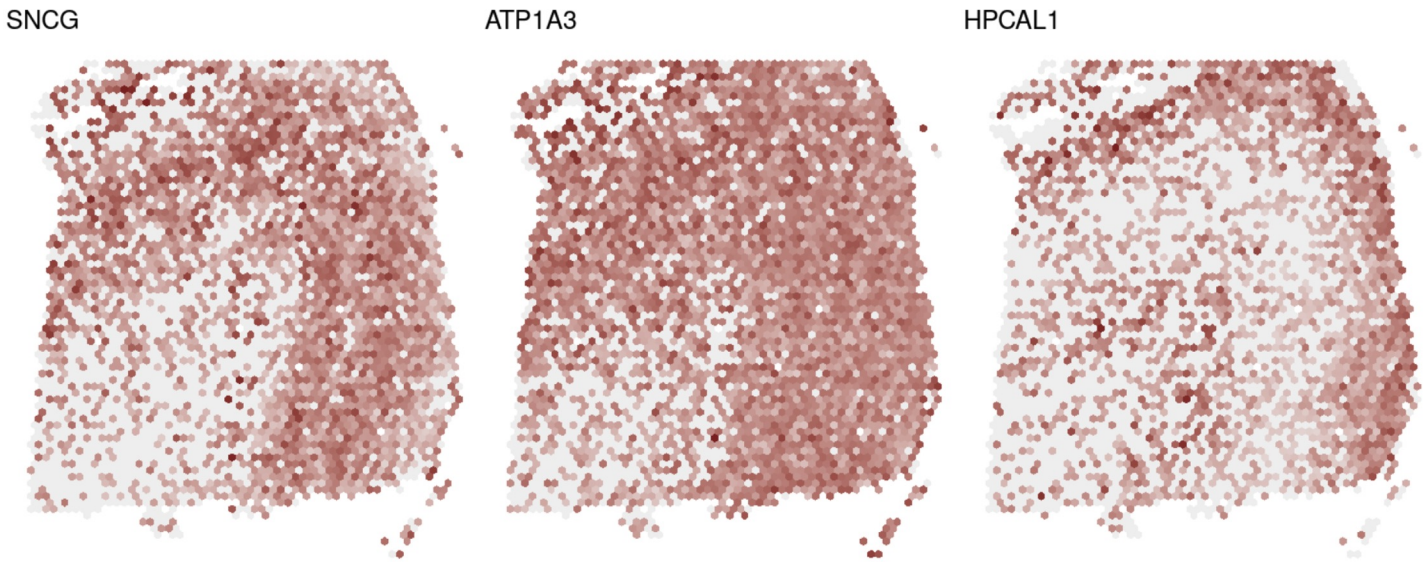
  (SpatialDE, nnSVG, SPARK-X, DESpace, C-DISE)

- Feature-set signatures

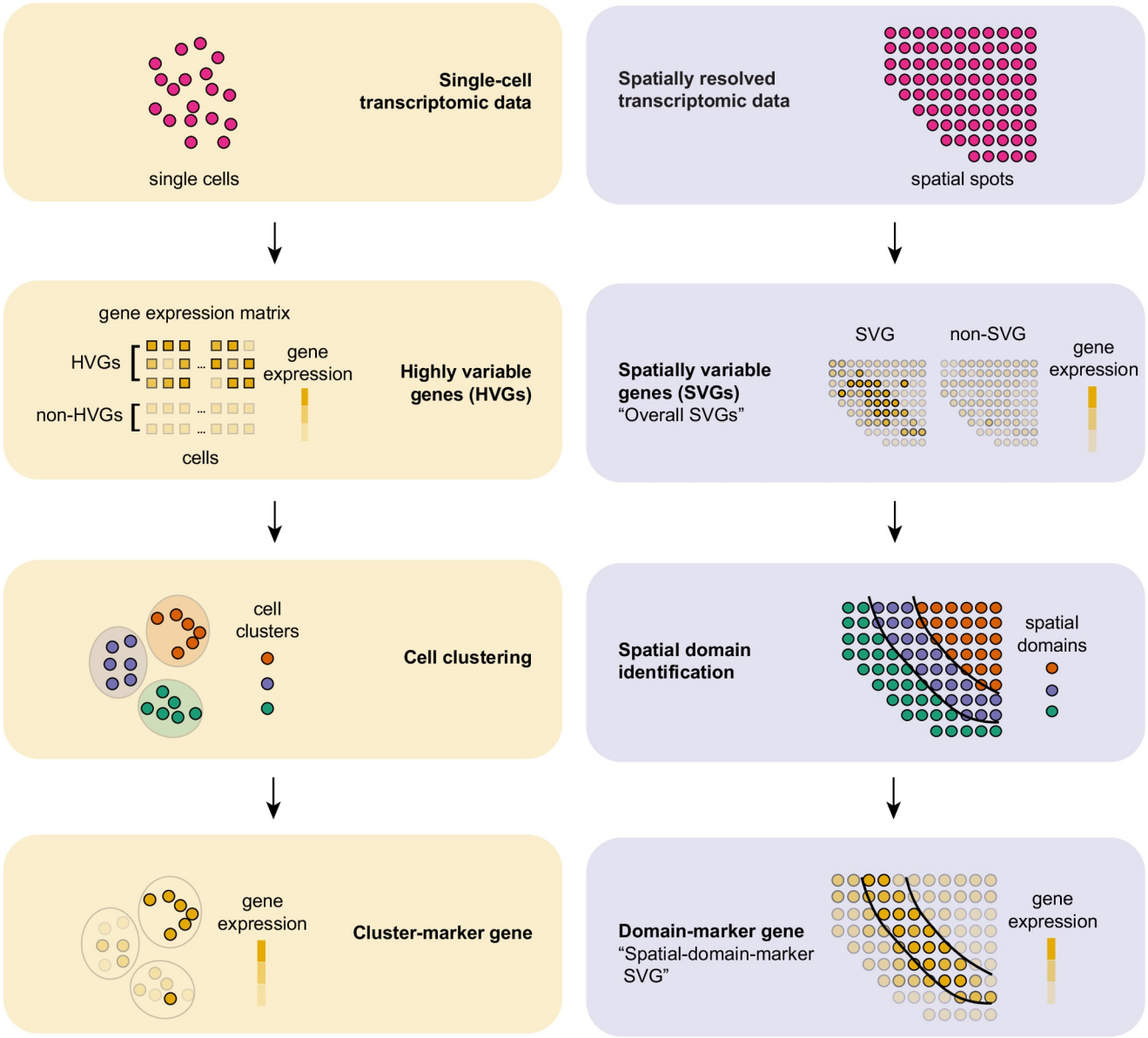
- Differential analysis with multi-sample and multi-condition

Spatially variable genes (SVGs)

genes whose expression profiles vary across tissue.

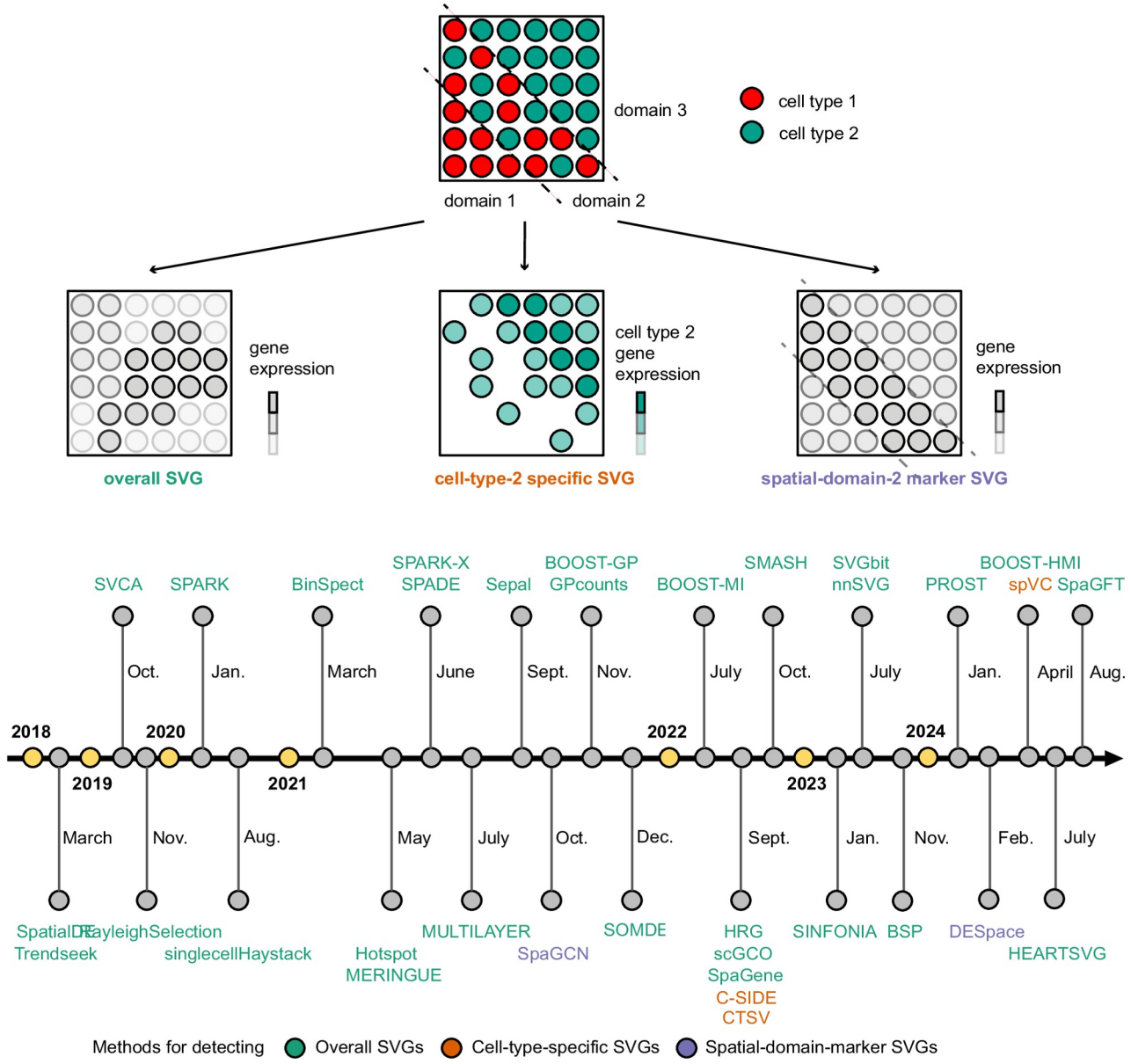


HVGs vs. SVGs



Yao et.al., 2025

SVGs approaches



Yan et.al., 2025

Overall SVG - SpatialDE

SpatialDE: identification of spatially variable genes

Valentine Svensson, Sarah A Teichmann & Oliver Stegle

Nature Methods **15**, 343–346 (2018) | Cite this article

- **Model assumption**: normalized gene expression $Y = (y_1, \ldots, y_n)$ follows an *n*-dimensional Gaussian distribution, containing:
  - a spatial covariance component
  - a non-spatial error variance component
- **Spatial covariance**: squared exponential covariance matrix based on coordinates of cells
- Null hypothesis: $H_0: \sigma_s{}^2 = 0$ (non spatial covariance component); tests via a **Likelihood ratio test**
- Computational cost scales **cubically** with the number of cells

$$\mathbf{Y} \sim \mathrm{MVN}\left(\boldsymbol{\mu},\ \sigma_s^2 \cdot \mathbf{K}(\mathbf{s}) + \delta \cdot \mathbf{I}\right)$$

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

Overall SVG - nnSVG

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

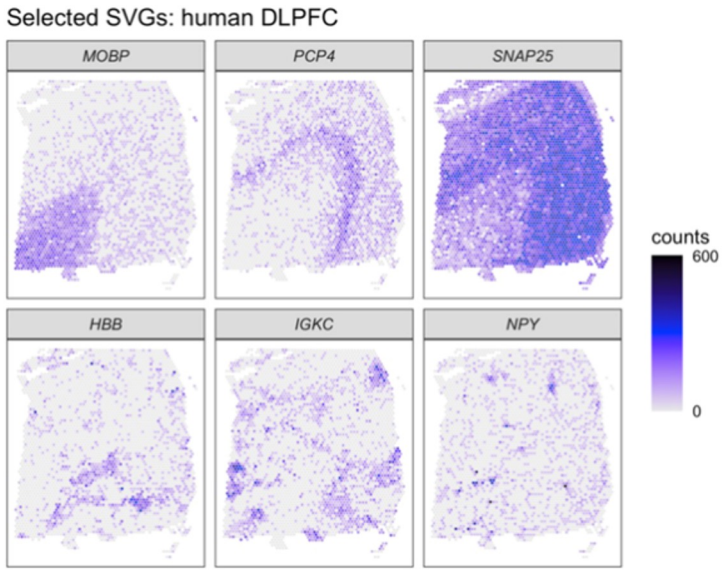
Lukas M. Weber, Arkajyoti Saha, Abhirup Datta, Kasper D. Hansen & Stephanie C. Hicks

Nature Communications **14**, Article number: 4059 (2023) | Cite this article

- More scalable: nearest-neighbor Gaussian process (SpatialDE: full Gaussian process)
- Spatial covariance: exponential covariance (SpatialDE: squared exponential covariance)
- Computational cost: ($O(n * m^3)$), n = number of spatial locations; m = number of nearest neighbors

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \widetilde{\Sigma}(\boldsymbol{\theta}, \tau^2))$$

$$C_{ij}(\boldsymbol{\theta}) = \sigma^2 \exp\left(\frac{-||\mathbf{s_i} - \mathbf{s_j}||}{l}\right)$$



Overall SVG - SPARK-X

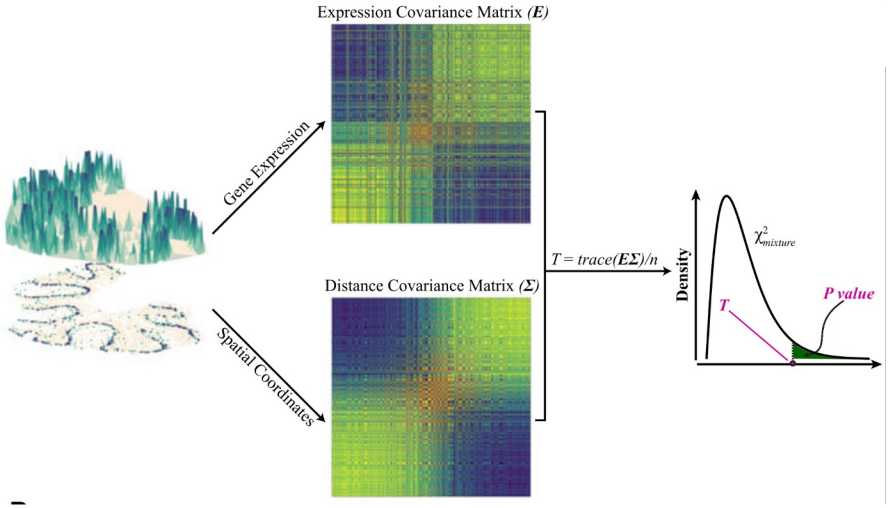
Home > Genome Biology > Article

SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies

Method | Open access | Published: 21 June 2021

Volume 22, article number 184, (2021) Cite this article

- Non-parametric spatial statistic
- Spatial modeling: performs kernel smoothing without specifying covariance
- Tests whether two similarity matrices are independent using Pearson correlation
  - One similarity matrix is based on the gene's expression
  - The other is based on the kernel-transformed spatial locations



Spatial-domain DE: DESpace

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

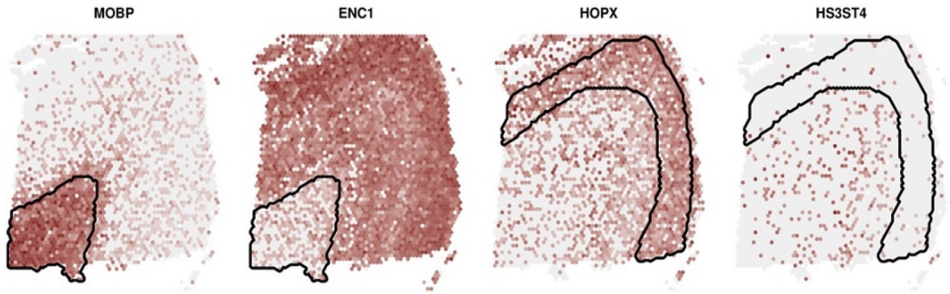
Peiying Cai, Mark D Robinson, Simone Tiberi ✉

Bioinformatics, Volume 40, Issue 2, February 2024, btae027,

- Key point: spatial domains taken as proxy for the actual spatial information.
  - Assumption: **spatial domains** successfully summarize spatial information.
- Fit a negative binomial (NB) model, with spatial domains as covariate.
- Null hypothesis: $H_0 : \quad \beta_{g1} = \cdots = \beta_{gC}$ tests via a **Likelihood ratio test**

$$x_{gi} \sim NB(\mu_{gi}, \phi_g),$$

$$log(\mu_{gi}) = log(M_i) + \beta_{gc}$$



Cell type-specific DE: C-SIDE

Cell type-specific inference of differential expression in spatial transcriptomics

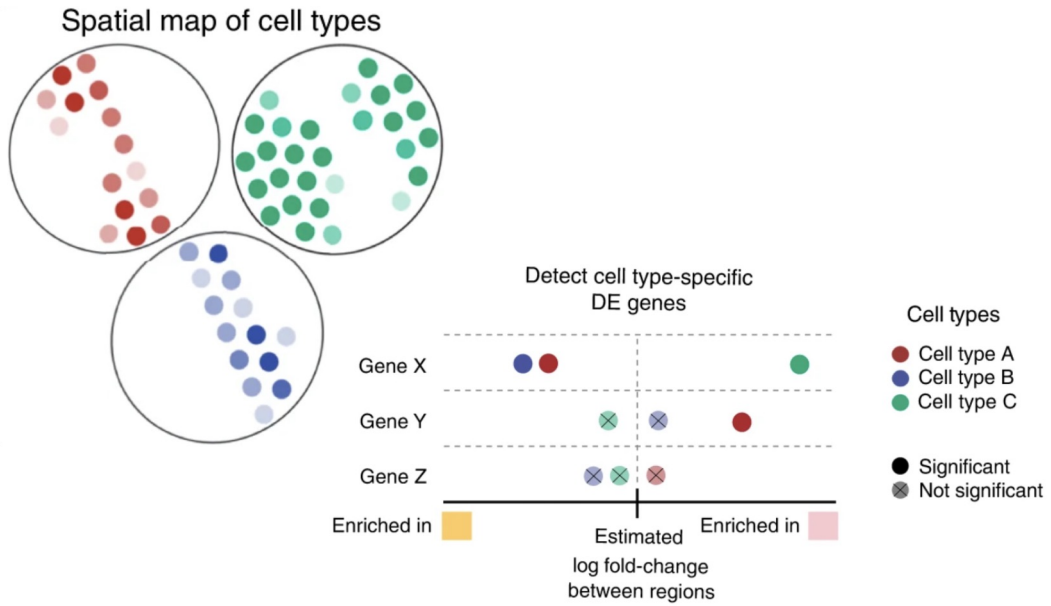
Dylan M. Cable, Evan Murray, Vignesh Shanmugam, Simon Zhang, Luli S. Zou, Michael Diao, Haiqi Chen, Evan Z. Macosko, Rafael A. Irizarry ✉ & Fei Chen ✉

Nature Methods **19**, 1076–1087 (2022) | Cite this article

- Model assumption: gene expression $Y = (y_1, \ldots, y_n)$ follows **Poisson** distribution
- Key idea:
  - For each cell type, C-SIDE learns a **smooth spatial curve** showing how gene expression changes across the tissue
  - Use L **smooth basis functions** to build the curve
- Test SVG specific to cell type k if $\beta_{k1} = \cdots = \beta_{kL} = 0$

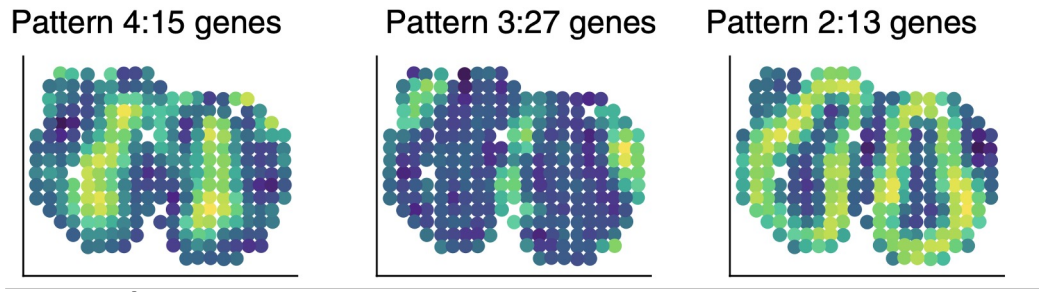
$$\log\left(\mu_i(\mathbf{s}_i)\right) = \gamma_0 + \log \ell_i + \log\left(\sum_{k=1}^{K} \eta_k(\mathbf{s}_i) w_{ik}\right) + \epsilon_i$$

$$\log\left(\eta_k(\mathbf{s}_i)\right) = \beta_{k0} + \sum_{\ell=1}^{L} \beta_{k\ell} b_\ell(\mathbf{s}_i),$$



Application of SVGs

- Identify informative genes for downstream analyses
- Downstream analysis:
  - Spatial domains
    - ❖ partition a tissue slice into regions
    - ❖ Cells/spots within the same domain have similar expression profiles
  - Spatial gene modules
    - ❖ Cluster overall SVGs into modules
    - ❖ Each module contains genes with similar spatial expression patterns



Feature-set analyses - AUCell

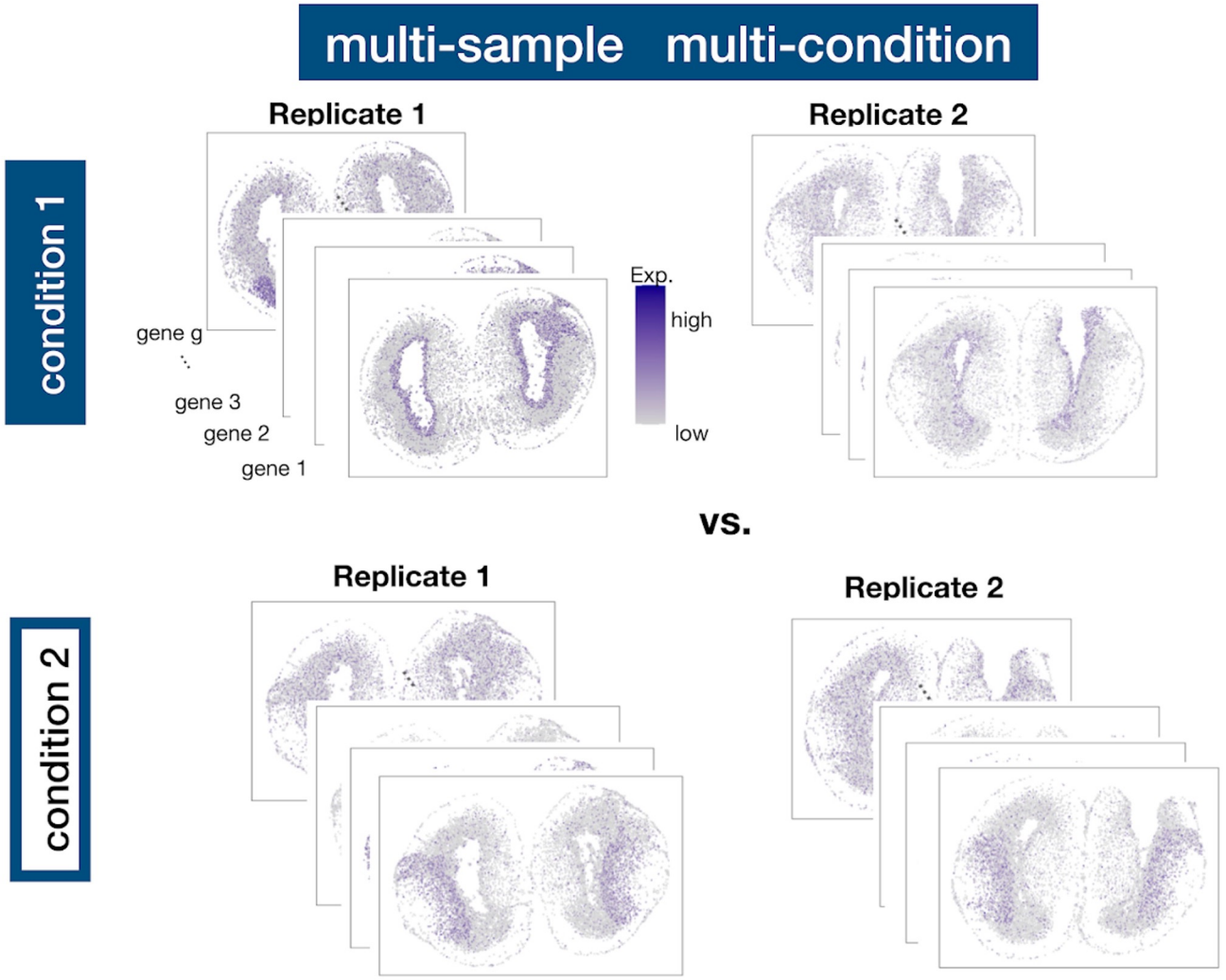
- Instead of focusing on single genes, we can also check pathways or gene modules
- AUCell identify cells with an active ‘gene set’
    - For each cell, rank all genes by expression
    - For each gene set, compute AUC (Area Under the Curve): measure how enriched the gene set is among the top ranked genes in that cell
- Signature scores summarize functional signals (e.g., immune activation, neuronal signaling)
- Helps link gene sets to cell states, differentiation, or other biological processes

SCENIC: single-cell regulatory network inference and clustering

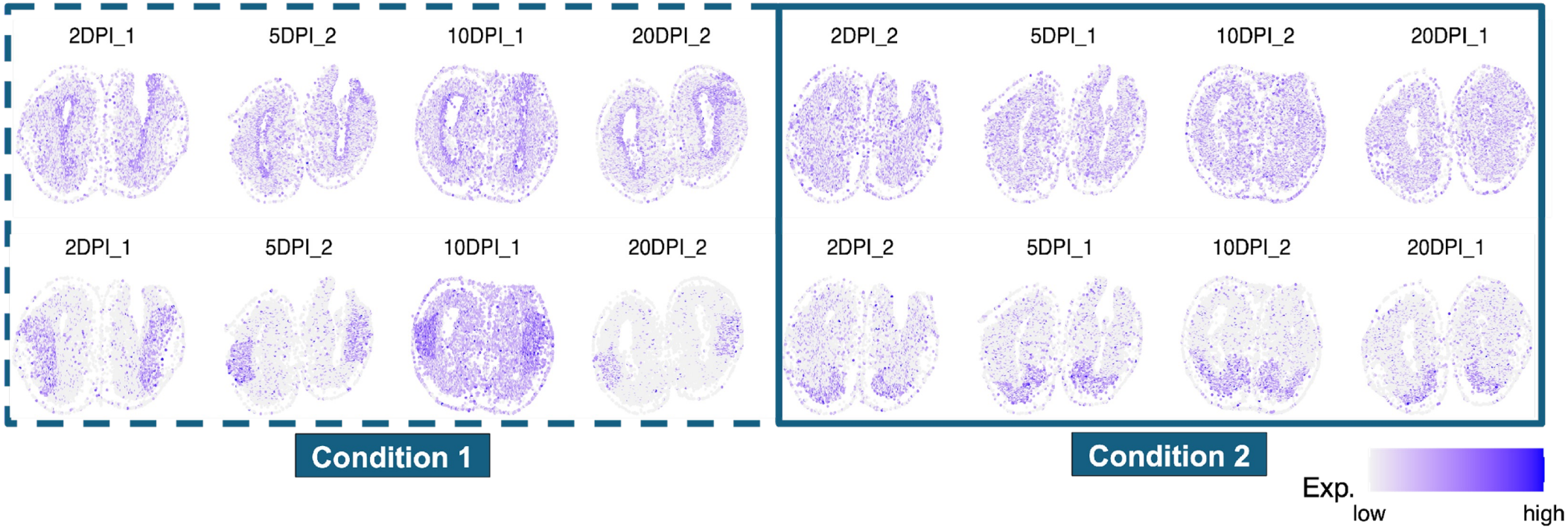
Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, Joost van den Oord, Zeynep Kalender Atak, Jasper Wouters & Stein Aerts

Nature Methods **14**, 1083–1086 (2017) | Cite this article

Differential analysis with multiple samples and multiple conditions

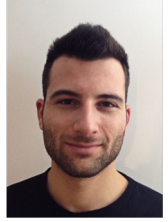


Differential spatial patterns (DSP)



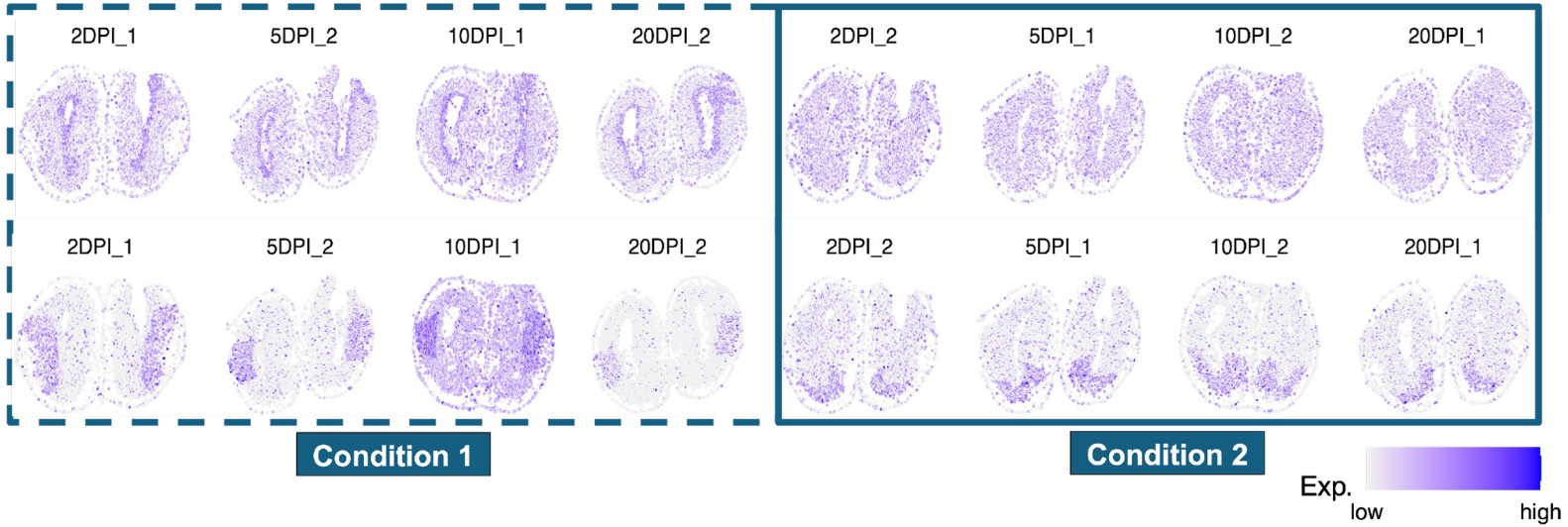
DSP genes are those whose spatial expression patterns change across groups, such as different treatment conditions or time phases.

DESpace

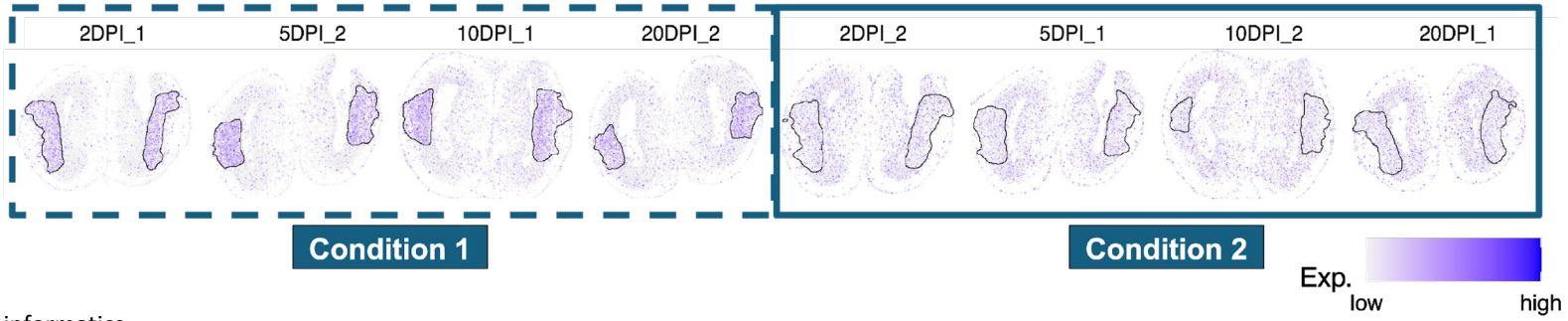



Mark Robinson Simone Tiberi

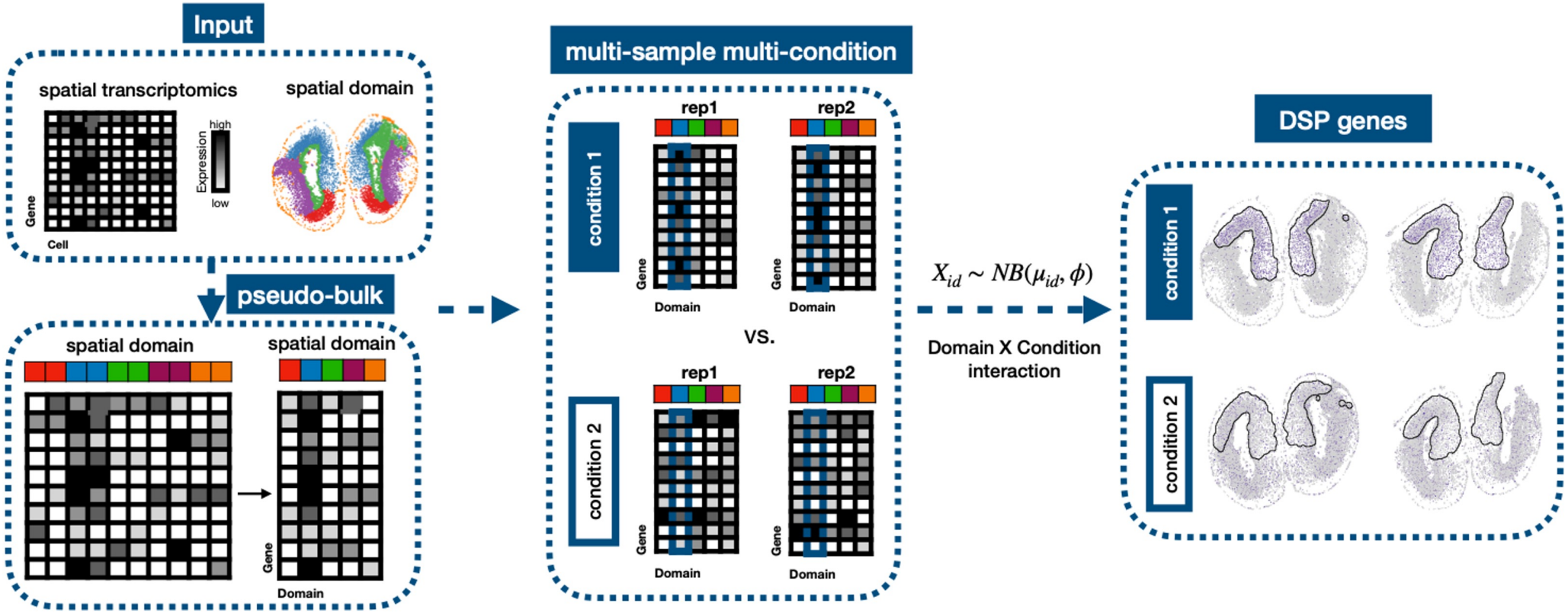
1. Global test: detect DSP genes.



2. Individual cluster test: identify the spatial clusters where gene abundance varies across conditions



Methodological details



$$
\begin{aligned}
X_{id} &\sim NB(\mu_{id}, \phi), && (1)\\
\log(\mu_{id}) &= \log(M_{id}) + \beta_d + \beta_{c_i} + \beta_{c_i d}, && (2)\\
&\text{for } i = 1, \ldots, N,\ d = 1, \ldots, D, && (3)\\
&\text{and } c_i = 1, \ldots, N_c, && (4)
\end{aligned}
$$

Methodological details

- Test for DSP via a quasi-likelihood F-test:

$$H_0 \quad : \quad \beta_{c_i d} = 0, \qquad (5)$$
$$\text{for } c_i = 1, \ldots, N_c, \text{ and } d = 1, \ldots, D;$$
$$H_1 \quad : \quad \text{otherwise.} \qquad (6)$$

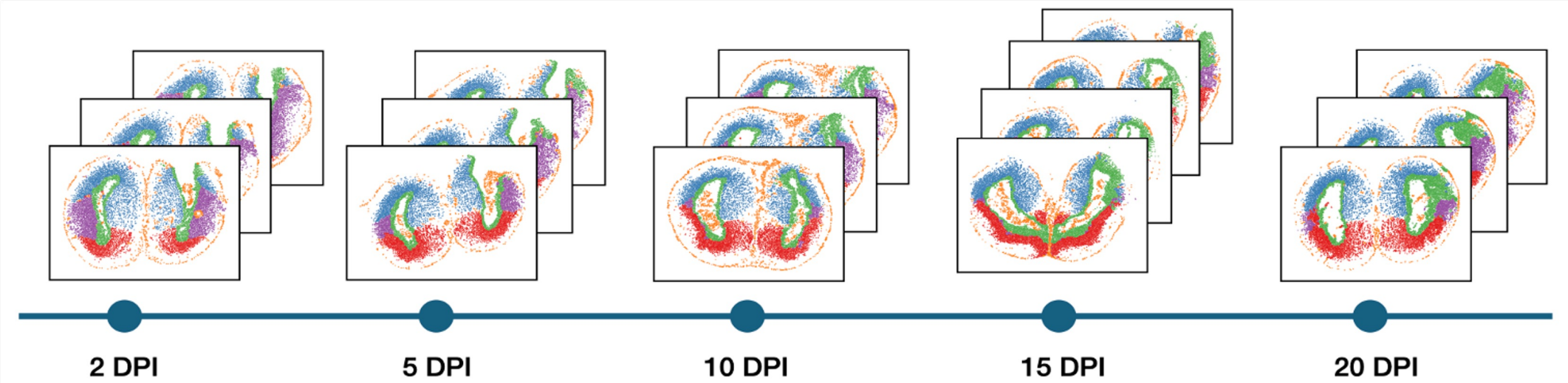
Under the null hypothesis, the cluster effect on gene expression are consistent across groups, while under the alternative hypothesis, the cluster effect varies between groups.

- To identify the **key** individual spatial domain, we test:

$$H_0 \quad : \quad \beta_{c_i d} = 0, \text{ for } c_i = 1, \ldots, N_c; \qquad (7)$$
$$H_1 \quad : \quad \text{otherwise.} \qquad (8)$$

Application to real data

- ARTISTA (Stereo-seq) dataset captures axolotl brain regeneration at single-cell resolution
- 16 samples from 5 time points, i.e., days post-injury (DPI)
- 5 consistent spatial clusters

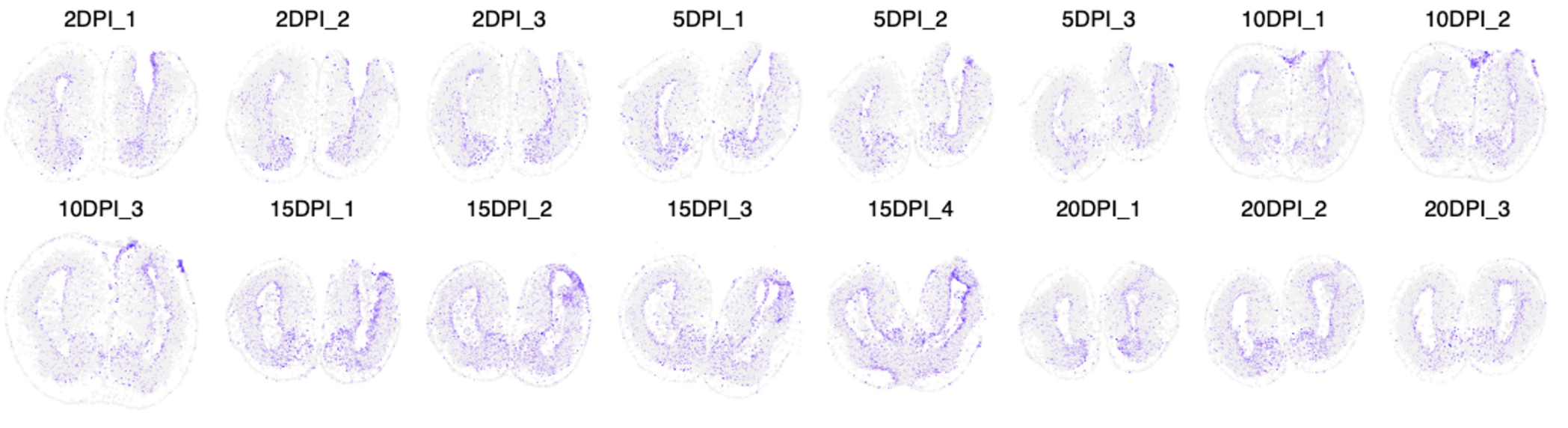


Data from Wei et al. 2022

Application to real data

DSP gene example

TNC: a glycoprotein in adult neurogenic niches, involved in tissue repair and regeneration



Take-home messages

- SVGs capture expression patterns that vary across tissue structure.
- Different SVG methods incorporate spatial information in distinct ways and could detect different types of structure.
- SVGs enable downstream analyses, such as spatial domain detection and spatial gene module identification.
- With multi-sample, multi-condition datasets, we can identify differential spatial patterns across conditions.

References

- Yan, Guanao, Shuo Harper Hua, and Jingyi Jessica Li. "Categorization of 34 computational methods to detect spatially variable genes from spatially resolved transcriptomics data." *Nature Communications* 16.1 (2025): 1141.
- Svensson, Valentine, Sarah A. Teichmann, and Oliver Stegle. "SpatialDE: identification of spatially variable genes." *Nature methods* 15.5 (2018): 343-346.
- Weber, Lukas M., et al. "nnSVG for the scalable identification of spatially variable genes using nearest-neighbor Gaussian processes." *Nature communications* 14.1 (2023): 4059.
- Zhu, Jiaqiang, Shiquan Sun, and Xiang Zhou. "SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies." *Genome biology* 22.1 (2021): 184.
- Cai, Peiying, Mark D. Robinson, and Simone Tiberi. "DESpace: spatially variable gene detection via differential expression testing of spatial clusters." *Bioinformatics* 40.2 (2024): btae027.
- Cable, Dylan M., et al. "Cell type-specific inference of differential expression in spatial transcriptomics." *Nature methods* 19.9 (2022): 1076-1087.
- Aibar, Sara, et al. "SCENIC: single-cell regulatory network inference and clustering." *Nature methods* 14.11 (2017): 1083-1086.
- Cai, Peiying, Mark D. Robinson, and Simone Tiberi. "DESpace2: detection of differential spatial patterns in spatial omics data." *bioRxiv* (2025): 2025-06.
- Wei, Xiaoyu, et al. "Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration." *Science* 377.6610 (2022): eabp9444.