



# From latent spaces to living systems: Lecture 3 & 4

# Dynamic tissue cartography

# Maps and Census as ways to record data and tell stories

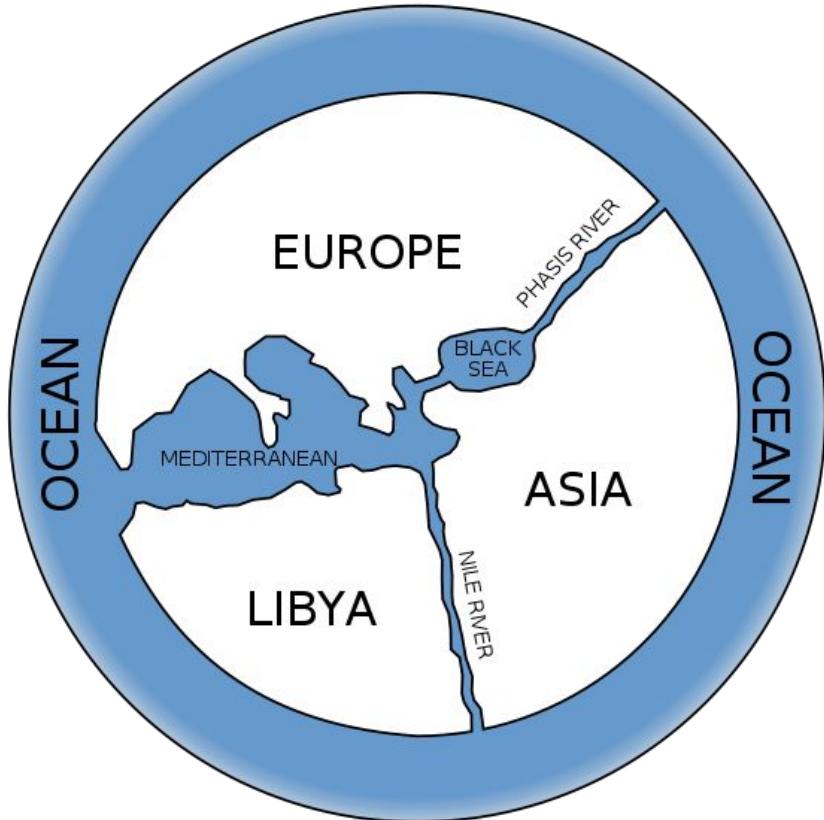


**Year Created:** c.17,000 BCE

**Country of Origin:** near Montignac, France, **Creator:** Unknown

**Area Depicted:** Area around Abauntz Lamizulo cave and animals such as red deer and ibex

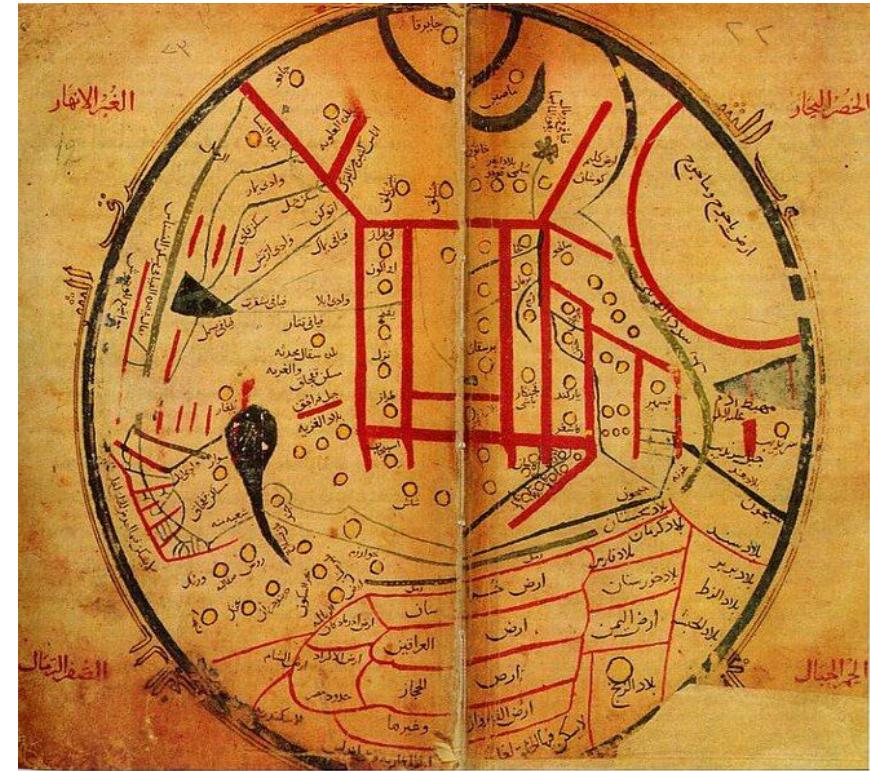
# Maps can tell many different stories



Year Created: c.610 – 546 BCE

Country of Origin: Ancient Greek city of Miletus

Creator: Anaximander



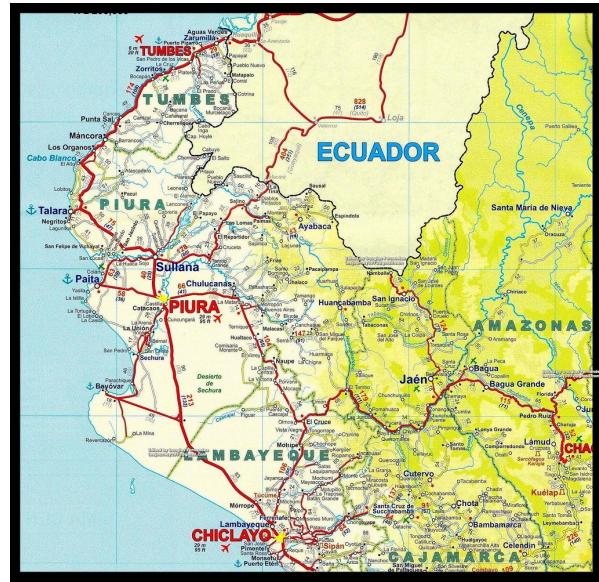
Kara-Khanid scholar Mahmud al-Kashgari compiled a *Dīwān Lughāt al-Turk* (*Compendium of the languages of the Turks*) in the 11th century. Source: Wikipedia

# Census records tell complementary stories



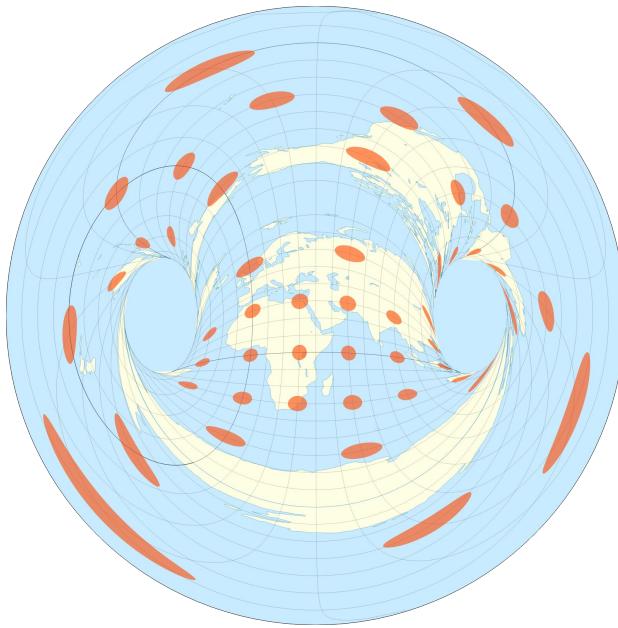
Quipus from Highlands of southern Peru  
Imperial Epoch (1300 d.C. – 1532 d.C.) Museo Larco

# Maps and Censuses



- Usually depict elements that are **known** (identifiable) with the appropriate scales and legends (rivers, factual information about number of cattle)
- Information can be spatial (map), temporal (census) or both (climate data)

# Maps and Censuses



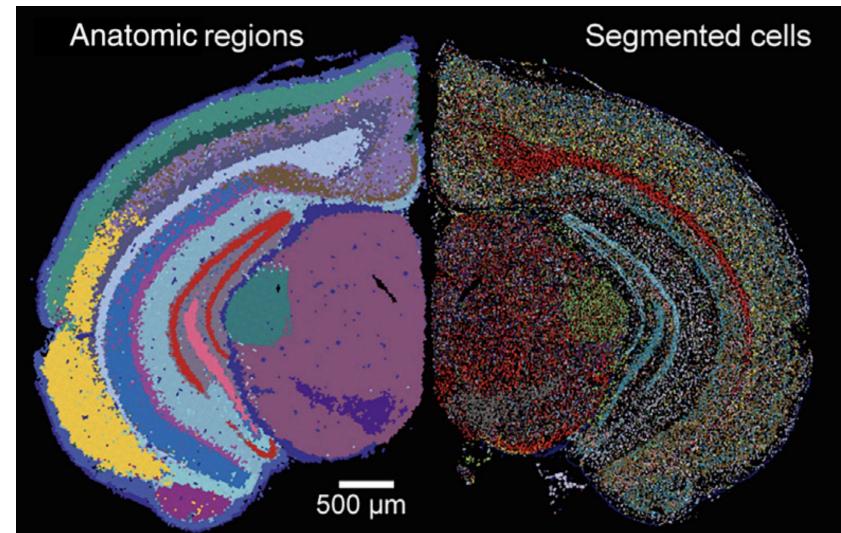
- Reflect a particular point of view and mislead (e.g. polling numbers)
- Subject to bias and error
- Require clear knowledge of the elements presented
- Need to be continuously updated to reflect changes

Figure: Hammer retroazimuthal projection.  
Source: Wikipedia

# Astronomy and Biology give us new Maps



Hubble NASA

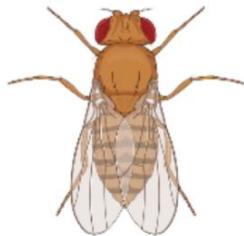


Mouse organogenesis spatiotemporal transcriptomic atlas. (Chen et al., 2022)

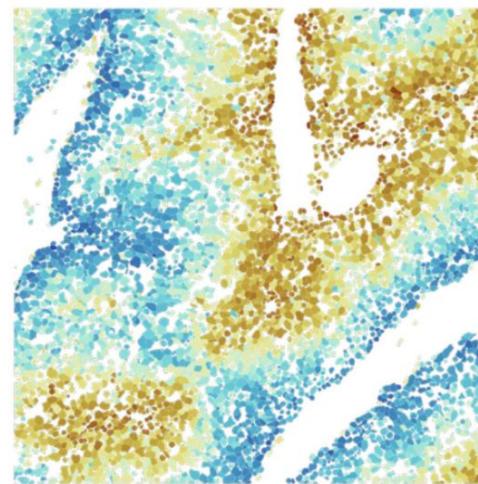
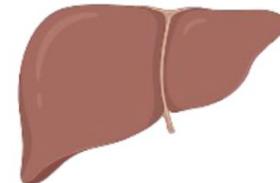
# To make Maps that represent Living Systems we need two answer at least two Questions

- **What are the units?**  
If we do not yet know the key components or players, how can we discover and define them?
- **How are they organized and how do they interact?**  
Given the players, what are the rules and patterns that govern their interactions **across spatial and temporal scales?**

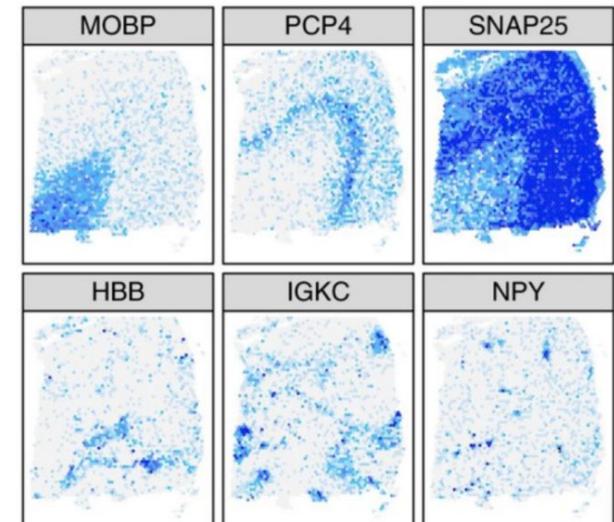
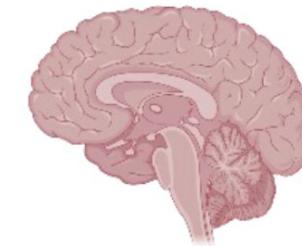
# Spatial transcriptomics: mapping living systems



BDGP



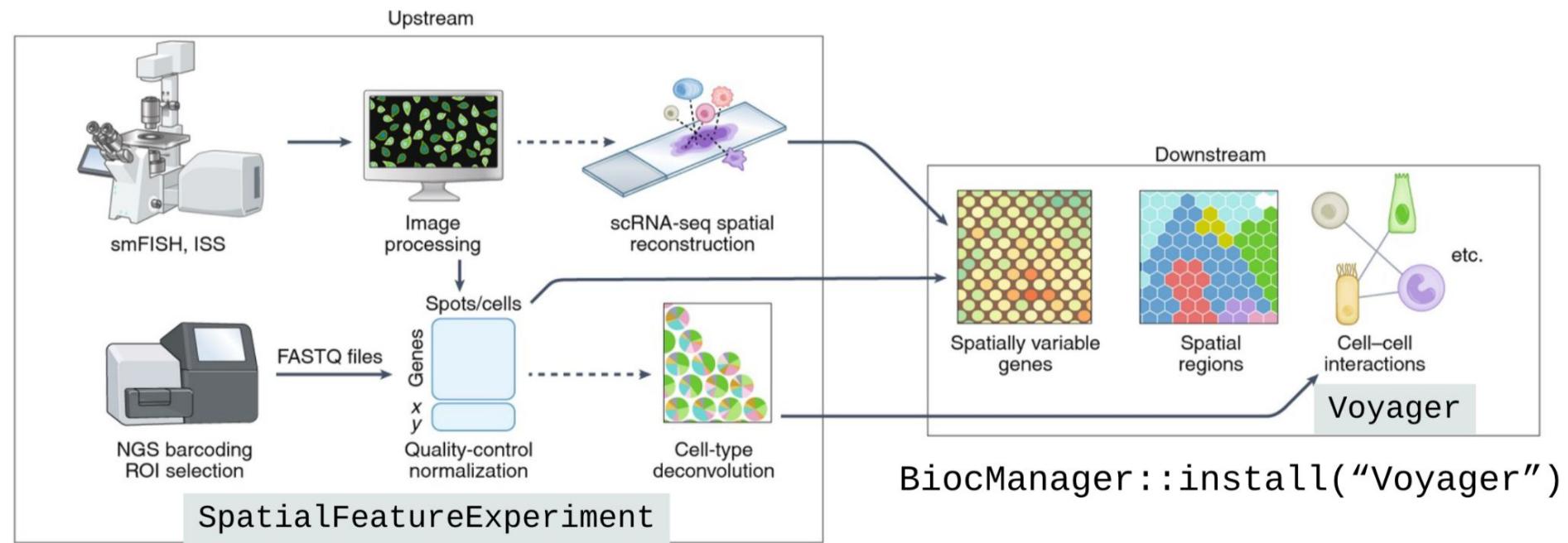
Voyager website



Maynard 2021 DLPFC

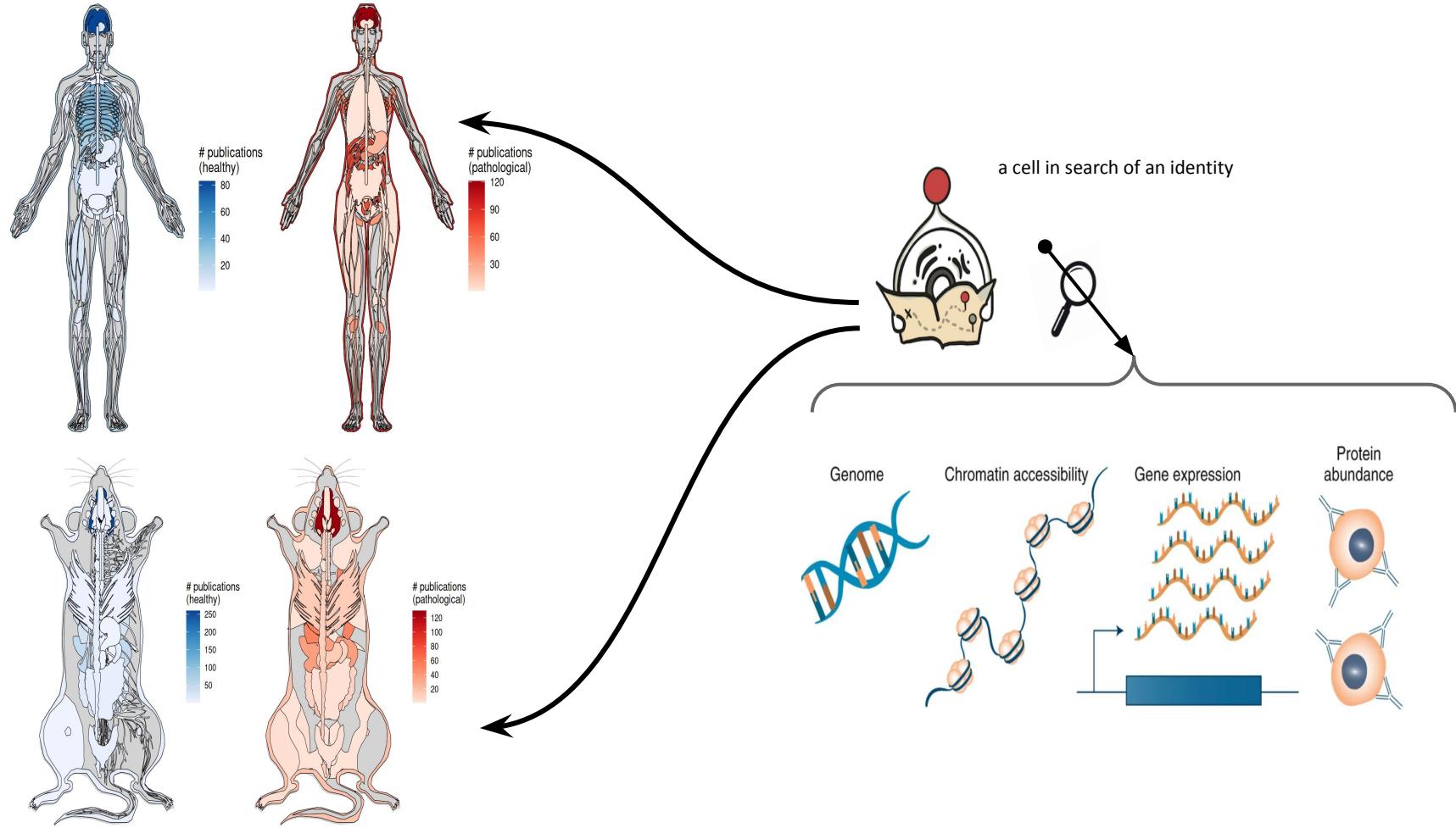
Figure: Lambda Moses

# Spatial transcriptomics: mapping living systems

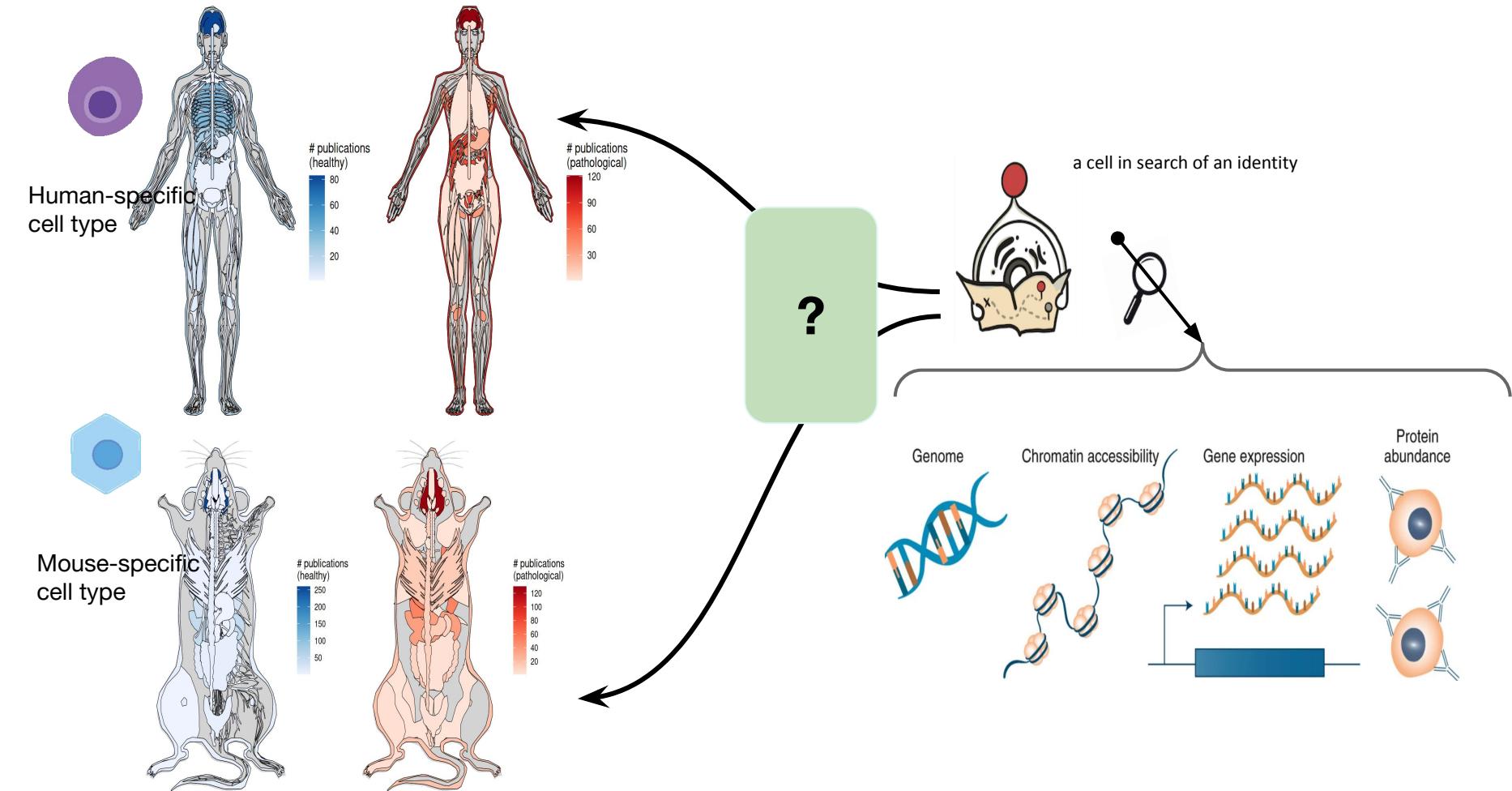


<https://pachterlab.github.io/voyager/>

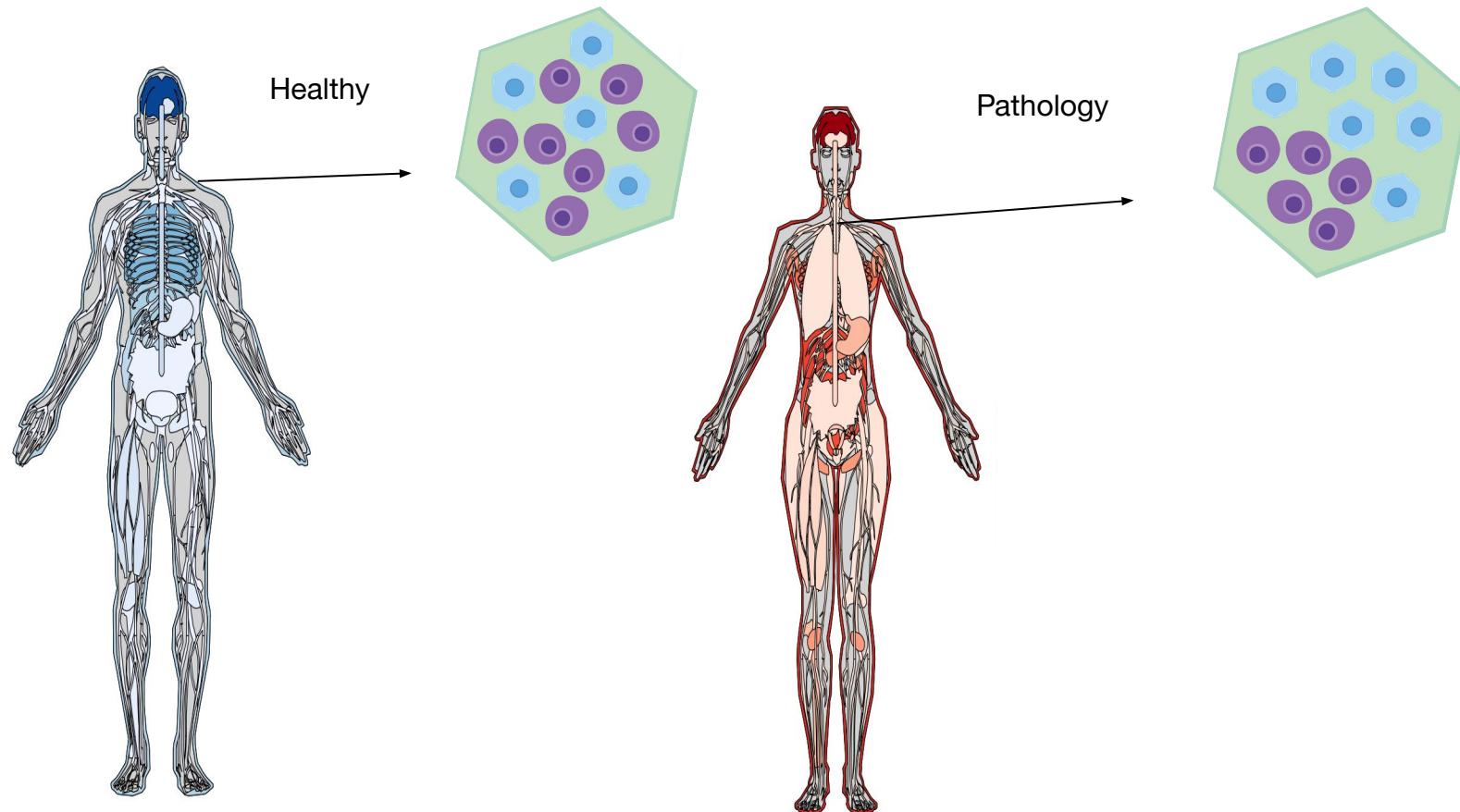
# Cellular heterogeneity across scales underlies healthy development and disease



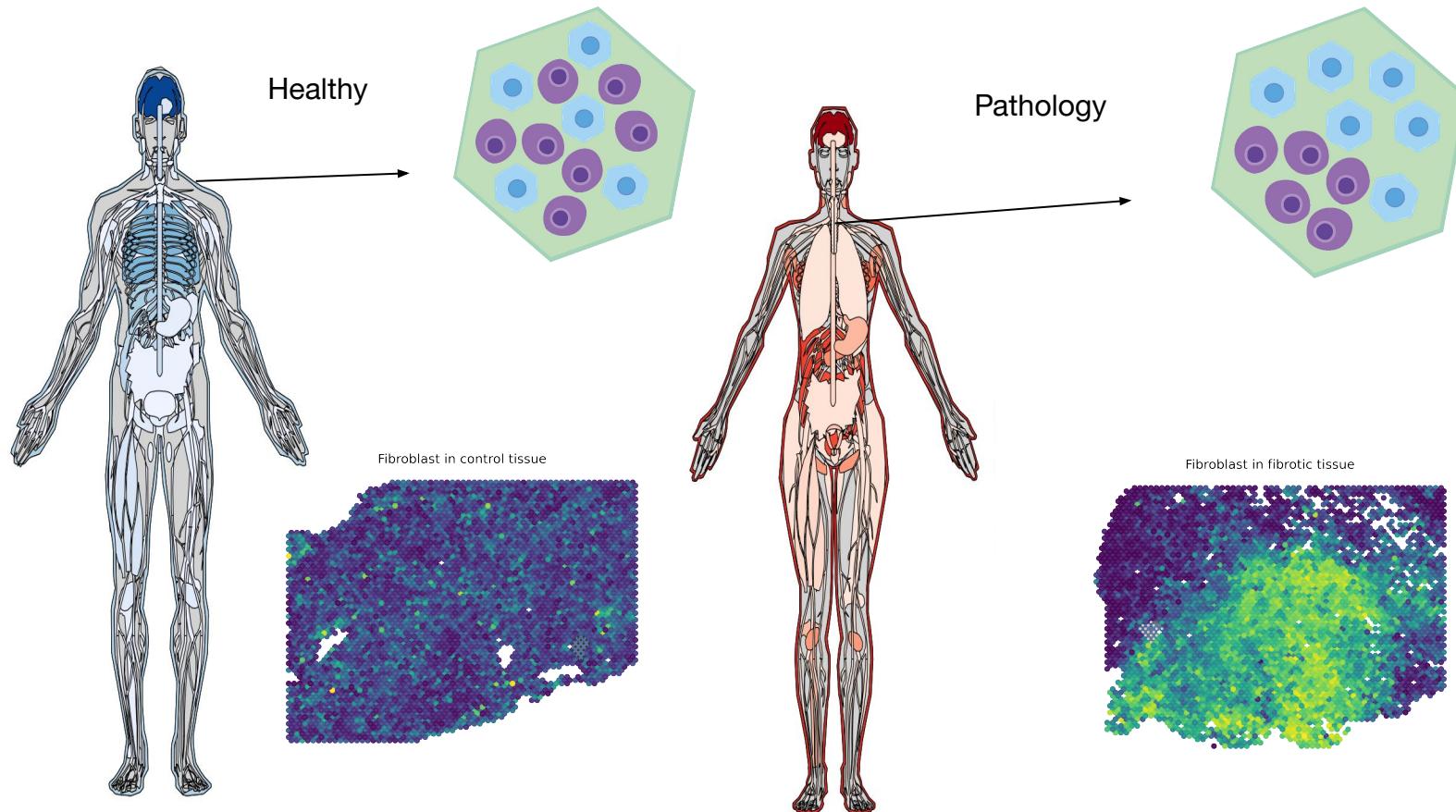
# Cellular heterogeneity across scales underlies healthy development and disease



# Spatial contexts can drive differences in health and disease contexts

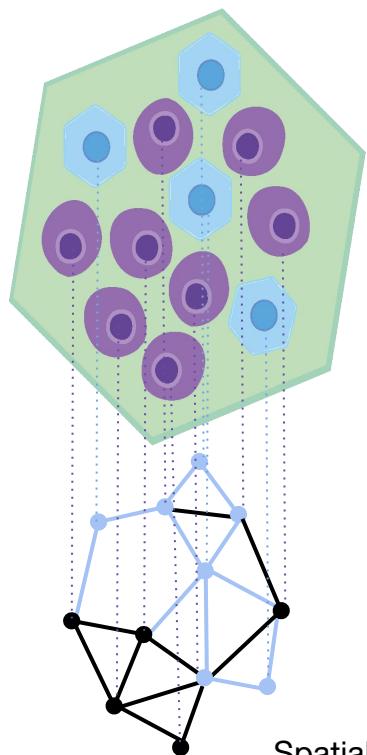


# Cellular heterogeneity across scales underlies healthy development and disease

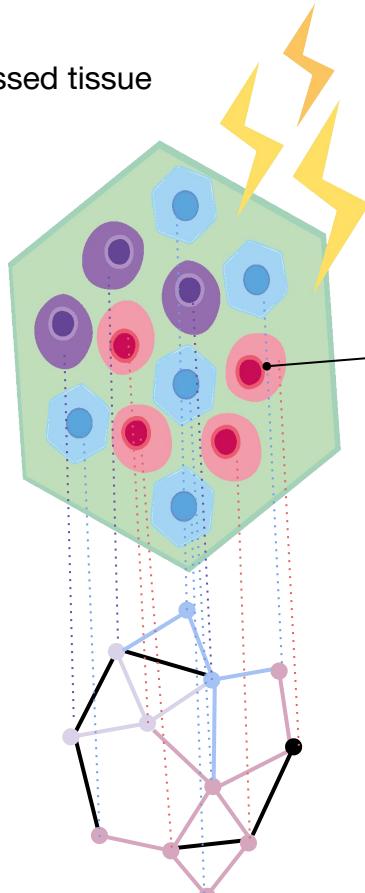


## Quantifying spatial differences

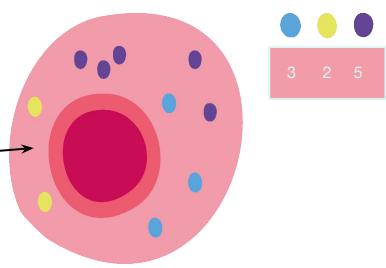
Healthy tissue



Stressed tissue



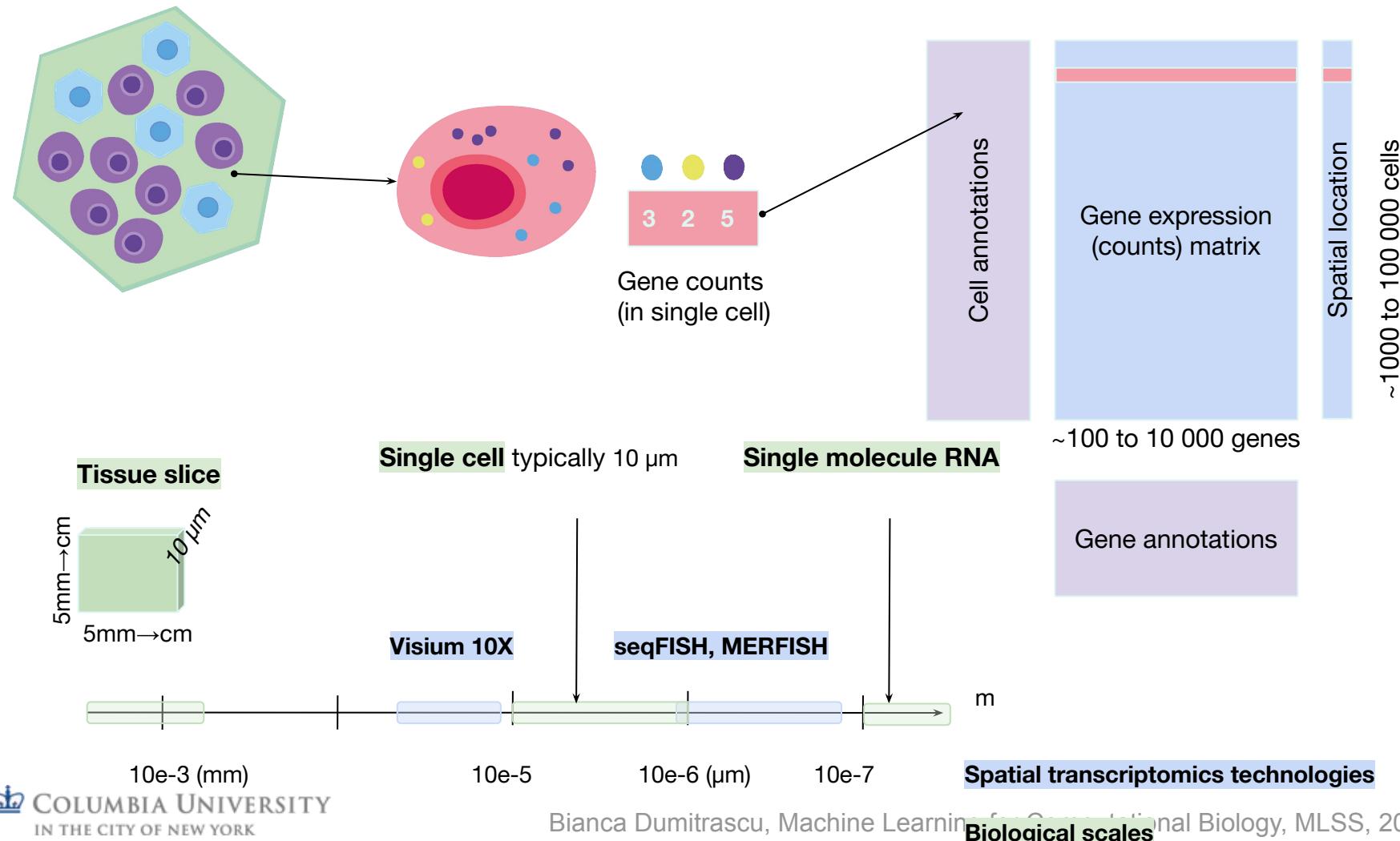
Cell  $i$  expressing genes



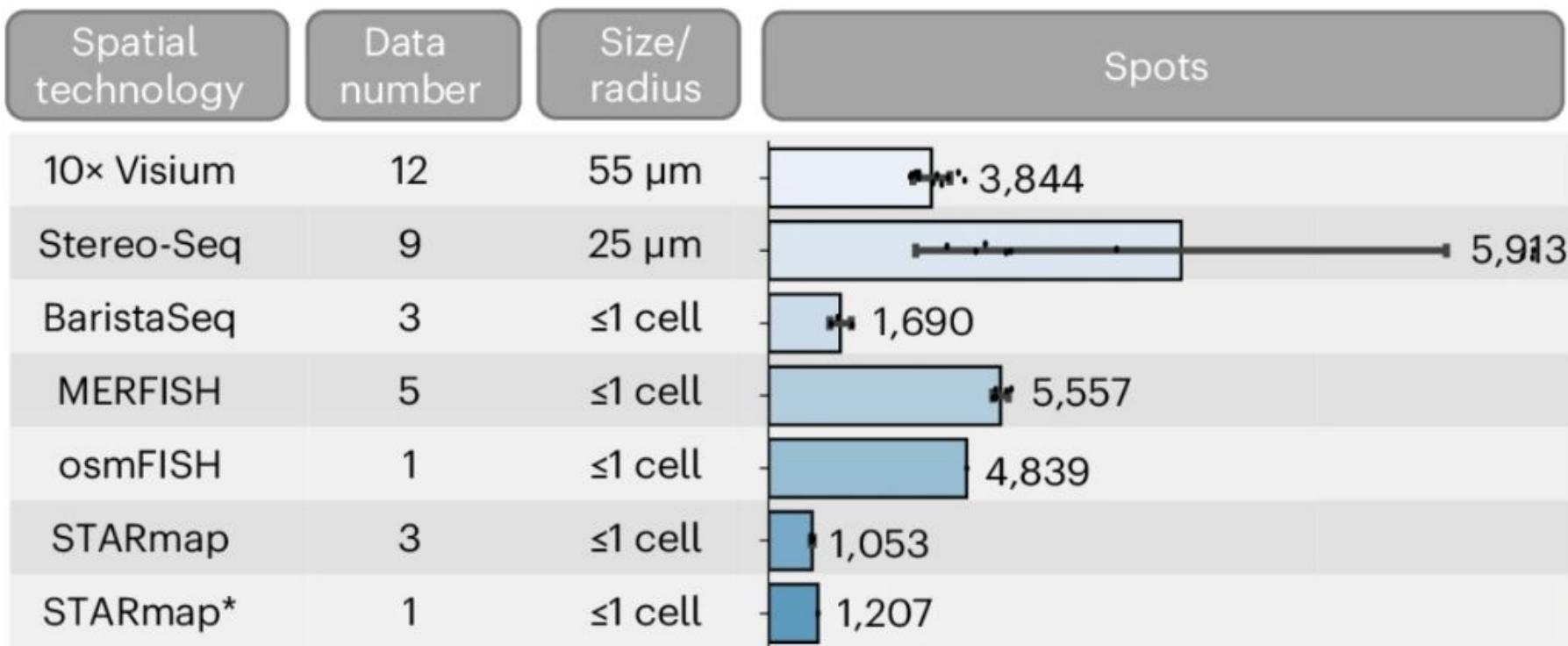
Spatial graph for each slice

# Cellular heterogeneity: a multiscale problem

**Scales:** Tissue → cellular → subcellular → molecular

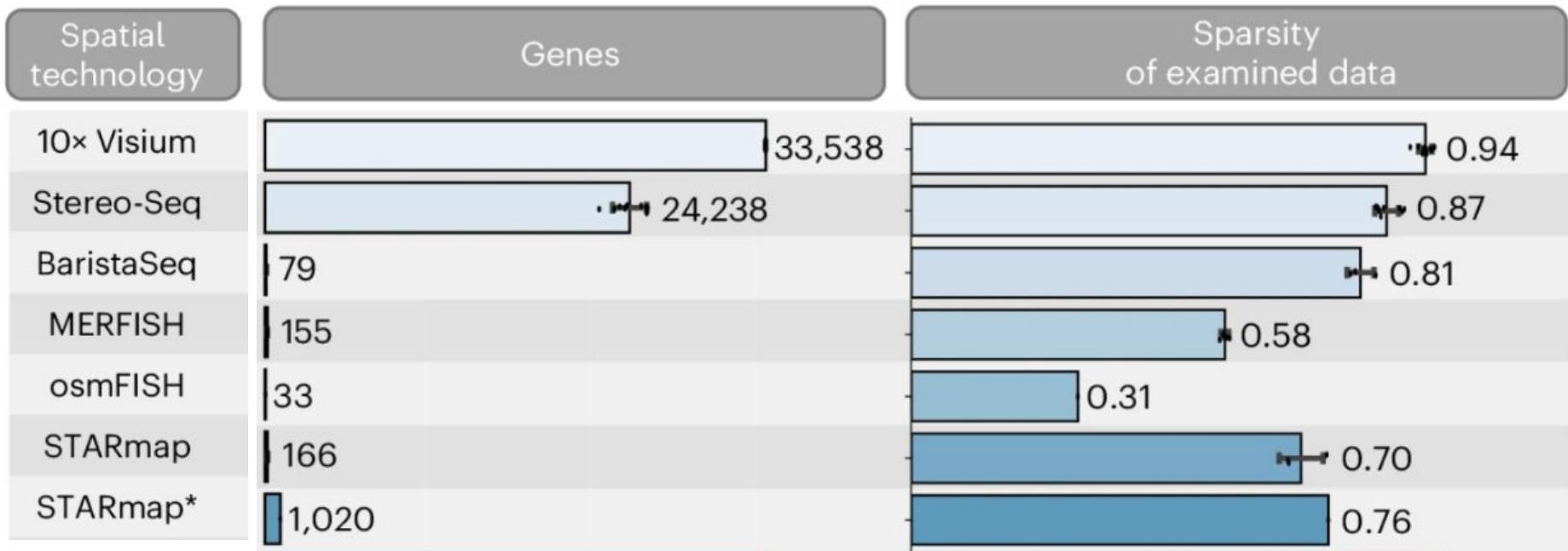


# Cellular heterogeneity: a multiscale problem with many tools to query it



From: Yuan, Z., Zhao, F., Lin, S. et al.  
Benchmarking spatial clustering methods with  
spatially resolved transcriptomics data. 2024

# Cellular heterogeneity: a multiscale problem with many tools to query it

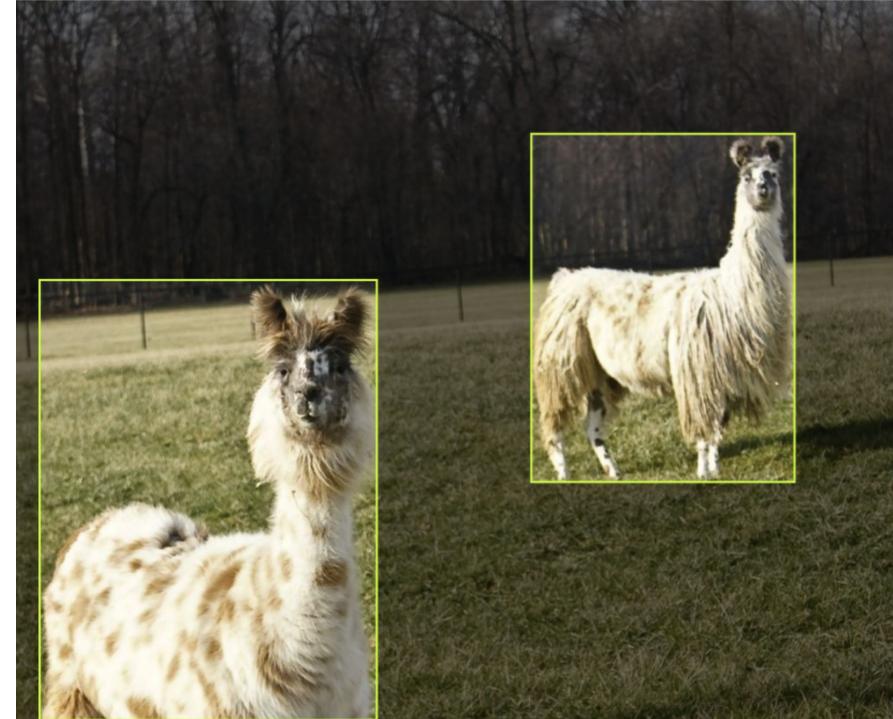


From: Yuan, Z., Zhao, F., Lin, S. et al.  
Benchmarking spatial clustering methods with  
spatially resolved transcriptomics data. 2024

# To make Maps that represent Living Systems we need two answer at least two Questions

- **What are the units? (This lecture: Spatial Patterns)**  
If we do not yet know the key components or players, how can we discover and define them?
- **How are they organized and how do they interact?**  
Given the players, what are the rules and patterns that govern their interactions **across spatial and temporal scales?**

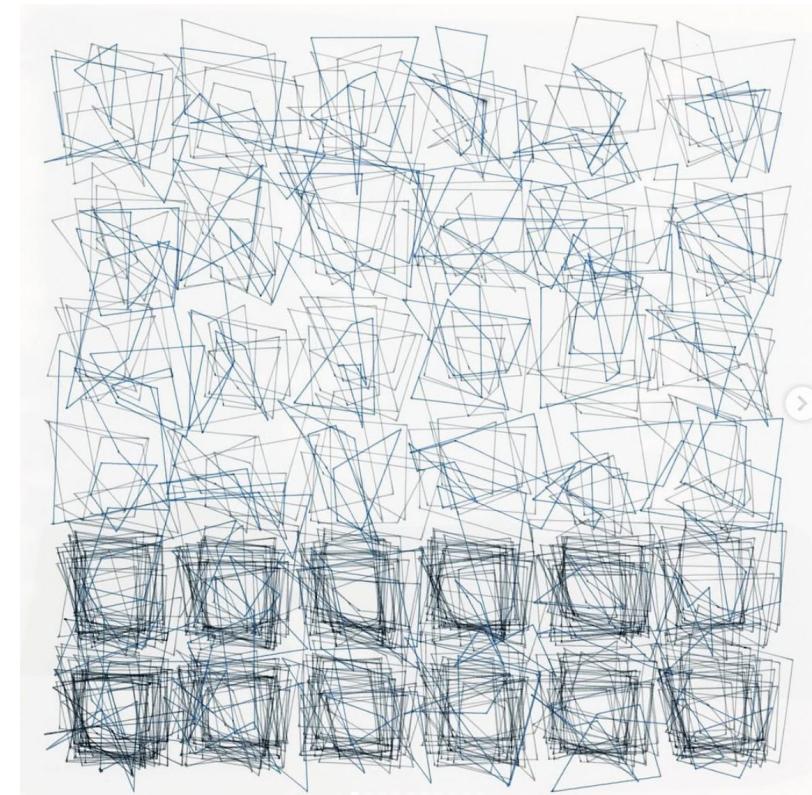
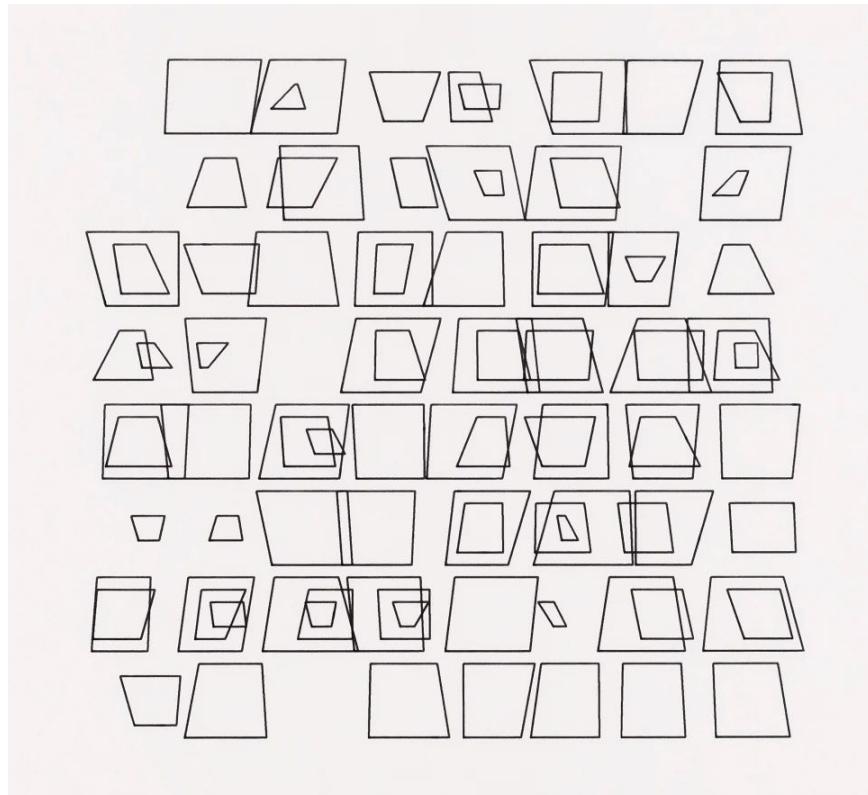
# Broader goal: quantify spatial patterns!



# Con: Spatial Patterns can be difficult to define



# Patterns can vary in space



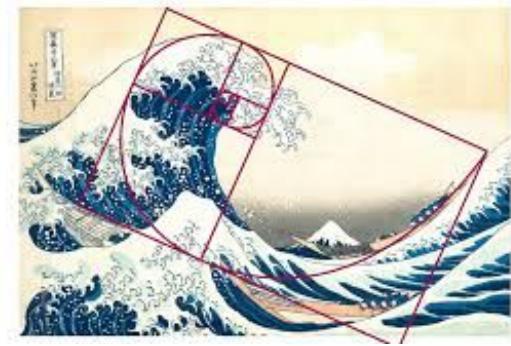
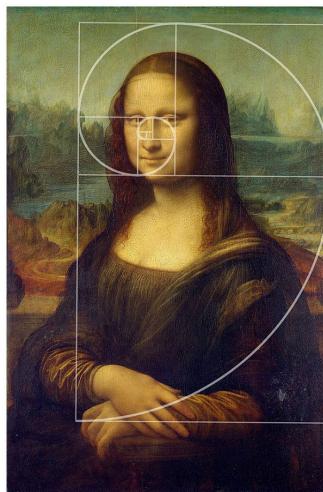
Vera Molnar, *Génèse du Trapèze*, 1974: art between the three cons: Conceptualists, Constructivists, and computers

# Patterns can segment space



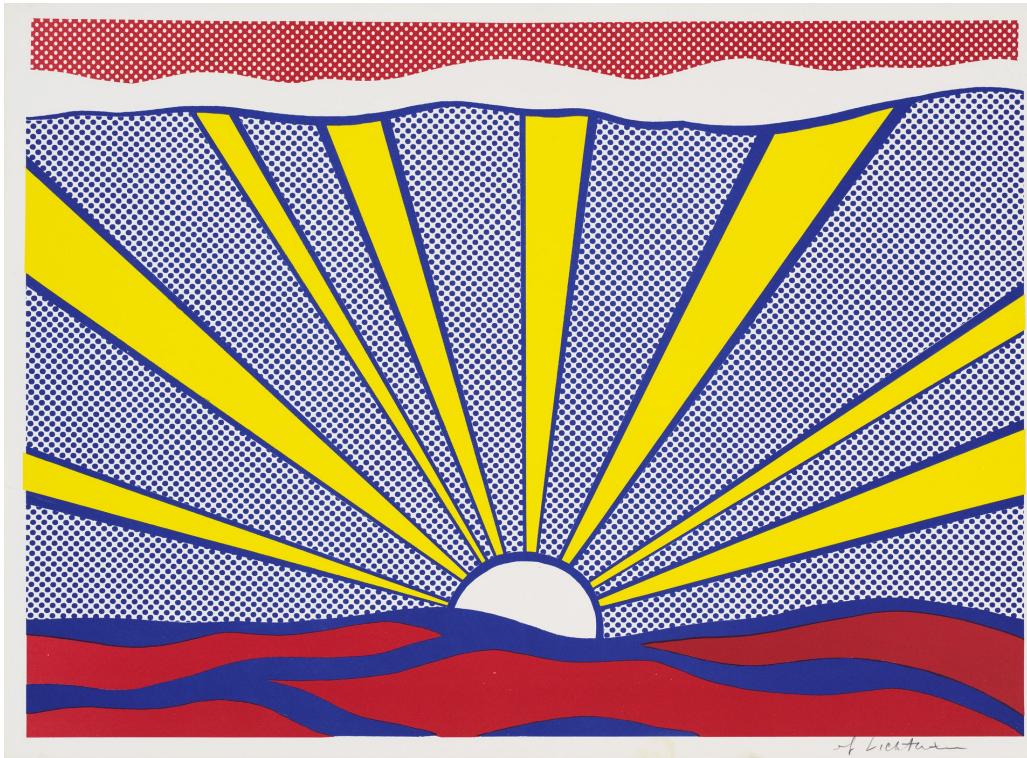
Copyright: (c) The estate of Sol LeWitt / Photo (c) Tate

# Patterns can be latent



Golden ratios

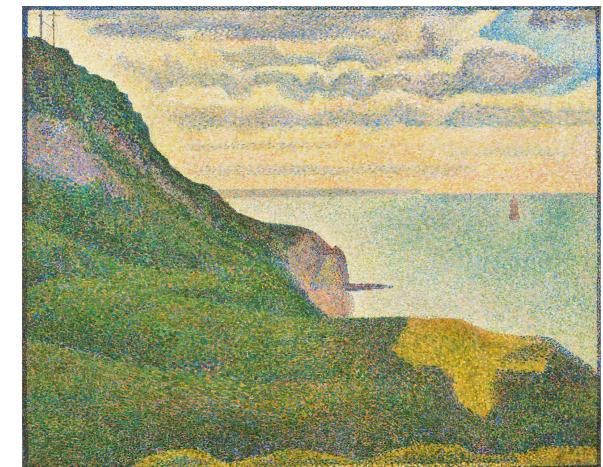
# Patterns can make sense semantically



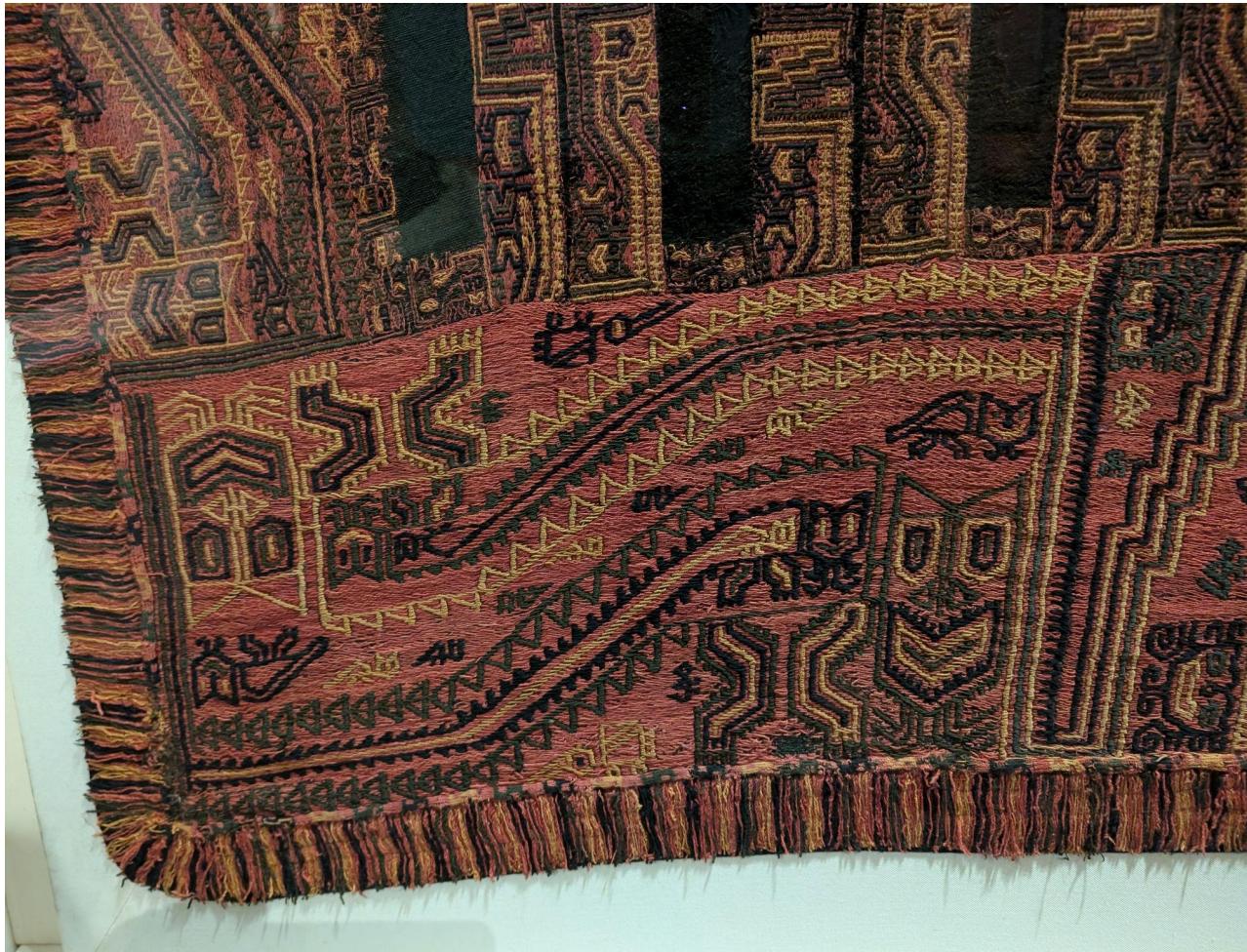
Sunrise (1965). Roy Lichtenstein.  
Whitney Museum of American Art, New York;



George Seurat cc 1884. Top: Farm people,  
Guggenheim Museum

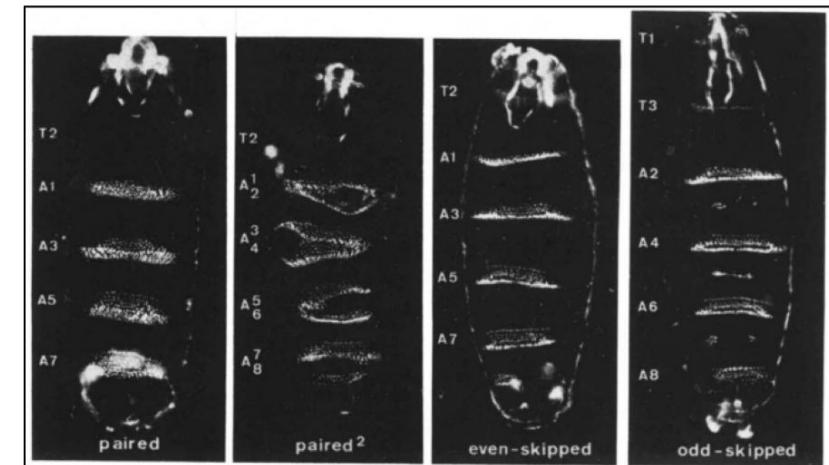


# Patterns can be present at different scales

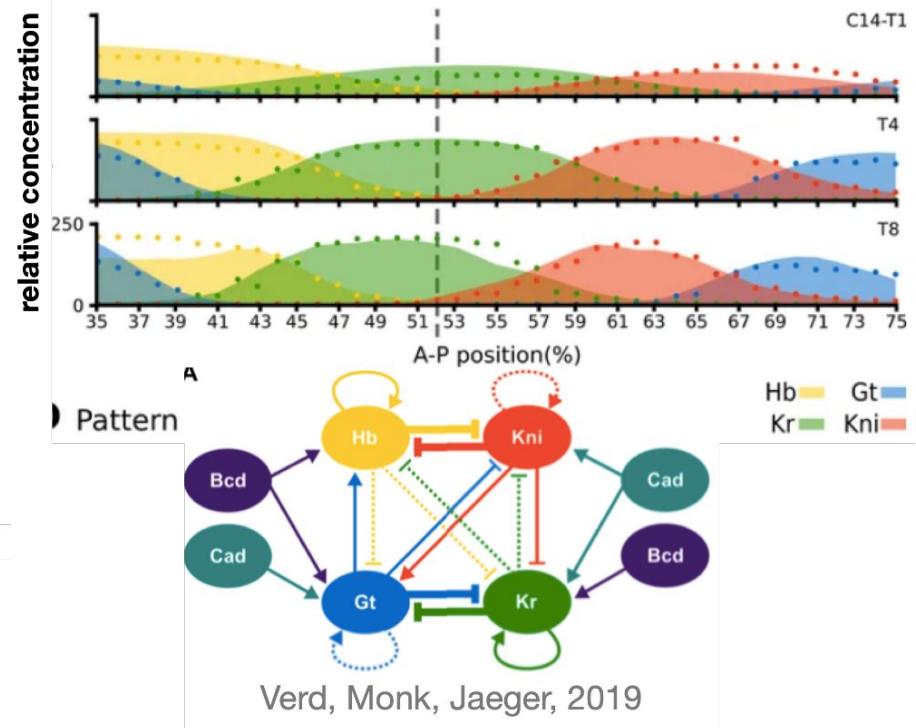


Paracas mantles. Museo Larco

# Biological Patterns: Fruit Fly Stripes

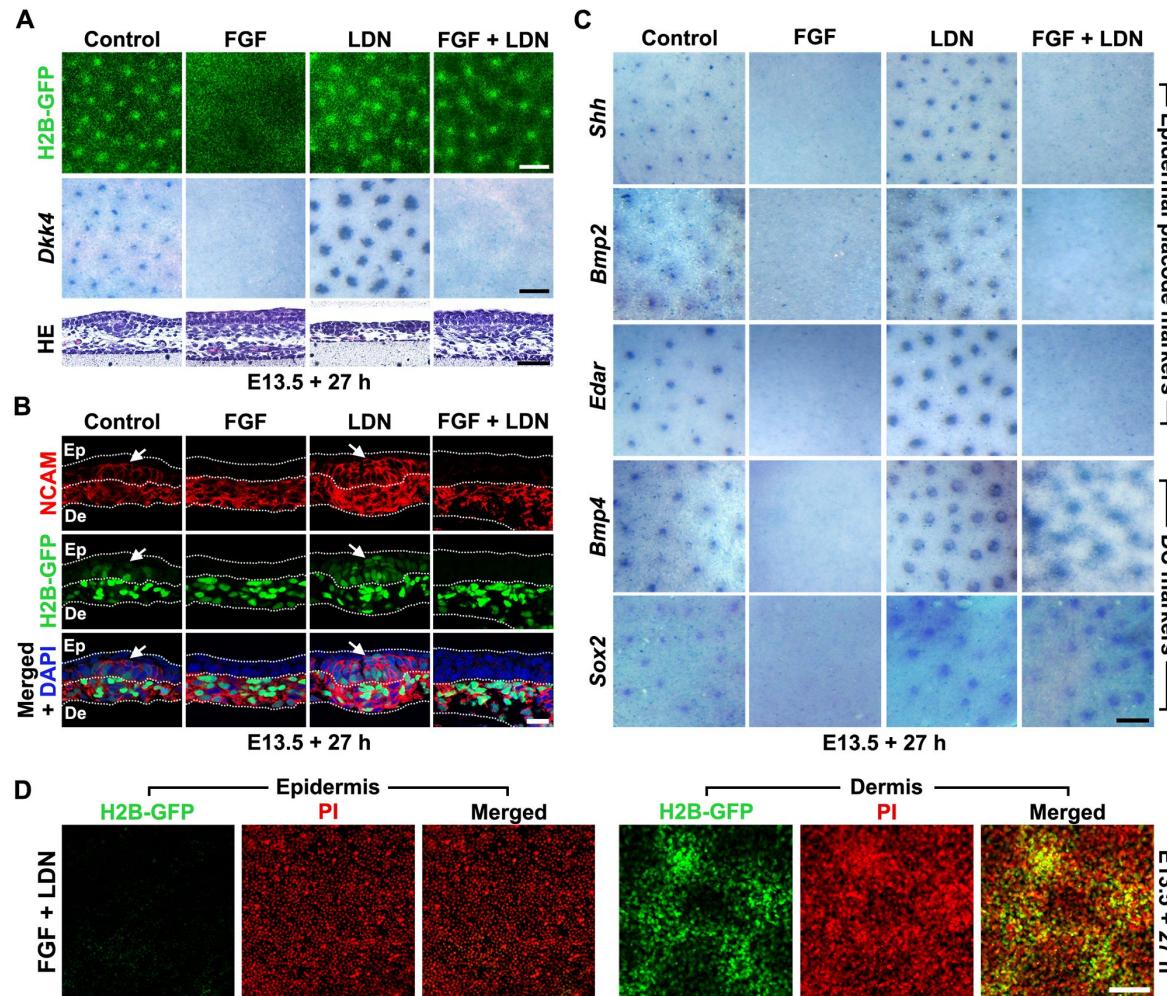


"Mutations affecting segment number and polarity in *Drosophila*"  
by C. Nüsslein-Volhard and E. Wieschaus, 1980



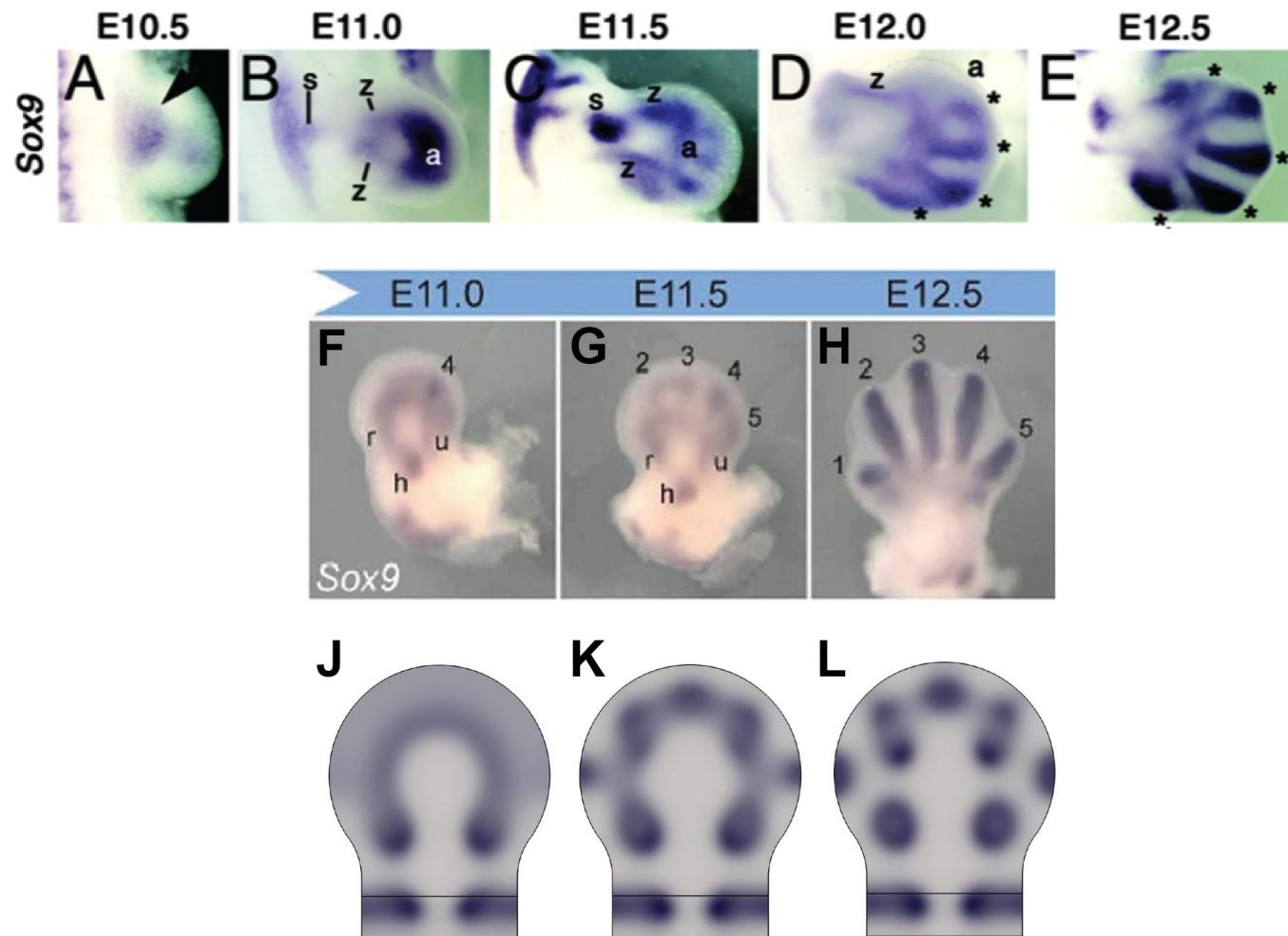
Verd, Monk, Jaeger, 2019

# Biological Patterns: Hair Follicles



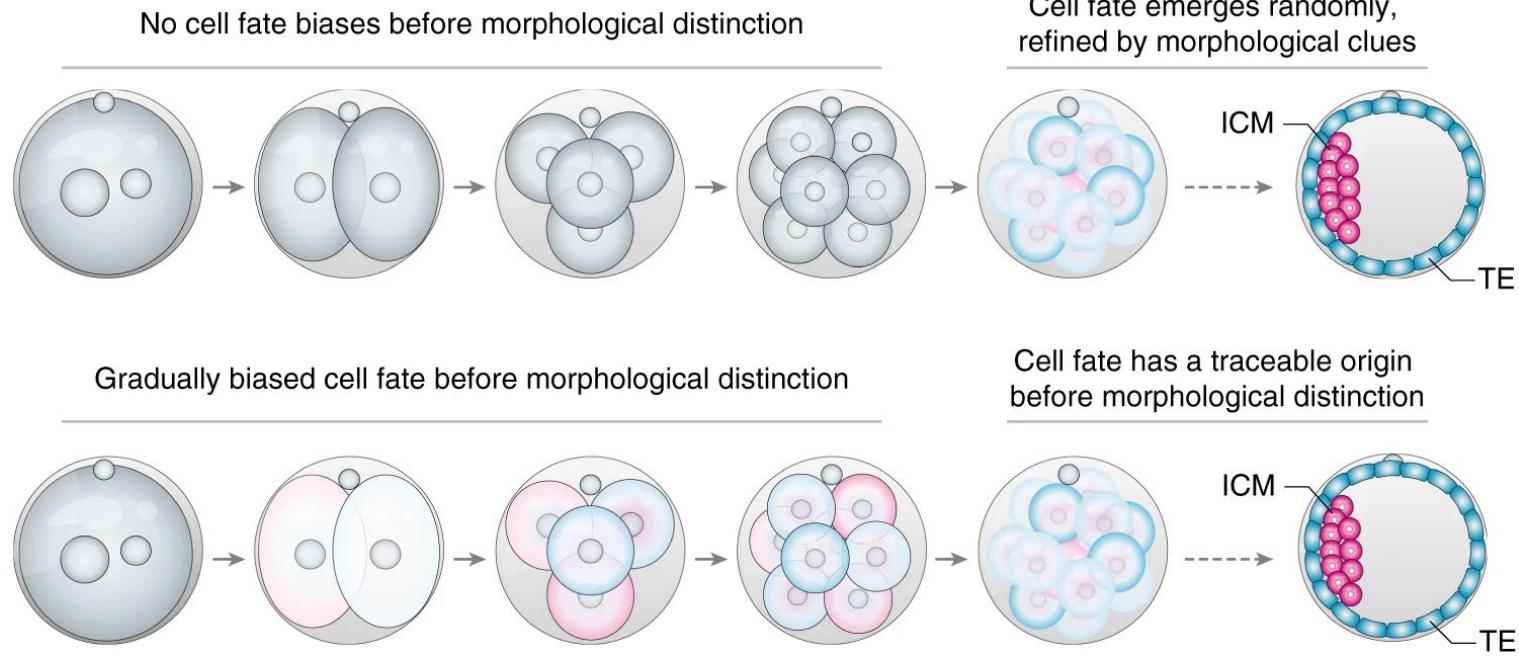
From: Hierarchical patterning modes orchestrate hair follicle morphogenesis. Glover et al. (2017)

# Biological Patterns: Digit Formation



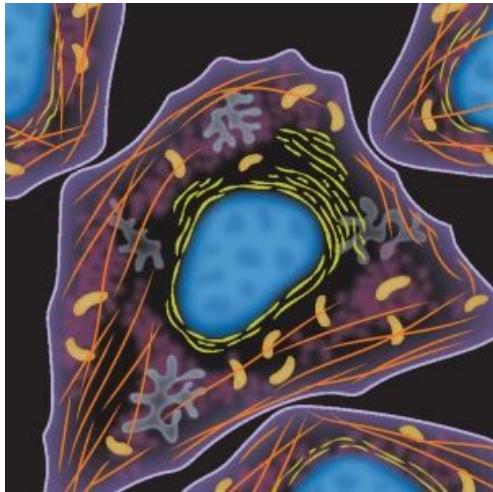
From: Digit patterning during limb development as a result of the BMP-receptor interaction. Badugu et al (2012)

# Biological Patterns: Early Embryo Symmetry Breaking

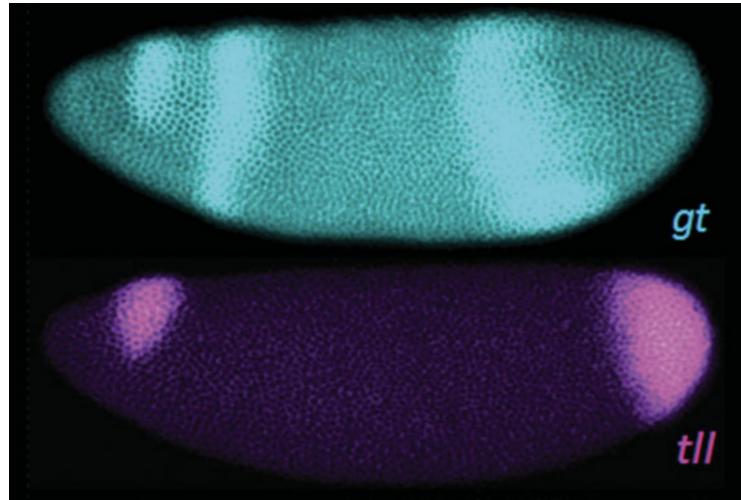


Adapted from: Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo. Chen et al. (2018)

# Examples of Spatial Patterns at the Single Cell Level



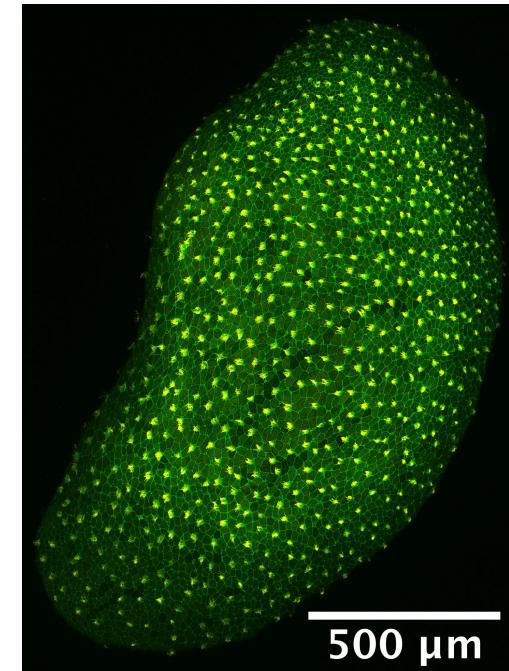
An artist's rendition of subcellular imaging. Credit: Debbie Maizels



Credit: S. Shvartzman & R. Baker

proteins are spatially distributed at subcellular level

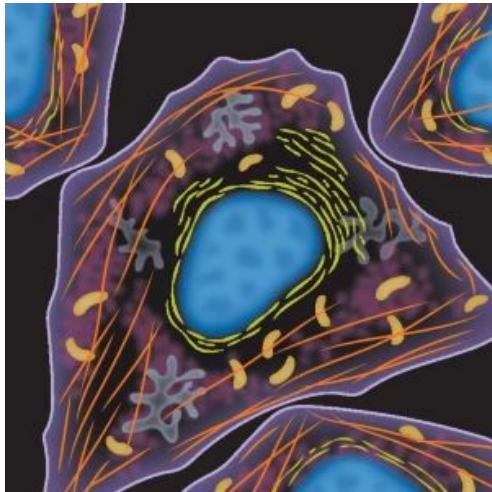
genes or gene programs can vary spatially, in a cell type dependent or cell type independent manner



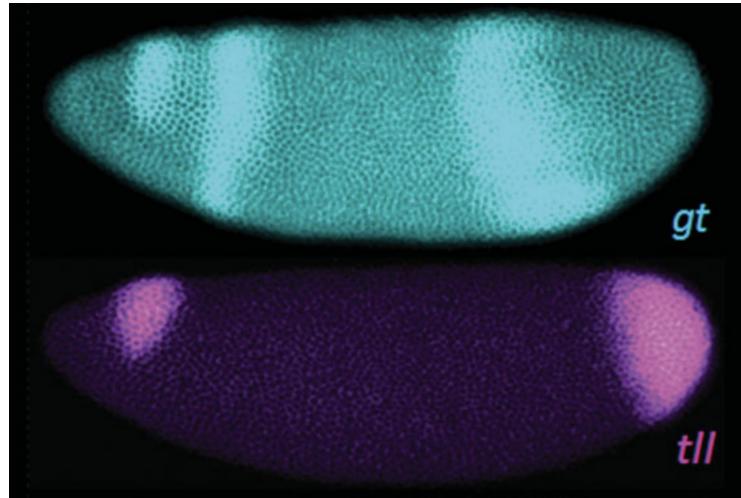
Credit: Frog embryos from Jakub Sedzinski

cell types themselves can form spatial units

# Examples of Spatial Patterns at the Single Cell Level



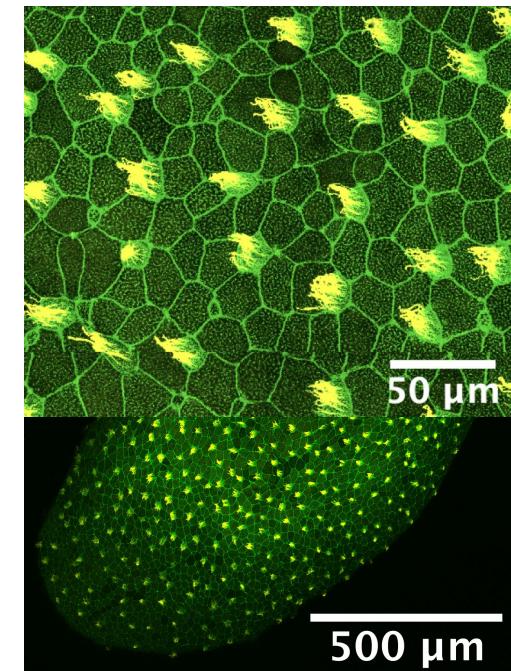
An artist's rendition of subcellular imaging. Credit: Debbie Maizels



Credit: S. Shvartzman & R. Baker

proteins are spatially distributed at subcellular level

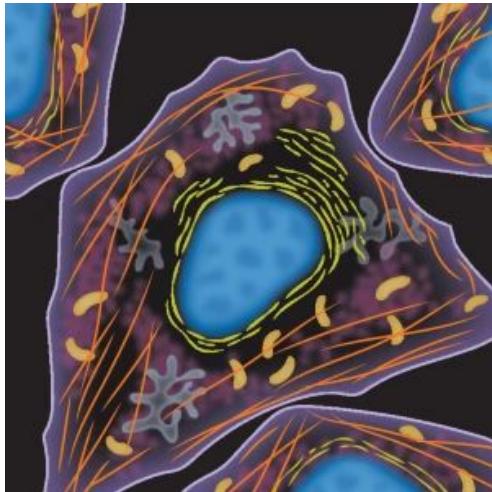
genes or gene programs can vary spatially, in a cell type dependent or cell type independent manner



Credit: Frog embryos from Jakub Sedzinski

cell types themselves can form spatial units

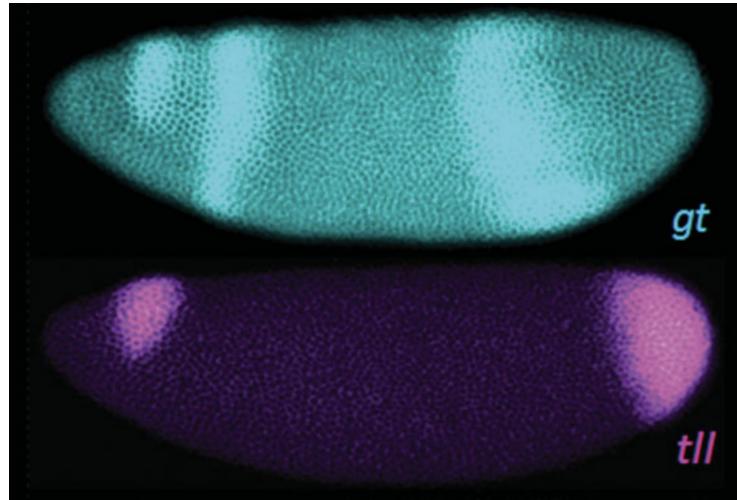
# Examples of Spatial Patterns at the Single Cell Level



An artist's rendition of subcellular imaging. Credit: Debbie Maizels

## Testing for randomness of spatial events

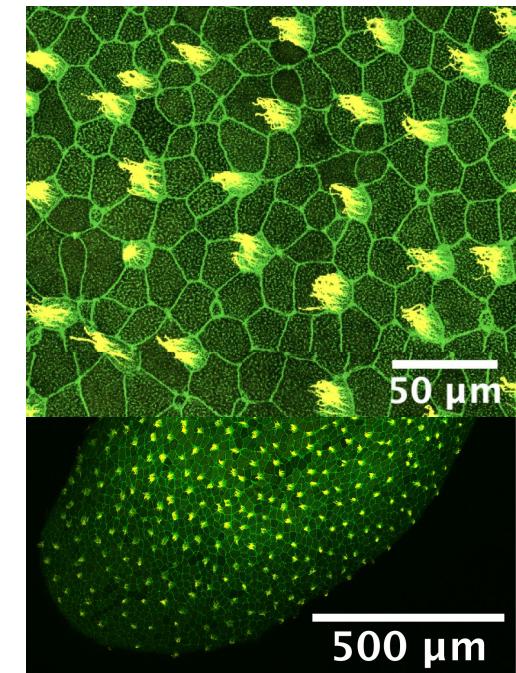
proteins are spatially distributed at subcellular level



Credit: S. Shvartzman & R. Baker

## Spatial Variable Gene Detection

genes or gene programs can vary spatially, in a cell type dependent or cell type independent manner



Credit: Frog embryos from Jakub Sedzinski

## Spatial Clustering

cell types themselves can form spatial units

# Computational tools for spatial data analysis

## Areal models:

- Moran's I
- Conditional Autoregressive Models

## In context

- (gene) SpatialIDE, nnSVG
- (gene programs )Spatial NMF

## GNNS

- GraphST

## Testing for randomness of spatial events

proteins are spatially distributed at subcellular level

## Spatial Variable Gene Detection

genes or gene programs can vary spatially, in a cell type dependent or cell type independent manner

## Spatial Clustering

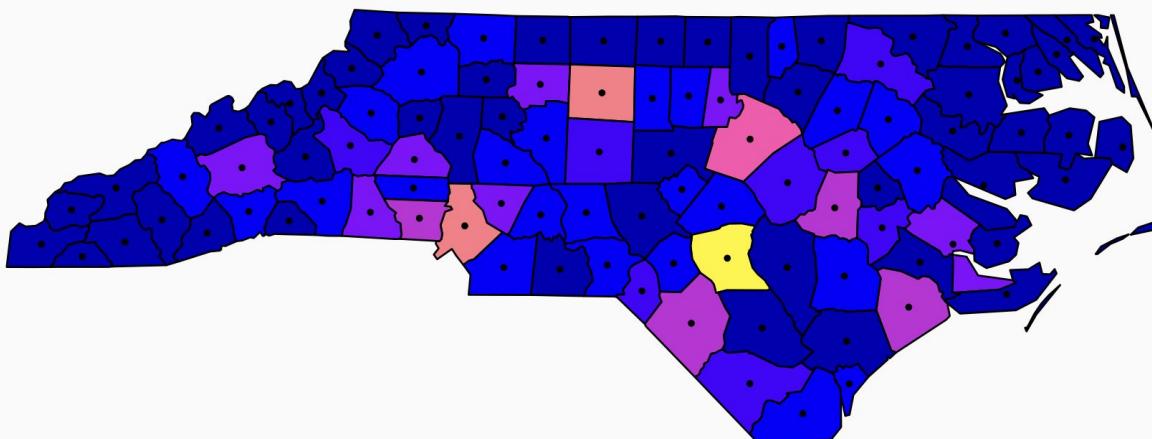
cell types themselves can form spatial units

# Modelling Spatial Correlation with Areal Models

## Areal models:

- Moran's I
- Conditional Autoregressive Models

# Modelling Spatial Correlation with Areal Models















































In plain English:  
**Encoding all the spatial marginals is expensive**, we want to define a model that uses **local information!**













































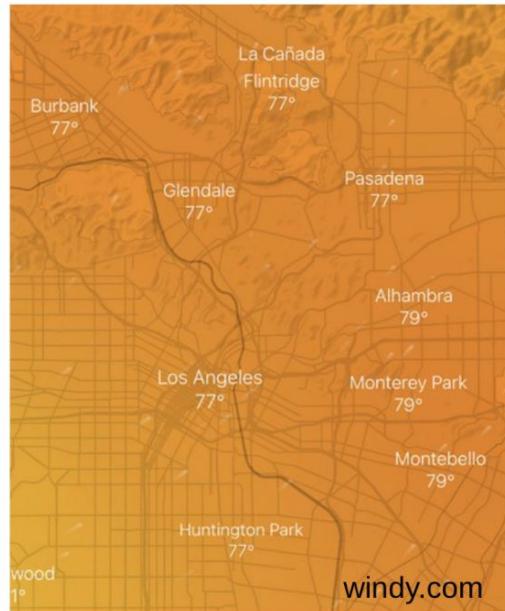




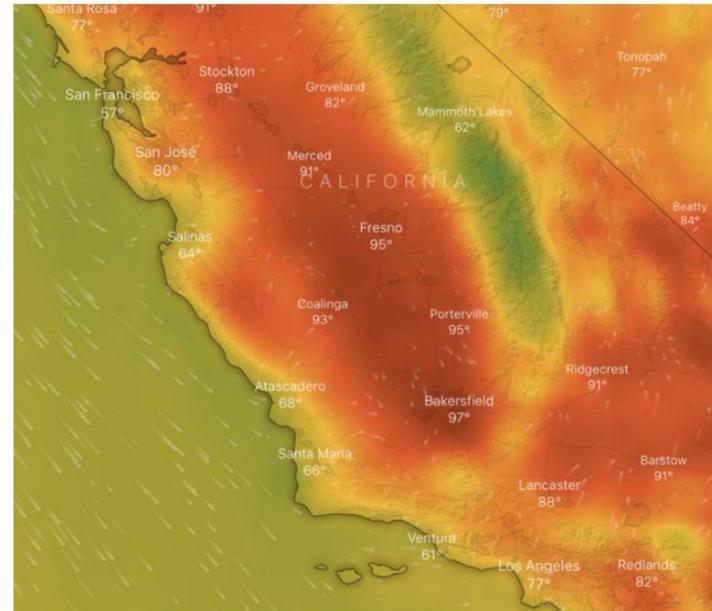
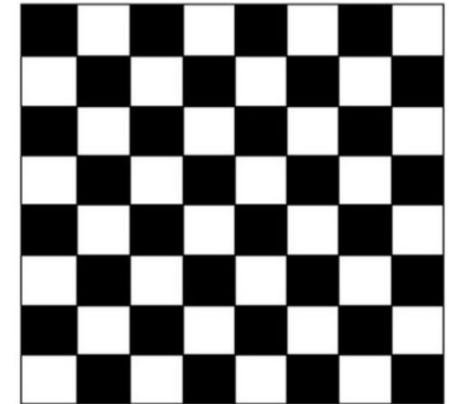
# Exploratory Data Analysis

Tobler's first law of geography: Everything is related to everything else.  
But near things are more related than distant things.

Positive

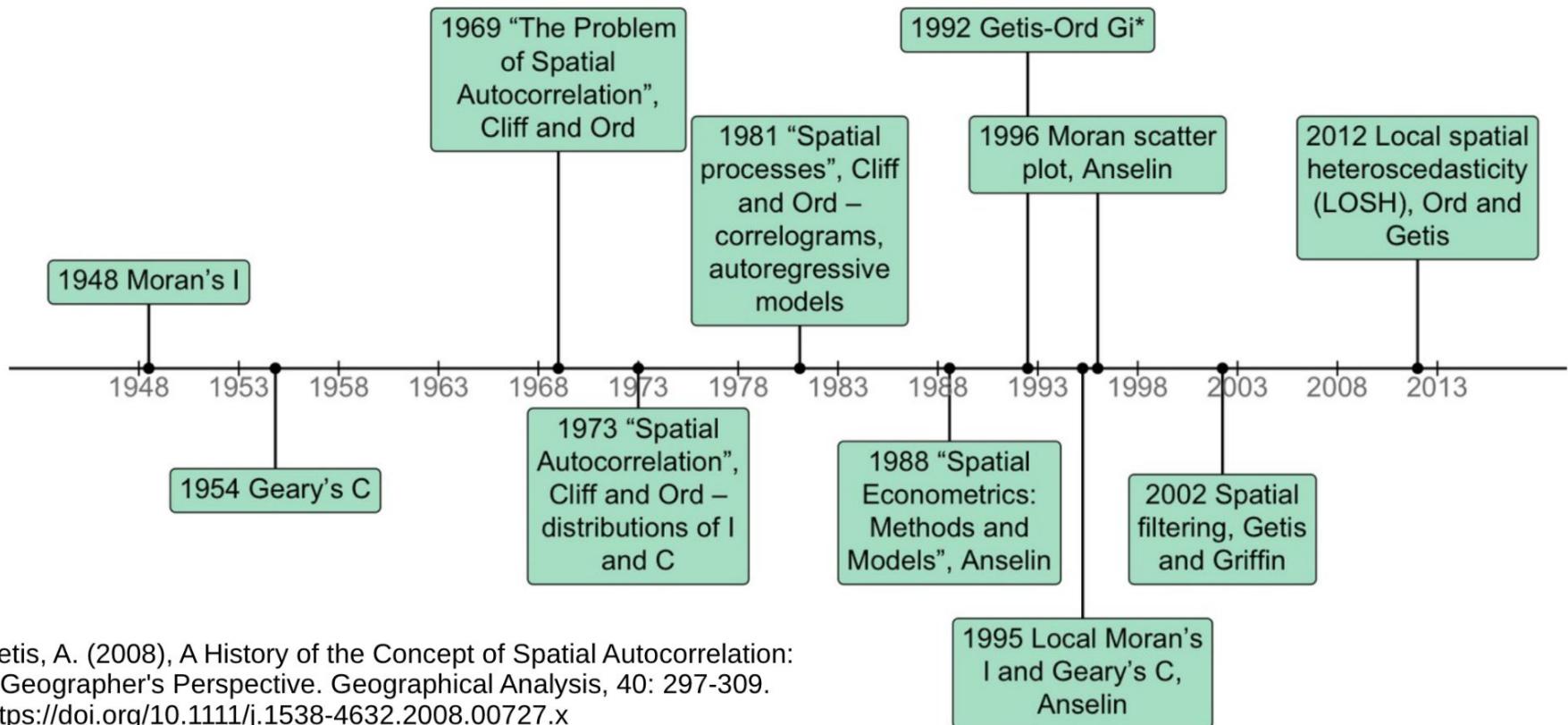


Negative



Slide credit: Lambda Moses

# Exploratory Data Analysis



Getis, A. (2008). A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective. *Geographical Analysis*, 40: 297-309.  
<https://doi.org/10.1111/j.1538-4632.2008.00727.x>

Slide credit: Lambda Moses

# Exploratory Data Analysis: Moran's I

## Pearson correlation: coexpression of 2 genes

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

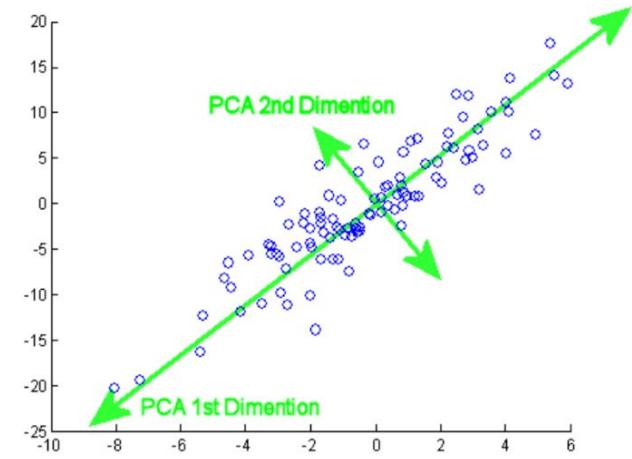
## Moran's I: spatial autocorrelation of 1 gene

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad \sum_{i=1}^n (x_i - \bar{x})^2$$

$w_{ij}$ : spatial weights indicating neighborhood



PCA: correlation among numerous genes



Slide credit: Lambda Moses

# Modeling Spatial Variability in Gene Space

## In Context

- (gene) SpatialIDE, nnSVG
- (gene programs )Spatial NMF

# Spatial Variable Gene Selection

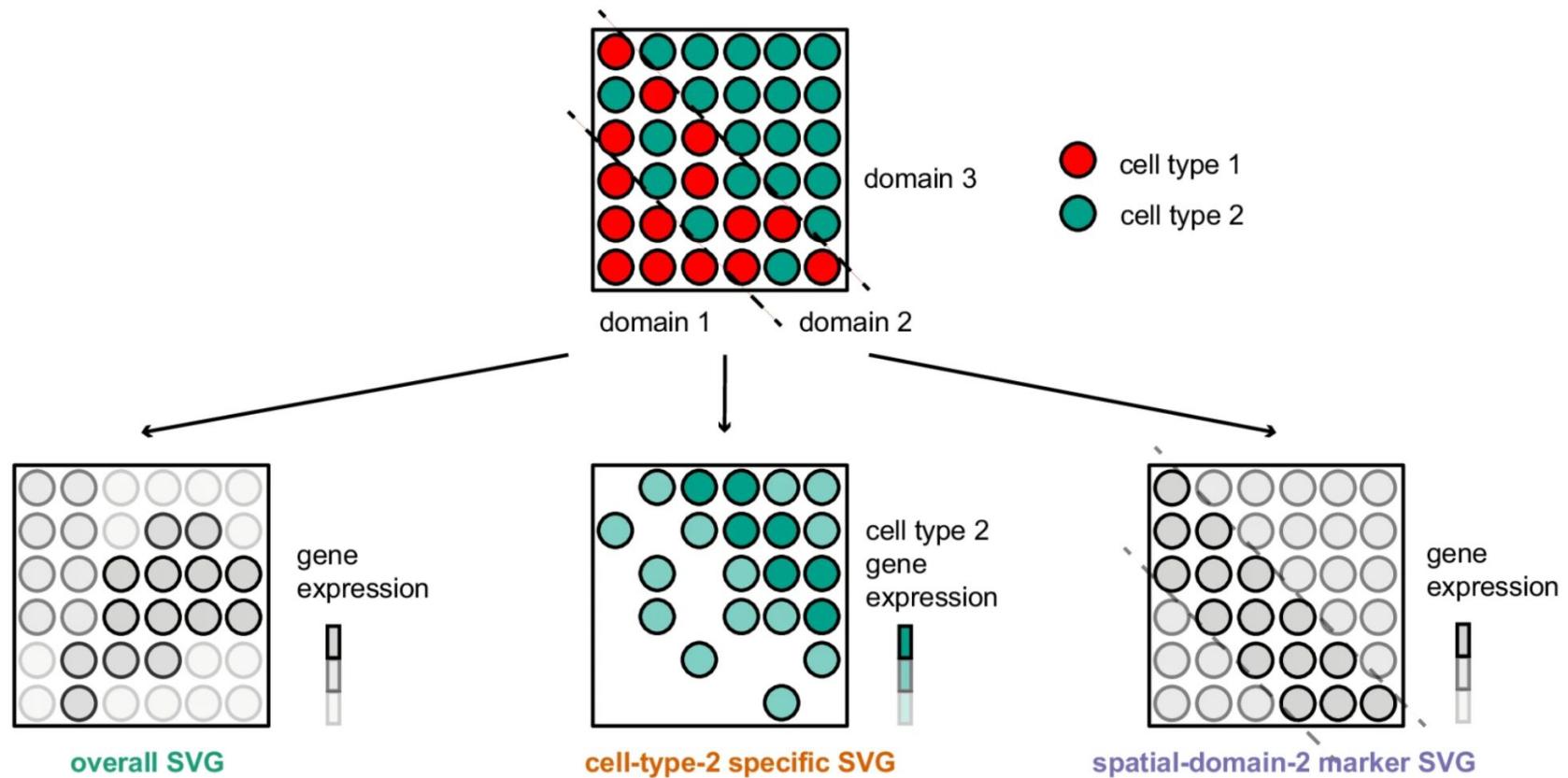
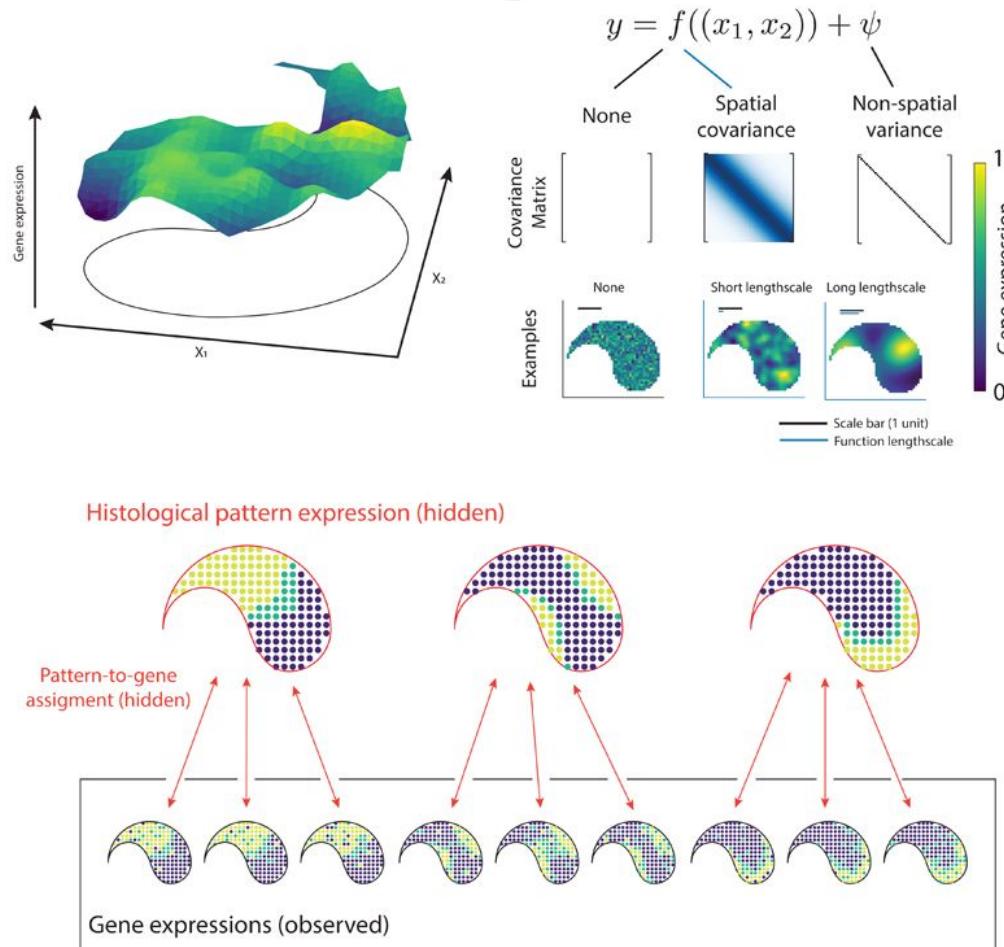


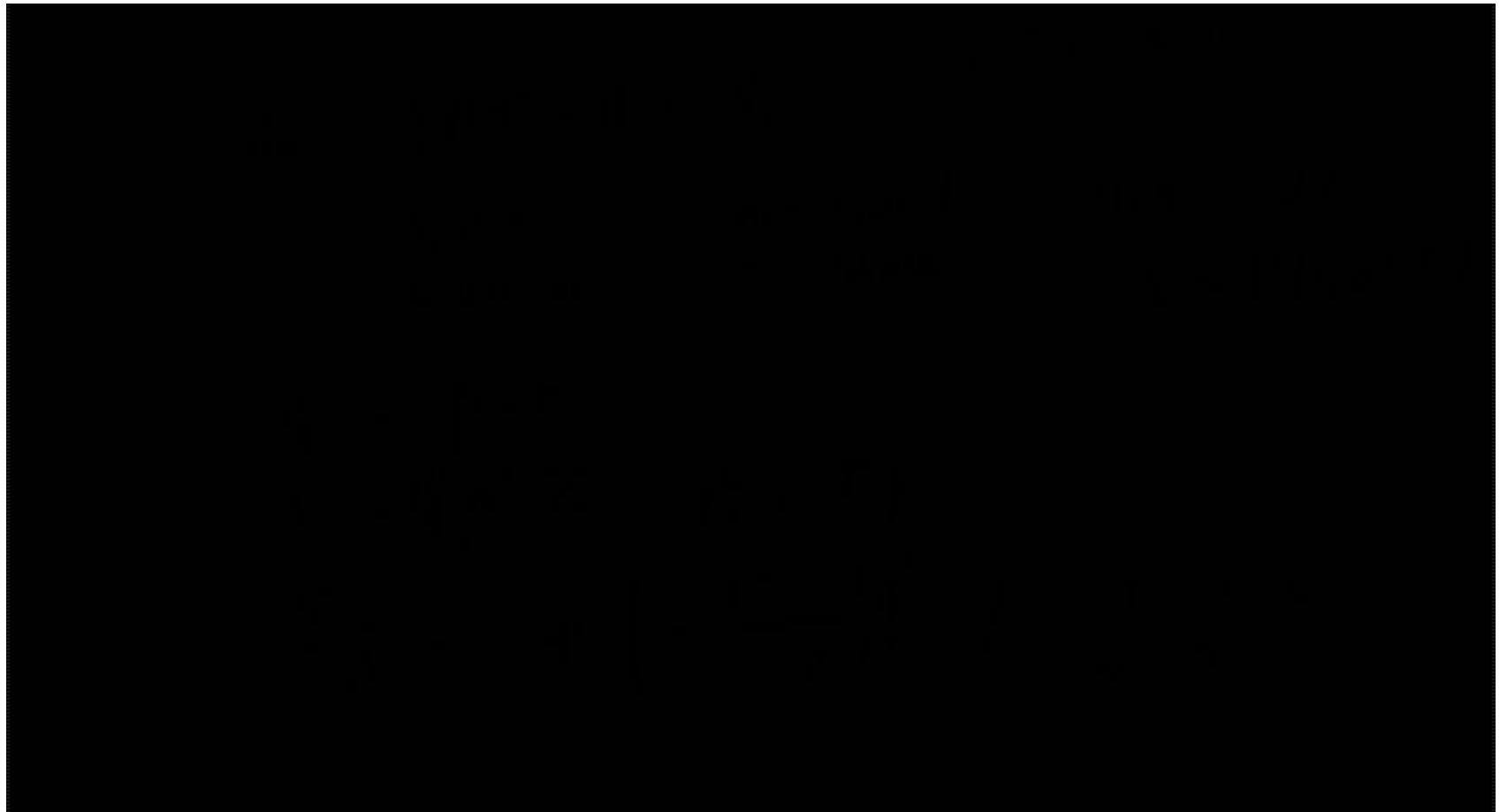
Figure from: Yan, G., Hua, S.H. & Li, J.J. Categorization of 34 computational methods to detect spatially variable genes from spatially resolved transcriptomics data. (2025)

# Spatial Variable Gene Selection: SpatialDE

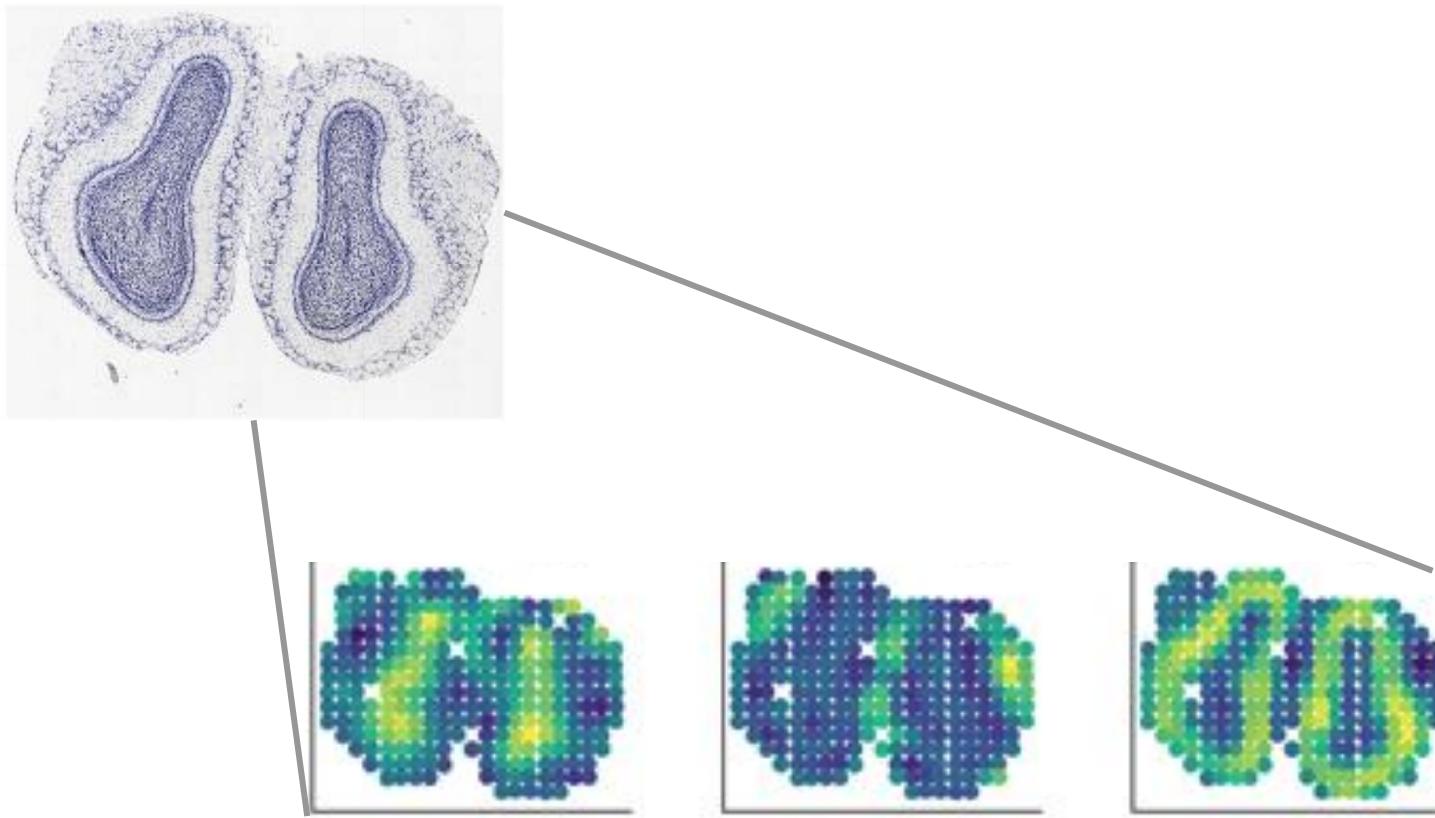


Adapted from Svensson. et al. SpatialDE: identification of spatially variable genes (2018).

# Spatial Variable Gene Selection: SpatialIDE



# Spatial Variable Gene Selection: SpatialDE



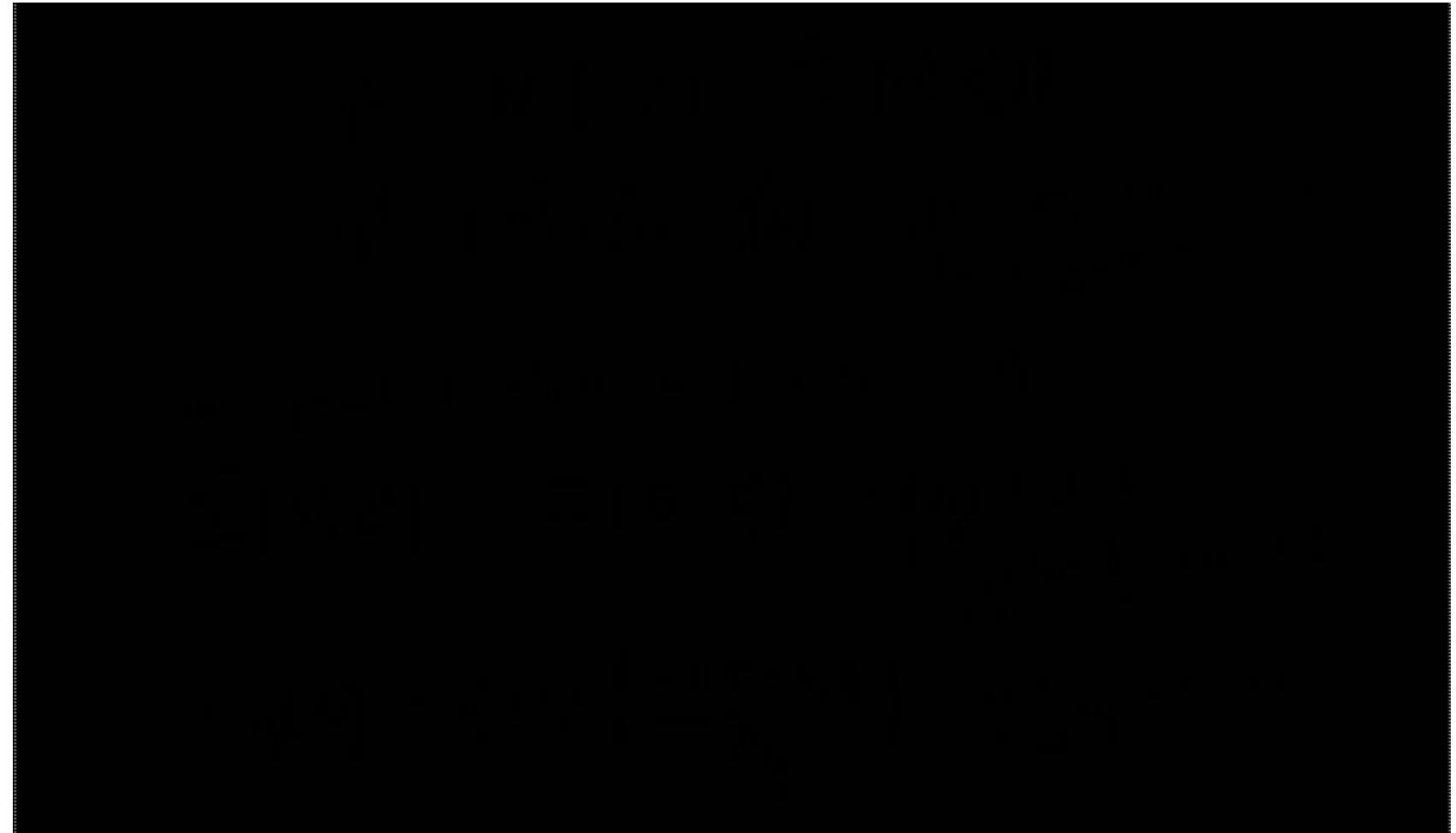
Adapted from Svensson. et al. SpatialDE: identification of spatially variable genes (2018).

# Spatial Variable Gene Selection: nnSVG

Motivation: why should all the genes share the same scale?

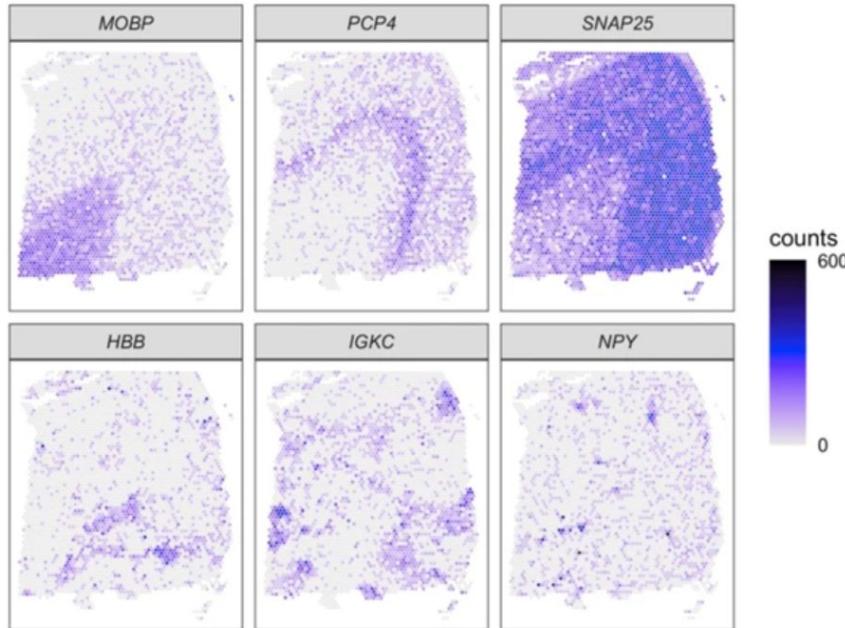
Based on Weber, L.M., Saha, A., Datta, A. et al. nnSVG for the scalable identification of spatially variable genes using nearest-neighbor Gaussian processes. (2023).

# Spatial Variable Gene Selection: nnSVG

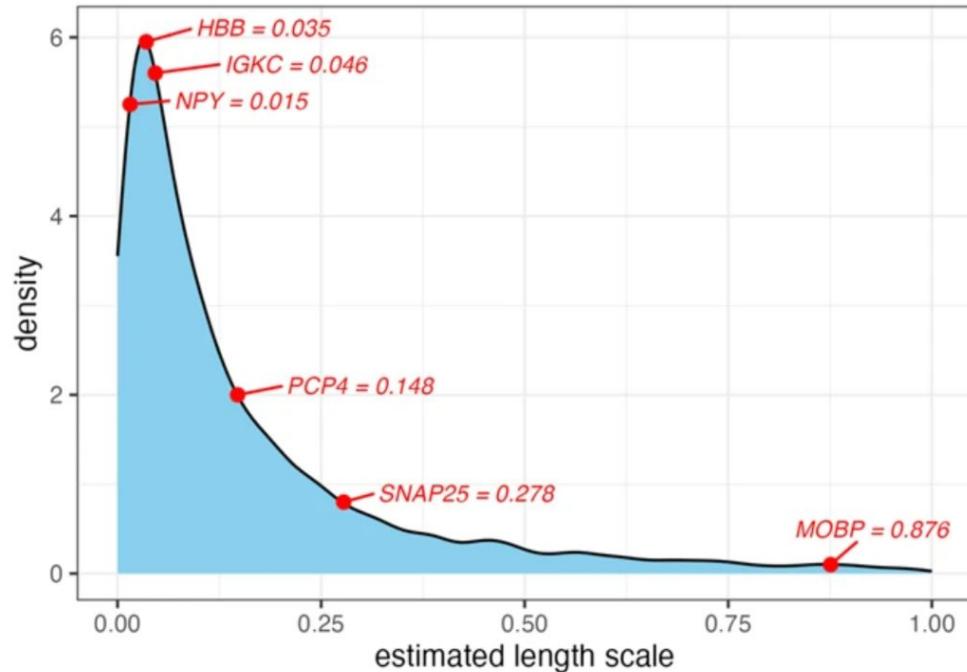


# Spatial Variable Gene Selection: nnSVG

Selected SVGs: human DLPFC



nnSVG length scales: human DLPFC



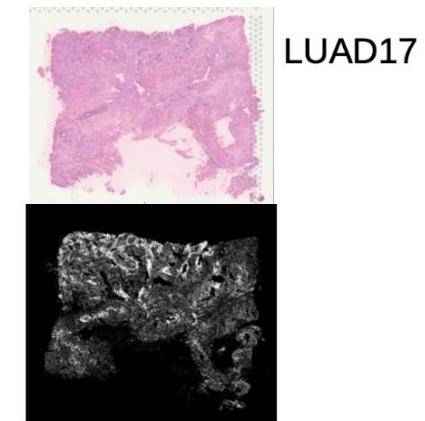
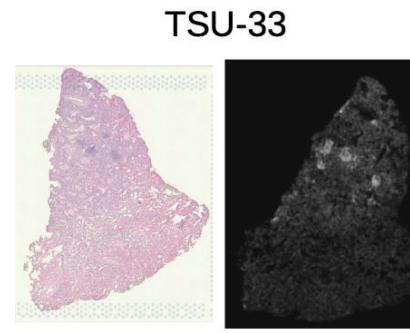
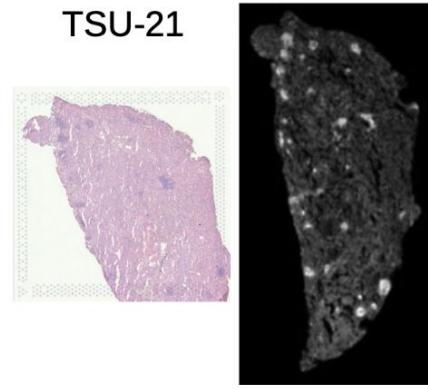
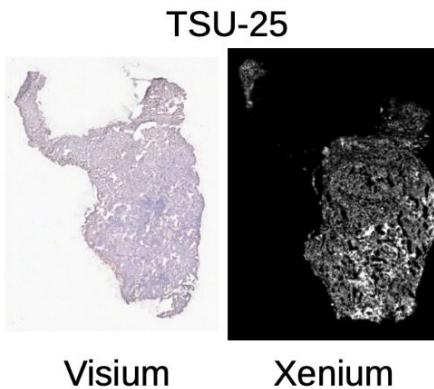
Adapted from Weber, L.M., Saha, A., Datta, A. et al. nnSVG for the scalable identification of spatially variable genes using nearest-neighbor Gaussian processes. (2023).

# SVG Case study: a problem of scale

LUAD samples across stages from Haga 2023 and Takano 2024

AIS Noguchi Type A AIS Noguchi Type B MIA Noguchi Type C

IA

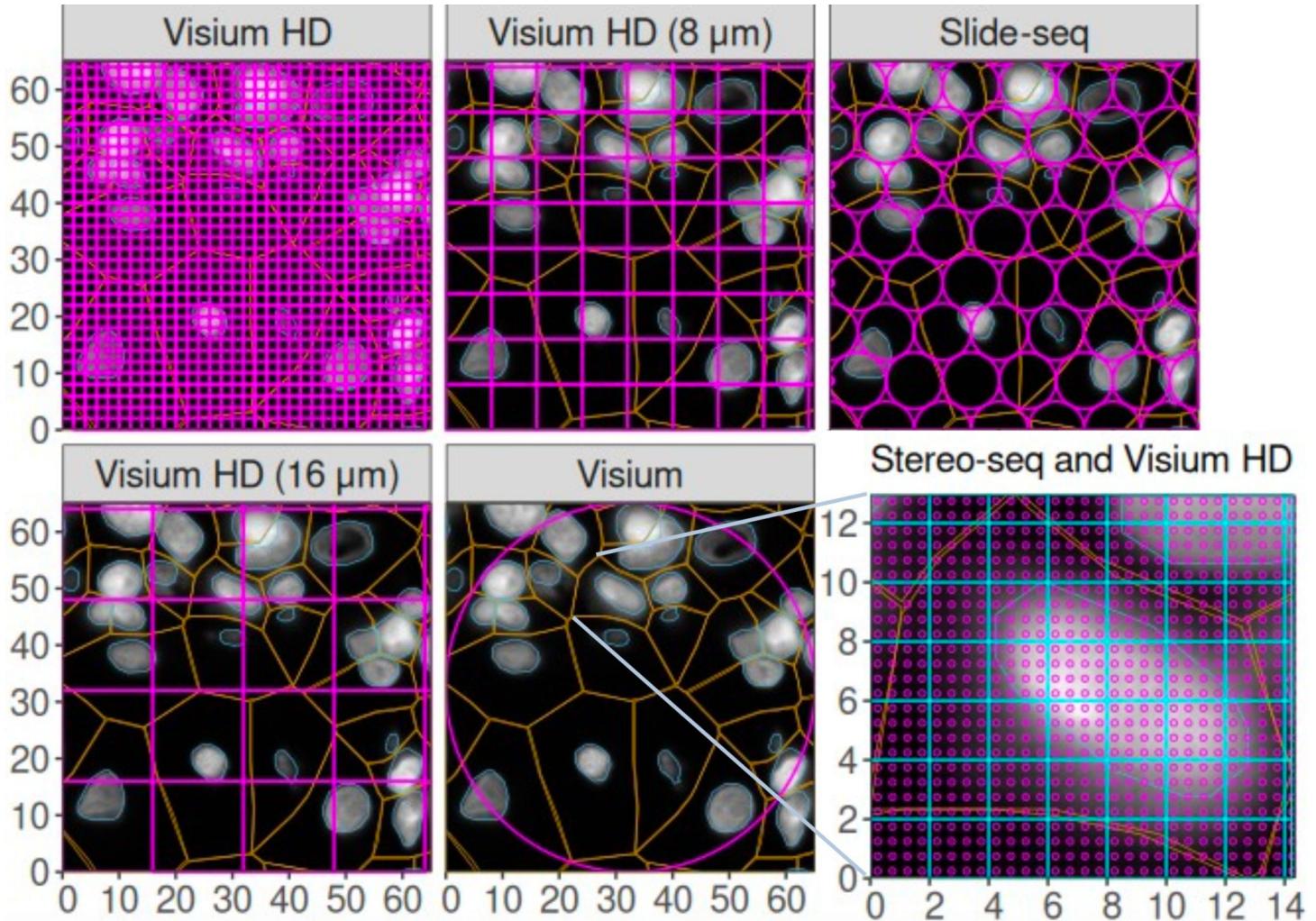


Early

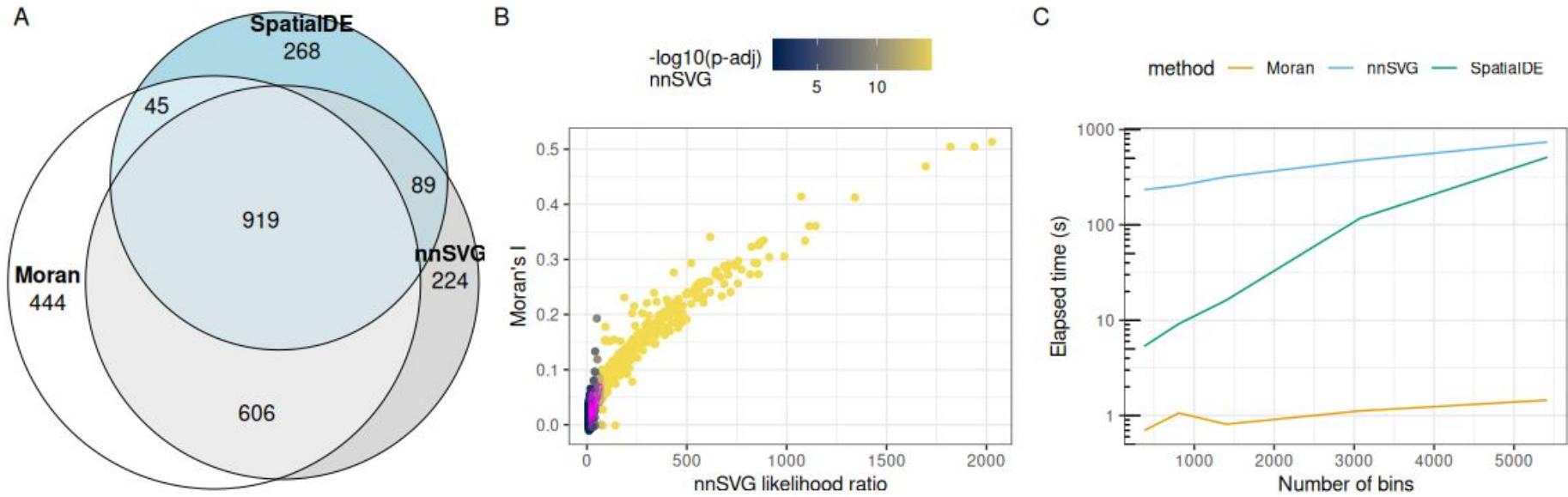
AIS: adenocarcinoma *in situ*  
MIA: minimally invasive adenocarcinoma  
IA: invasive adenocarcinoma

Advanced

# SVG Case study: a problem of scale



# SVG Case study: a problem of scale



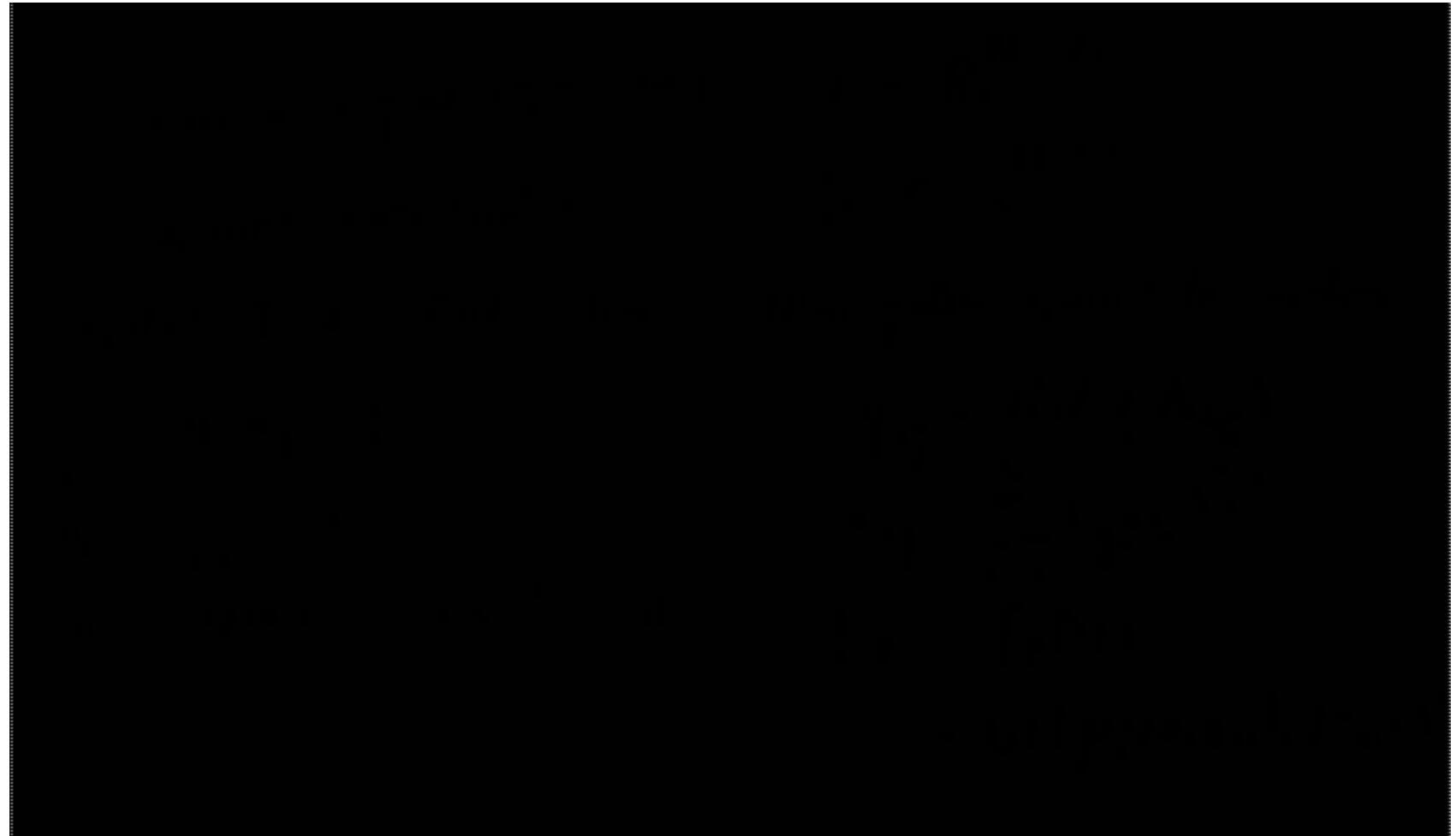
# Spatially Variable Gene Expression Programs

based on Spatial Nonnegative Matrix  
Factorization, Townes & Engelhardt (2023)

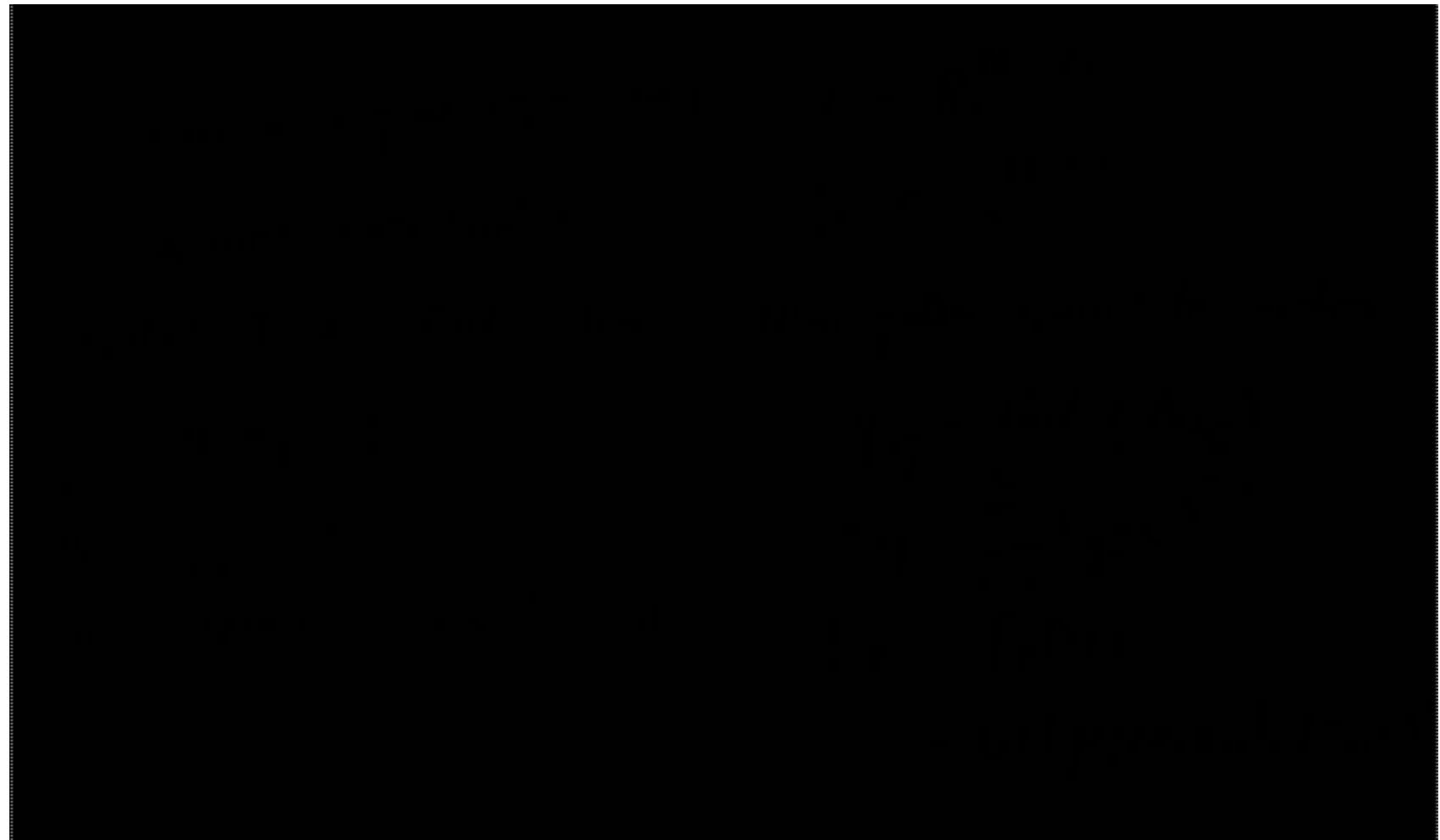
# Spatially Variable Gene Expression Programs

# Spatially Variable Gene Expression Programs

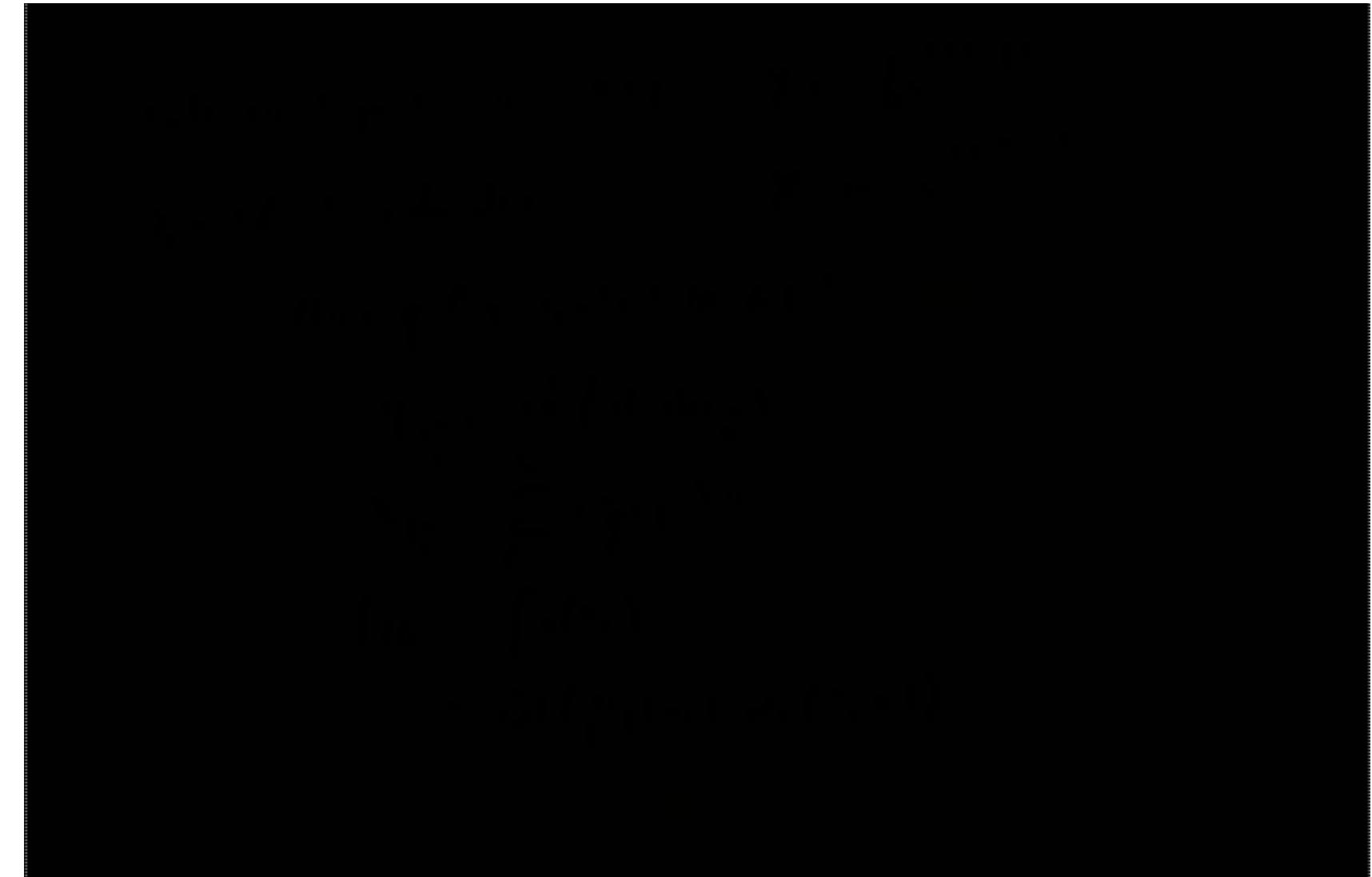
# Spatially Variable Gene Expression Programs



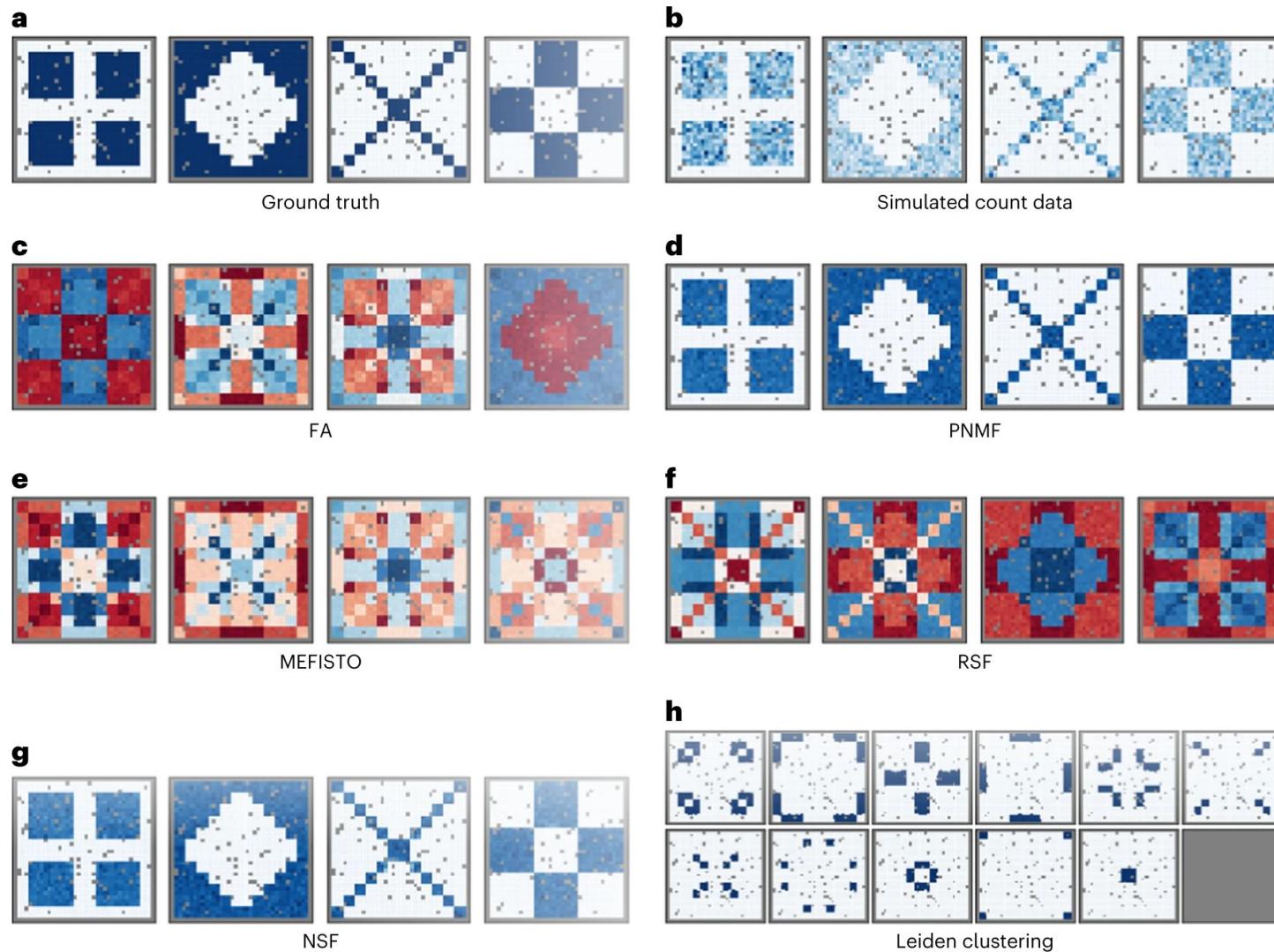
# Spatially Variable Gene Expression Programs



# Spatially Variable Gene Expression Programs



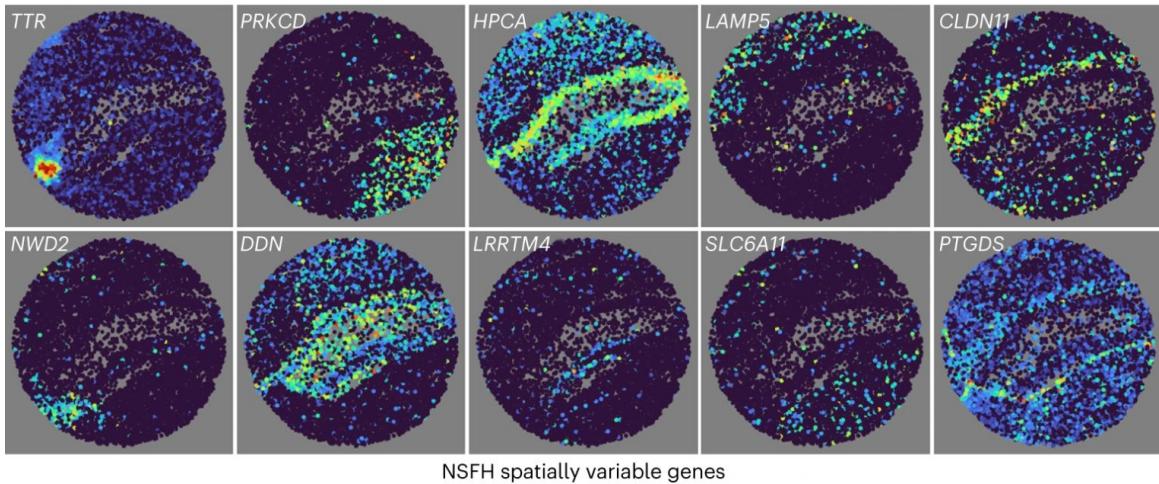
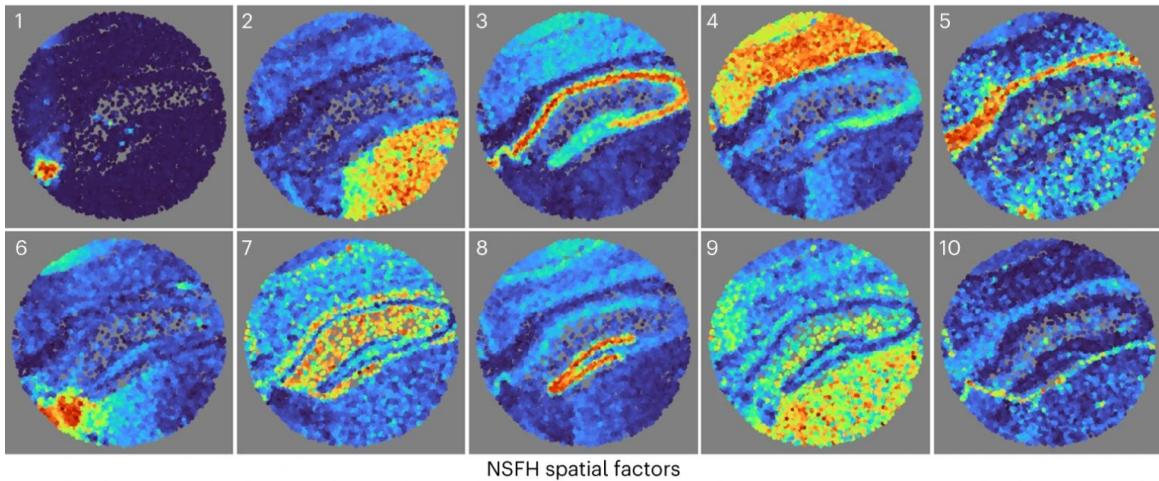
# Spatially Variable Gene Expression Programs



Spatial Nonnegative Matrix Factorization, Townes & Engelhardt (2023)

Bianca Dumitrascu, Machine Learning for Computational Biology, MLSS, 2025

# Spatially Variable Gene Expression Programs



Slide-seqV2 mouse  
hippocampus gene  
expression data

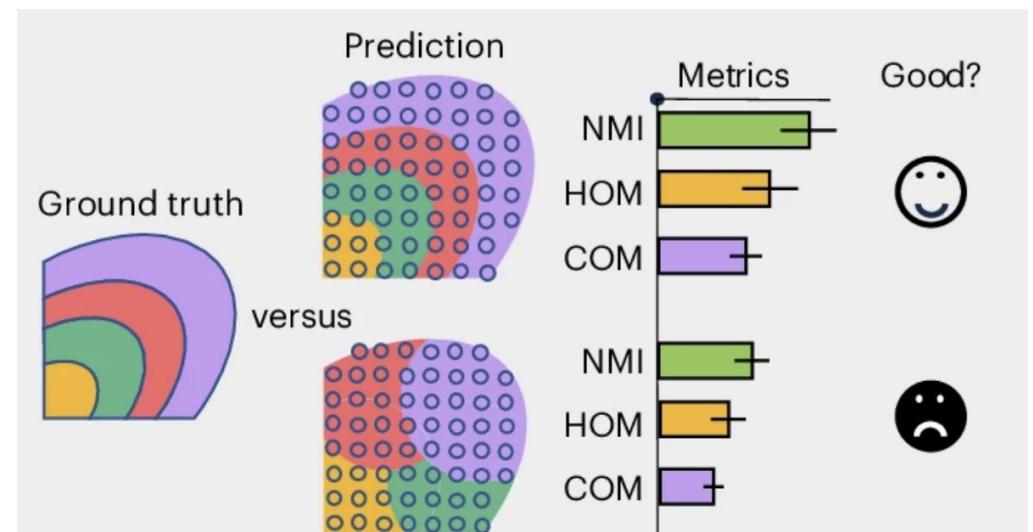
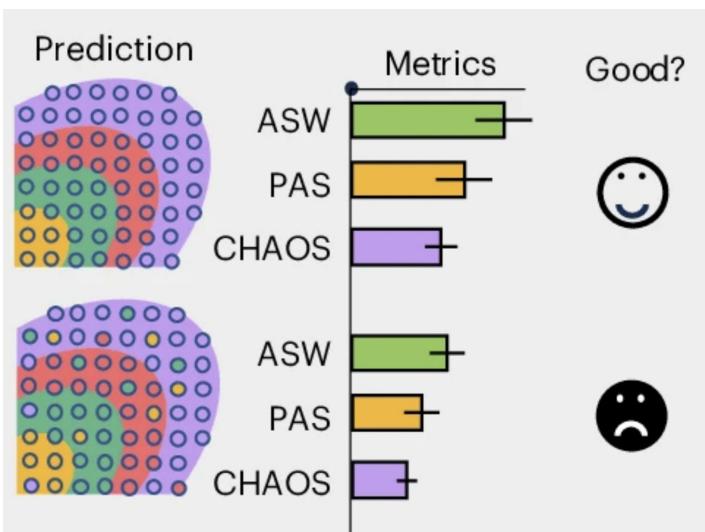
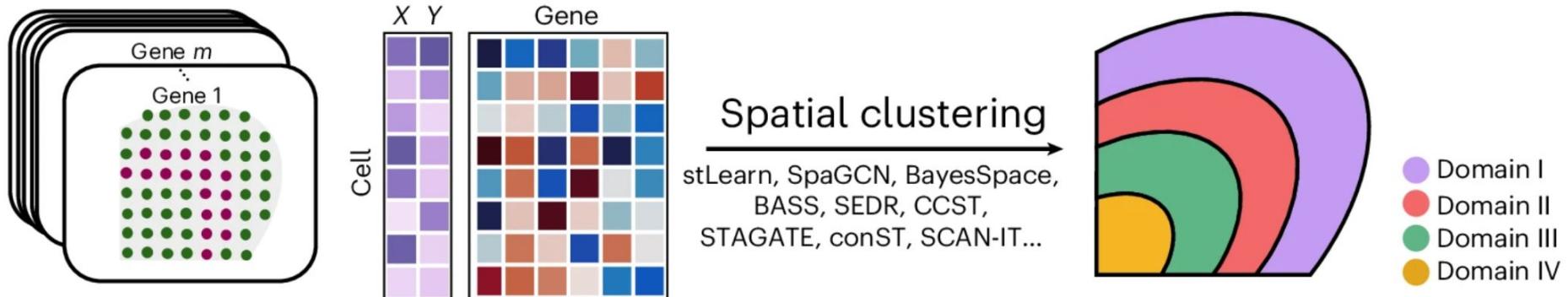
Spatial Nonnegative Matrix Factorization, Townes & Engelhardt (2023)

# Modelling spatial neighbourhoods

## Spatial clustering via Graph Neural Networks (GNNs)

- GraphST

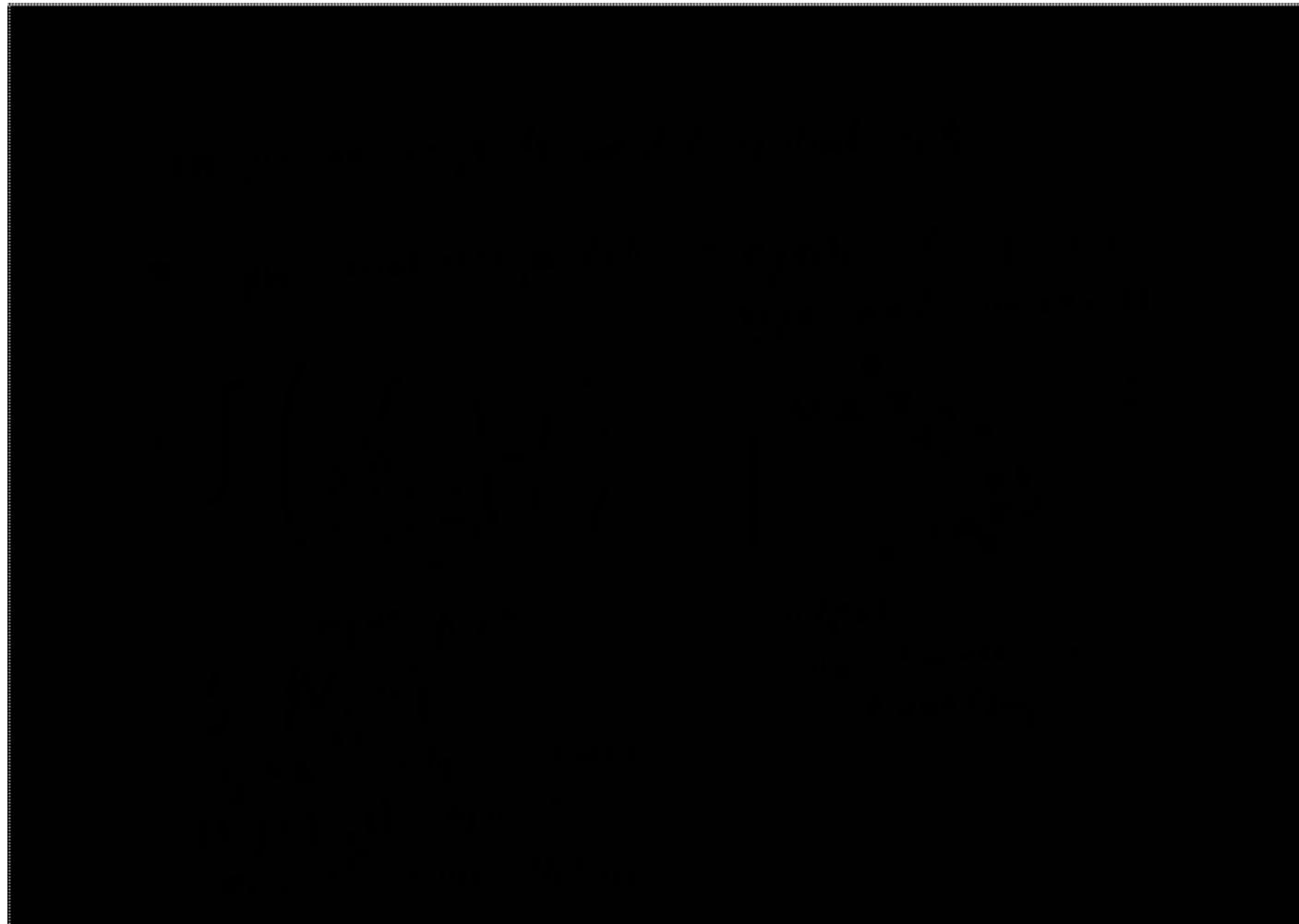
# Spatial Clustering



From: Yuan, Z., Zhao, F., Lin, S. et al.  
Benchmarking spatial clustering methods with  
spatially resolved transcriptomics data. 2024

Bianca Dumitrescu, Machine Learning for Computational Biology, MLSS, 2025

# GraphST and GNNs for Spatial Transcriptomics

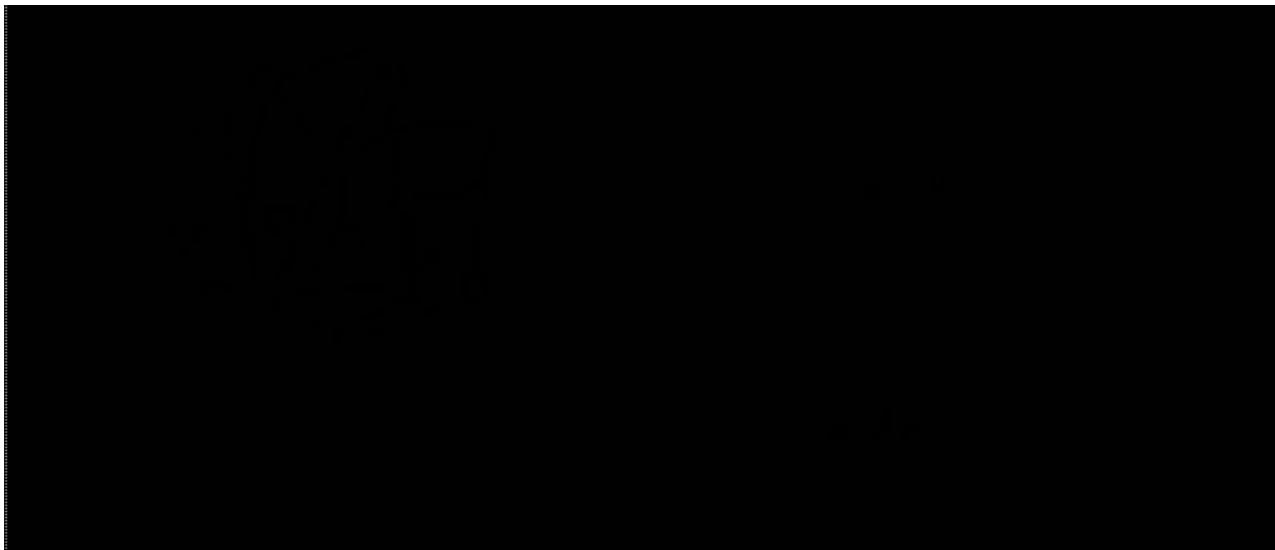


based on: Long, Y., Ang, K.S., Li, M. *et al.* Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. (2023).

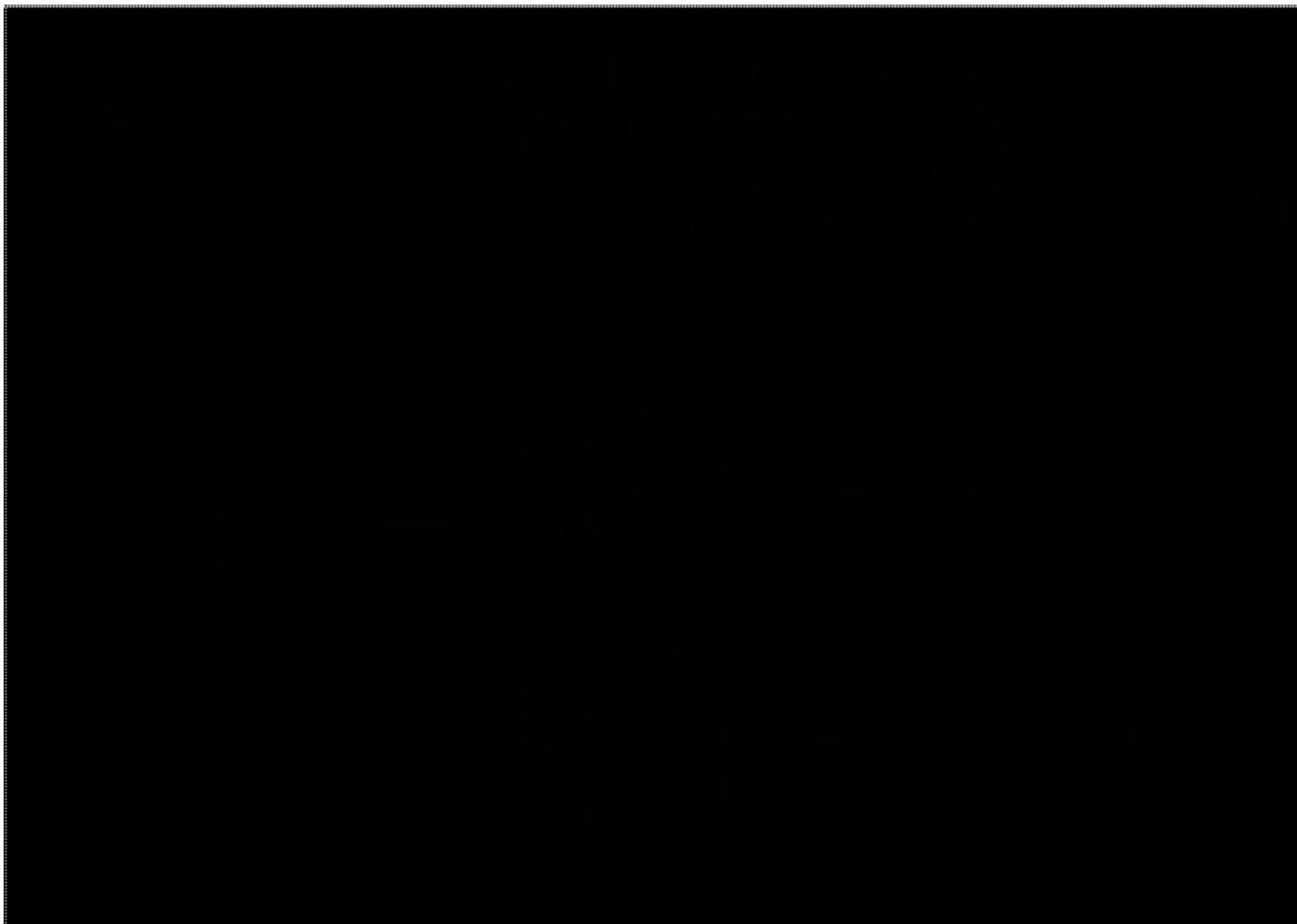
# GraphST and GNNs for Spatial Transcriptomics



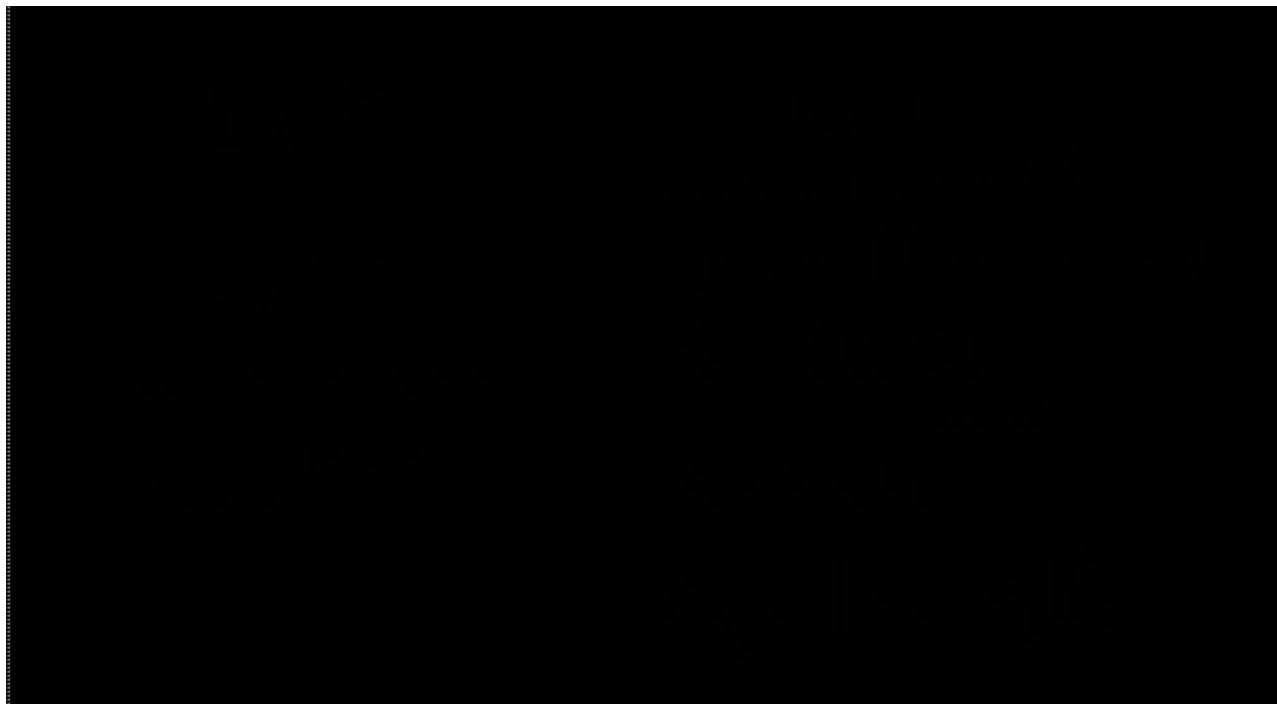
# GraphST and GNNs for Spatial Transcriptomics



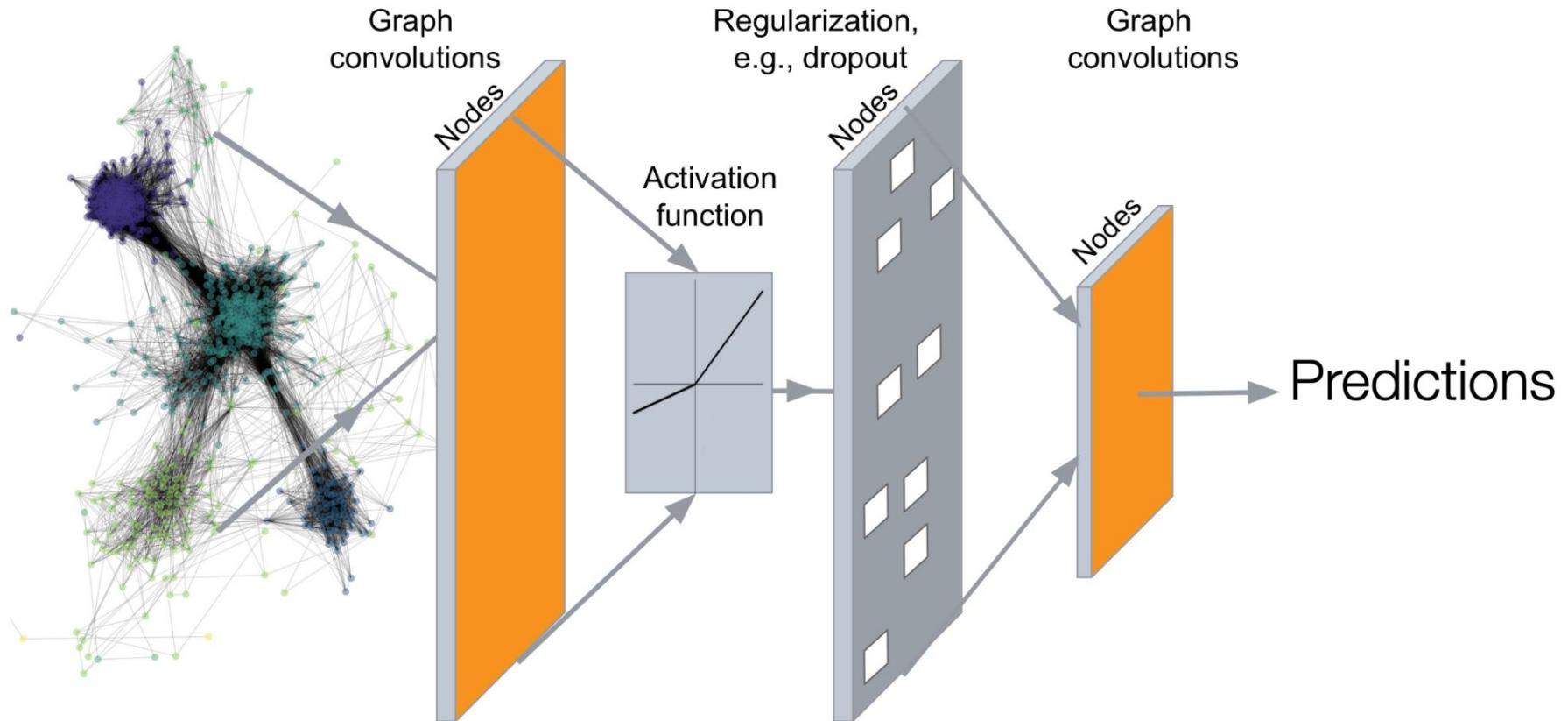
# GraphST and GNNs for Spatial Transcriptomics



# GraphST and GNNs for Spatial Transcriptomics

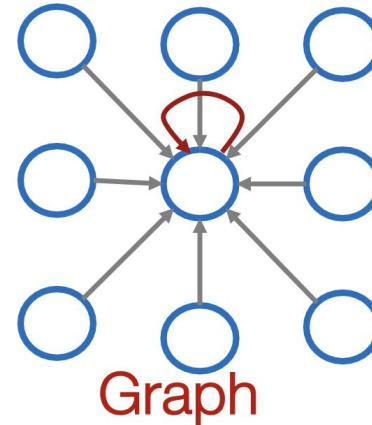
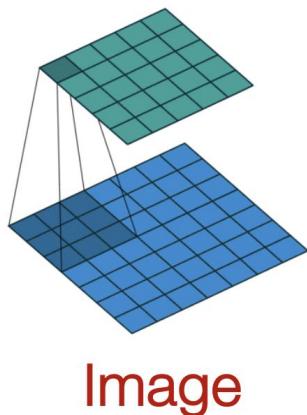


# GraphST and GNNs for Spatial Transcriptomics

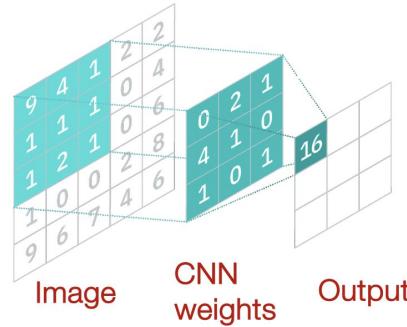
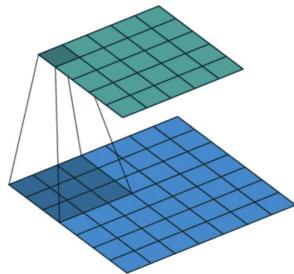


based on <https://web.stanford.edu/class/cs224w/>

# Image Convolutions



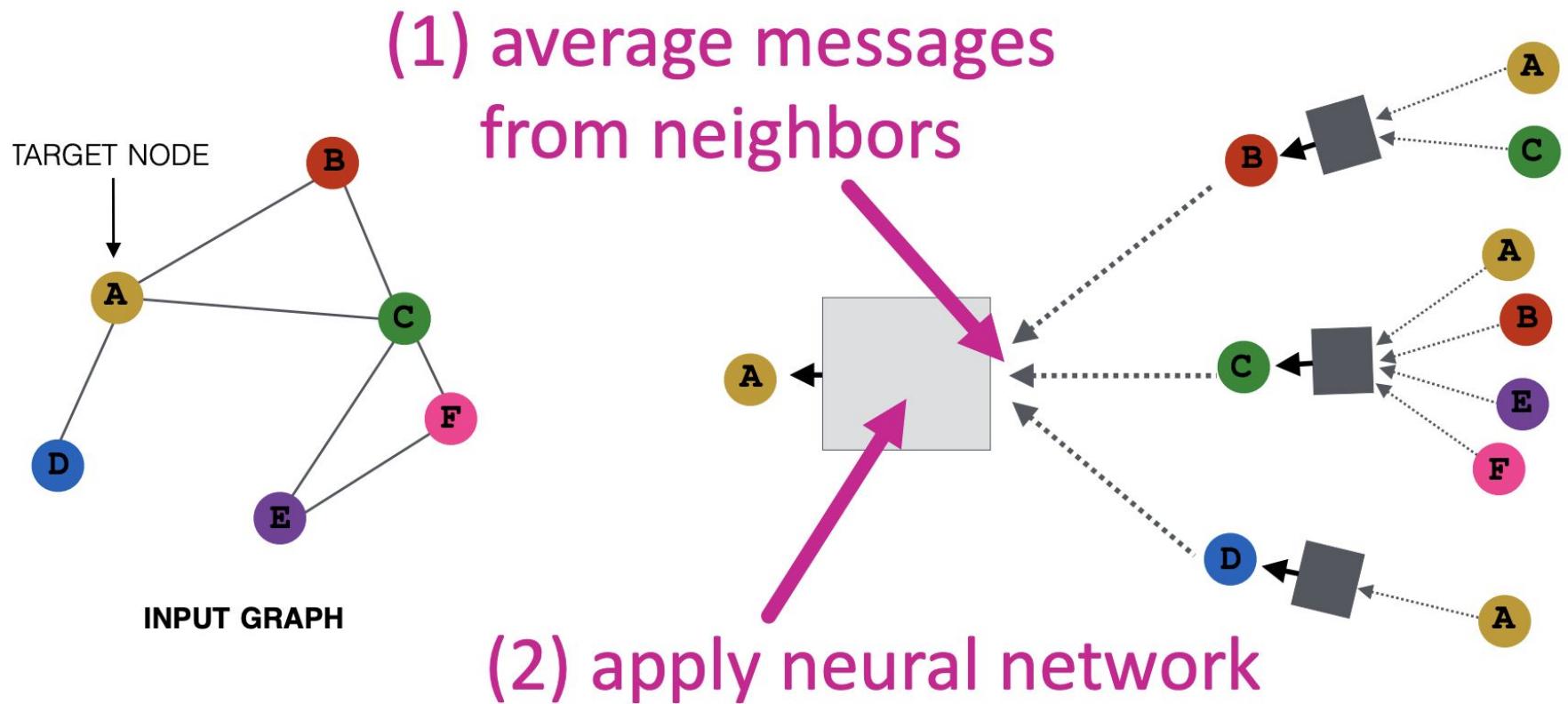
Convolutional neural network (CNN) layer with  
3x3 filter:



$$\text{CNN formulation: } h_v^{(l+1)} = \sigma(\sum_{u \in N(v) \cup \{v\}} W_l^u h_u^{(l)}), \quad \forall l \in \{0, \dots, L-1\}$$

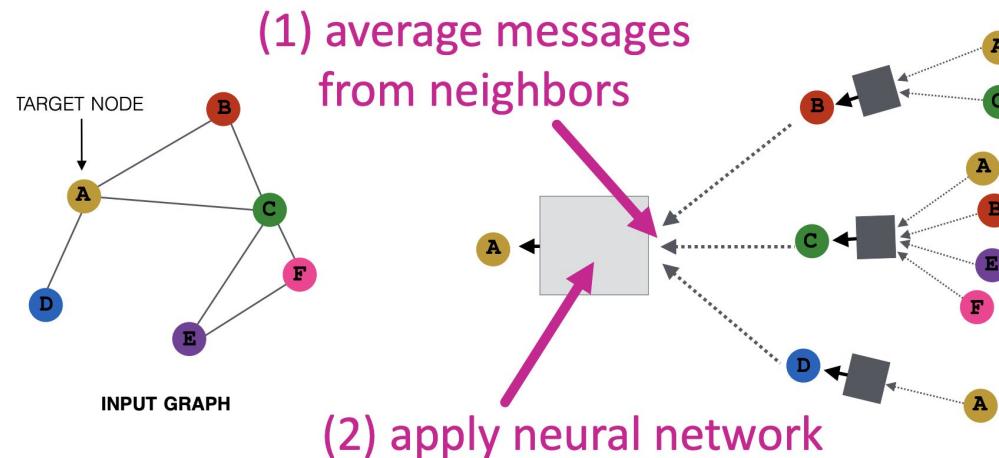
based on <https://web.stanford.edu/class/cs224w/>

# GraphST and GNNs for Spatial Transcriptomics



based on <https://web.stanford.edu/class/cs224w/>

# GraphST and GNNs for Spatial Transcriptomics



Trainable weight matrices  
(i.e., what we learn)

$$\begin{aligned} h_v^{(0)} &= x_v \\ h_v^{(k+1)} &= \sigma(W_k \sum_{u \in N(v)} \frac{h_u^{(k)}}{|N(v)|} + B_k h_v^{(k)}) \\ z_v &= h_v^{(K)} \end{aligned}$$

Final node embedding

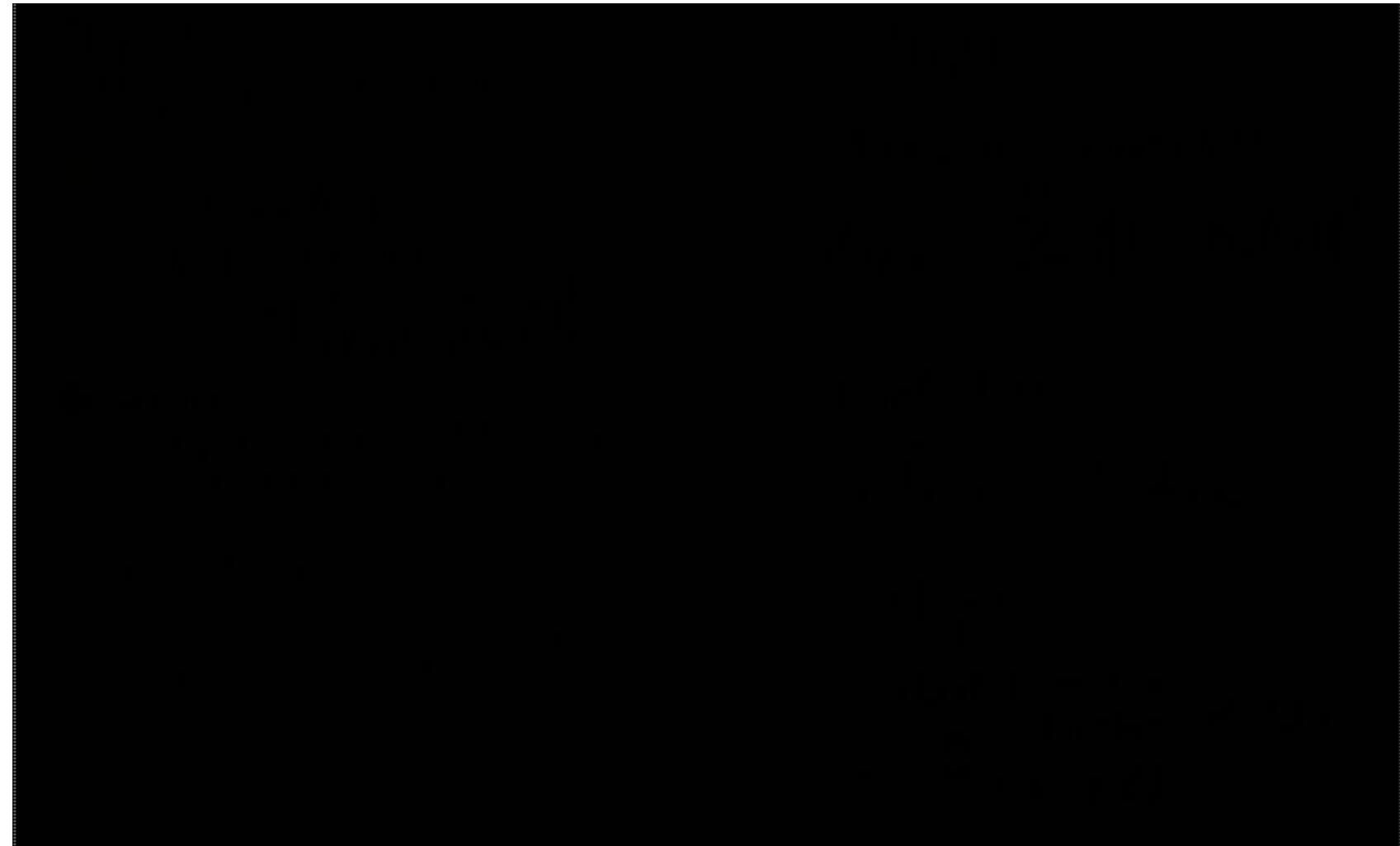
based on <https://web.stanford.edu/class/cs224w/>

# GraphST

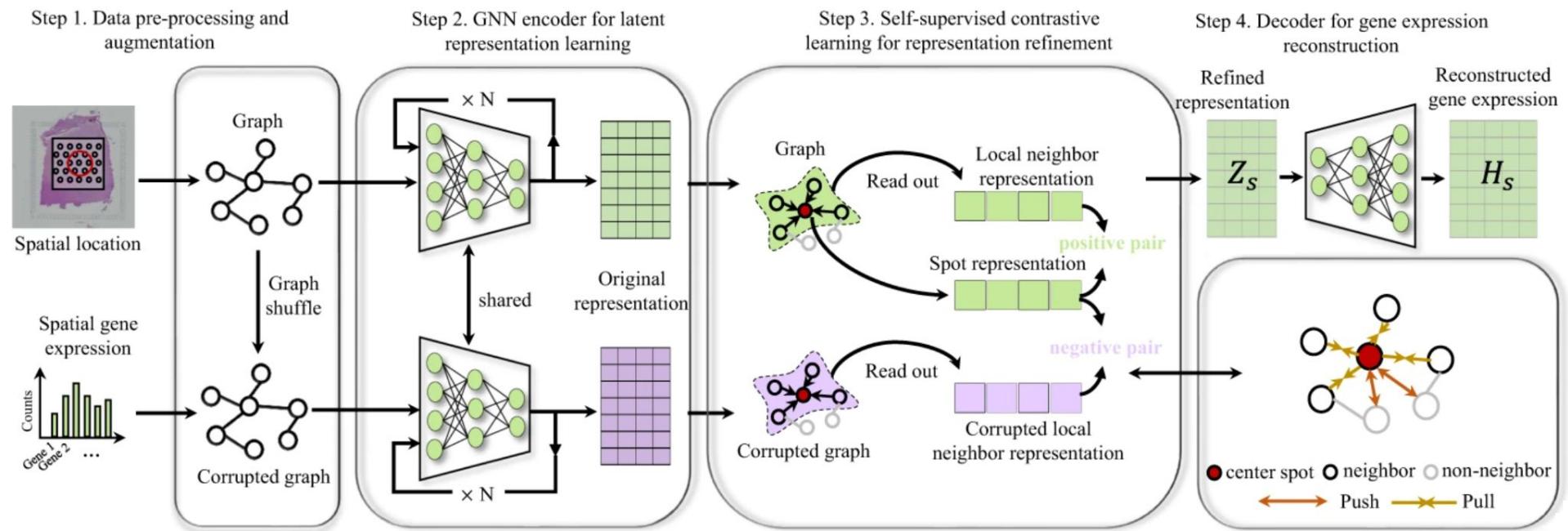
The diagram shows the computation of a node's embedding over time steps. A grey box contains the equations for  $h_v^{(0)}$ ,  $h_v^{(k+1)}$ , and  $z_v$ . A pink bracket labeled "Trainable weight matrices (i.e., what we learn)" spans the  $W_k$  term in the second equation and the  $B_k$  term in the third. A green arrow labeled "Final node embedding" points from the  $z_v = h_v^{(K)}$  equation to the bottom right.

$$h_v^{(0)} = x_v$$
$$h_v^{(k+1)} = \sigma(W_k \sum_{u \in N(v)} \frac{h_u^{(k)}}{|N(v)|} + B_k h_v^{(k)})$$
$$z_v = h_v^{(K)}$$

# GraphST



# GraphST

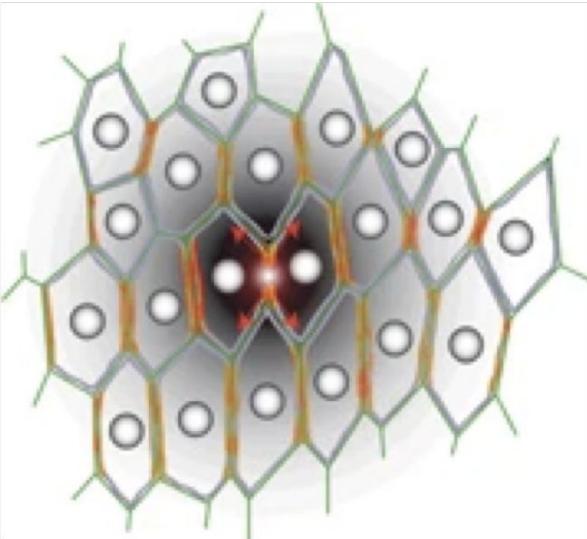


adapted from: Long, Y., Ang, K.S., Li, M. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. (2023).

# To make Maps that represent Living Systems we need two answer at least two Questions

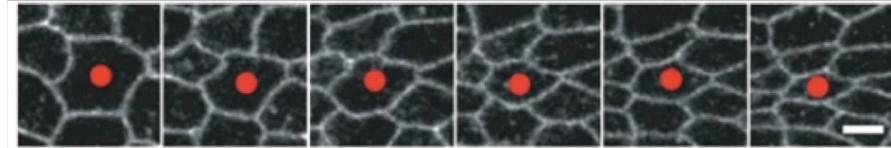
- **What are the units?**  
If we do not yet know the key components or players, how can we discover and define them?
- **How are they organized and how do they interact?**  
Given the players, what are the rules and patterns that govern their interactions **across scales?**

# Case study: material science meets biology meets machine learning



<https://www.nature.com/articles/nrm2606>

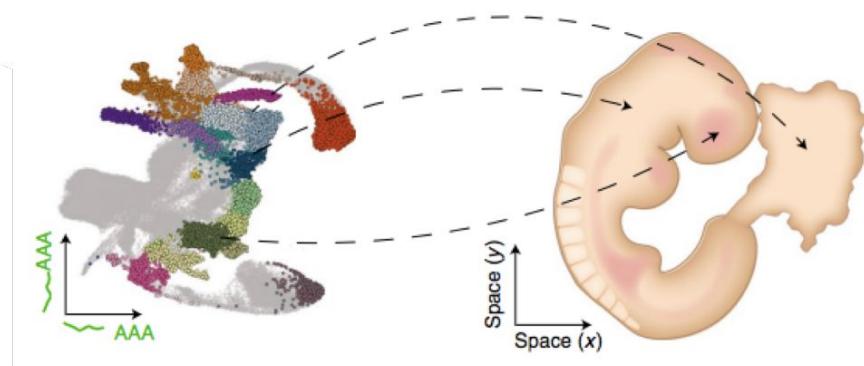
Tissues are shaped by mechanical forces,  
\* What are the origins and the nature of these forces?  
\* How do tissues fold to create 3D shapes?



Martin, Kaschube, & Wieschaus.  
Pulsed contractions of an actin–myosin network  
drive apical constriction.  
*Nature* 457, 495–499 (2009)

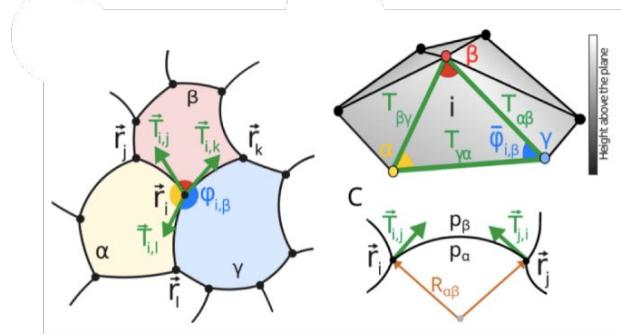
# Case study: material science meets biology meets machine learning

**Quantifying how multiple modalities interact:**  
How do mechanics and genomics interact in the context of early boundary formation?

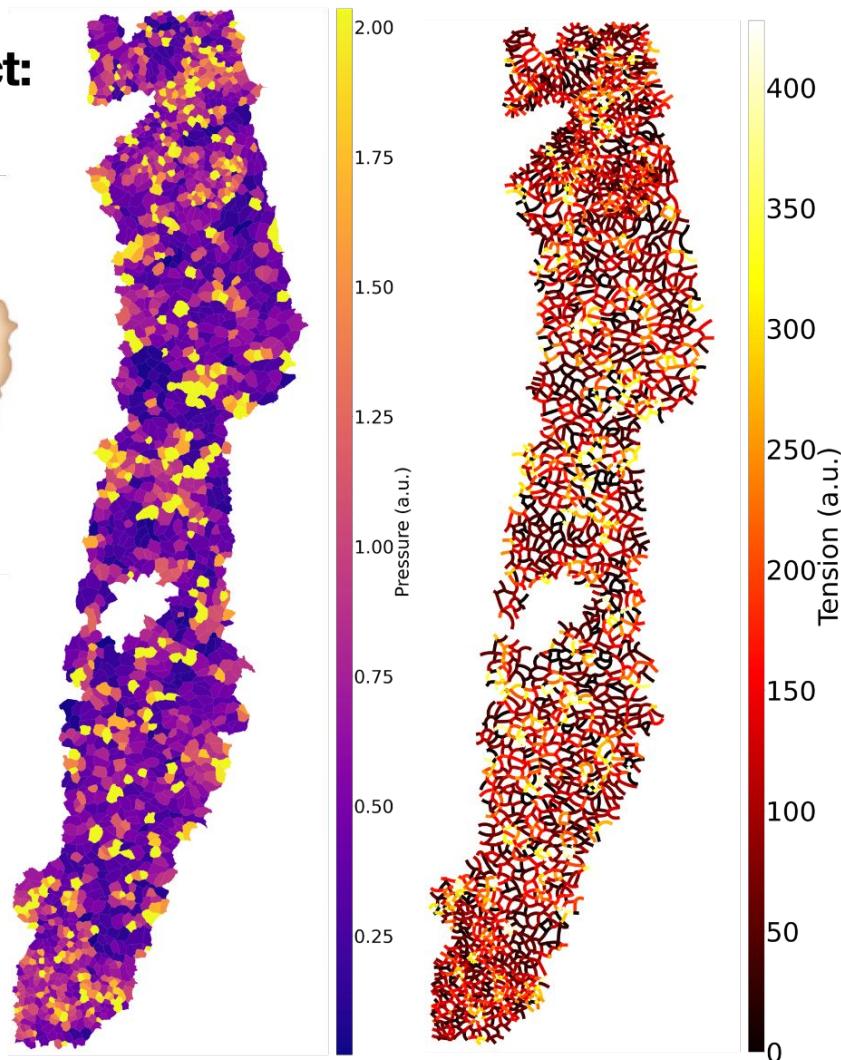


Arguelaget, Cuomo, Stegle ,Marioni (2021)

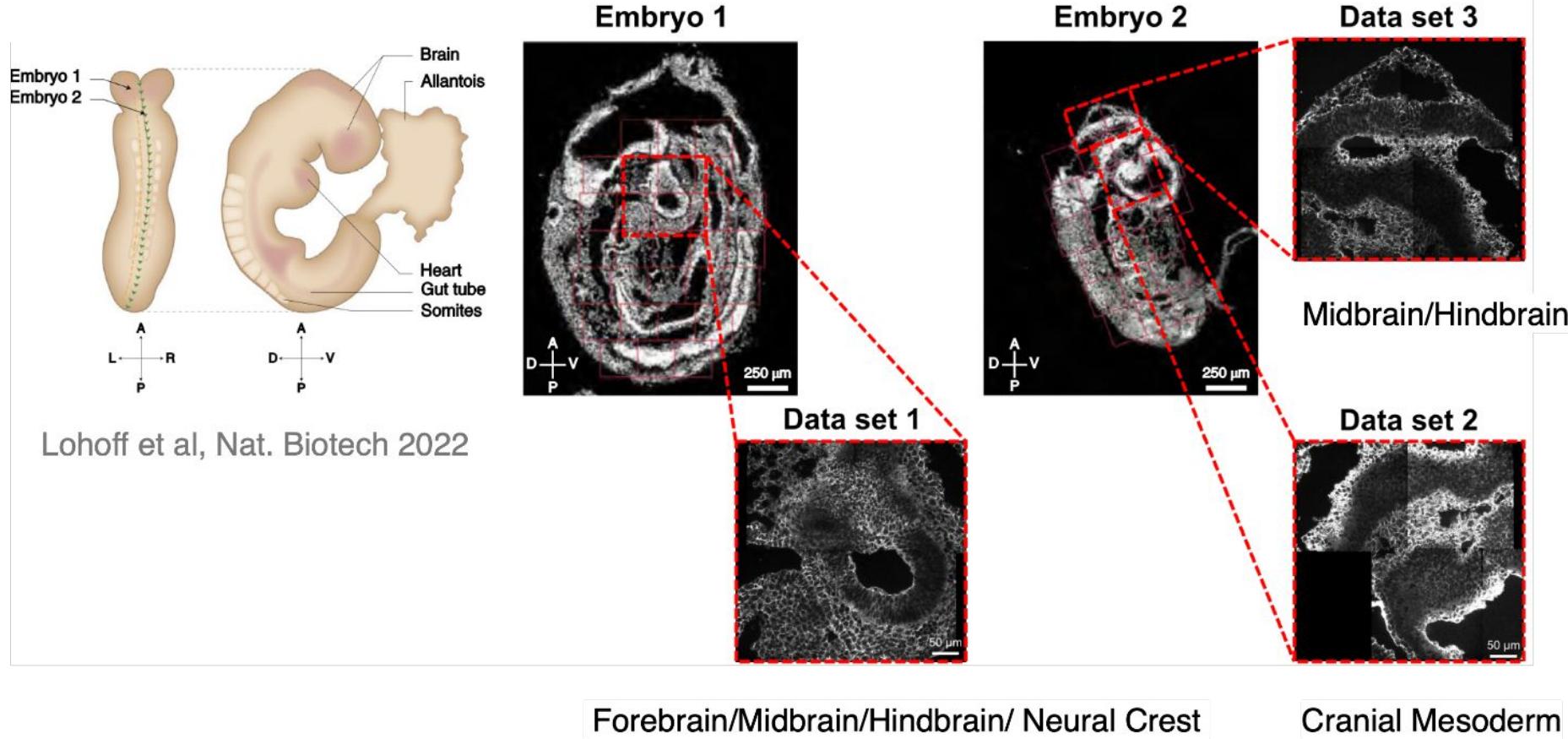
Spatial Mouse, Lohoff et al, Nat. Biotech 2022



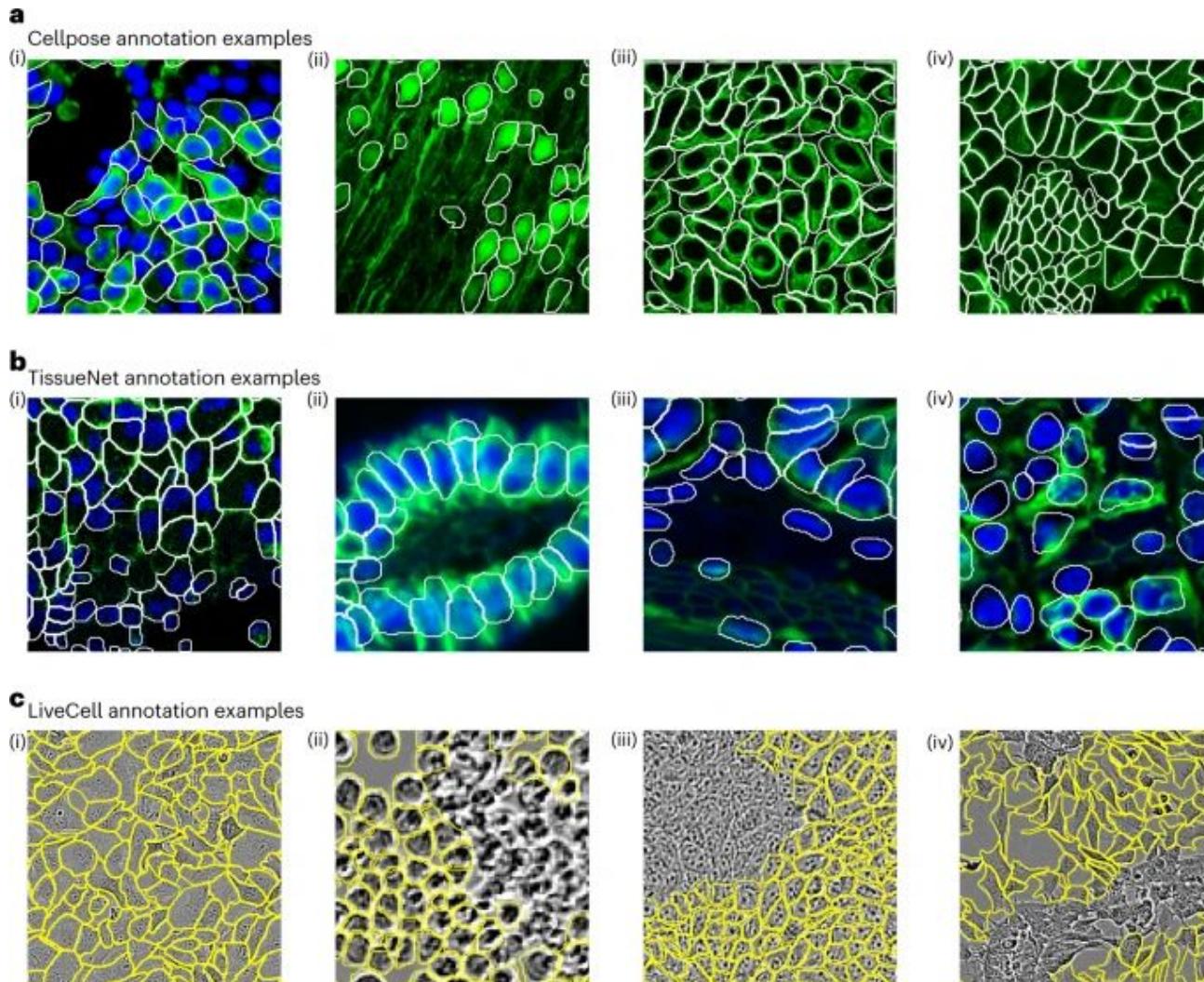
Noll, Streichan, Schraiman, Phys Rev X. 2020



# Case study: material science meets biology meets machine learning



# Ingredient: segmentation

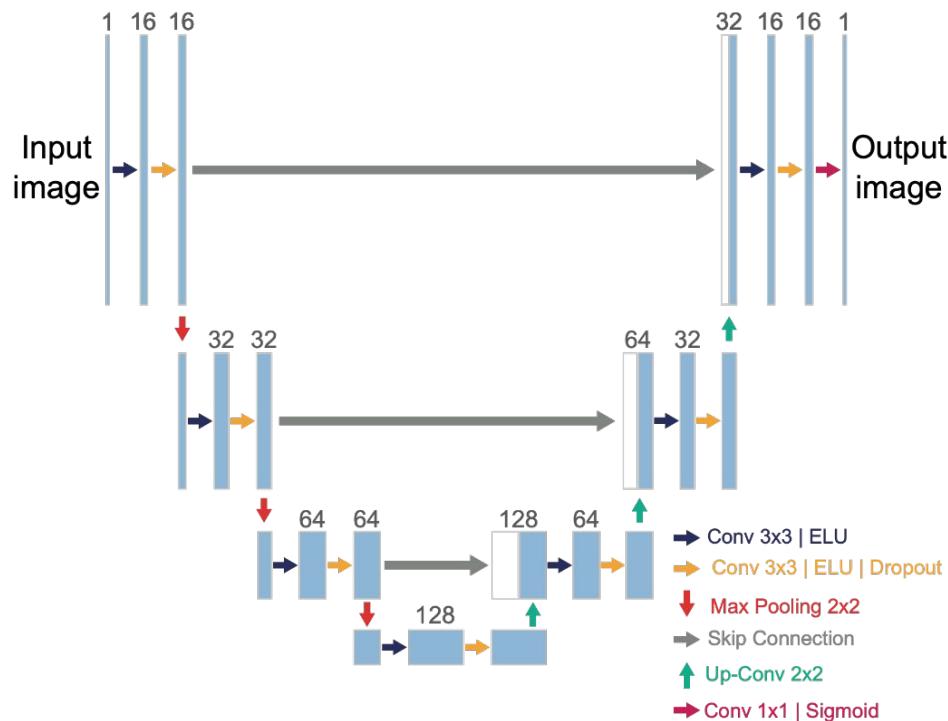


Cellpose 2.0: how to train your own model. Pachitariu & Stringer, 2022

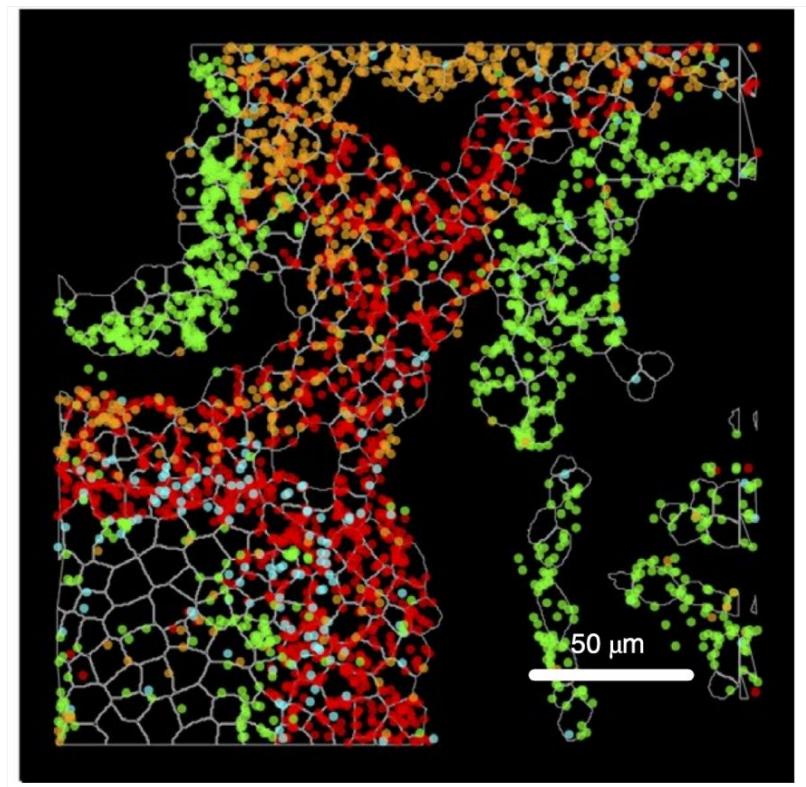
# Scale: cellular organization and data processing

## Ingredient: segmentation

modified U-Net: neural network for segmentation

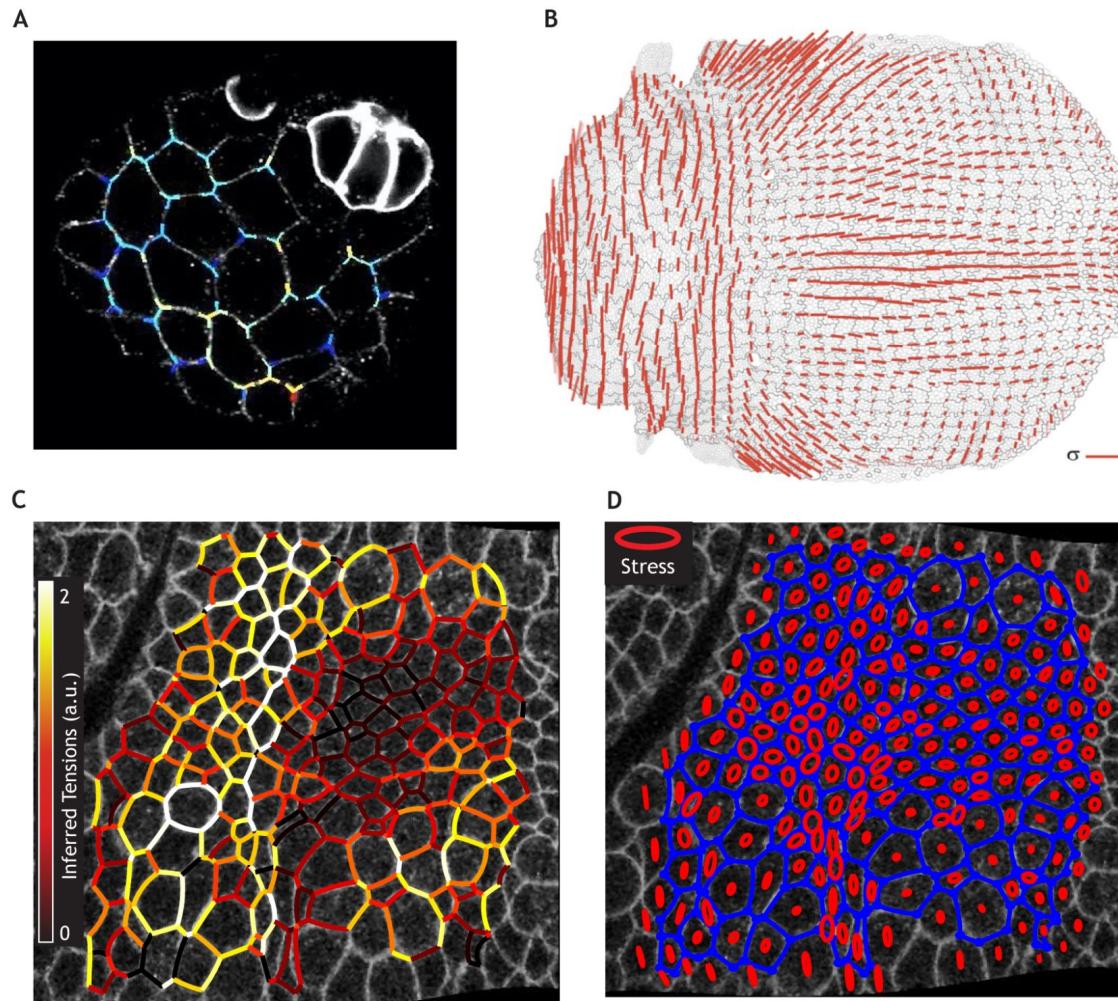


Ronneberger, Fischer, Brox, MICCAI, 2015



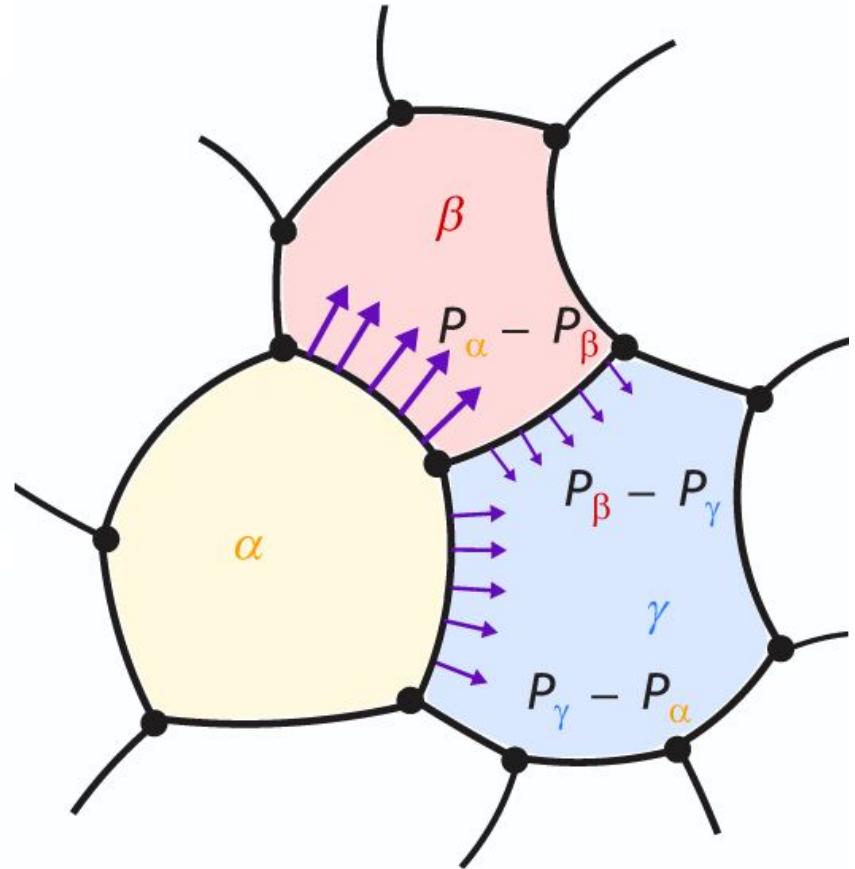
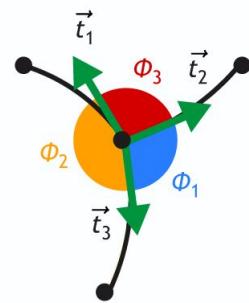
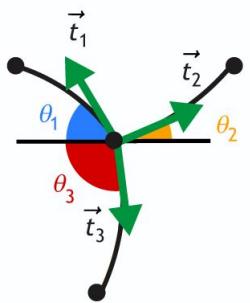
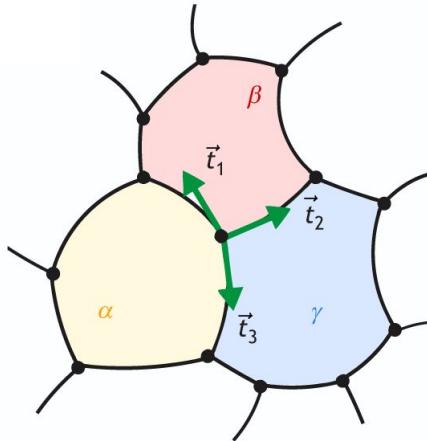
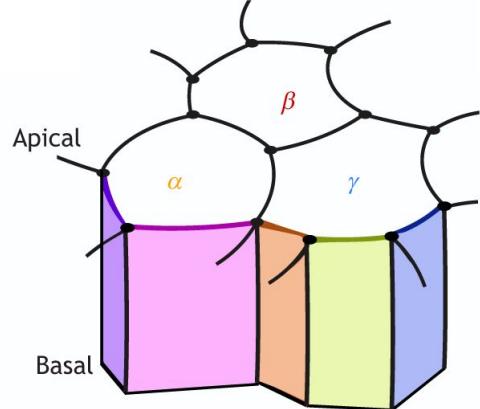
Lohoff et al., Nature Biotech., 2022 (seqFISH)

# Ingredient: inferring tension without measuring it



Examples of stress inference in development. Adapted from: Inferring cell junction tension and pressure from cell geometry. Roffay et al, 2021

# Ingredient: inferring tension without measuring it

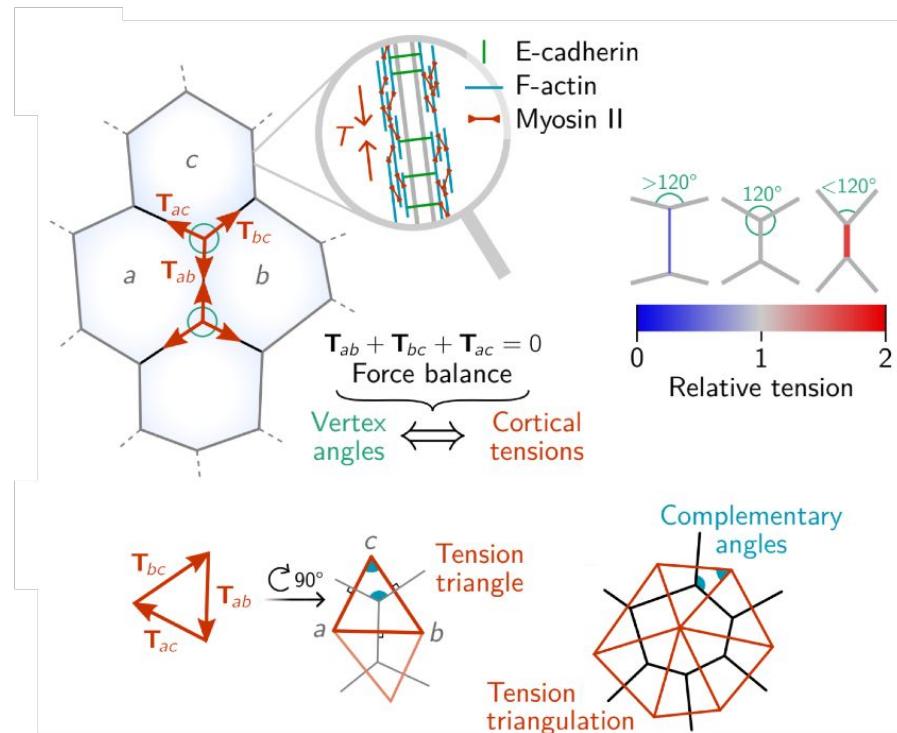
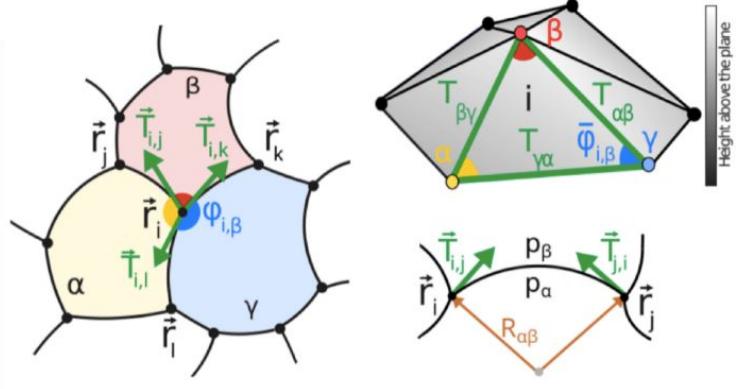


Adapted from: Inferring cell junction tension and pressure from cell geometry. Roffay et al, 2021

# Scale: morphological and mechanical phenotypes

## Ingredient: estimating tension without measuring it

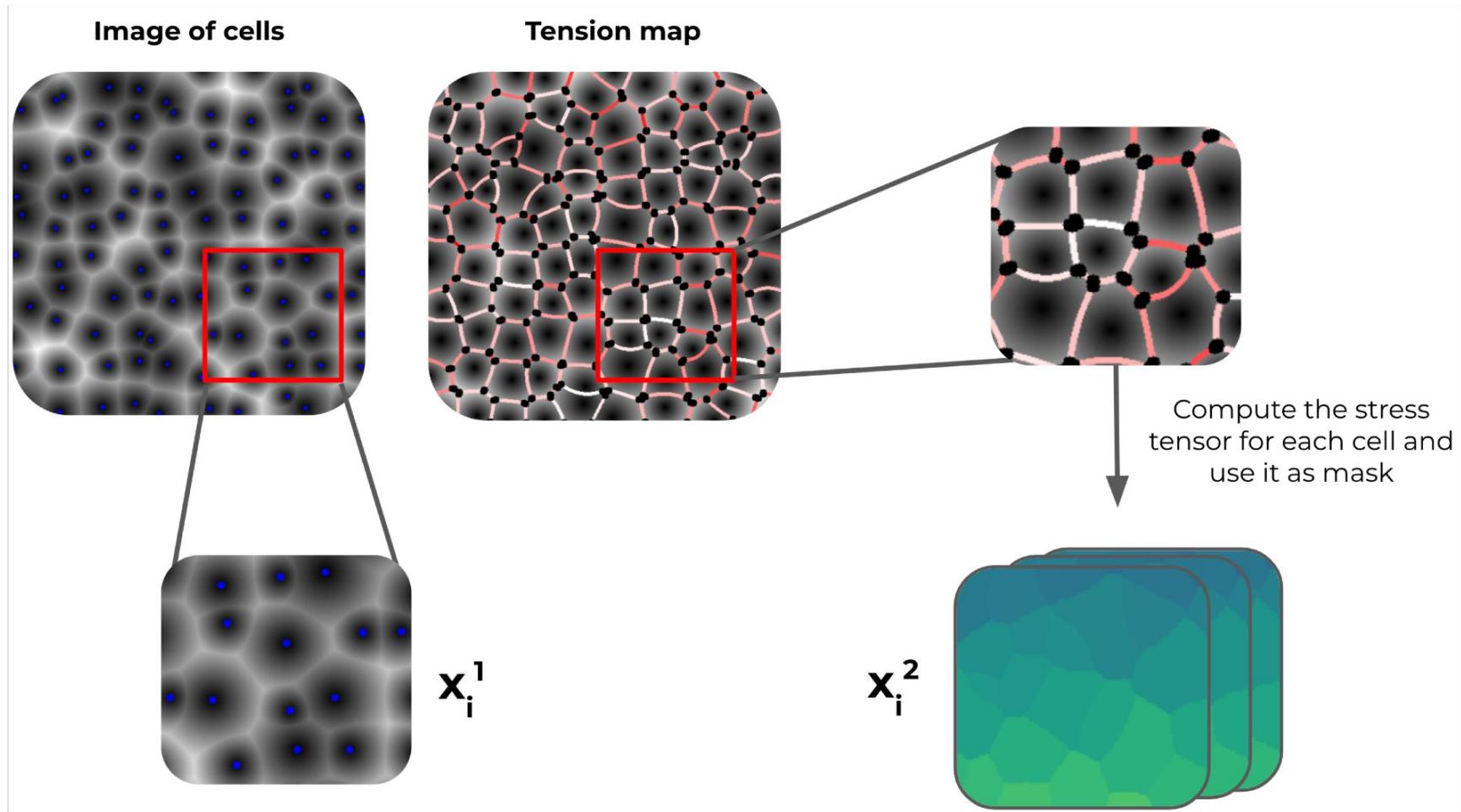
### Variational Methods of Stress Inference



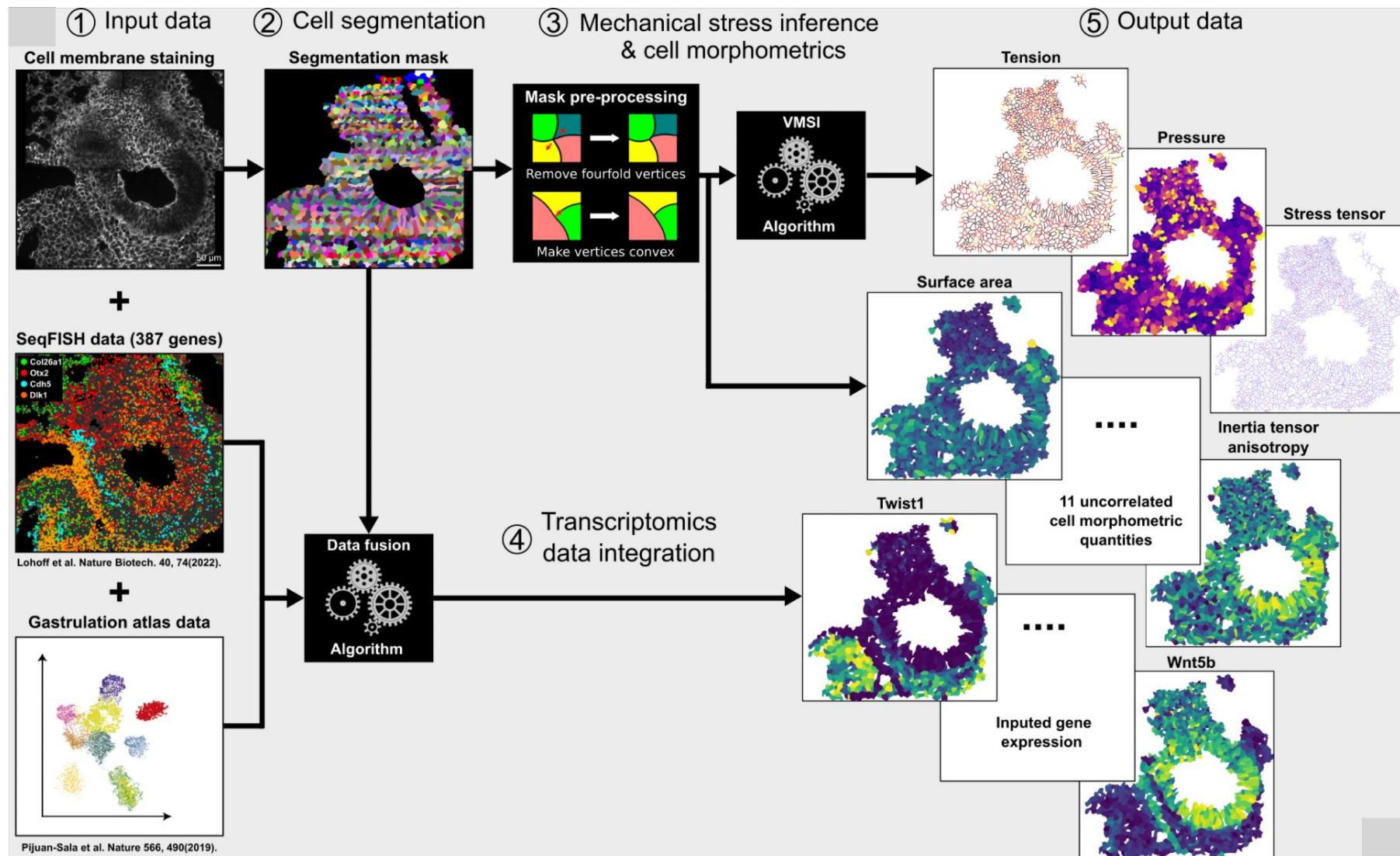
Noll, Streichan, Schraiman, Phys Rev X. 2020

Brauns, Claussen, Wieschaus, Schraiman, arxiv 2023

# Inferring tension from images: a pseudo-gene toy case



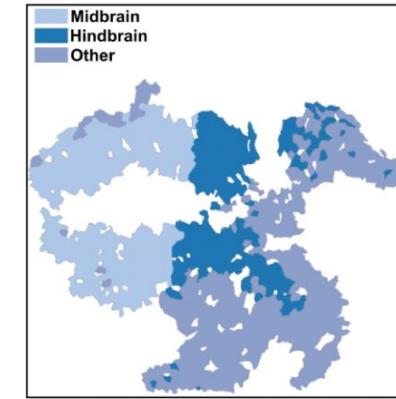
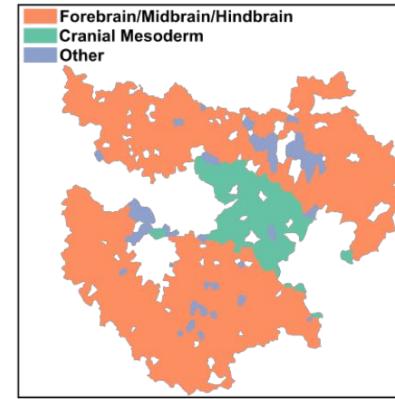
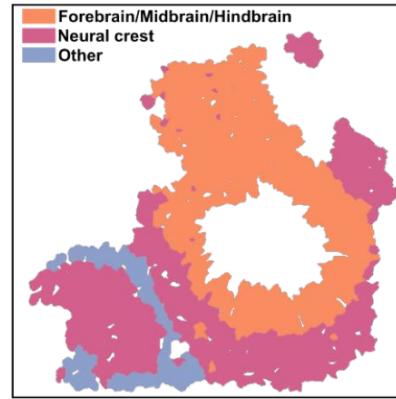
# A machine learning pipeline for multi-scale integration



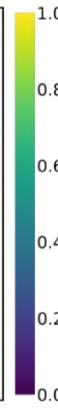
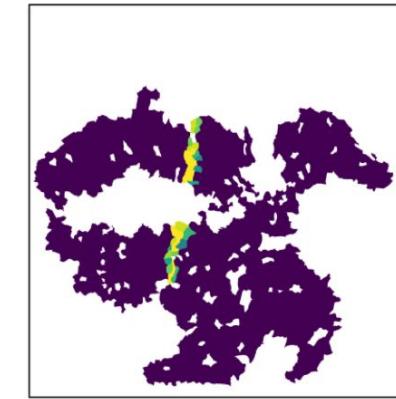
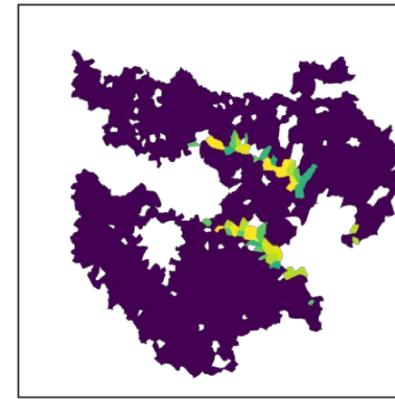
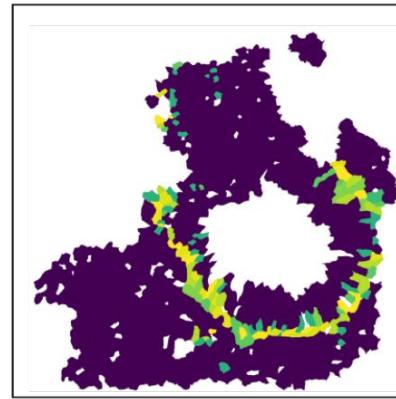
# Cell type boundaries recapitulate mechanical properties

Gene expression identified boundaries recapitulate tension driven tissue compartments

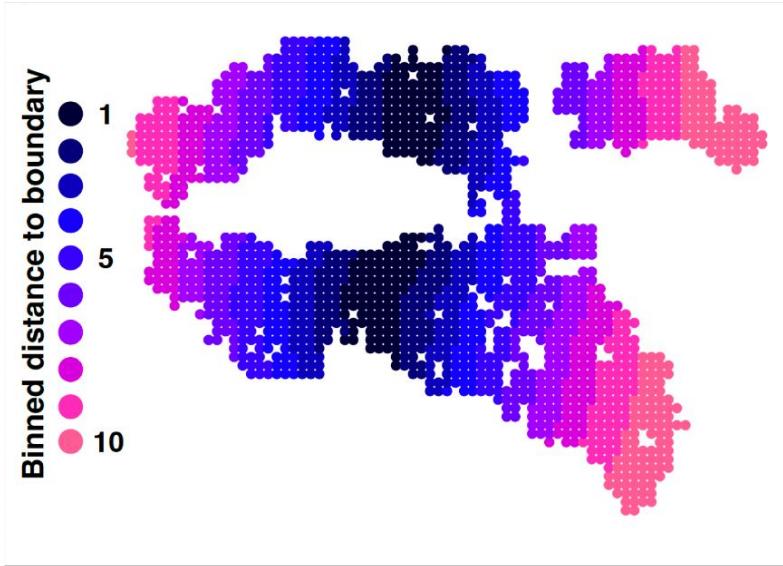
Cell regions



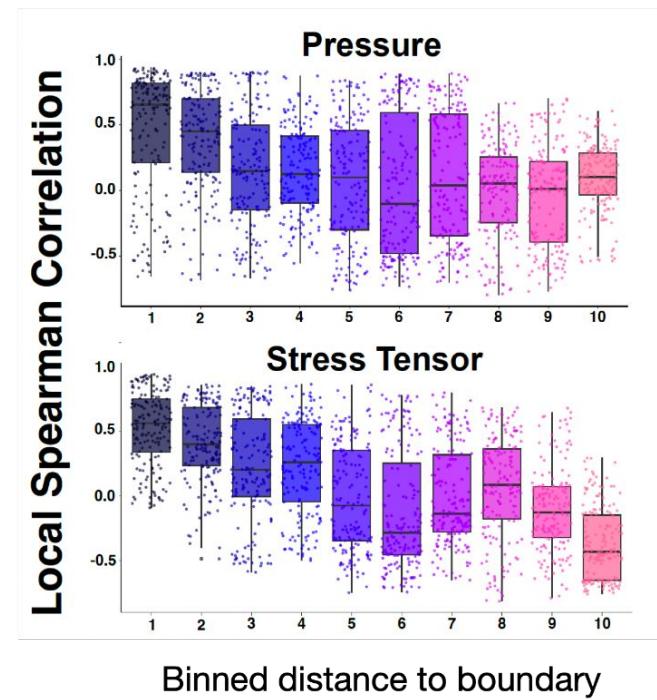
Boundaries



# Gene expression boundaries recapitulate tension defined compartments

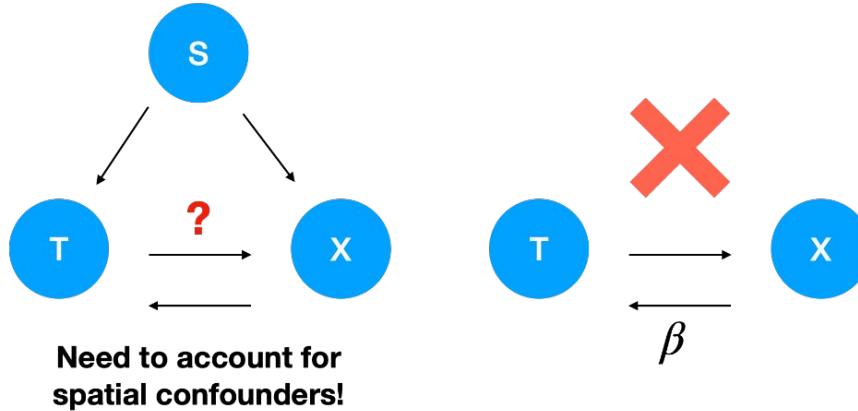


Midbrain-hindbrain region



# Spatial regression testing framework

Geoadditive structural equation (gSEM) models identify groups of genes associated with tension



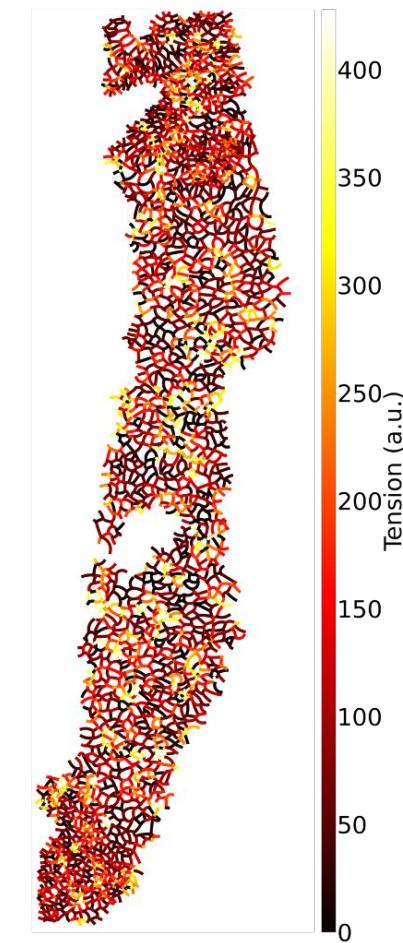
$$x_i^g = f^x(s_i) + \epsilon_i^x, \epsilon_i^x \sim N(0, \sigma^2 I)$$

$$t_i = f^t(s_i) + \epsilon_i^t, \epsilon_i^t \sim N(0, \sigma^2 I)$$

$$r_i^{x,g} = x_i^g - \hat{f}^x(s_i)$$

$$r_i^t = t_i - \hat{f}^t(s_i)$$

$$r_i^{x,g} = \beta_g r_i^t + \epsilon_i,$$

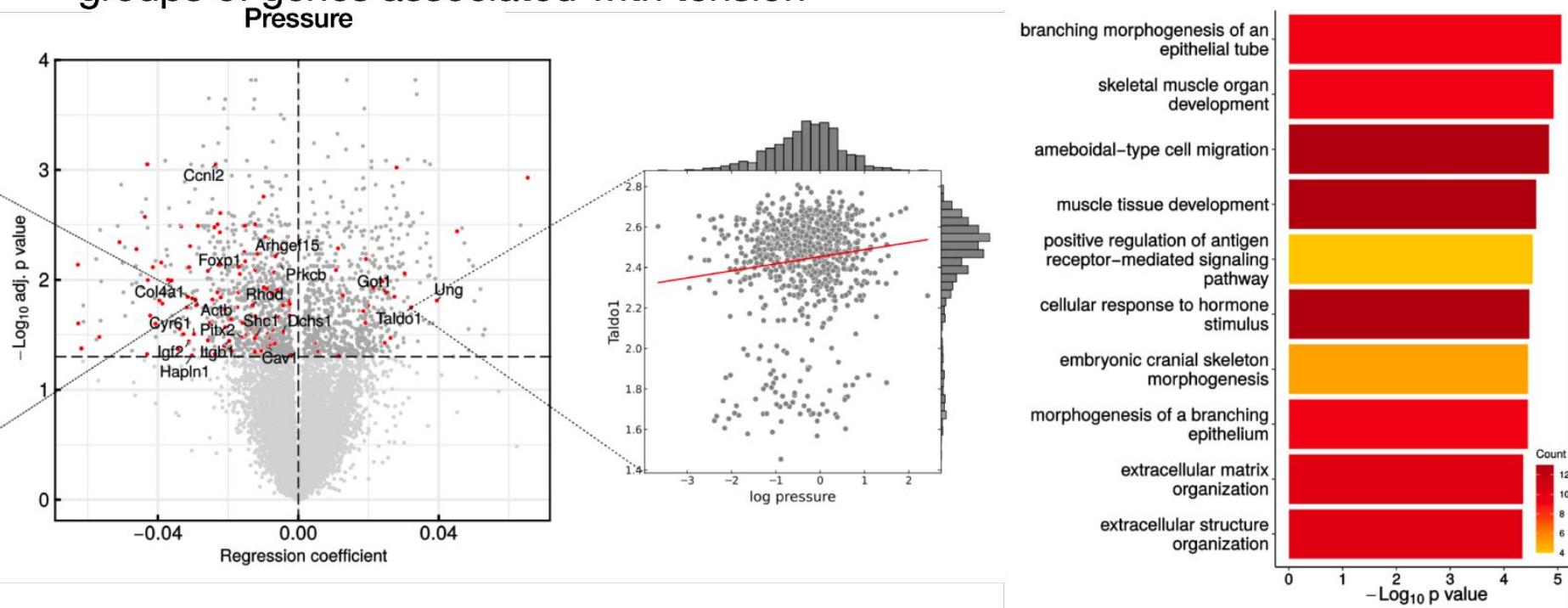


gSEM: Thaden and Kneib (2018)

Structural Equation Models for  
Dealing With Spatial Confounding

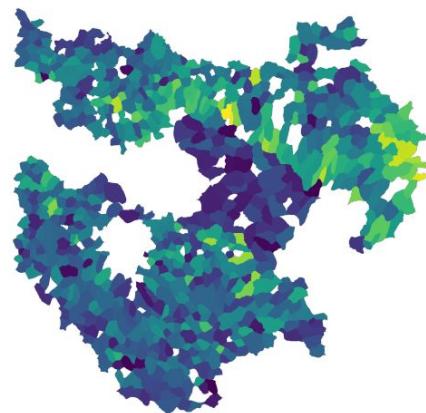
# Spatial regression testing framework

Geoadditive structural equation (gSEM) models identify groups of genes associated with tension

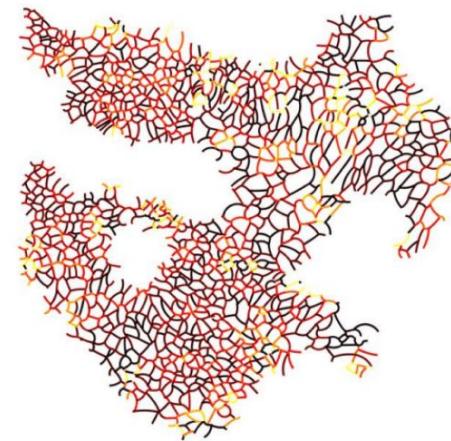
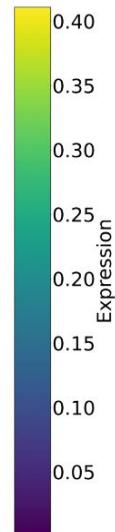


# Spatial gene variability informs tissue mechanical stability

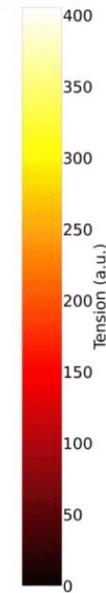
Geoadditive structural equation (gSEM) models identify groups of genes associated with tensional differences



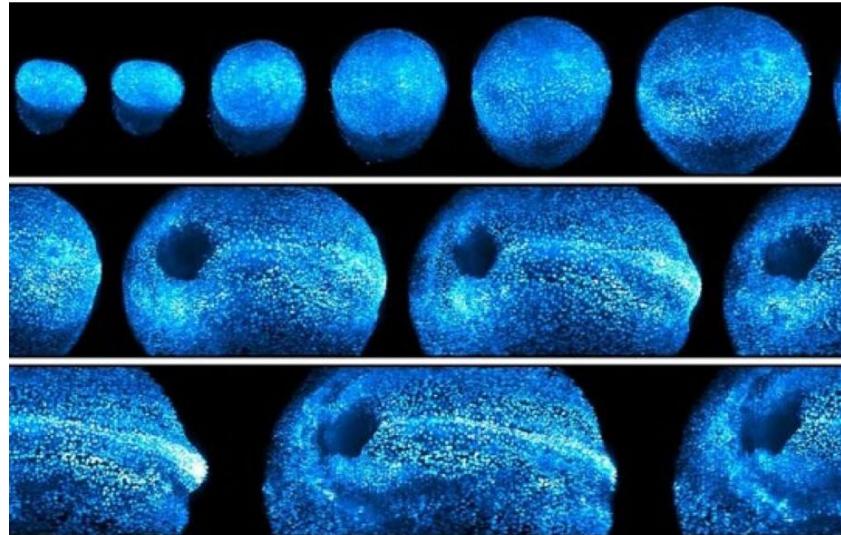
Ephb2 expression



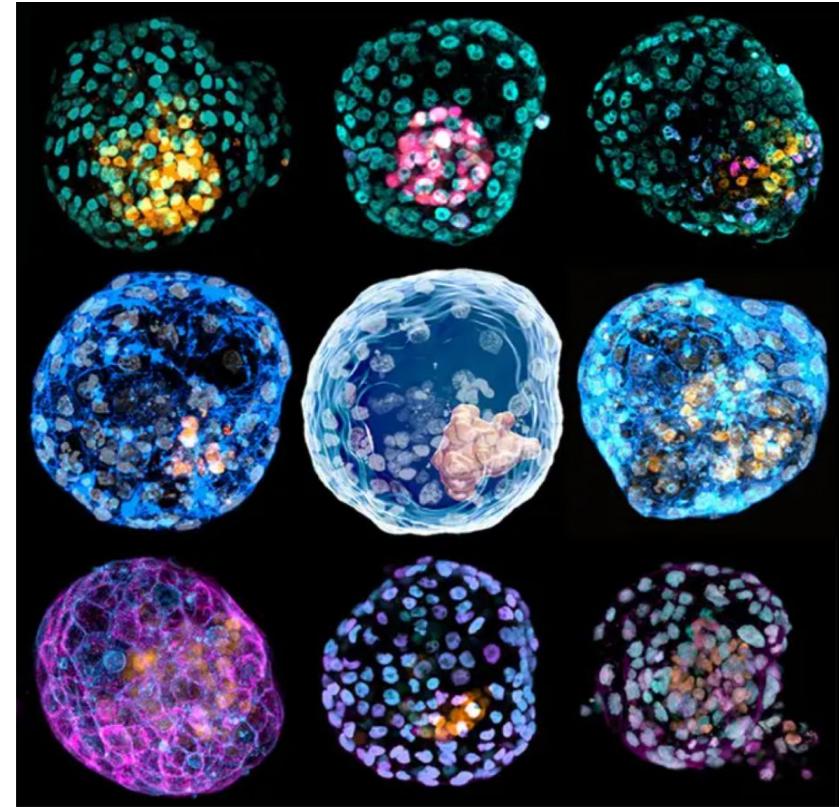
Tension across cell boundaries



# Integrating tension & genomics in a spatial context: Going forward



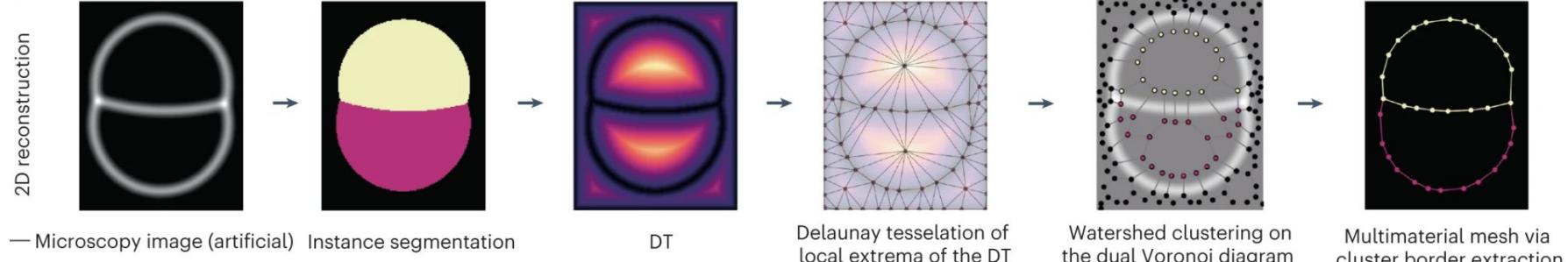
K McDole et al., Cell 2018



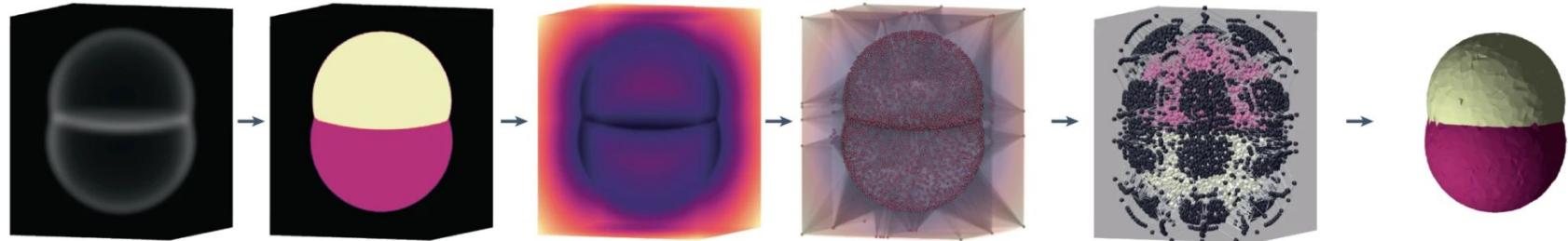
X. Liu et al., Nature 2021 (iBlastoids)

# Integrating tension & genomics in a spatial context: Going forward

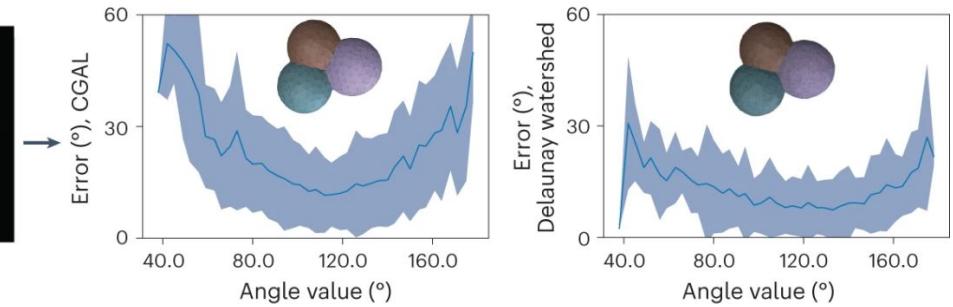
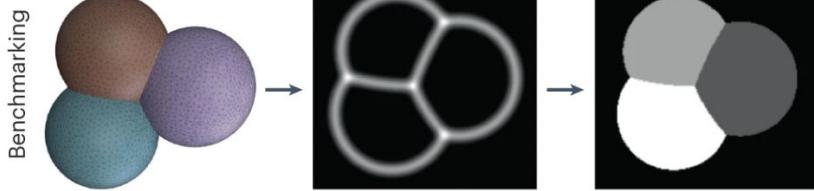
a



3D reconstruction



b

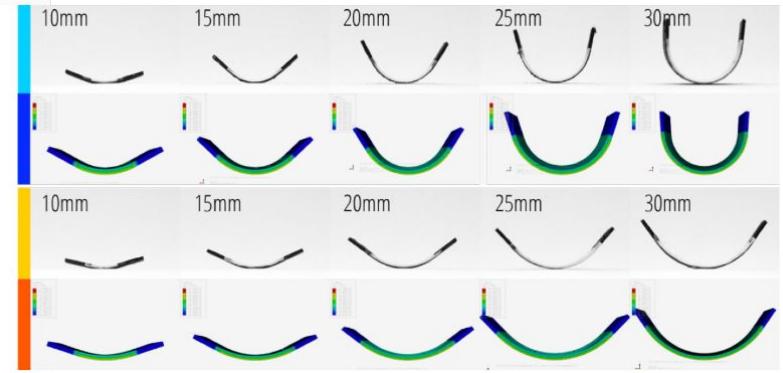


Multimaterial mesh generation algorithm. From: Embryo mechanics cartography: inference of 3D force atlases from fluorescence microscopy. Ichbiah et al, 2023

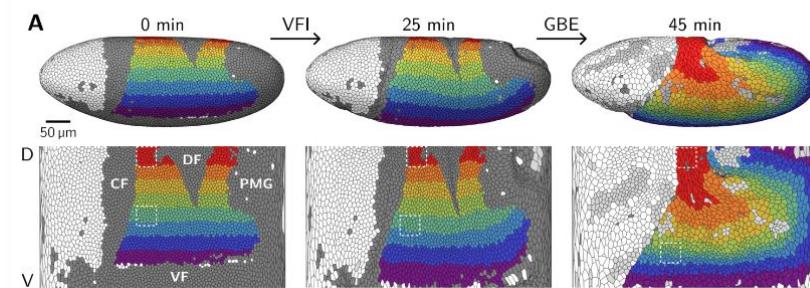
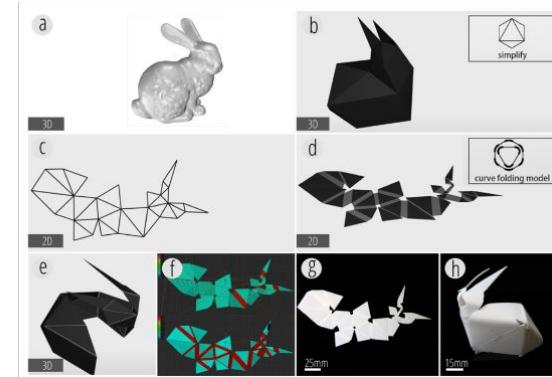
# Beyond representation learning: digital twins + modeling and inverting the dynamics of emergent systems



Source: Jo Nakashima



Thermorph: Democratizing 4D Printing of Self-Folding Materials and Interfaces; An et. al, 2018



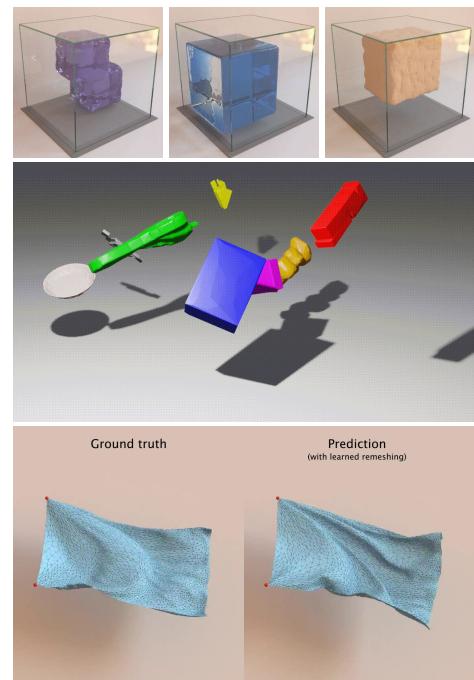
Brauns, Claussen, Wieschaus, Schraiman, arxiv 2023

## Next:

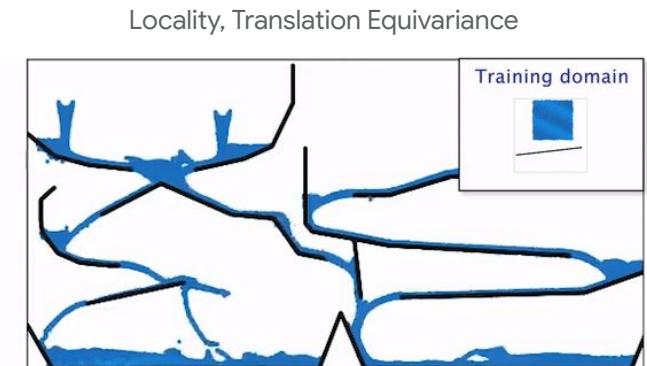
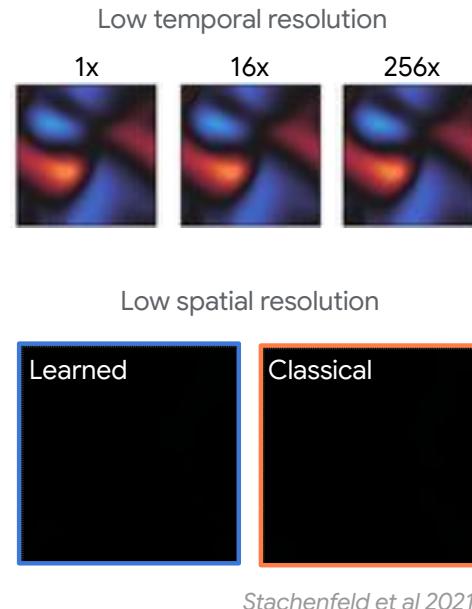
- \* Embryos as **materials**: where do the forces that fold embryos come from? Can we distinguish between active and passive forces?
- \* Embryos as **materials**: inverse design - which tensions are required to generate a final shape/ final tensional distribution?

# Beyond representation learning: digital twins + modeling and inverting the dynamics of emergent systems

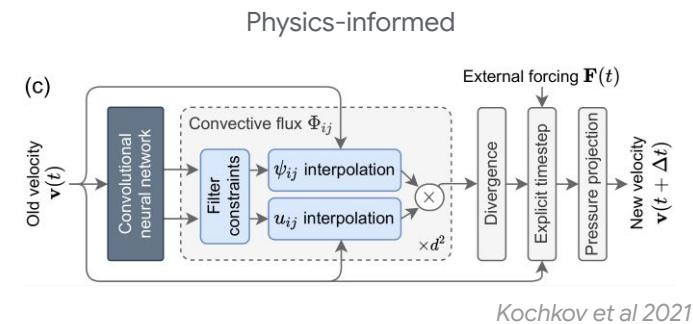
## AI for simulation



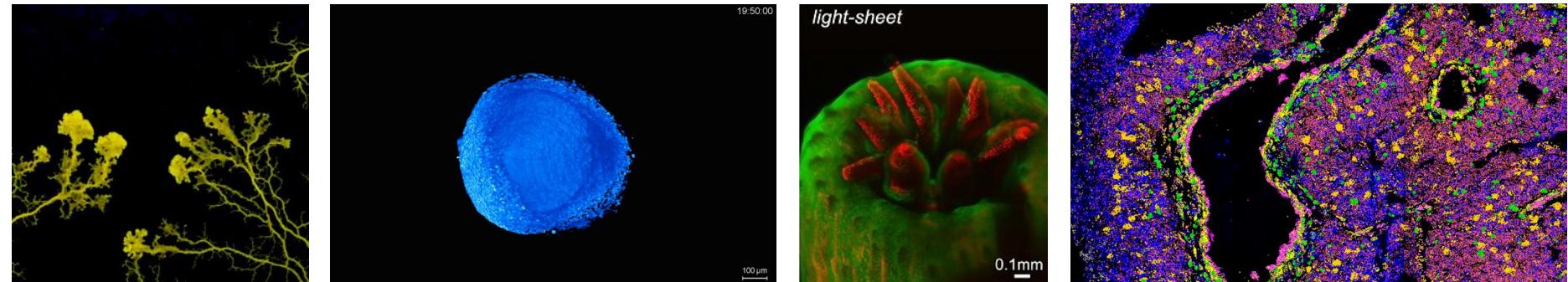
Mrowca et al (2018); Sanchez-Gonzalez et al. (2020); Pfaff et al (2021); Stachenfeld et al. (2022); Allen\*, Rubanova\* et al. (2023); Wu et al (2023)



Pfaff et al 2021



# Beyond representation learning: digital twins + modeling and inverting the dynamics of emergent systems



# Thank you!



More @ <https://www.morpho-lab.com/>; illustration by Elena Bansh

**\*\* Accepting postdoc and grad student applications ! \*\***  
**Get in touch: bmd2151@columbia.edu**



# From detecting patterns to interrogating how they are formed

Two historical ideas: Turing and Wolpert

# From detecting patterns to interrogating how they are formed

Latent ODE models

# From detecting patterns to interrogating how they are formed

Wombling

# From detecting patterns to interrogating how they are formed Conway's Game of Life

# Neural Cellular Automata

# Experimental design: building better atlases