

# The Single-Cell Spatial Transcriptomic Analysis (ScSTA) Cookbook

Arun Das

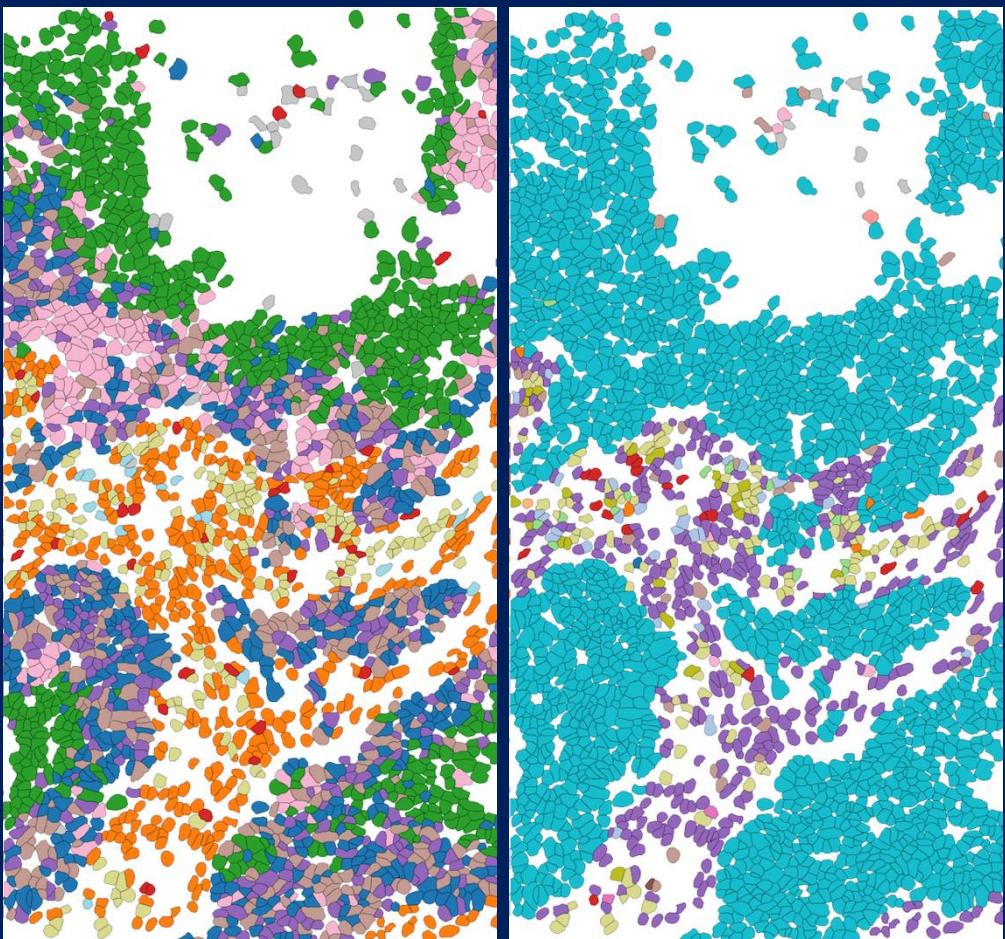
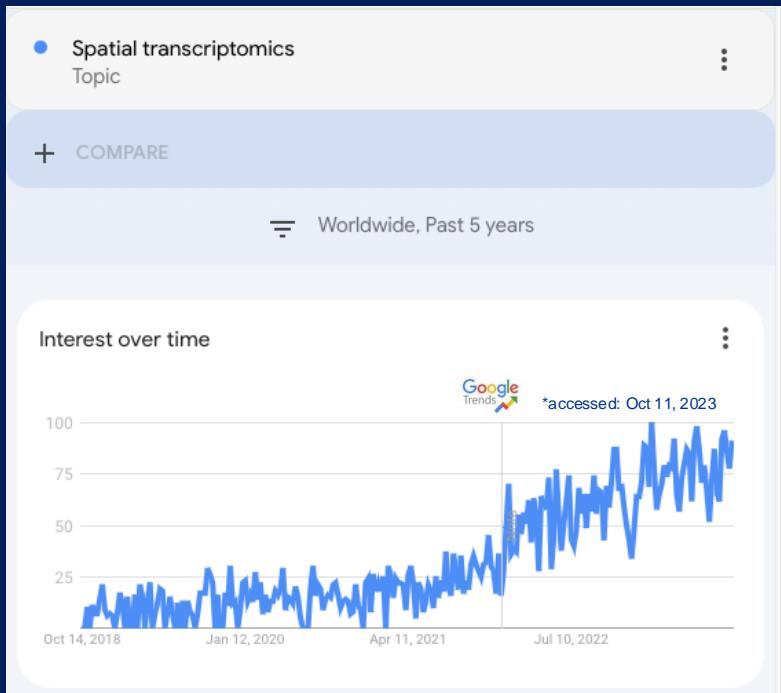
Postdoctoral Associate

Hillman Cancer Center

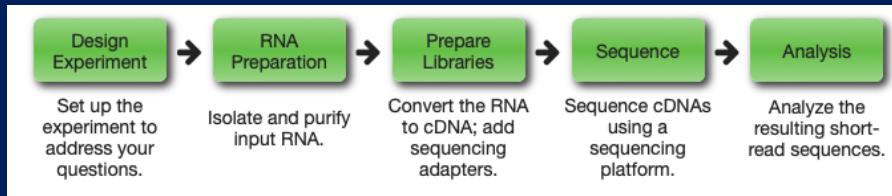
University of Pittsburgh Medical Center

“Spatial transcriptomics is an attempt to quantify mRNA expression of large number of genes within the spatial context of tissues and cells.”

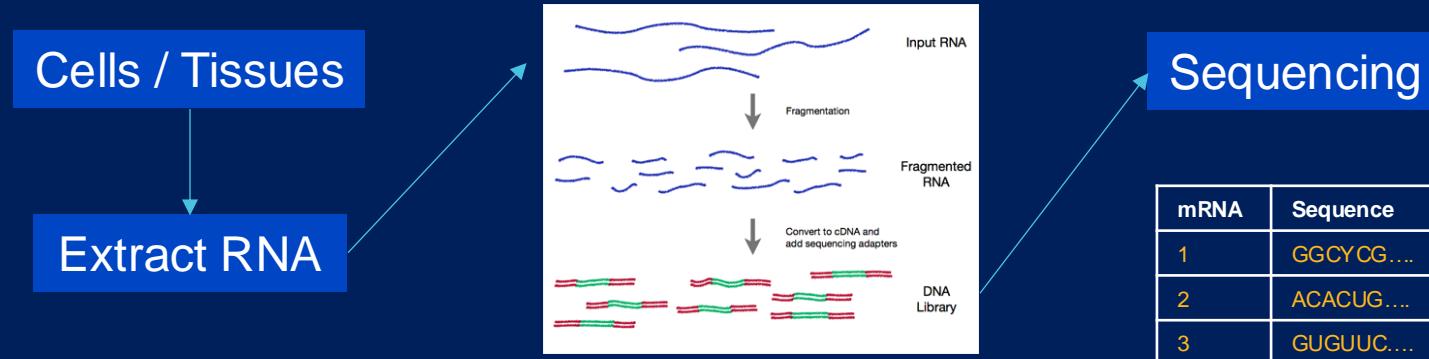
Pachter et al. 2022



# The 'How' – Sequencing based.



We lost the spatial context though!



# The ‘How’ – Imaging based - FISH.

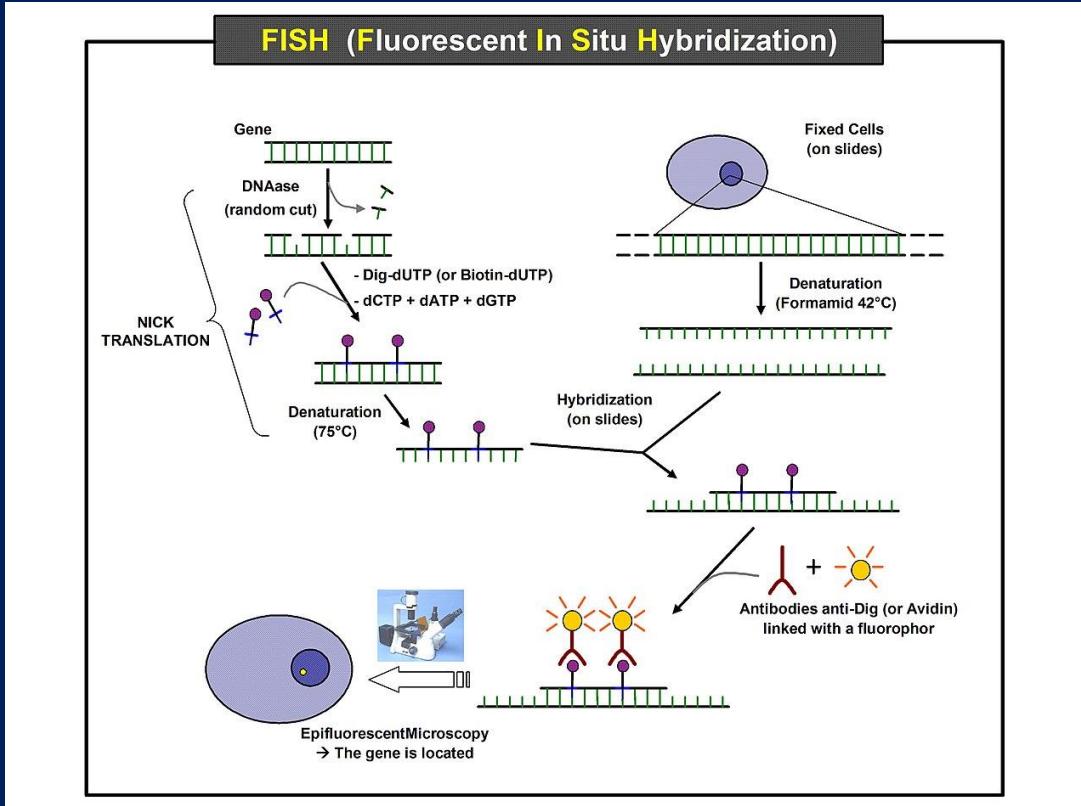
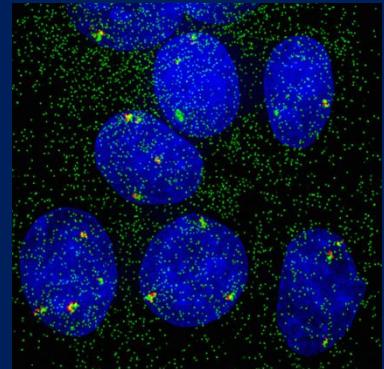
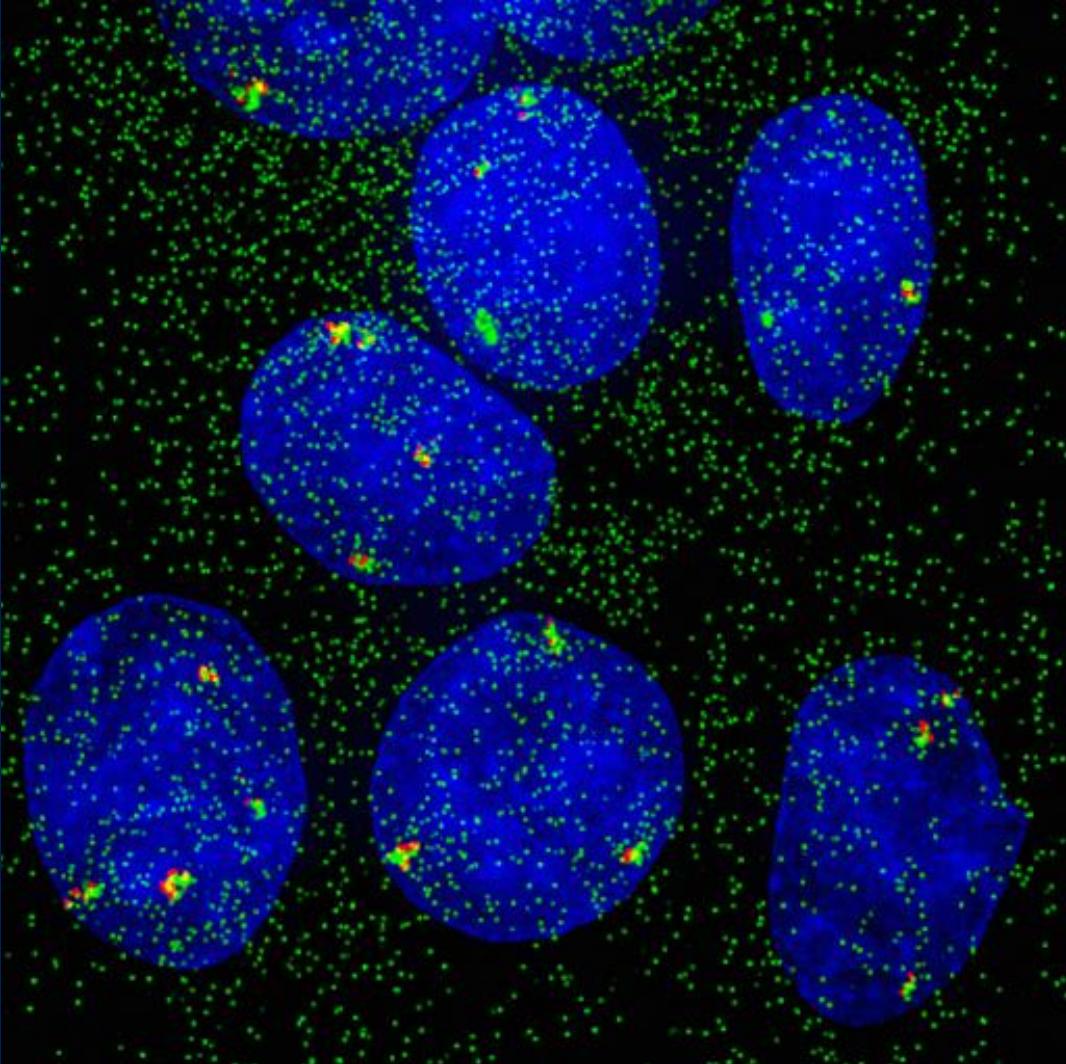


Image from Wikipedia



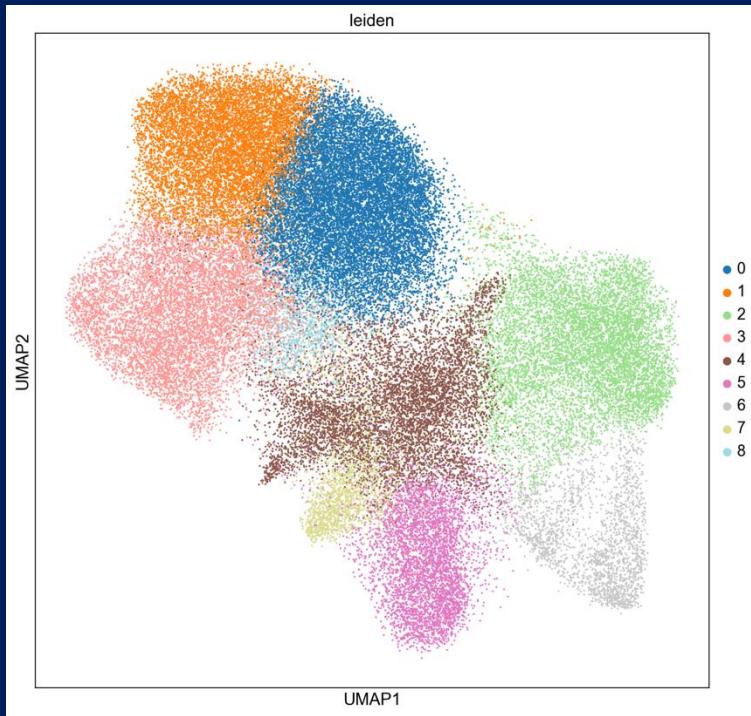
GREB1

Stossi et al. 2020

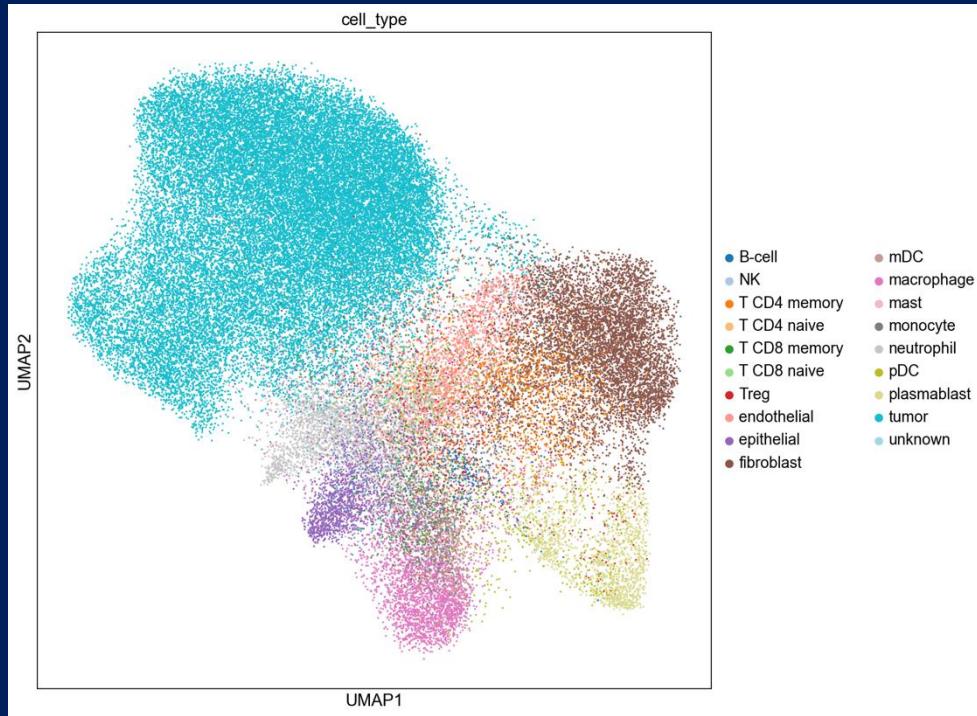


Cell id	G1	g2
1	100	200
2	10	400
..		

# Why do we need to study gene expressions?



# Why do we need to study gene expressions?



# Why do we need to study gene expressions?

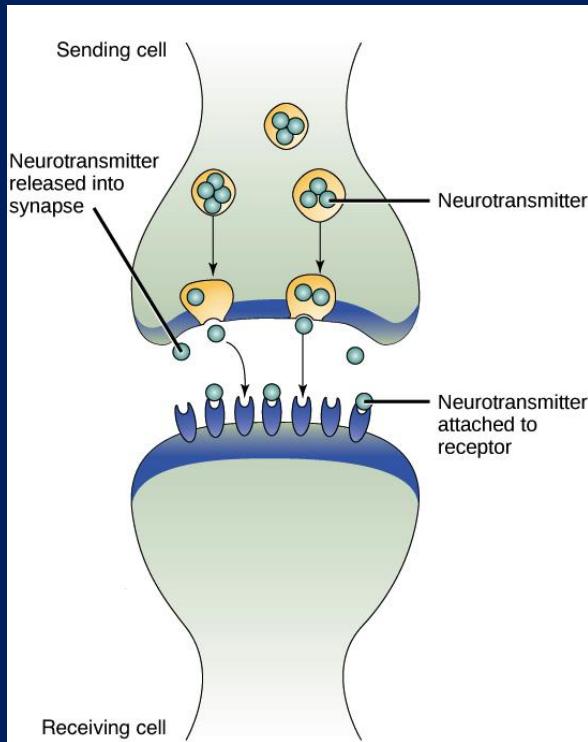
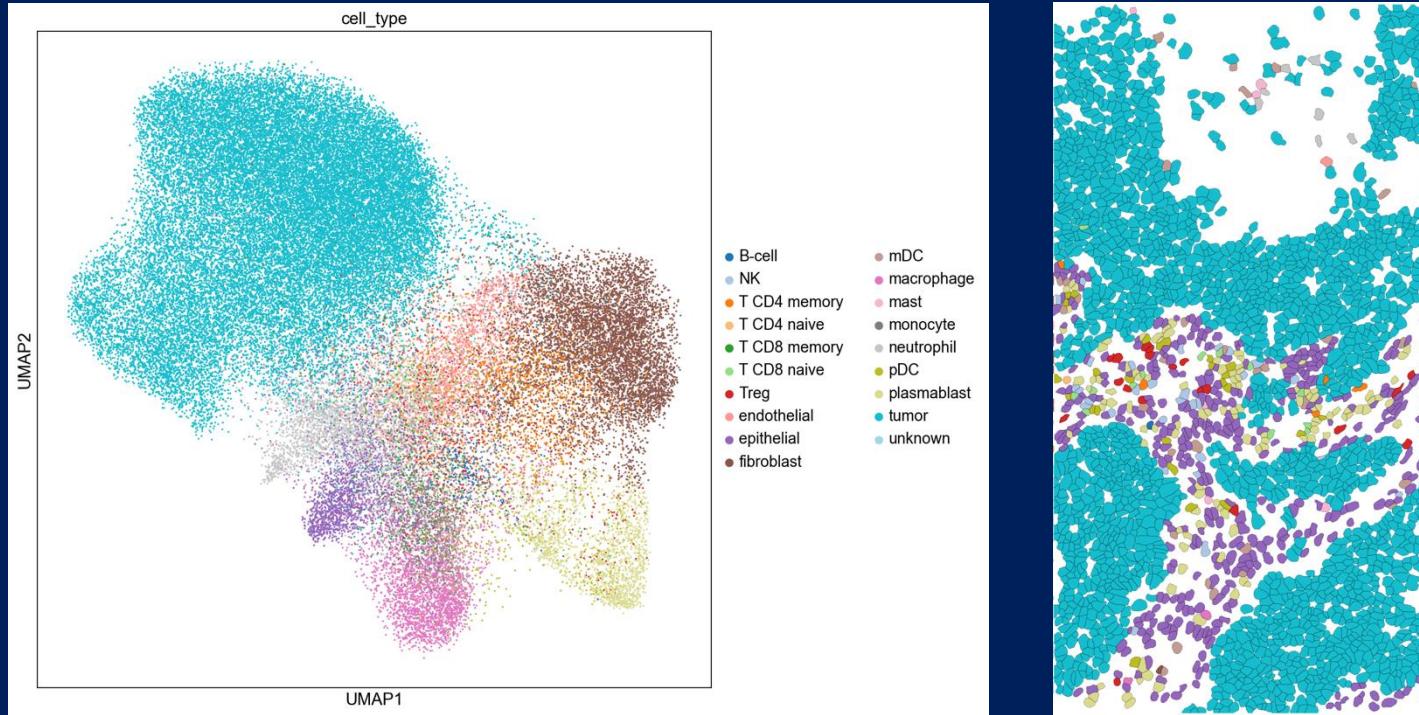
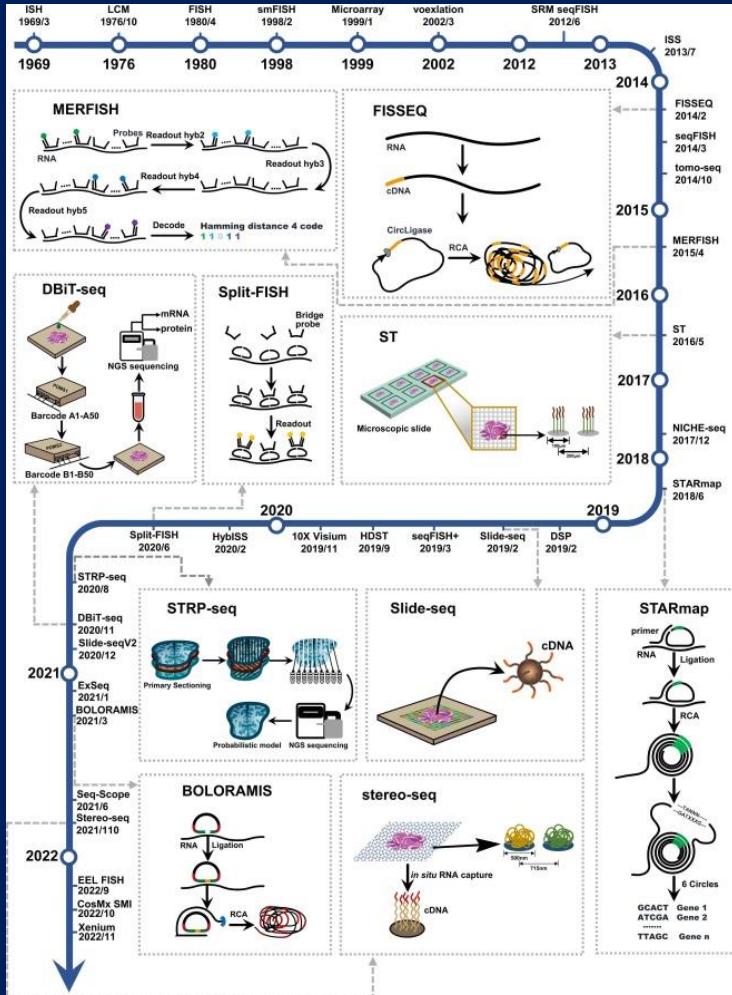
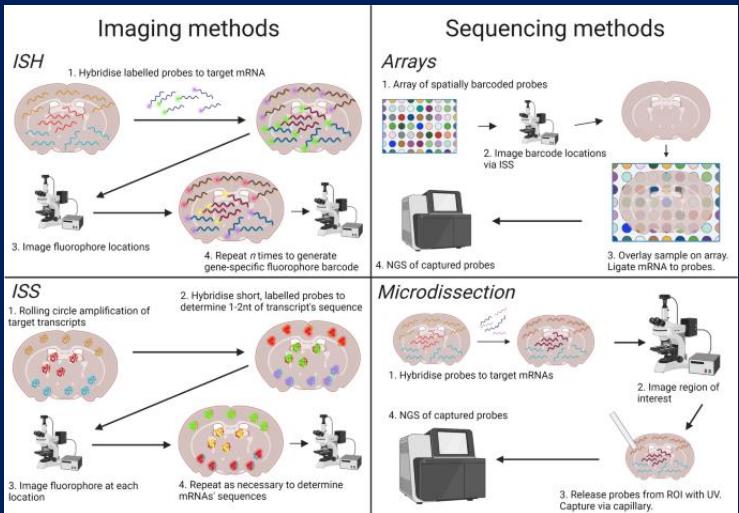


Image from Khan Academy

# Why do we need to study the “**spatial context**” of gene expressions?



# Current State-of-the-Art in Spatial Transcriptomics



# Why Single-cell Resolution?



A screenshot of the Single Cell Portal website. The header includes the logo and name 'Single Cell PORTAL' with the tagline 'Reducing barriers and accelerating single-cell research'. A central callout box states 'Featuring 611 studies 37,161,643 cells'. Below the header are search and filter options for studies and genes, along with a 'Browse collections' button. At the bottom, it shows '611 total studies found'.

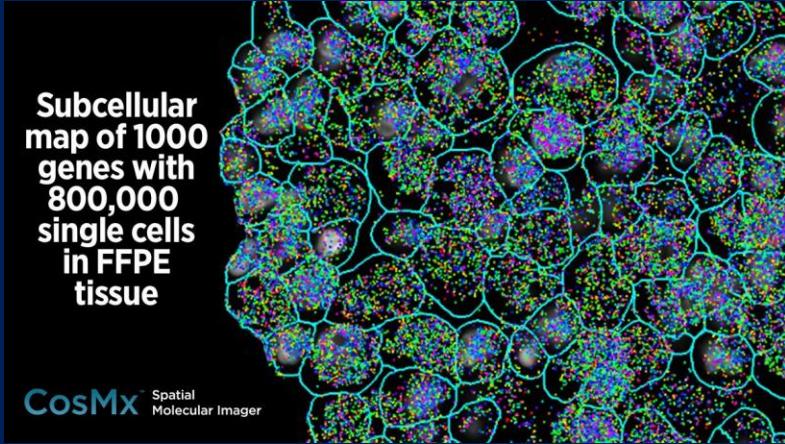
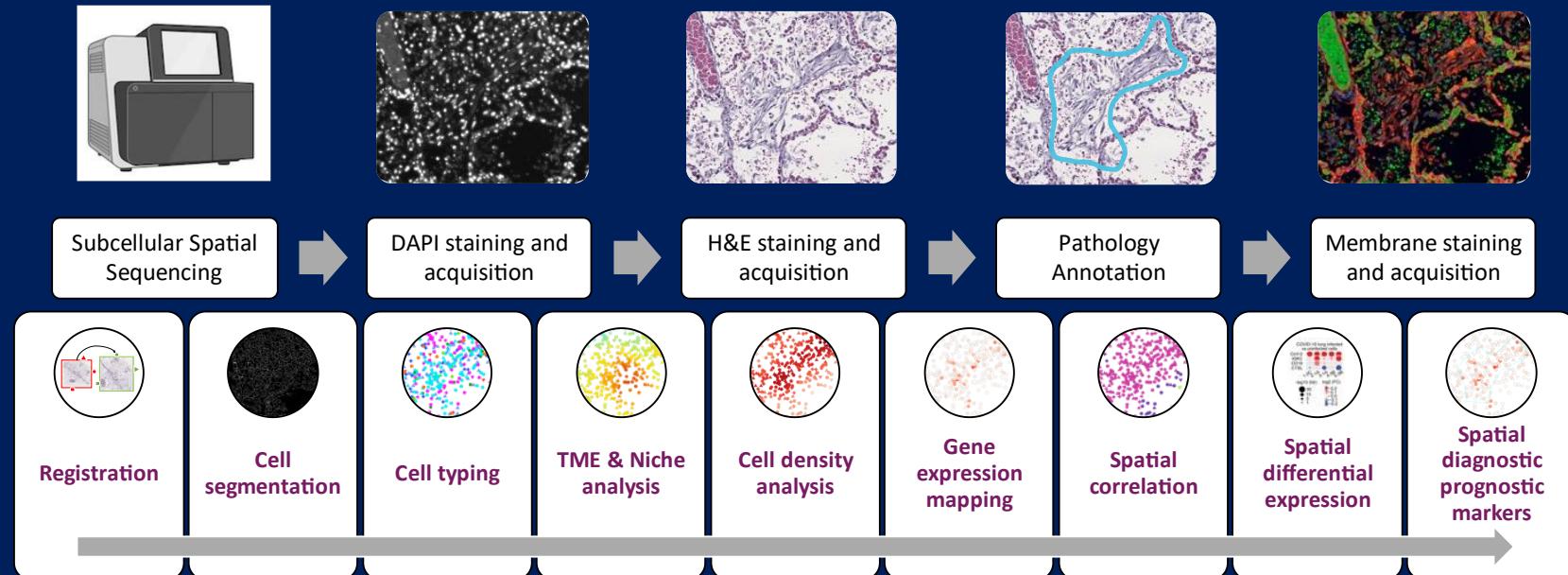


Image from Nanostring

# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



JOURNAL OF  
**MEDICAL VIROLOGY**

RESEARCH ARTICLE

## **Molecular and immune signatures, and pathological trajectories of fatal COVID-19 lungs defined by *in situ* spatial single-cell transcriptome analysis**

Arun Das, Wen Meng, Zhentao Liu, Md Musaddaqul Hasib, Hugh Galloway, Suzane Ramos da Silva, Luping Chen, Gabriel L. Sica, Alberto Paniz-Mondolfi, Clare Bryce, Zachary Grimes, Emilia M. Sordillo, Carlos Cordon-Cardo, Karla Paniagua Rivera, Mario Flores, Yu-Chiao Chiu, Yufei Huang✉, Shou-Jiang Gao✉ ... See fewer authors ▾

First published: 10 August 2023 | <https://doi.org/10.1002/jmv.29009>

Arun Das and Wen Meng contributed equally to this study.

Das, A, Meng, W, Liu, Z, et al. Molecular and immune signatures, and pathological trajectories of fatal COVID-19 lungs defined by *in situ* spatial single-cell transcriptome analysis. *J Med Virol*. 2023; 95:e29009. doi:10.1002/jmv.29009

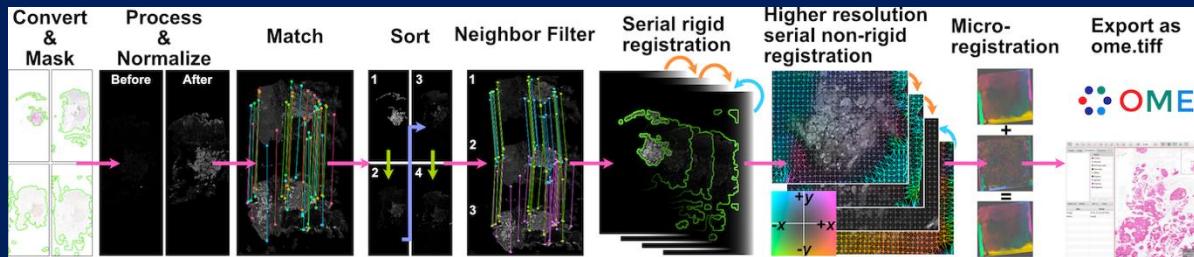
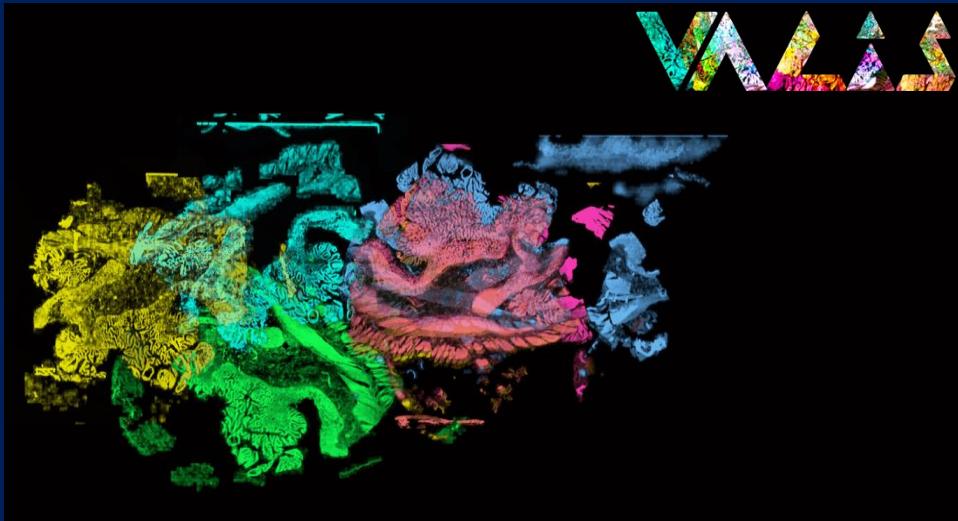
<https://zenodo.org/records/7636104>

# Image Registration



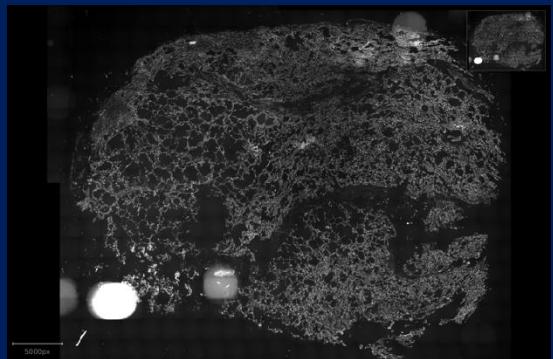
- Define a measure to quantify the similarity between two images.
- Apply affine transformations to morph the moving image.
- Optimize the similarity metric by tuning the affine transformations.

# Image Registration

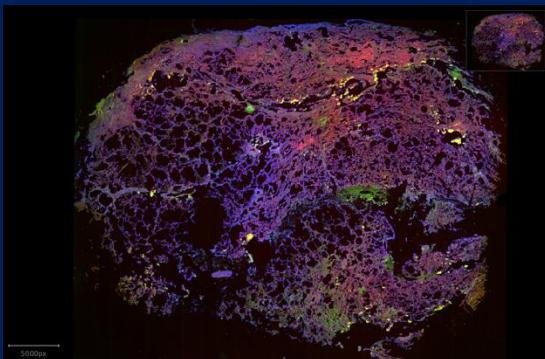


<https://valis.readthedocs.io/en/latest/>

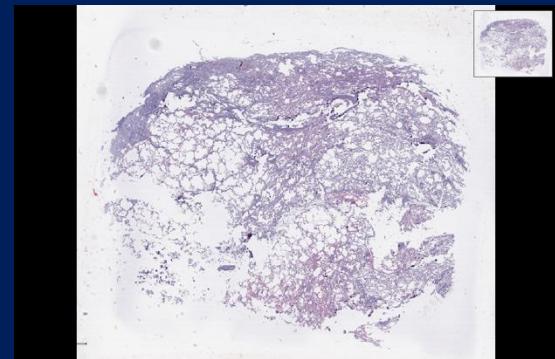
DAPI



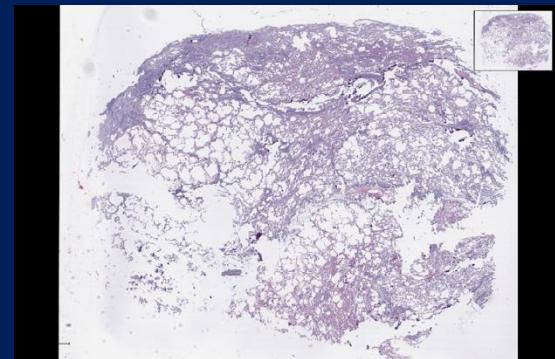
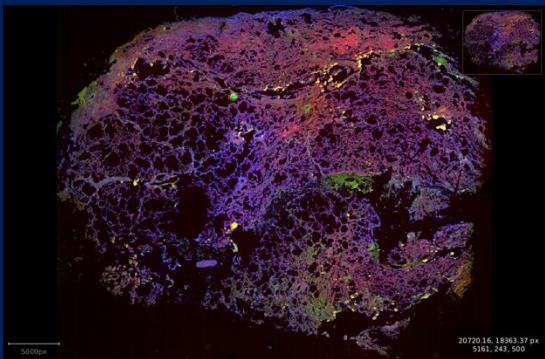
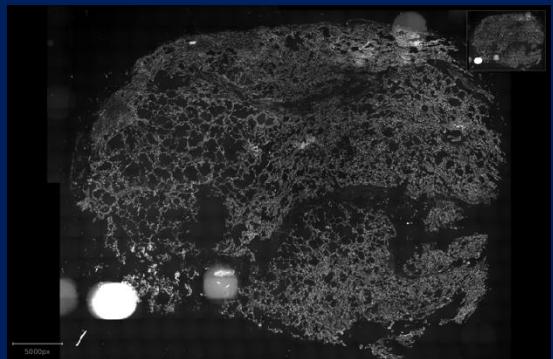
Membrane Staining



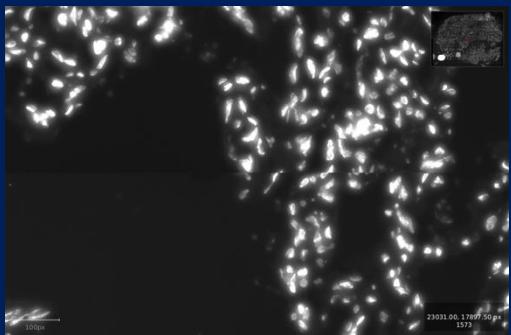
H&E Imaging



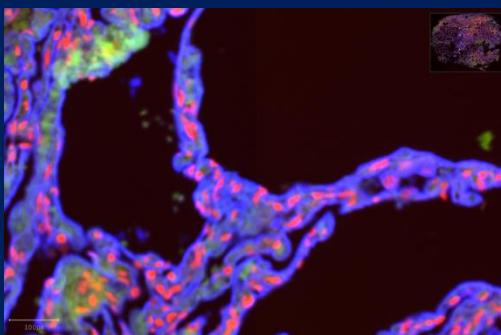
Spot any differences?



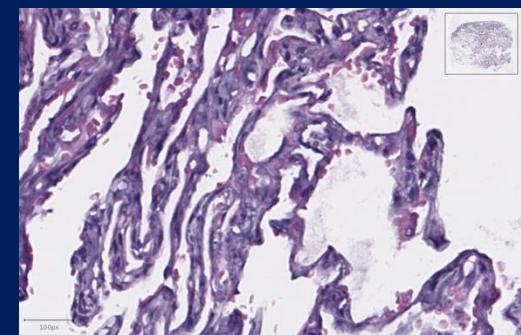
DAPI



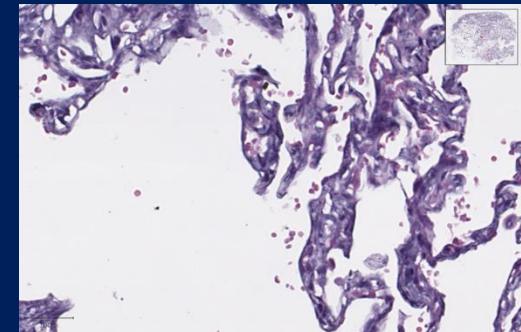
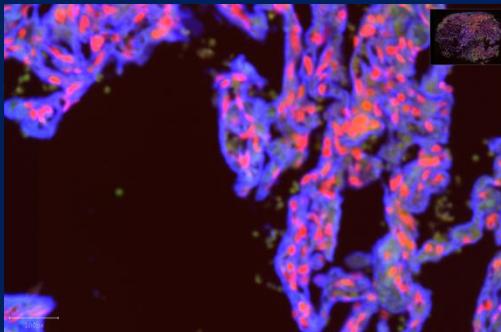
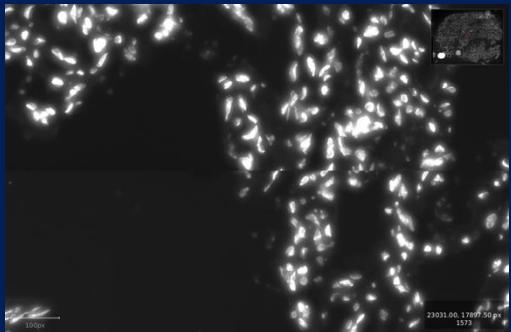
Membrane Staining



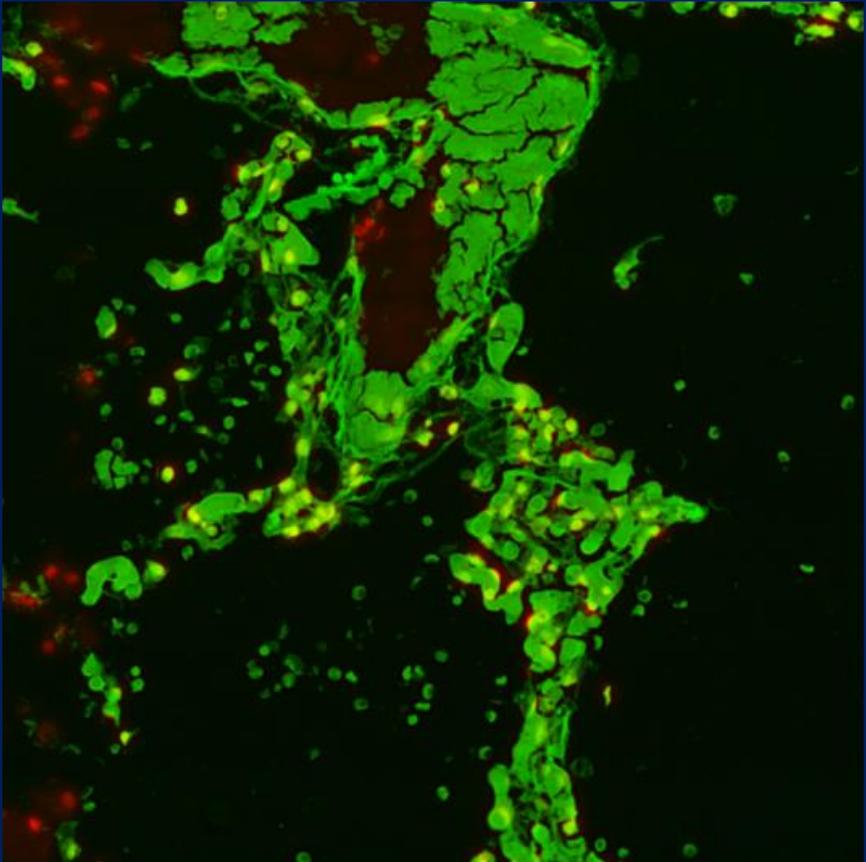
H&E Imaging



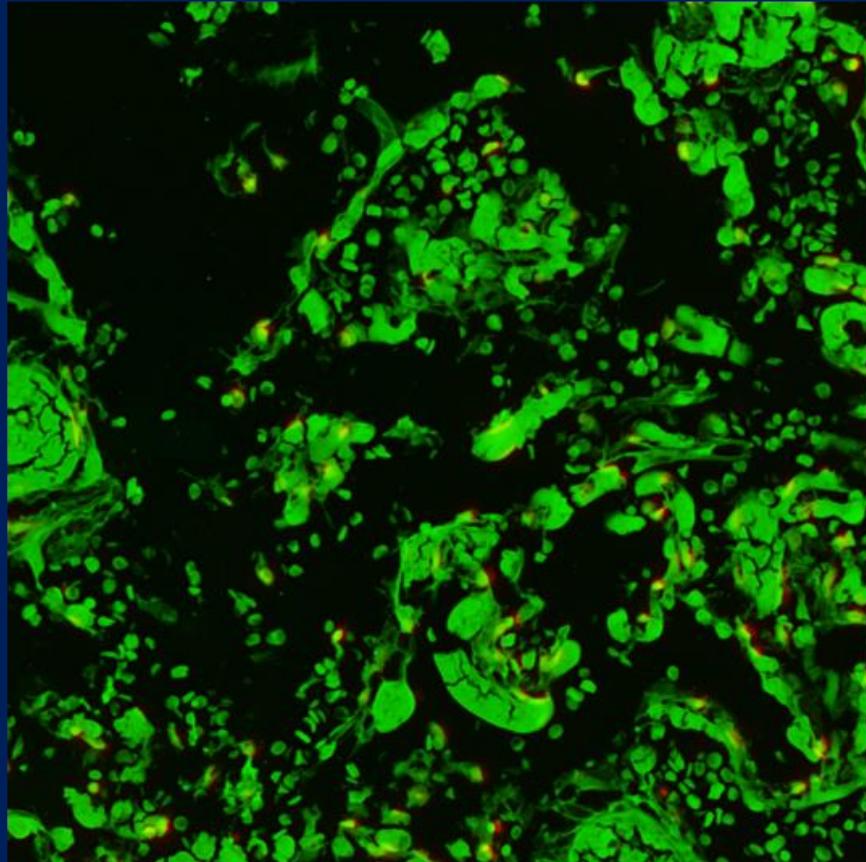
Spot any differences now?

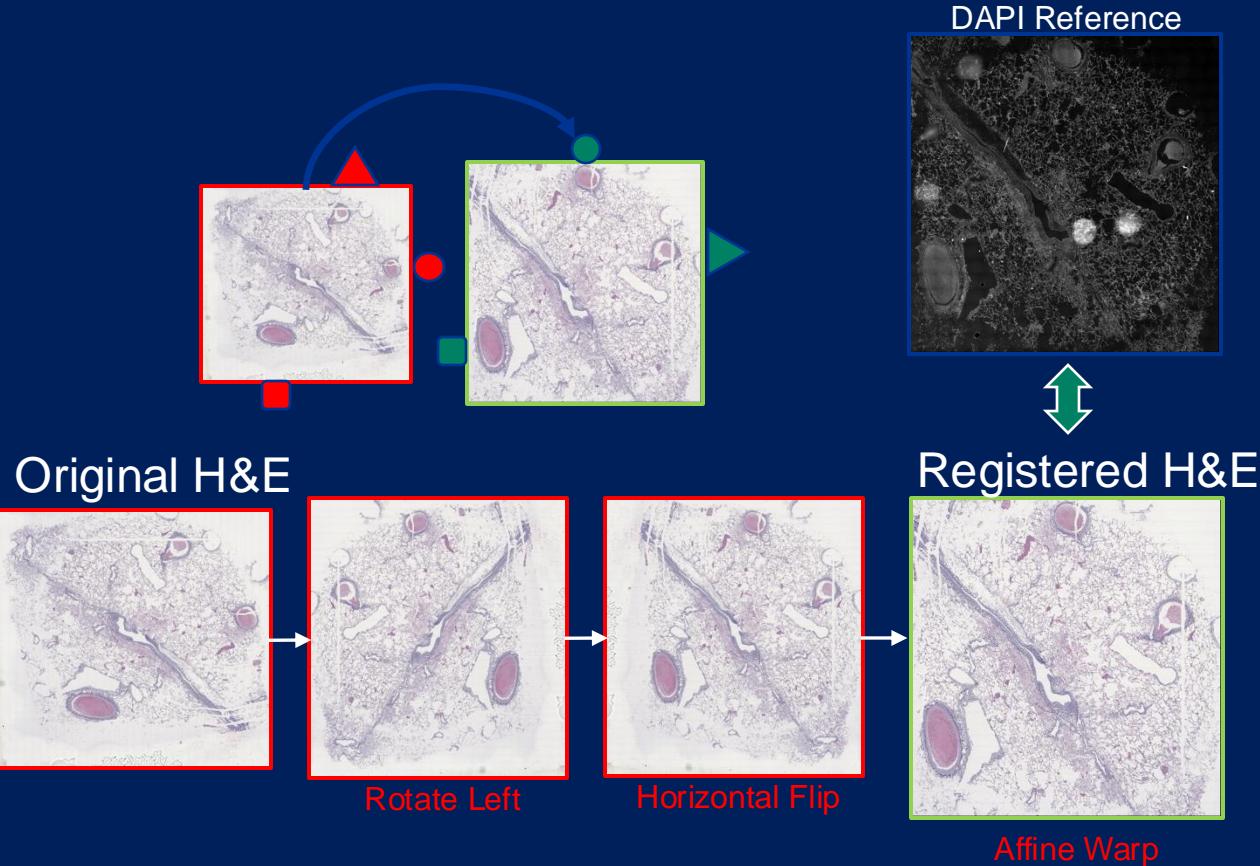


1-2C MATLAB\_VALIS Registration H&E  
Red Channel: DAPI, Blue Channel: Zero  
Green Channel: gray image of H&E

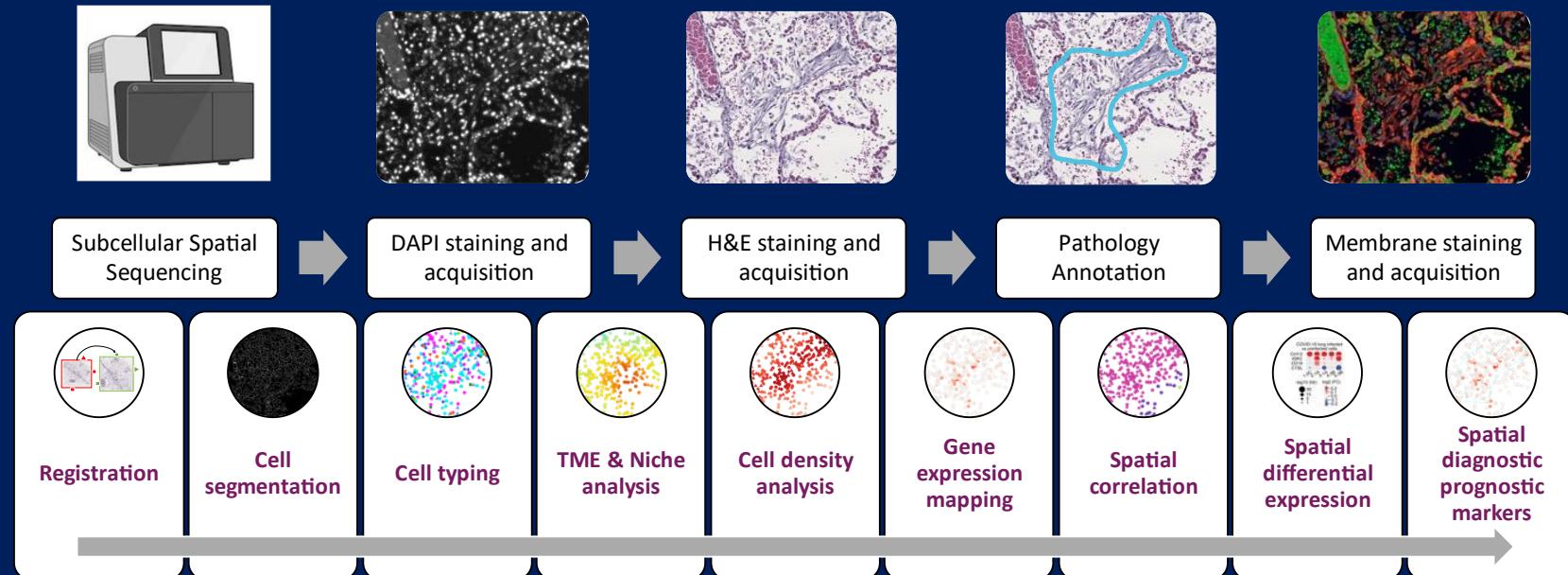


4-3B MATLAB\_VALIS Registration H&E  
Red Channel: DAPI, Blue Channel: Zero  
Green Channel: gray image of H&E





# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



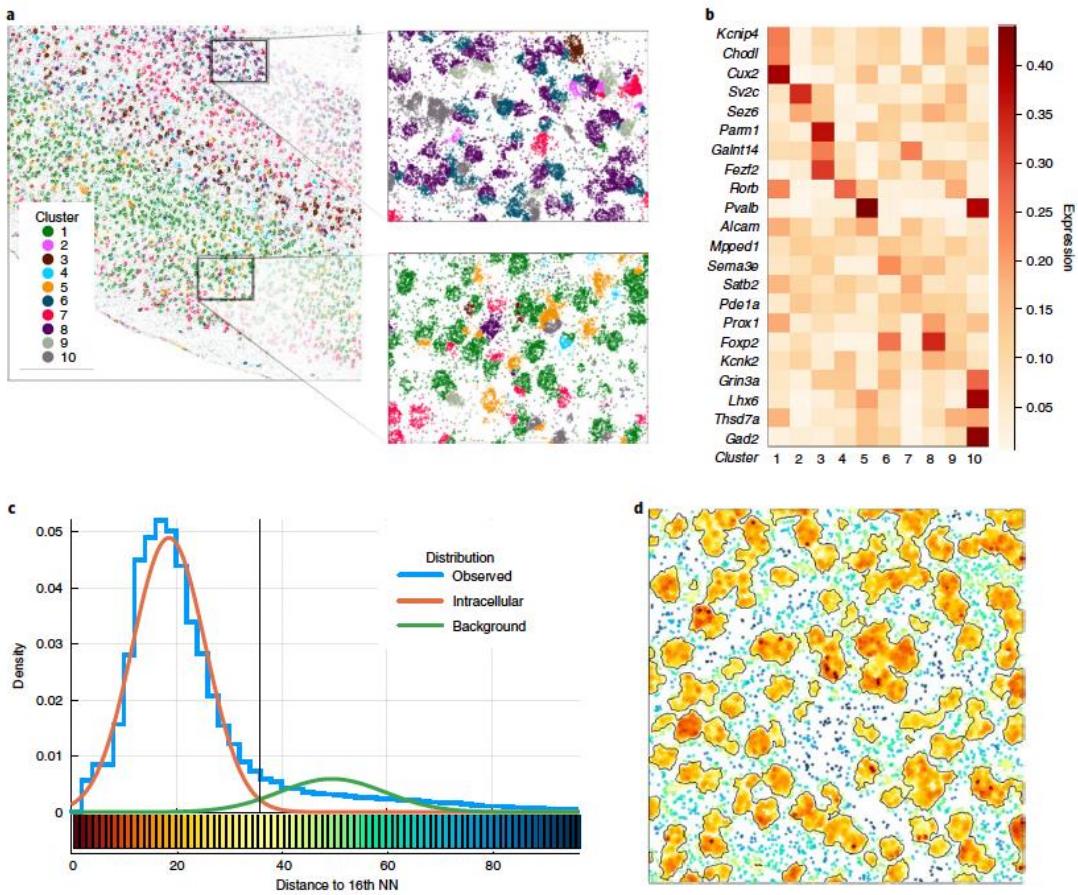
# Cell Segmentation

- Image-based – CellPose
- Transcriptomics-based – Baysoar

# Baysor Cell Segmentation

- Segments cells based on the joint likelihood of transcriptional composition and cell morphology.
- Can take DAPI images as a prior to guide transcripts as cellular clusters.

<https://github.com/kharchenkolab/Baysor>

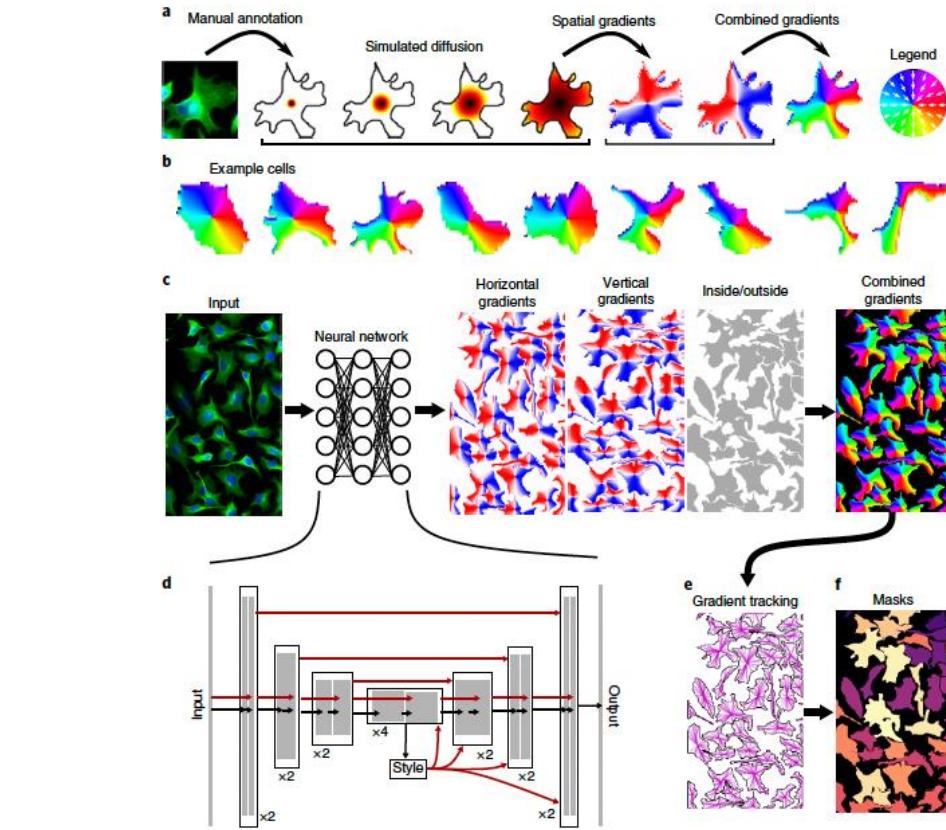


**Fig. 2 | Application of an MRF framework for segmentation-free cell-type inference and background filtration.** **a,b**, Individual molecules were clustered into major cell types by modeling the tissue as a mixture of multinomial distributions with an MRF prior. Cluster labels per molecule are shown in **a** with expression vectors for each of the clusters shown in **b**. **c**, The MRF approach is used to separate background from intracellular signal by modeling distance to its  $k$ -th nearest molecule ( $x$  axis,  $k=16$ ) as a mixture of two normal distributions. Fitted intracellular and background distributions for the Allen smFISH dataset are shown in red and green, respectively. The vertical black line shows the optimal separation point. **d**, Molecules from a subset of the Allen smFISH dataset are shown as dots colored by their distance to the  $k$ -th NN, with the color key shown on the bottom of **c**. The black contours mark regions above 50% probability of being intracellular.

# CellPose Cell Segmentation

- Segments cells based on the joint likelihood of transcriptional composition and cell morphology.
- Can take DAPI images as a prior to guide transcripts as cellular clusters.

<https://www.cellpose.org>

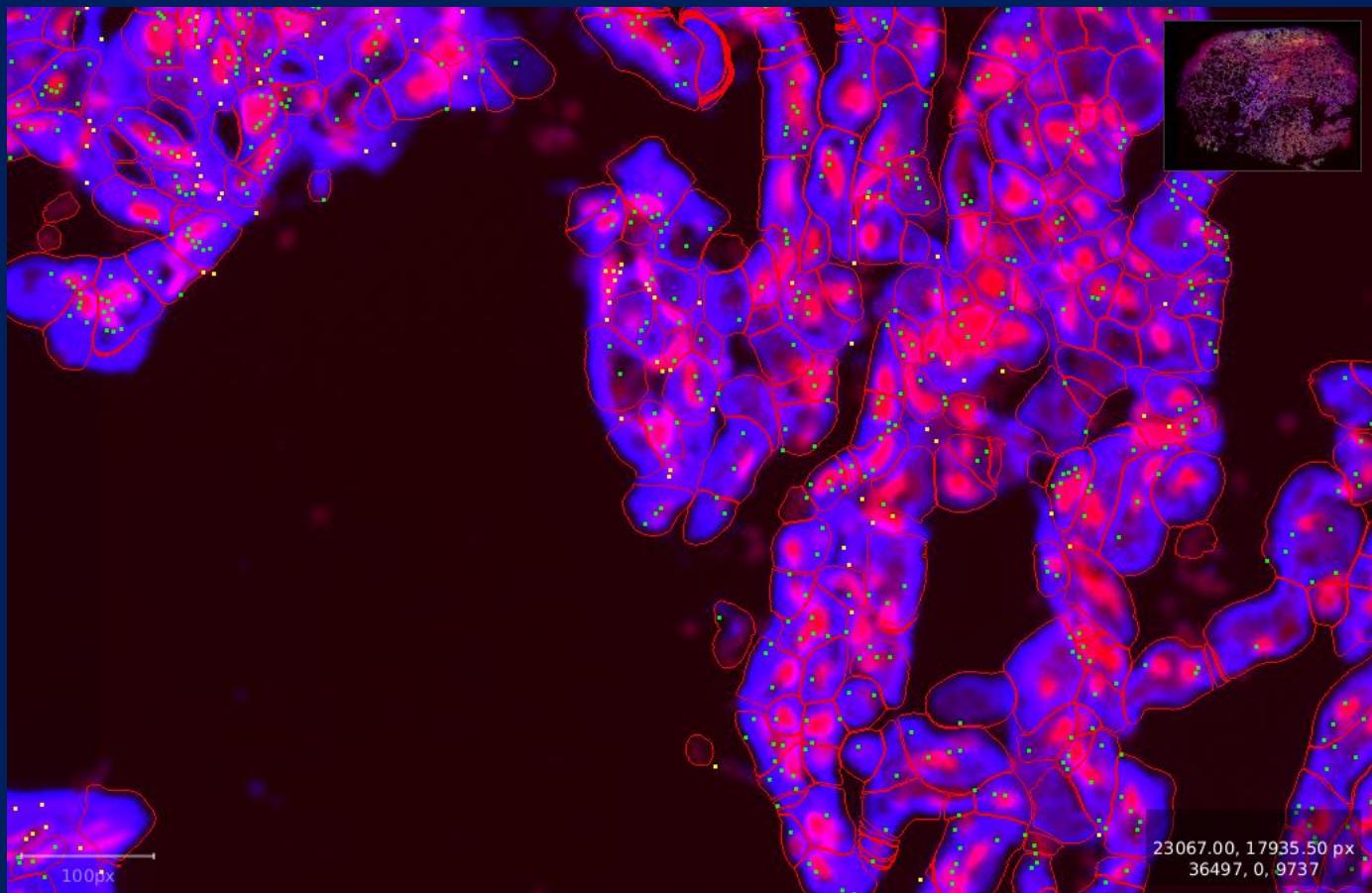
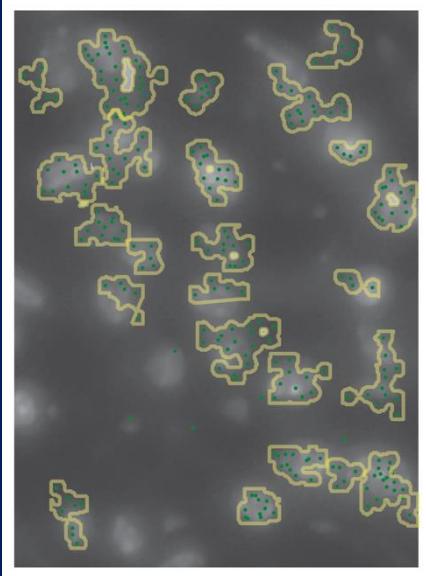


**Fig. 1 | Model architecture.** **a**, Procedure for transforming manually annotated masks into a vector flow representation that can be predicted by a neural network. A simulated diffusion process starting from the center of the mask is used to derive spatial gradients that point toward the center of the cell, potentially indirectly around corners. The  $x$  and  $y$  gradients are combined in a single direction from  $0^\circ$  to  $360^\circ$ . **b**, Example spatial gradients for cells from the training dataset. **c**, A neural network is trained to predict the horizontal and vertical gradients, as well as whether a pixel belongs to any cell. The three predicted maps are combined into a gradient vector field. **d**, The details of the neural network that contains a standard backbone U-Net<sup>3</sup> to downsample and then upsample the feature maps, with skip connections between layers of the same size and global skip connections from the image styles, computed at the lowest resolution, to all successive computations. **e**, At test time, the predicted gradient vector fields are used to construct a dynamical system with fixed points whose basins of attraction represent the predicted masks. Informally, every pixel ‘tracks the gradients’ toward their eventual fixed point. **f**, All the pixels that converge to the same fixed point are assigned to the same mask.

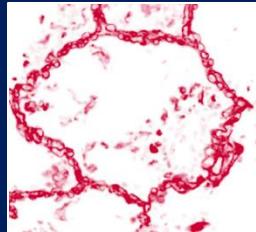
# CellPose

Baysor

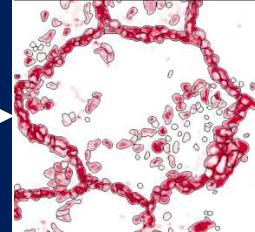
Figure S4



Registered  
Membrane Staining

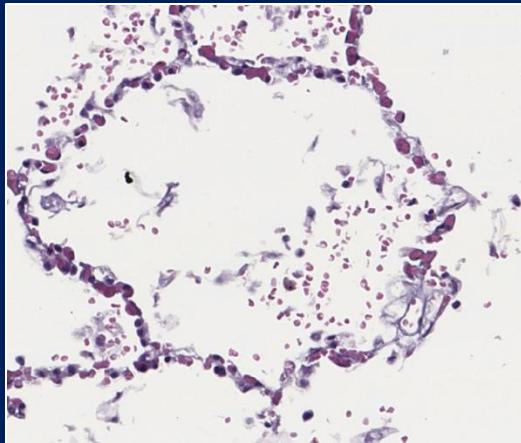


Cell Segmentation

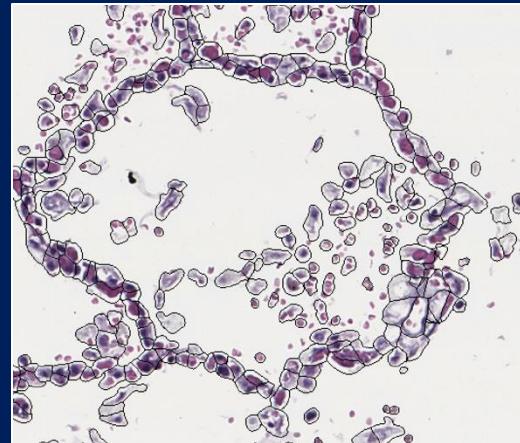


CellPose

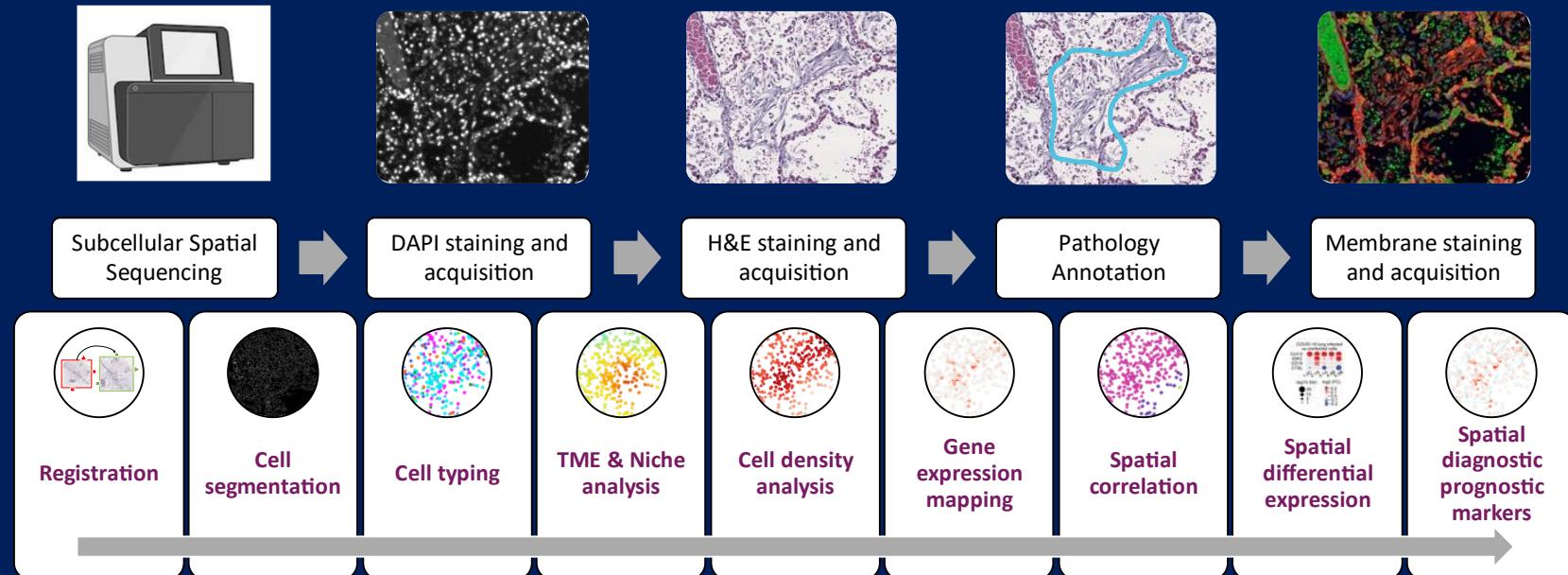
Registered H&E image



Cell Segmentation Overlay



# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



# Cell Typing

- One of the most important and time-consuming steps is cell typing.
- Ideally, we can identify cell types based on protein staining.
- Cell-based: Cell types can be defined based on reference marker genes.
- Cluster-based: After applying dimensionally reduction techniques on the gene expression, we can cluster the reduced dimensions and identify clusters that enrich marker genes.

# Cell Typing

- Reference Profile based: Another prominent way to cell type is by transferring cell type labels from scRNA-seq to spatial.
- “Insitutype” is an algorithm that is designed to carry out cell type transfer.

# Cell Typing

Published online 10 May 2022

Nucleic Acids Research, 2022, Vol. 50, No. 14 e80  
<https://doi.org/10.1093/nar/gkac320>

**Cell type identification in spatial transcriptomics data can be improved by leveraging cell-type-informative paired tissue images using a Bayesian probabilistic model**

Asif Zubair<sup>1</sup>, Richard H. Chapple<sup>1</sup>, Sivaraman Natarajan<sup>1</sup>, William C. Wright<sup>1</sup>, Min Pan<sup>1</sup>, Hyoeng-Min Lee<sup>1</sup>, Heather Tillman<sup>1</sup>, John Easton<sup>1</sup> and Paul Geelhoer<sup>1,2\*</sup>

<sup>1</sup>Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA and

<sup>2</sup>Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

Received December 16, 2021; Revised April 13, 2022; Editorial Decision April 19, 2022; Accepted April 21, 2022

nature communications

**ARTICLE** <https://doi.org/10.1038/s41467-022-28803-w> **OPEN**

Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data

Aleksandr Ianevski<sup>1,2</sup>, Anil K. Giri<sup>1,3\*</sup> & Tero Aittokallio<sup>1,2,3,4,5</sup>

*Bioinformatics Advances*, 2023, vba030  
<https://doi.org/10.1093/bioadv/vba030>  
Advance Access Publication Date: 13 March 2023  
Original Paper

**Gene expression**  
**scAnnotate: an automated cell-type annotation tool for single-cell RNA-seqencing data**

Xiangling Ji<sup>1</sup>, Danielle Tsao<sup>1</sup>, Kailun Bai<sup>1</sup>, Min Tsao<sup>1</sup>, Li Xing<sup>2,\*</sup> and Xuekui Zhang<sup>3,1\*</sup>

<sup>1</sup>Department of Mathematics and Statistics, University of Victoria, Victoria V8P 5C2, Canada and <sup>2</sup>Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon S7N 5C9, Canada

\*To whom correspondence should be addressed.  
Associate Editor: Thomas Lengauer

Received on November 7, 2022; revised on February 28, 2023; editorial decision on March 5, 2023; accepted on March 10, 2023

Received: 17 June 2022

Accepted: 25 November 2022

Published online: 10 December 2022

Rongbo Shen<sup>1,2</sup>, Lin Liu<sup>2,3</sup>, Jihen Wu<sup>1,2</sup>, Ying Zhang<sup>2,3</sup>, Zhiyuan Yuan<sup>1,2</sup>, Junfu Guo<sup>2,3</sup>, Fan Yang<sup>1</sup>, Chao Zhang<sup>2</sup>, Bihao Chen<sup>2,3</sup>, Wanwan Feng<sup>1,4</sup>, Chao Liu<sup>2</sup>, Jing Guo<sup>2</sup>, Guochen Fan<sup>2</sup>, Yong Zhang<sup>2,3</sup>, Yuxiang Li<sup>2,3</sup>, Xun Xu<sup>1,2\*</sup> & Jianhua Yao<sup>2,3</sup>



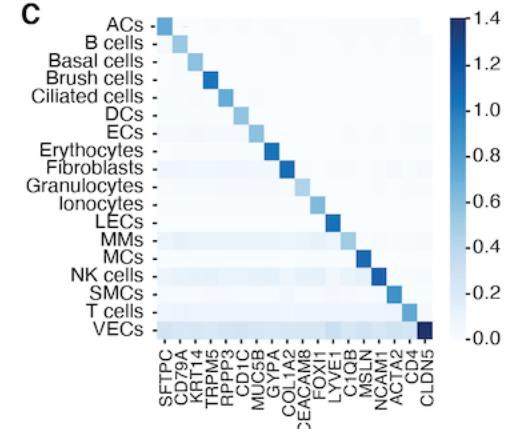
Automated methods for cell type annotation on scRNA-seq data

Giovanni Pasquini<sup>a,c</sup>, Jesus Eduardo Rojo Arias<sup>b</sup>, Patrick Schäfer<sup>a</sup>, Volker Busskamp<sup>a,c,\*</sup>

<sup>a</sup> Technische Universität Dresden, Center for Molecular and Cellular Bioengineering (CMCB), Center for Regenerative Therapies Dresden (CRTD), Dresden 01307, Germany

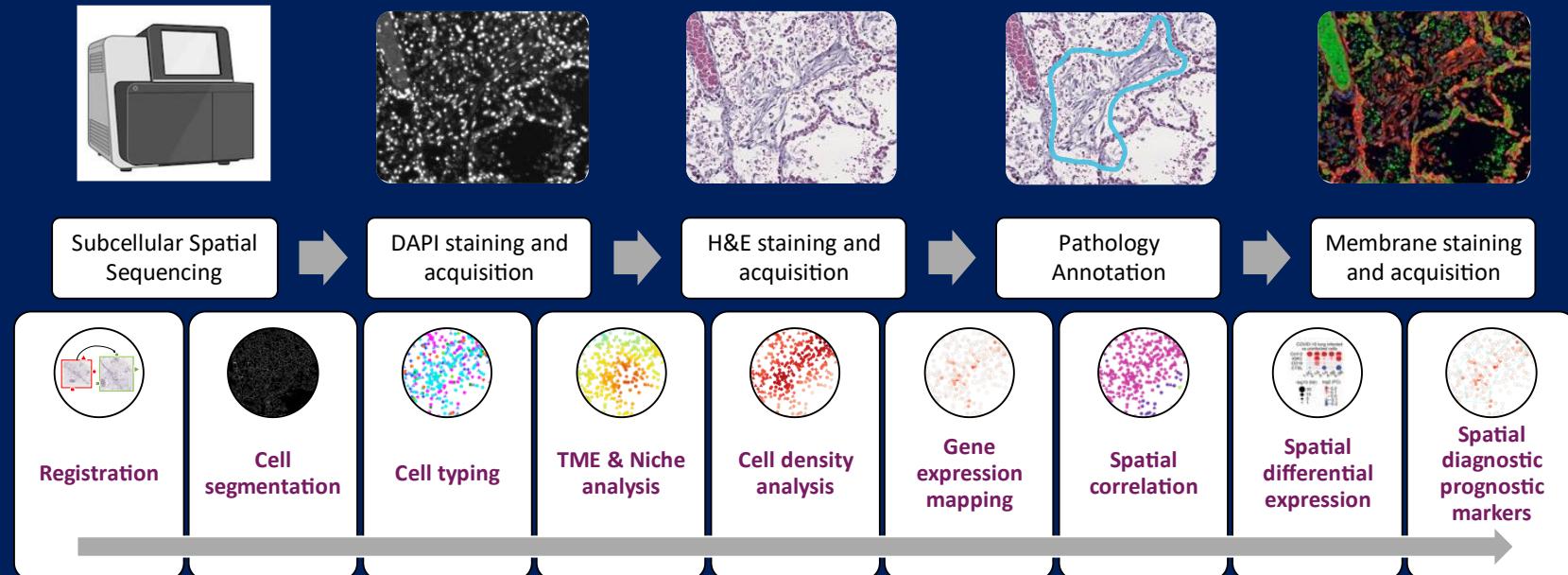
<sup>b</sup> Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK

<sup>c</sup> Universitäts-Augenklinik Bonn, University of Bonn, Department of Ophthalmology, Bonn 53127, Germany



Das et al. 2023, <https://doi.org/10.1002/jmv.29009>

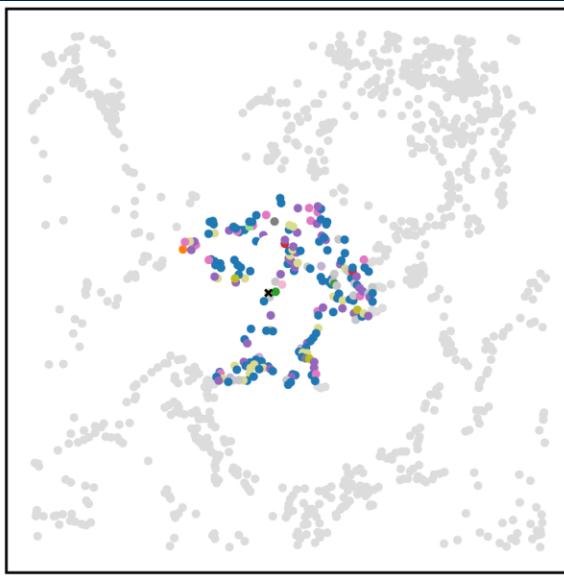
# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



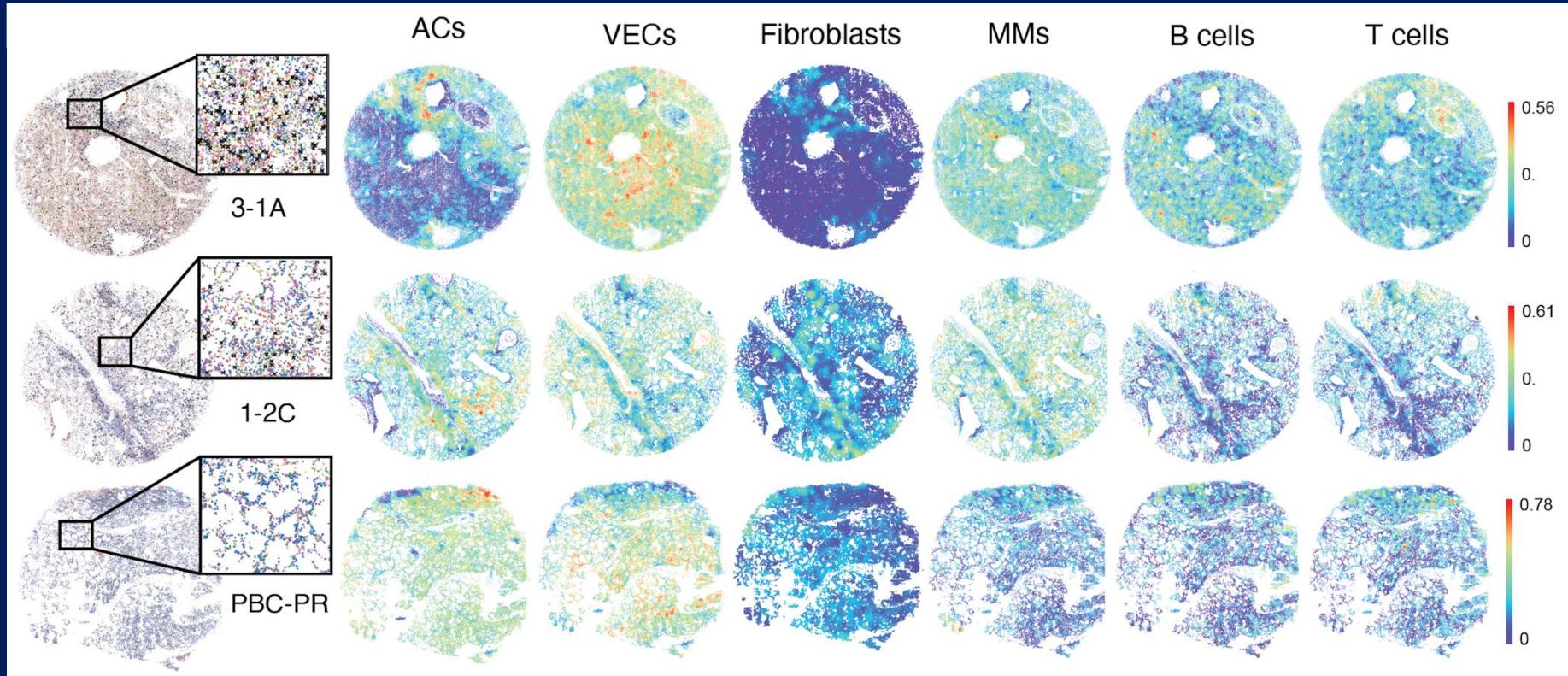
# Niche Analysis

## Algorithm:

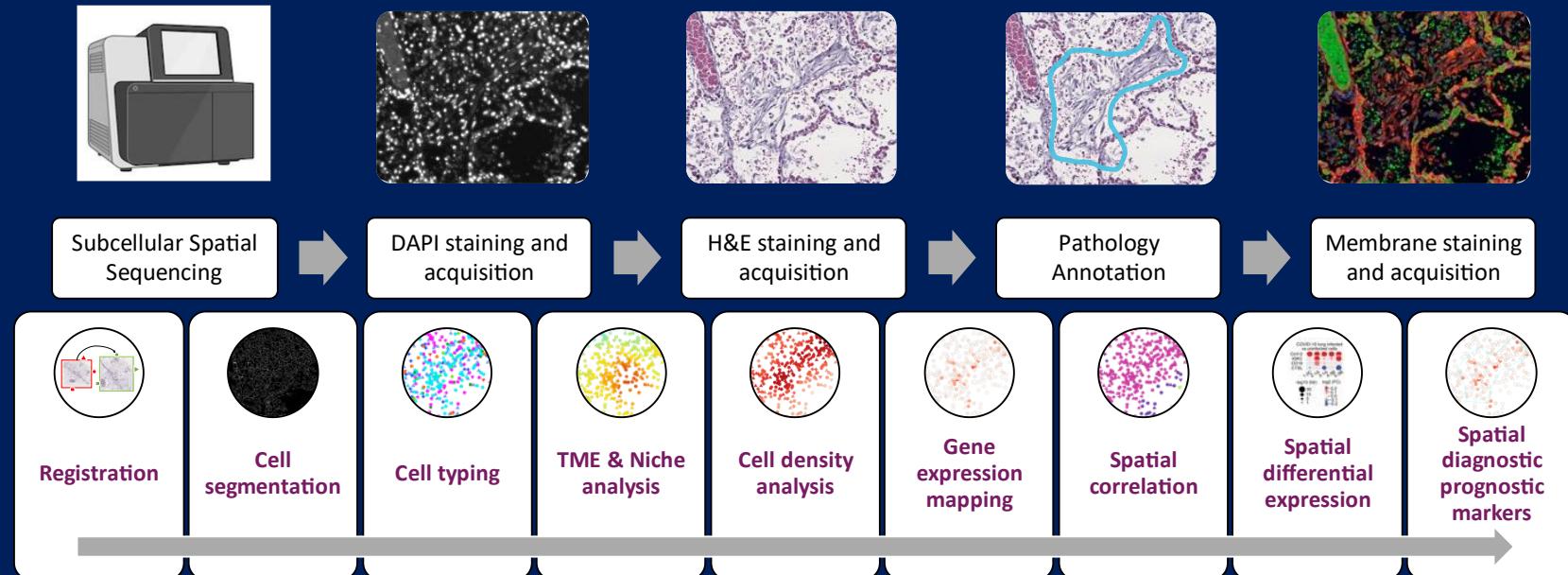
1. Pick one cell.
2. Count the number of each cell type in the neighborhood.
  - We now have a cell type count vector for which is the local ***Neighborhood Cell-type Composition (NCTC)***.
3. Repeat steps 1-2.



Cell Type	Percentage
ACs	0.455
Fibroblasts	0.185
VECs	0.130
SMCs	0.070
MMs	0.070
Granulocytes	0.025
T cells	0.015
NK cells	0.010
ECs	0.010
CCs	0.010
DCs	0.005
Basal cells	0.005
MCs	0.005
B cells	0.005

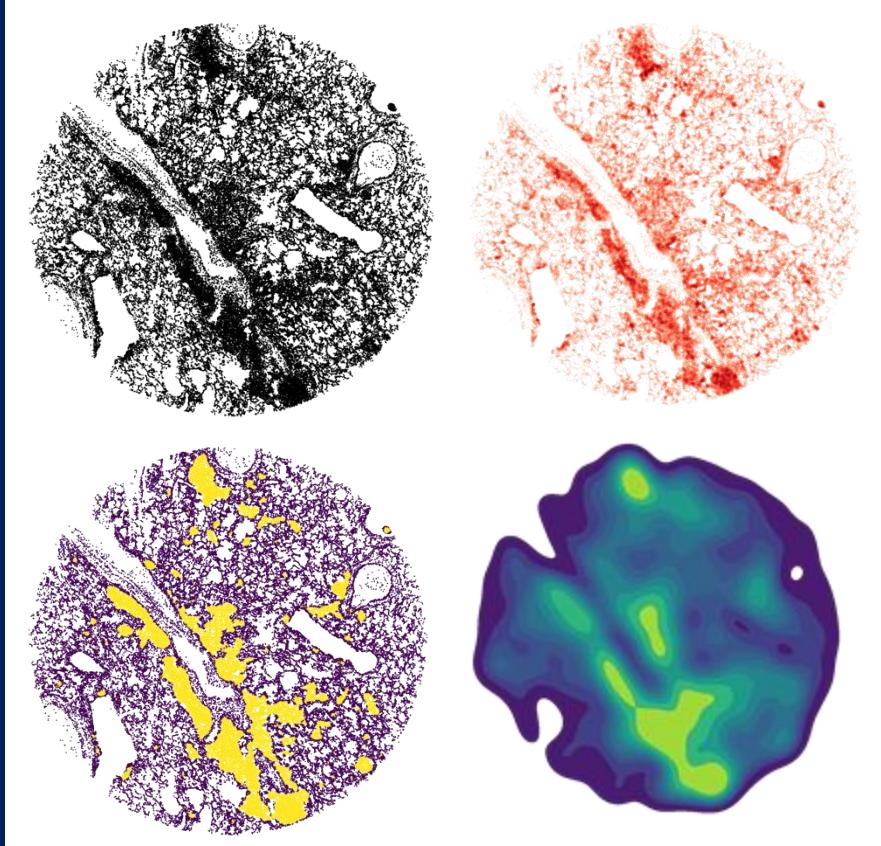


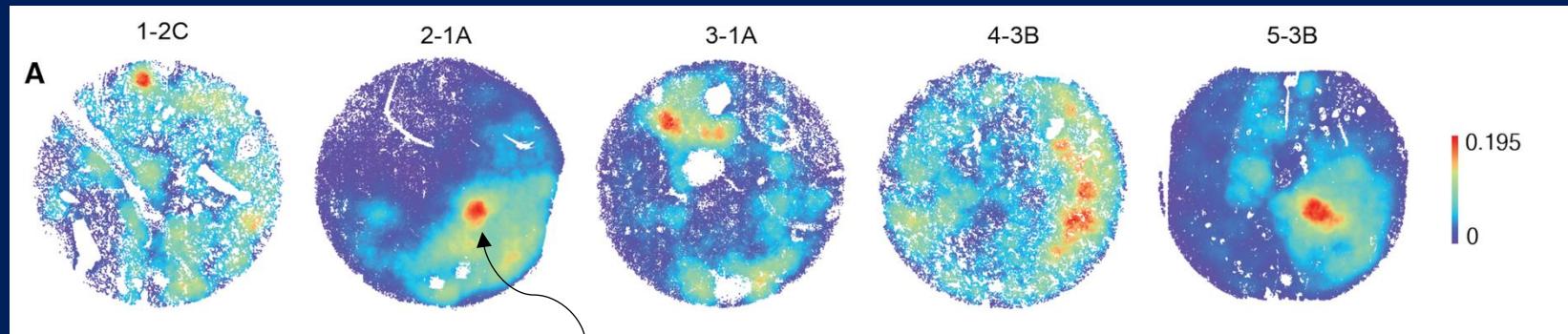
# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



# Cell Density Analysis

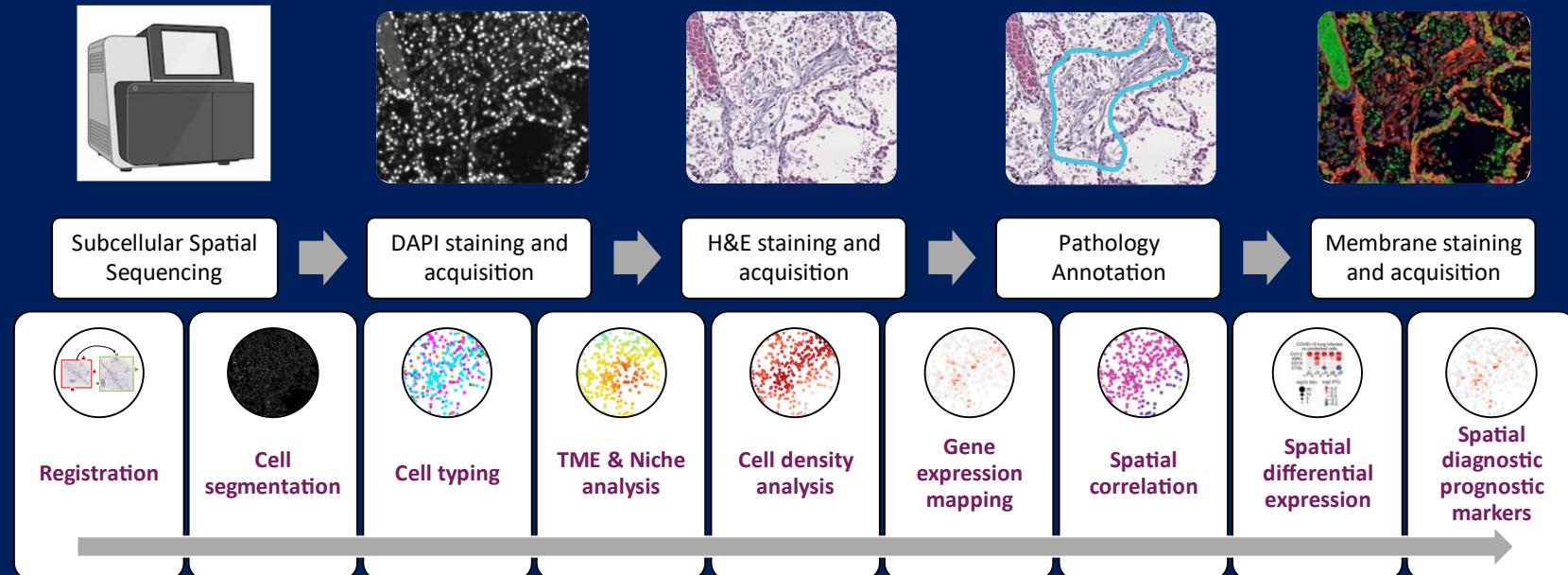
Instead of fixing the number of cells like we did for NCTC, take a count all the cells inside a circular region of fixed radius.



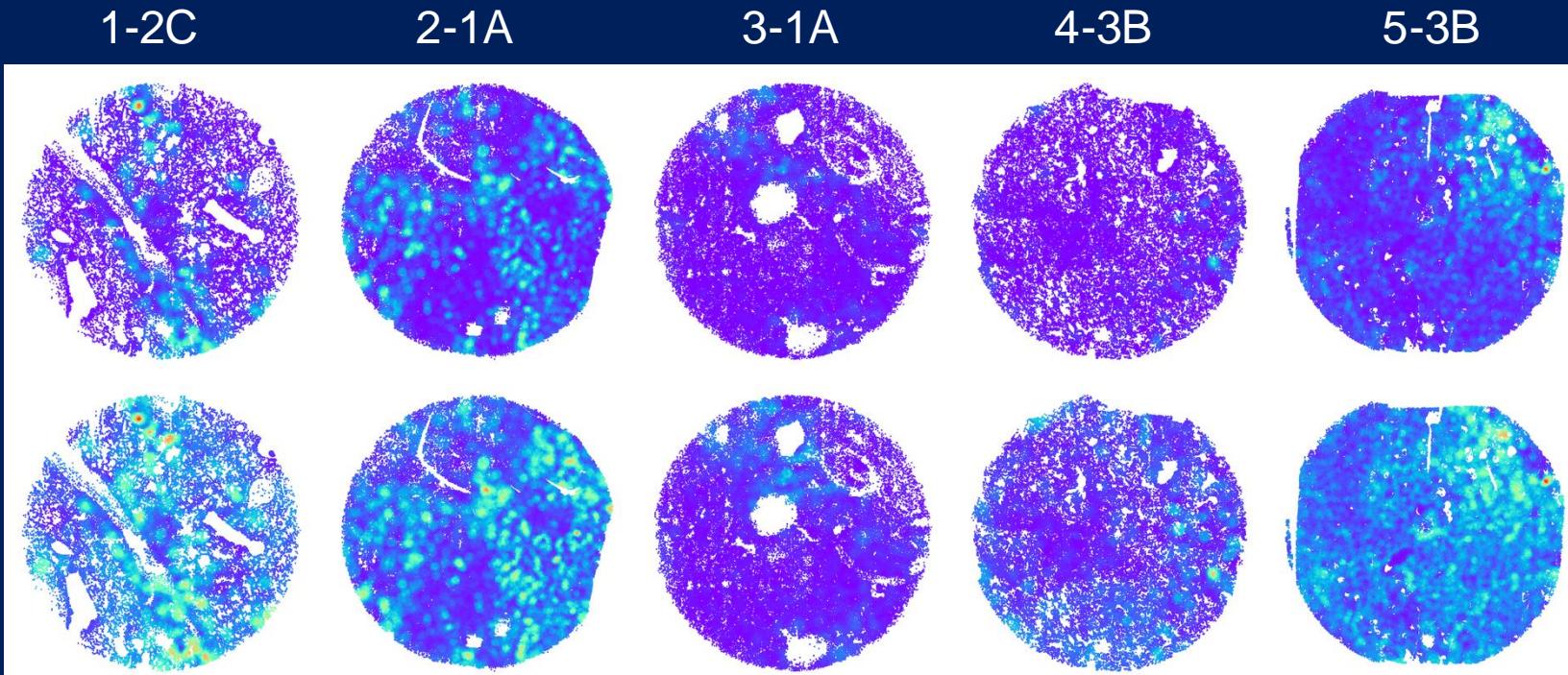


High infection regions

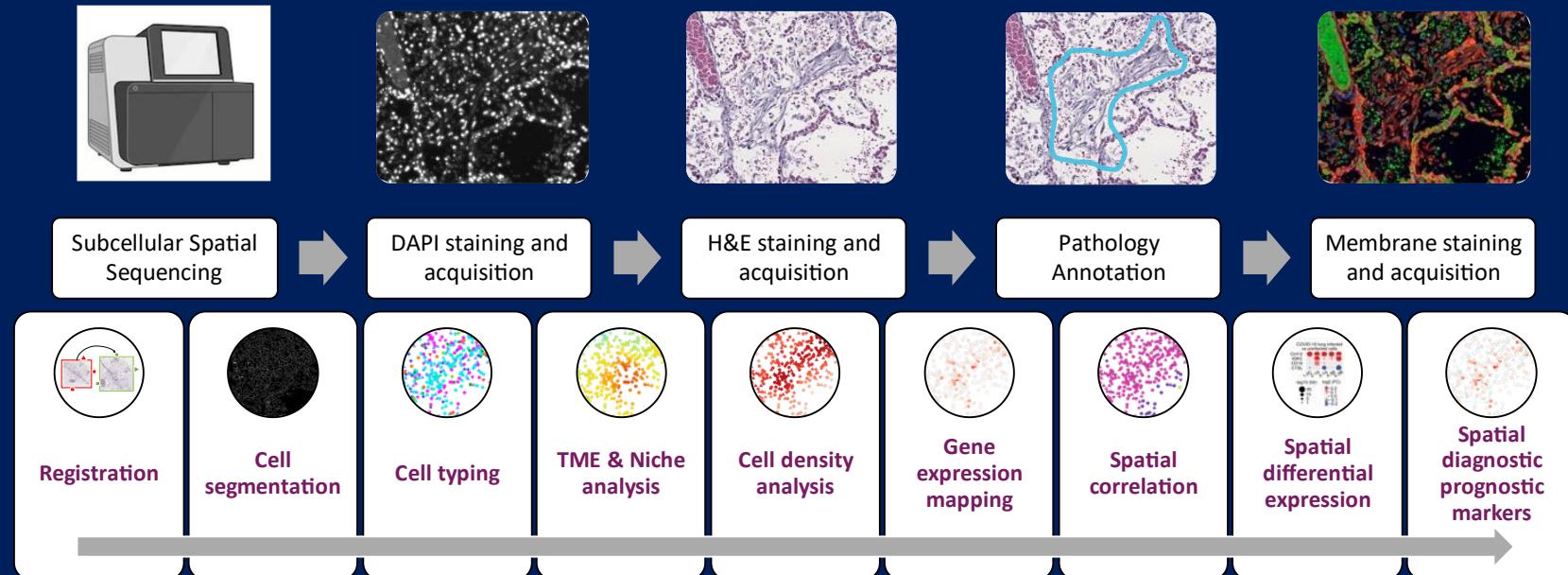
# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



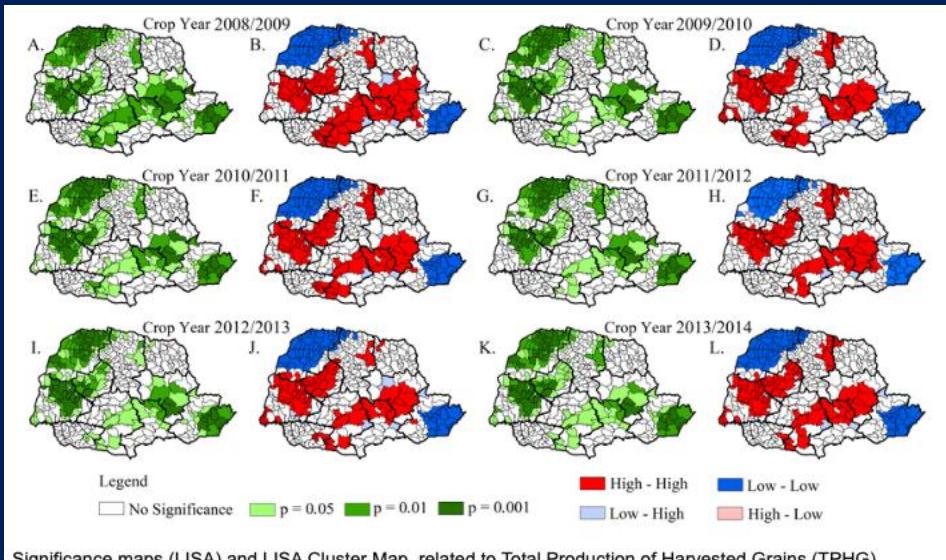
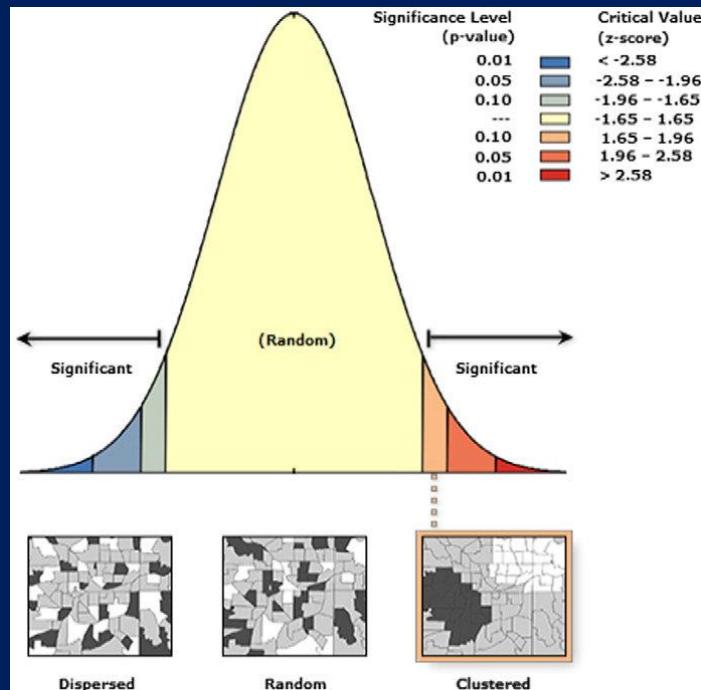
# Gene Expression Mapping



# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



# Spatial Correlation Analysis



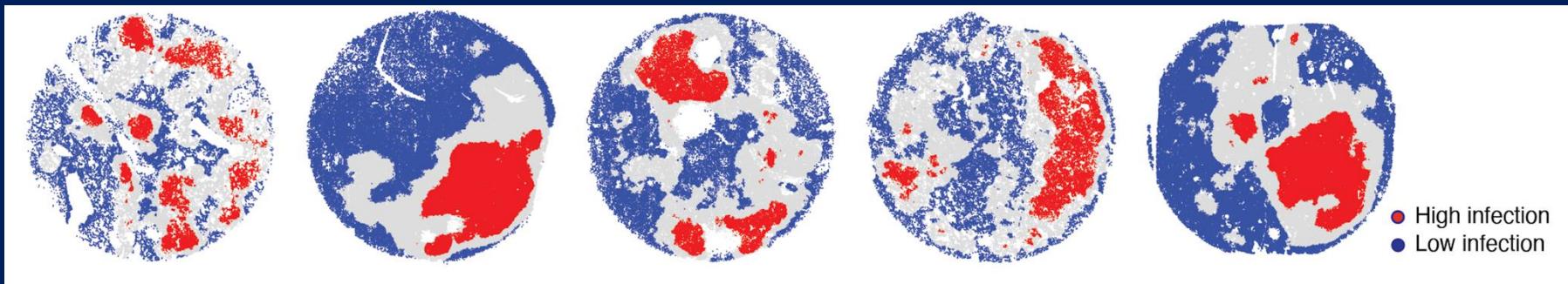
Significance maps (LISA) and LISA Cluster Map, related to Total Production of Harvested Grains (TPHG) (soybean, corn 1 st and 2 nd harvests and wheat) for the 2008/2009 harvest years (A and B); 2009/2010 (C and D); 2010/2011 (E and F); 2011/2012 (G and H); 2012/2013 (I and J) and 2013/2014 (K and L).

# Spatial Correlation Analysis

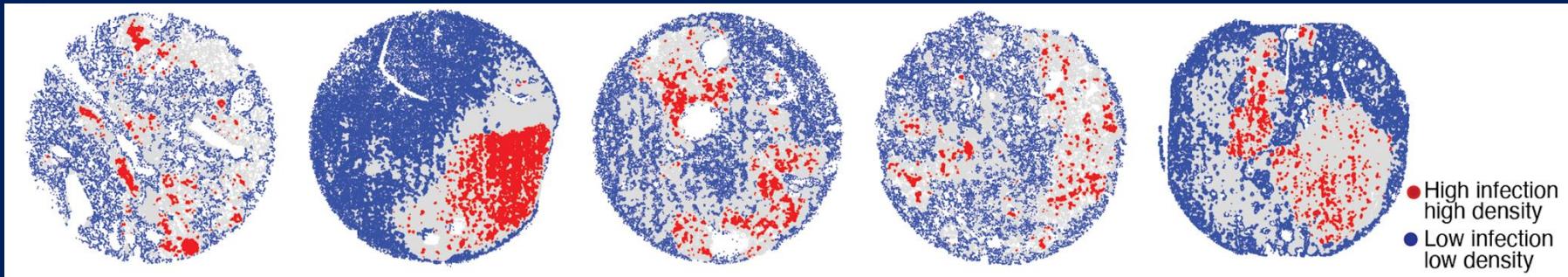
- What kind of questions can we ask?
- Can we identify high-density spatial clusters?
- What about spatial hotspots of infection?

# Spatial Correlation Analysis

Moran's I Autocorrelation



Moran's I Cross-correlation

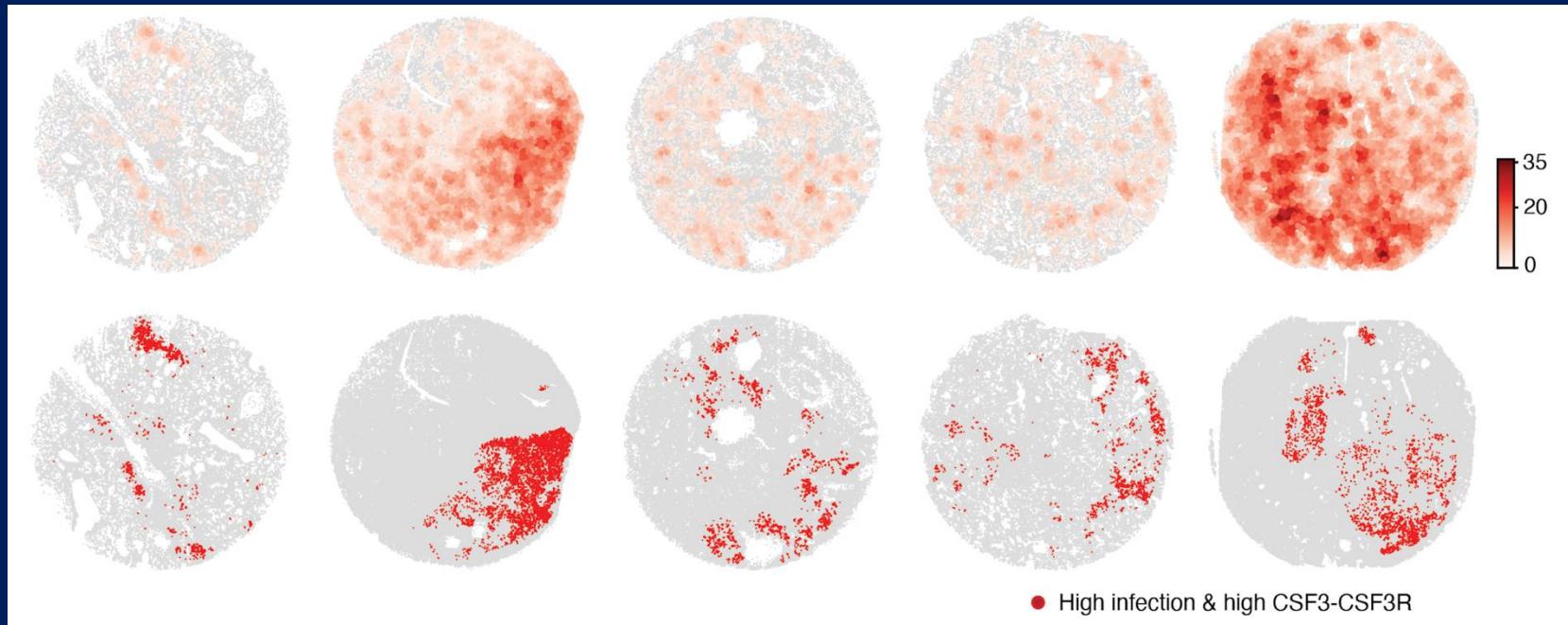


# Spatial Correlation Analysis

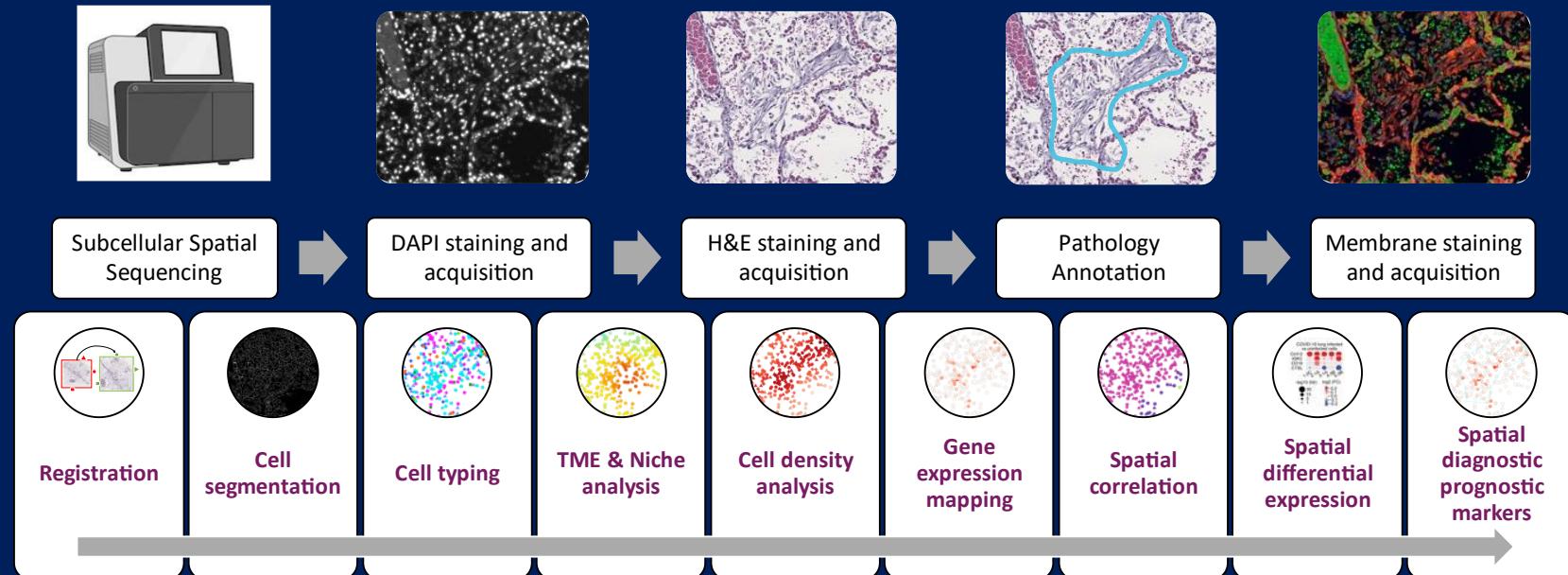
Bivariate Moran's I analysis to summarize high-infection regions with high entry protein expression in each tissue sample.



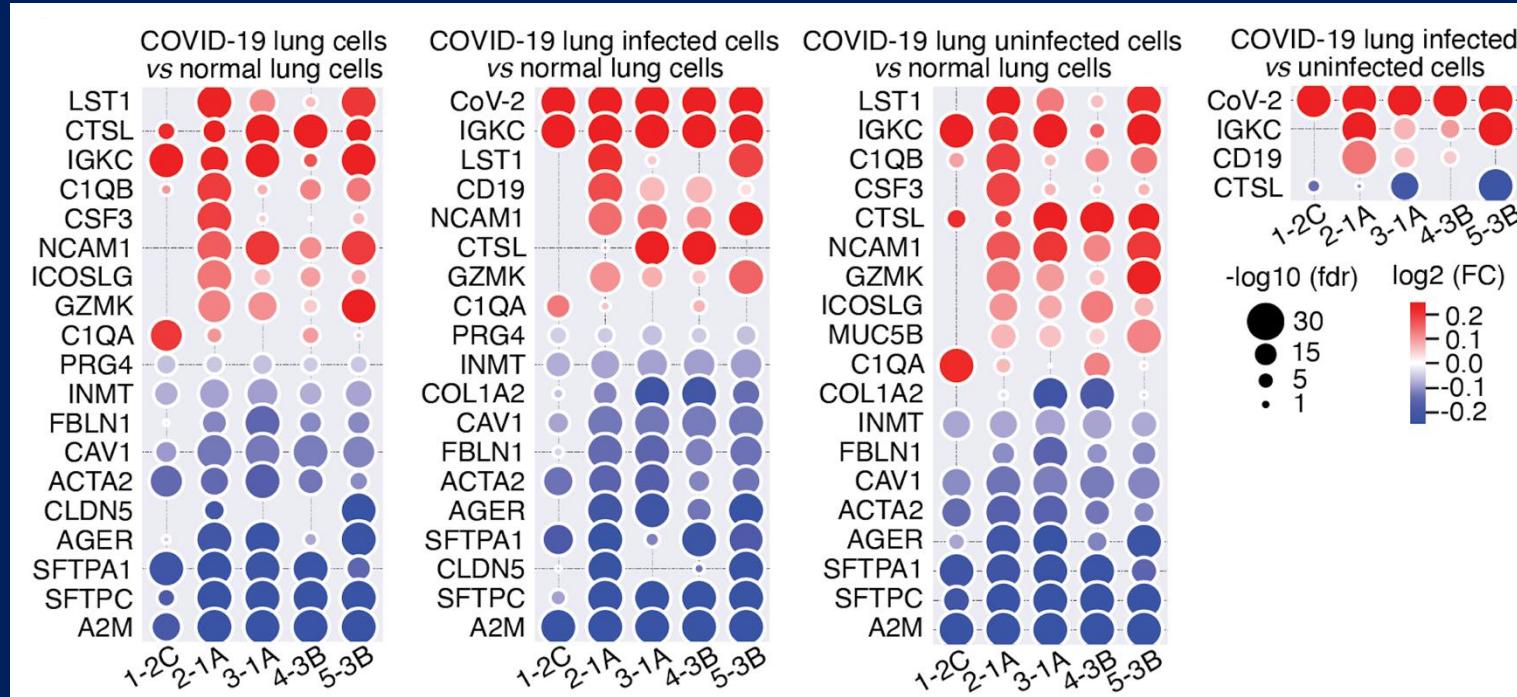
# Spatial Correlation Analysis



# Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



# Differential Expression Analysis



# Let's Code:

[https://github.com/arundasan91/ScSTA\\_tutorial](https://github.com/arundasan91/ScSTA_tutorial)

<https://tinyurl.com/das-spatial>