

The Single-Cell Spatial Transcriptomics Analysis (ScSTA) Cookbook

Arun Das

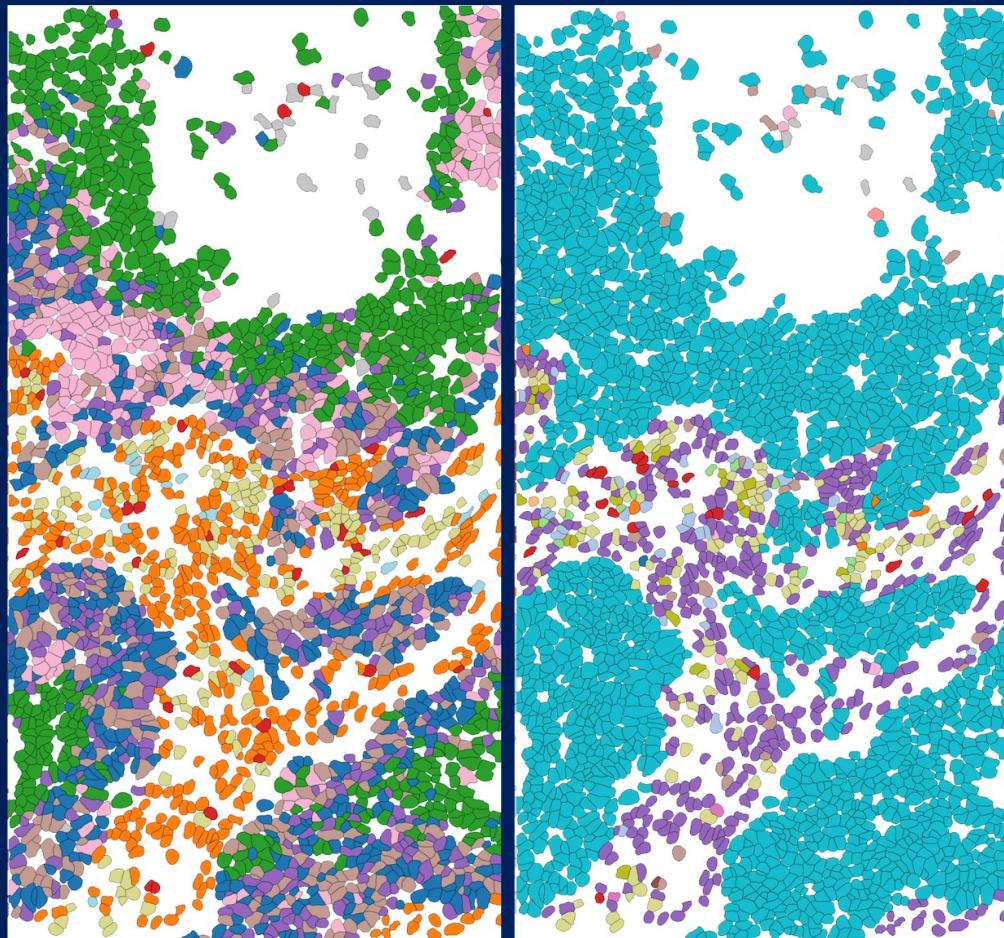
Postdoctoral Associate

Hillman Cancer Center

University of Pittsburgh Medical Center

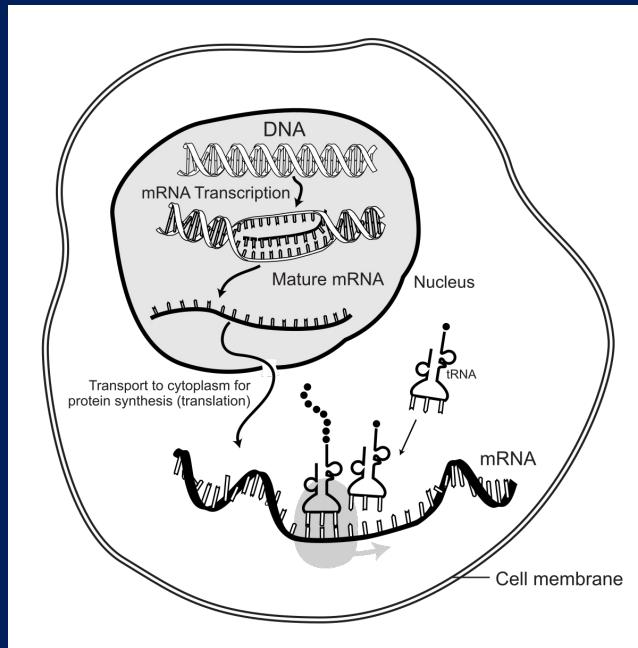
“Spatial transcriptomics is an attempt to quantify mRNA expression of large number of genes within the spatial context of tissues and cells.”

Pachter et al. 2022



'Why' quantify mRNA's and the 'How" it is done.

We need to answer this!



mRNA carries genetic information of DNA from nucleus to cytoplasm for protein synthesis

The ‘Why’.

- What information does DNA hold?
- What about proteins?

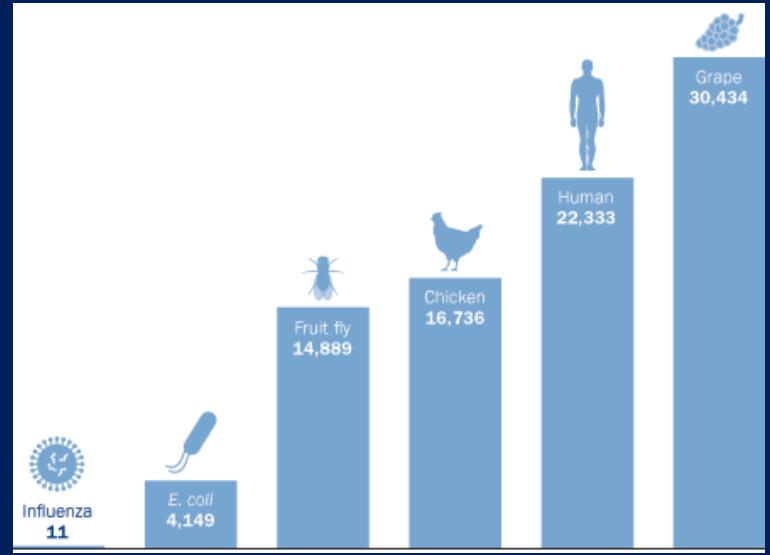


Image from Science News

The ‘How’.

- Can we read the order of the nucleotide bases (A, U, G, C) in an RNA molecule?
- Can we image the RNA in some way?
- What is the throughput?

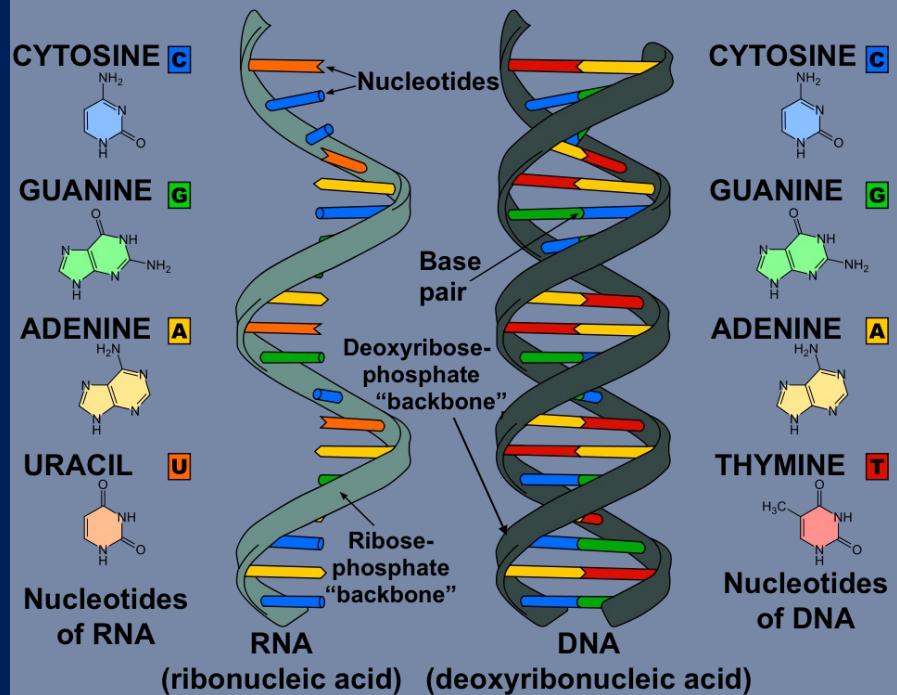


Image from Khan Academy

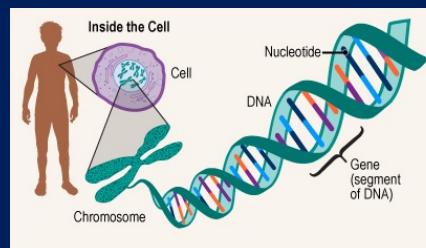
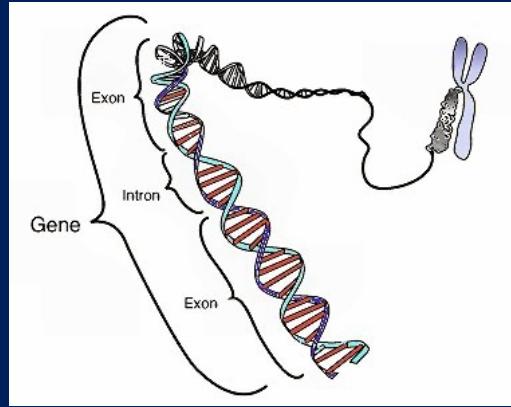


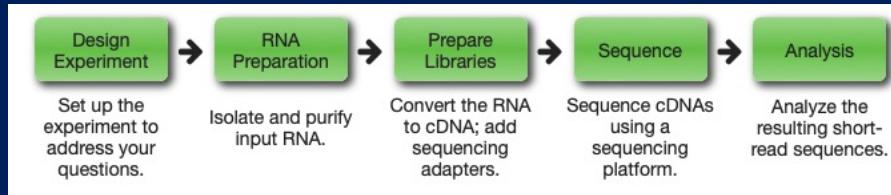
Image from KidsHealth.org

The ‘How’.

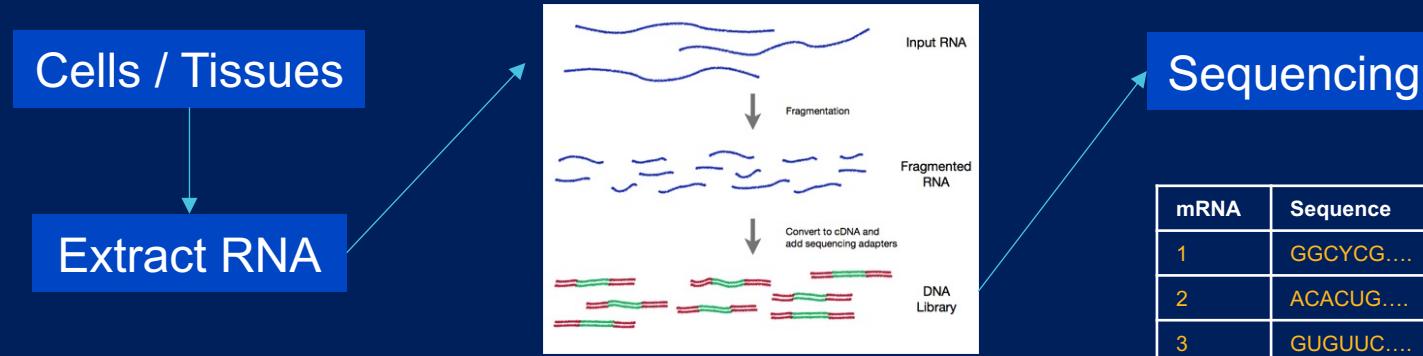


If there is a way to read the sequence of letters, that would be enough to identify genes...

The 'How' – Sequencing based.



We lost the spatial context though!



The 'How' – Imaging based - FISH.

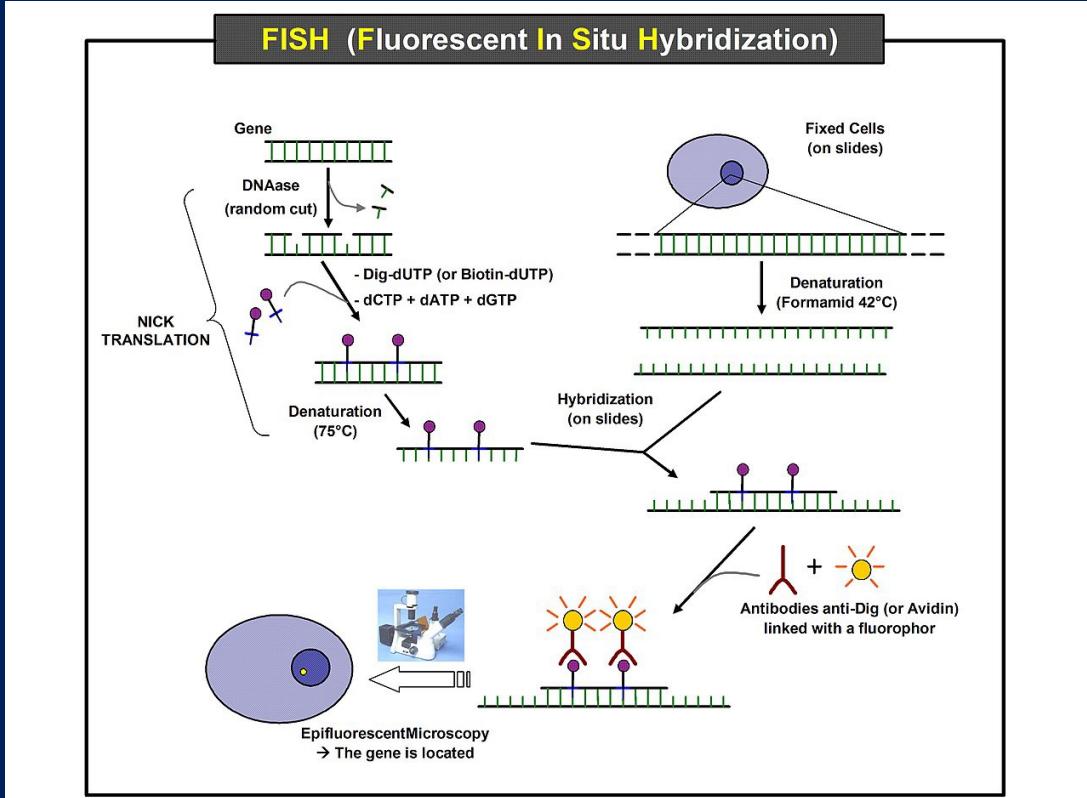
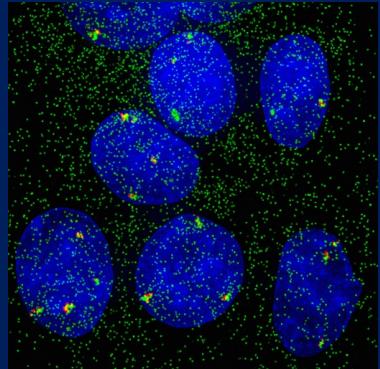
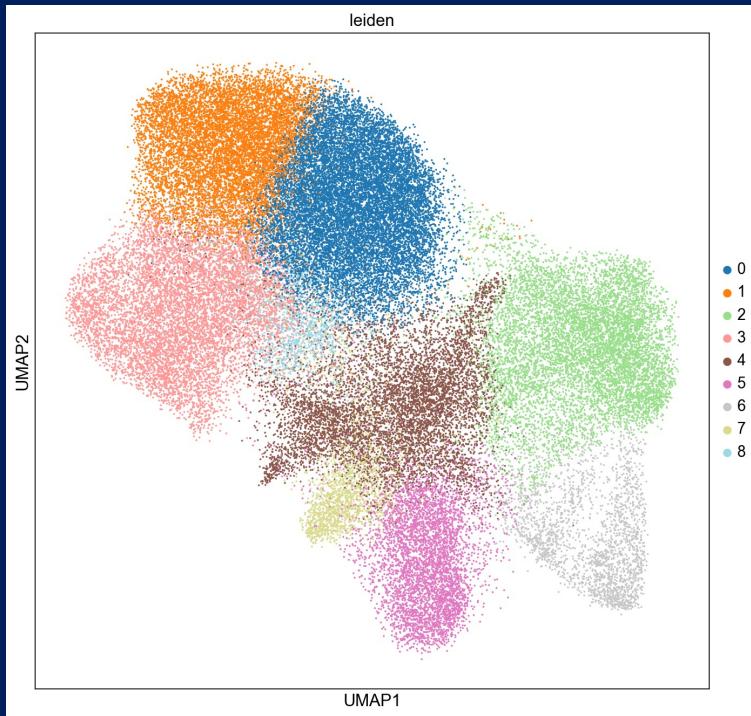


Image from Wikipedia

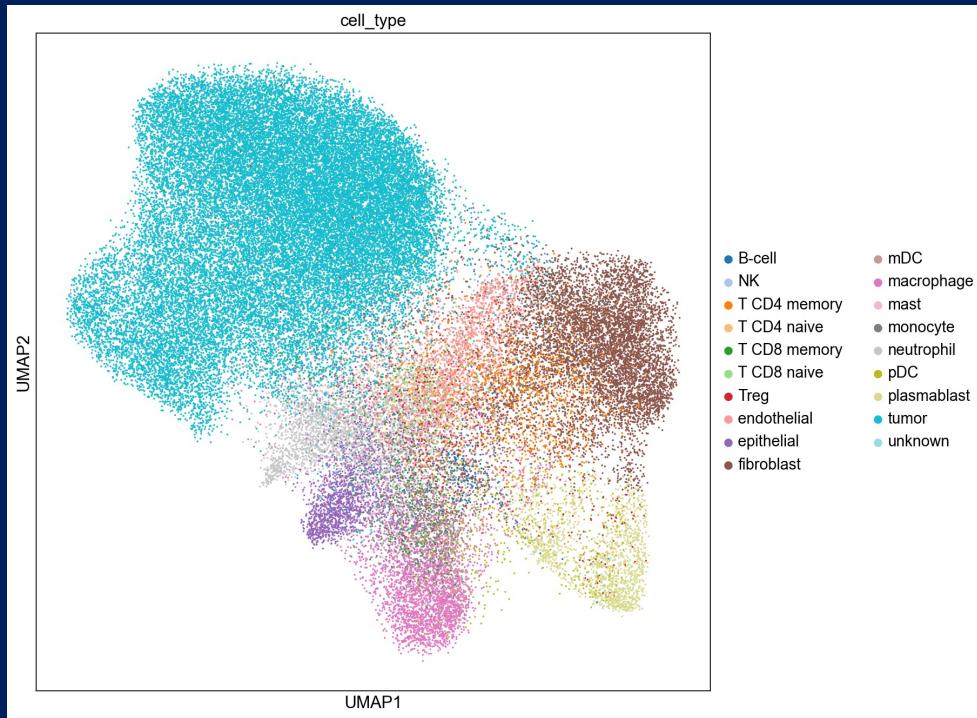


Stossi et al. 2020

Why do we need to study gene expressions?



Why do we need to study gene expressions?



Why do we need to study gene expressions?

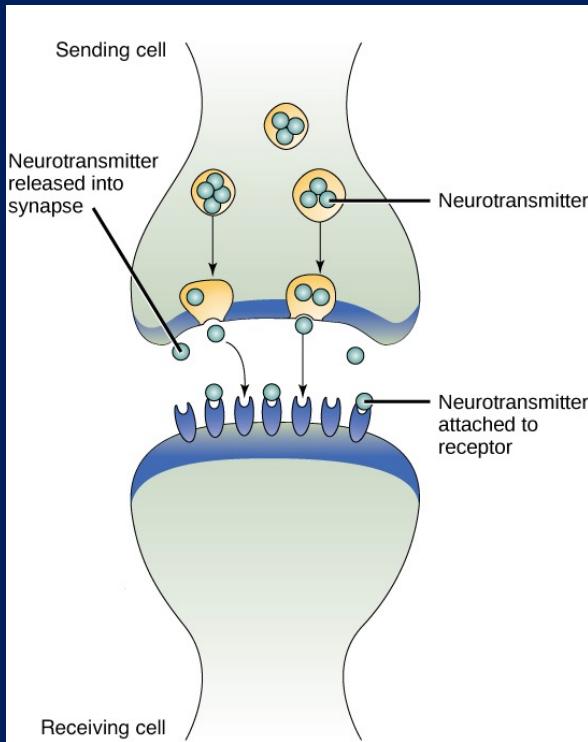
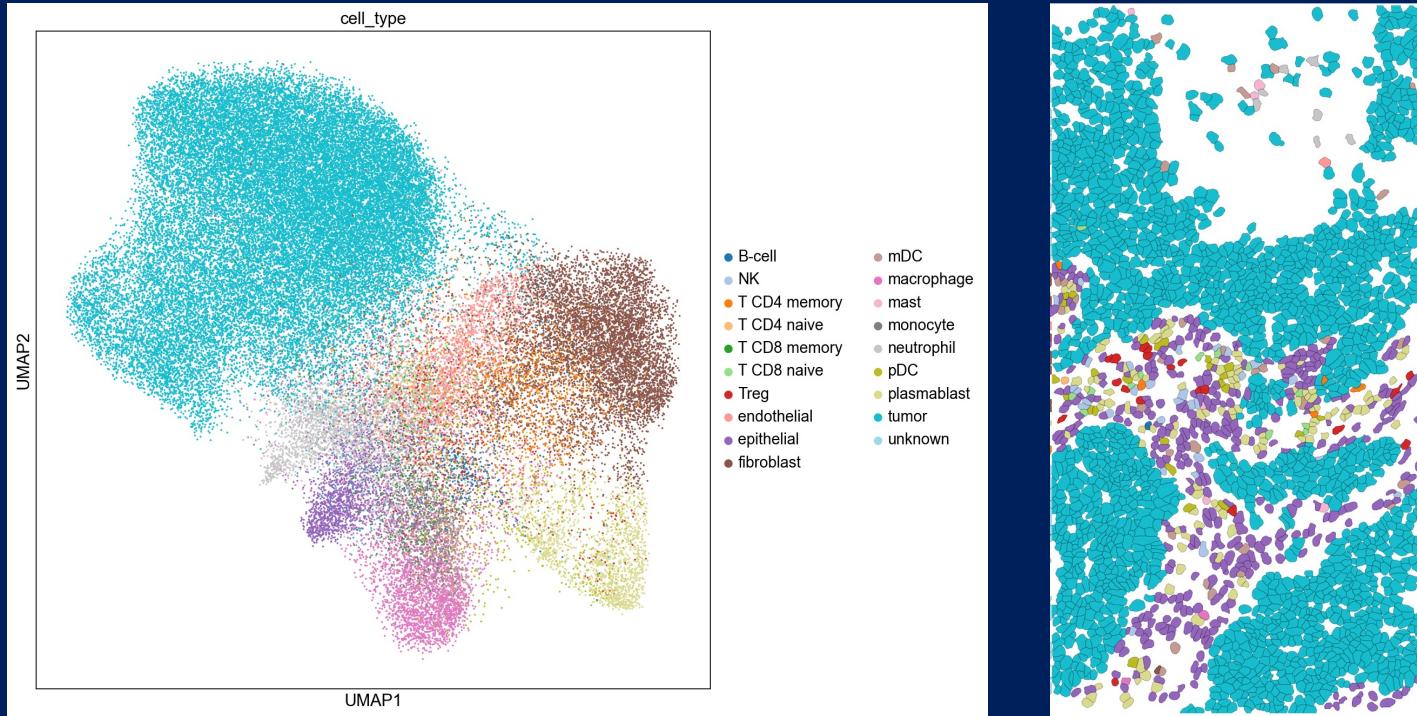


Image from Khan Academy

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



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

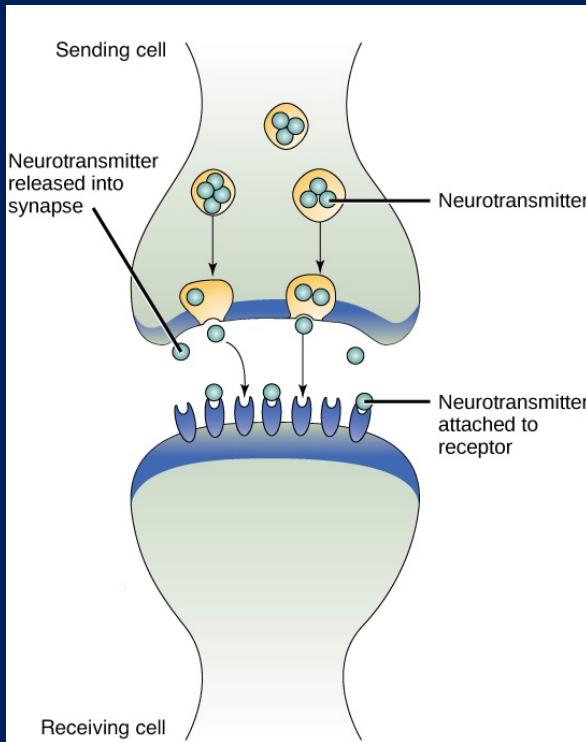
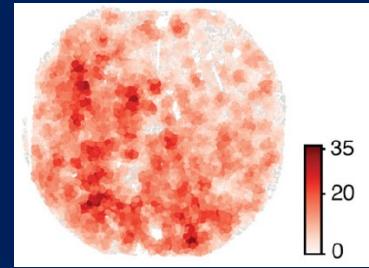
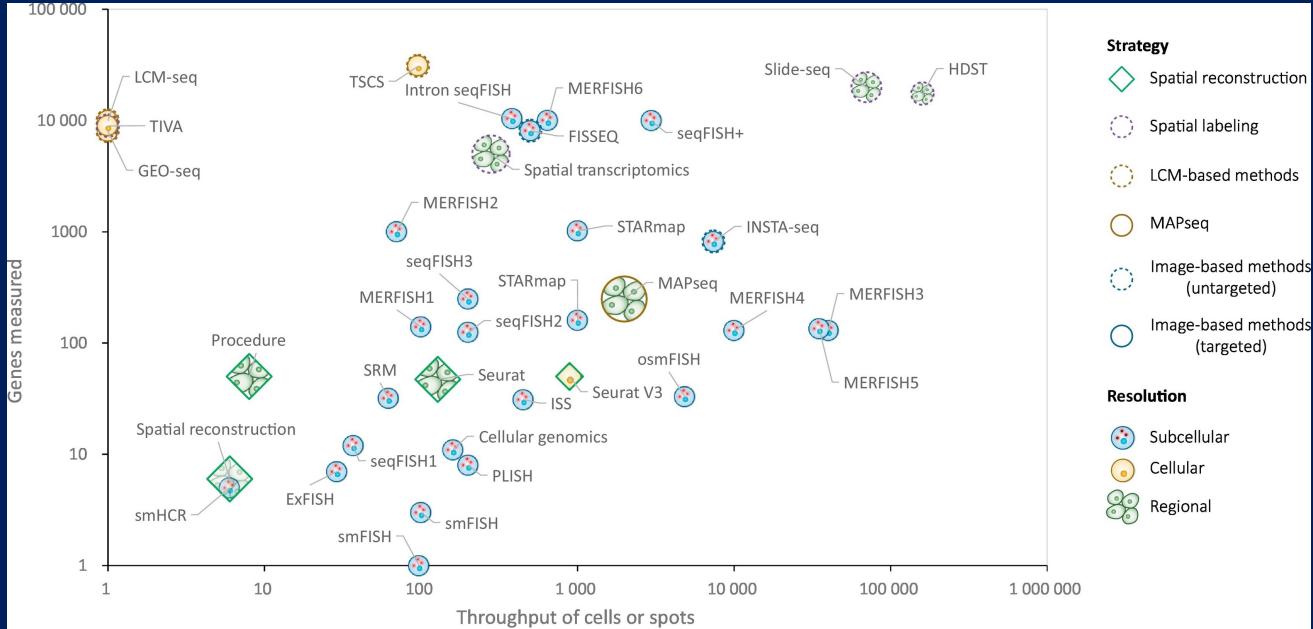
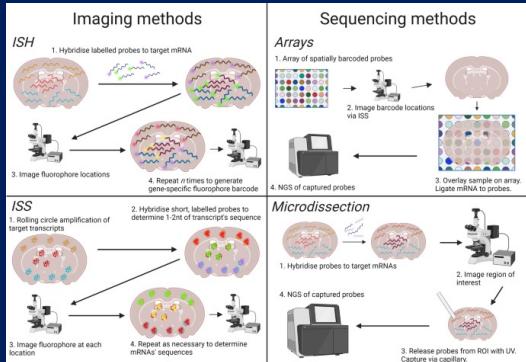


Image from Khan Academy



Current State-of-the-Art in Spatial Transcriptomics



Trends In Biotechnology

Liao et al. 2020

Current State-of-the-Art in Spatial Transcriptomics

High-performance multiplexed fluorescence *in situ* hybridization in culture and tissue with matrix imprinting and clearing

Jeffrey L. Kuttler^{a,b,c}, Junjie Hao^{b,d,e}, Dhananjay Bambhani-Mukku^{c,f}, Tian Lu^b, Catherine Dulac^b, and Xiaowei Zhuang^{a,g}^aHarvard Medical Institute, Harvard University, Cambridge, MA 02138; ^bDepartment of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; ^cDepartment of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138; and ^dDepartment of Physics, Harvard University, Cambridge, MA 02138

Contributed by Xiaowei Zhuang, October 25, 2016 (with review for October 10, 2016); reviewed by Guadalupe Danzer and Tsakjip Ha

This article is licensed under a Creative Commons Attribution 4.0 International license

Highly multiplexed single-molecule FISH has emerged as a promising technique for measuring gene expression at the single-cell level because of its ability to directly image and profile numerous RNA species within their native cellular contexts. However, the “gold standard” for transcriptome analysis—quantitative PCR—achieves this level of multiplexing by assigning error-robust barcodes to each target transcript. In contrast, MERFISH achieves this level of multiplexing by assigning error-prone barcodes with oligonucleotide tags that contain a representation of these barcode sequences. By using this approach, we have imaged 140 and 160 RNA species in multiplexed cell cultures and tissue samples, respectively, using a single-color FISH approach. In addition, we embedded samples in polystyrene, anchored RNAs to the polymer matrix, and cleared cellular proteins and fats, while maintaining the integrity of the RNA. We show that the efficacy of this approach, we measured the copy number of 130 RNA species in complex tissue samples with a resolution of 10–20 nm (μ MERFISH). We observed a reduction both in the background because of the reduction in the number of RNA species and in the overall detectable loss in RNA. This process led to an improvement throughout the extension of MERFISH into four color channels. We further demonstrated MERFISH measurements of complex tissue samples using a four-color FISH approach. We also used the clearing approach. We envision that this method will improve the performance of MERFISH and its hybridization-based techniques in both cell culture and tissues.

time clearing | fluorescence *in situ* hybridization | multiplexed imaging | single-cell transcriptomics | brain

Single-molecule FISH (merFISH) is a powerful technology that allows for the simultaneous measurement of multiple RNA species in single cells (1, 2). In this approach, RNAs are labeled via the hybridization of fluorescently labeled oligonucleotide probes, producing hybridization signals from individual RNAs, which reveal both the abundance and the spatial distribution of those RNAs. By applying this approach to both cell culture and tissue, we can measure the spatial distribution of RNA species at the single-cell level in both cell culture and tissue, but also in complex tissue samples. This approach reduces natural noise in gene expression and its role in cellular response (3), the intracellular spatial organization of RNAs and its role in posttranslational regulation (4, 5), and the spatial distribution of gene expression within complex tissues and its role in the molecular mechanisms of disease (6, 7).

To extend the benefits of this technique to systems-level transcriptome analysis, we have developed several approaches to increase the multiplexing of MERFISH (i.e., the number of different RNA species that can be measured simultaneously) without increasing the cost (8–13). Most of these approaches take advantage of color multiplexing, which has allowed us to measure up to 160 RNA species in complex tissue using the recently introduced multiplexed error-robust FISH (MERFISH).

Author contributions: J.L.K., J.H., D.M., T.L., C.D., and X.Z. designed research; J.L.K., J.H., D.M., and T.L. performed research; J.L.K., J.H., and T.L. analyzed data; and J.L.K., J.H., D.M., and X.Z. wrote the paper.
Reviewers: G.O., University of Texas Southwestern Medical Center, and T.H., The Johns Hopkins University School of Medicine.
Conflict of interest statement: X.Z., J.L.K., J.H., and T.L. are involved in patients applied for grants related to this work. J.L.K., J.H., and T.L. are involved in patients applied for grants related to this work. J.L.K., J.H., and T.L. are involved in patients applied for grants related to this work.

*J.L.K. and J.H. contributed equally to this work.

Correspondence should be addressed to Xiaowei Zhuang (e-mail: xzhuang@seas.harvard.edu).
Citation: Kuttler, J.L., Hao, J.J., Bambhani-Mukku, D., Lu, T., Dulac, C., and Zhuang, X.W. (2017) High-performance multiplexed fluorescence *in situ* hybridization in culture and tissue with matrix imprinting and clearing. *Proc Natl Acad Sci USA* 114(46):14465–14470. <https://doi.org/10.1073/pnas.1610991114>

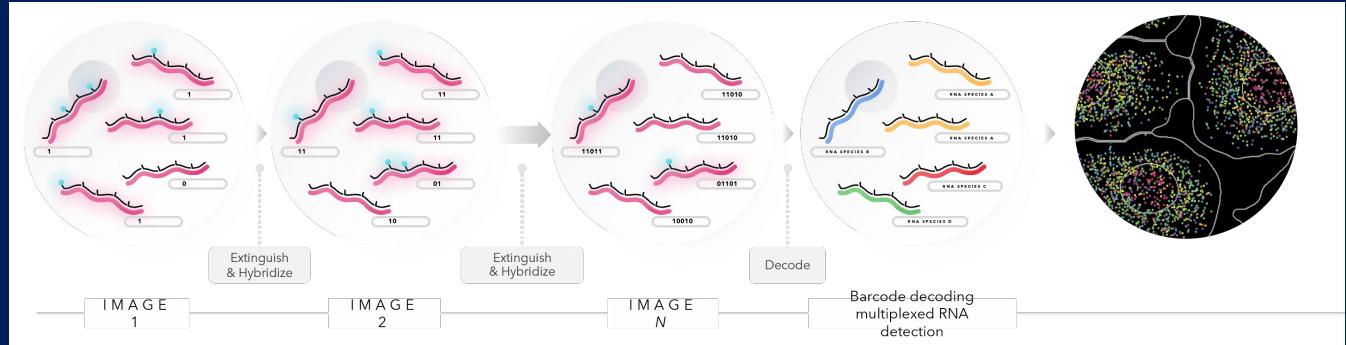


Image from Vizgen.com

MERFISH

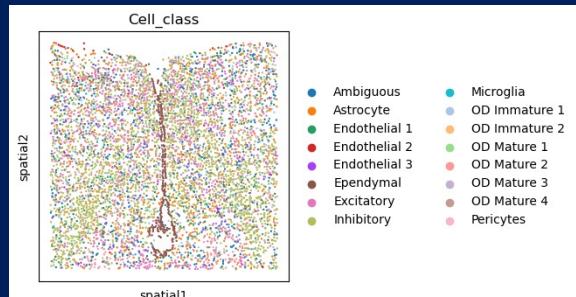


Image from ScanPy

University of
Pittsburgh.**UPMC**
LIFE CHANGING MEDICINE

Current State-of-the-Art in Spatial Transcriptomics

CosMx™ SMI for Single-Cell Imaging

- Quantification of up to 1000 RNAs or up to 100 proteins.
- Being updated to 6000 RNAs.

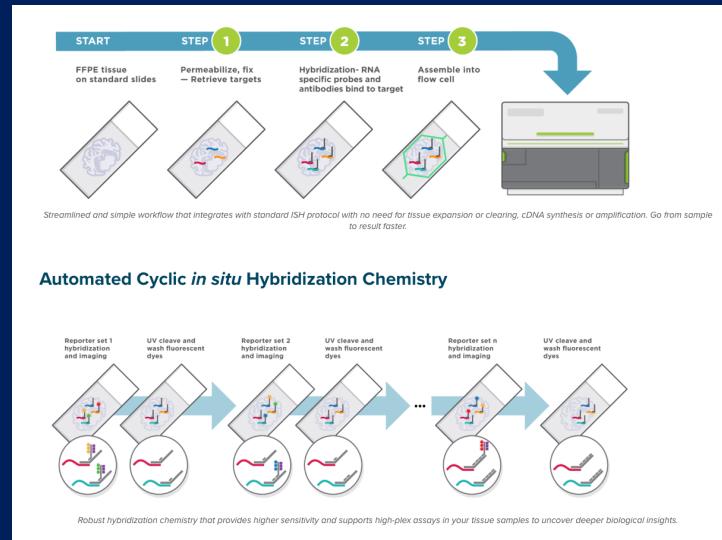


Image from Nanostring.com

Why Single-cell Resolution?



A screenshot of the Single Cell PORTAL website. The header features the 'Single Cell PORTAL' logo and navigation links for 'Help & resources', '+ Create a study', and 'Sign in'. A central callout box highlights 'Featuring 611 studies' and '37,161,643 cells'. Below the header is a search bar with fields for 'Search studies' and 'Search genes', and filters for 'Search by filters' (organ, species, disease, cell type) and 'Search by text'. A 'Browse collections' button and a 'Download' link are also present. At the bottom, it shows '611 total studies found'.

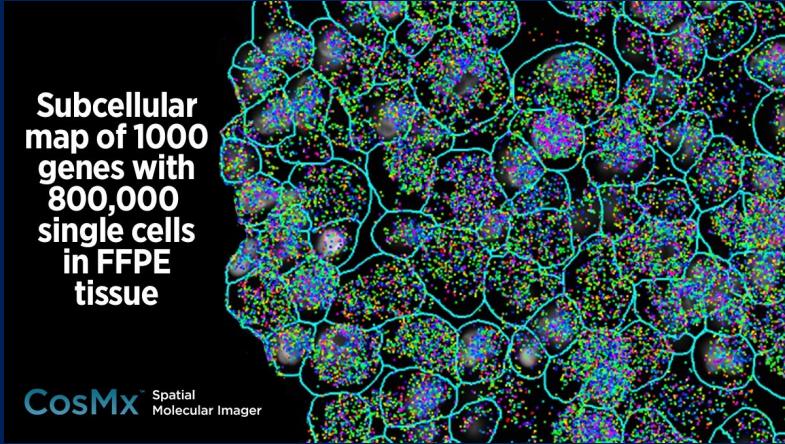
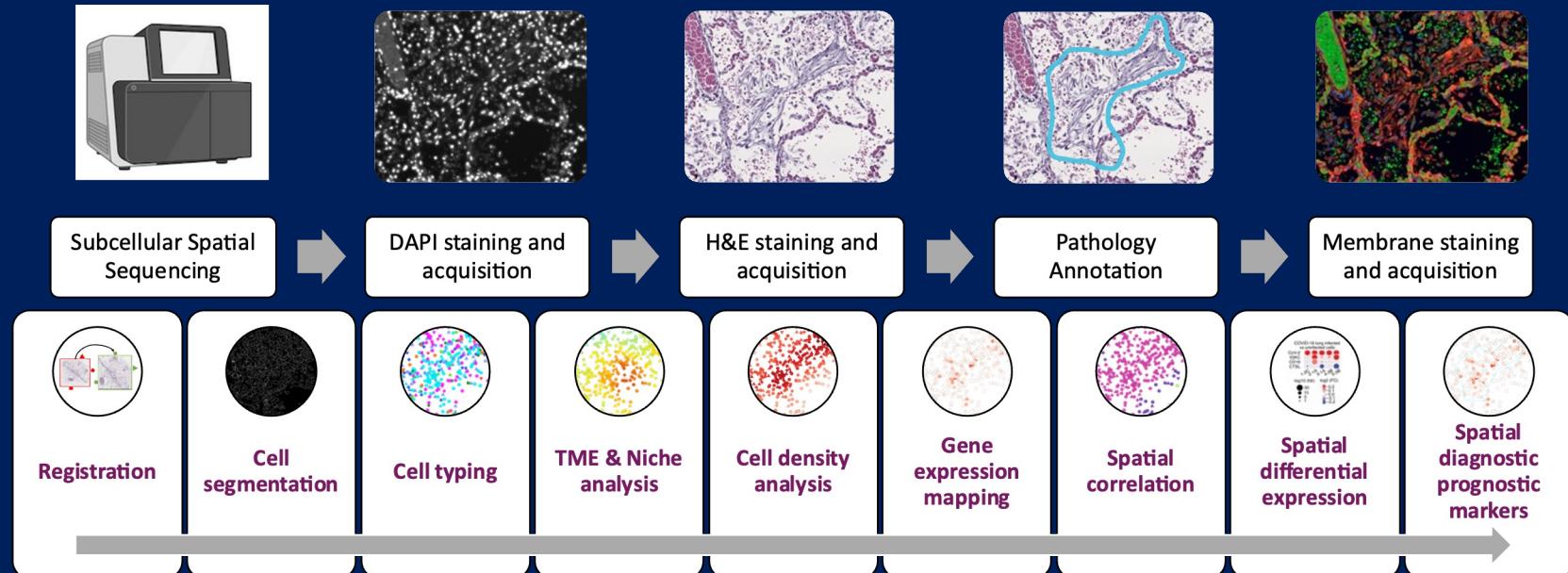


Image from Nanostring

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

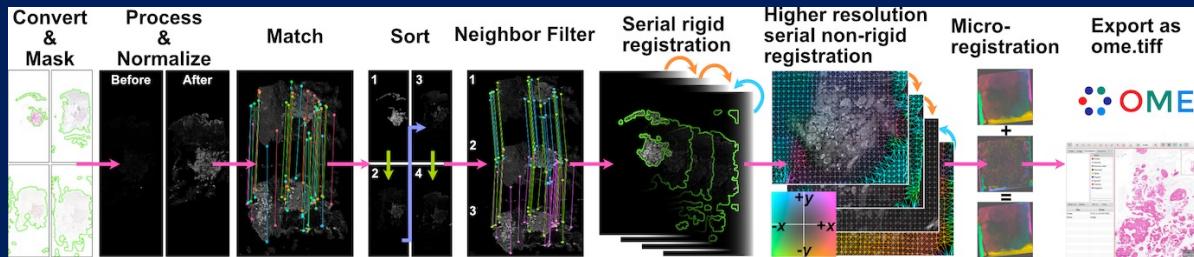
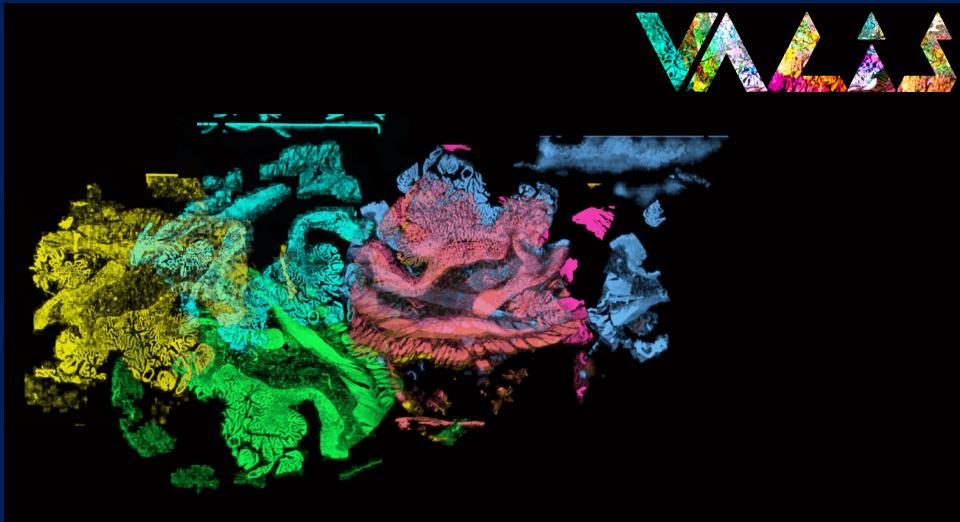
<https://zenodo.org/records/7636104>

Image Registration



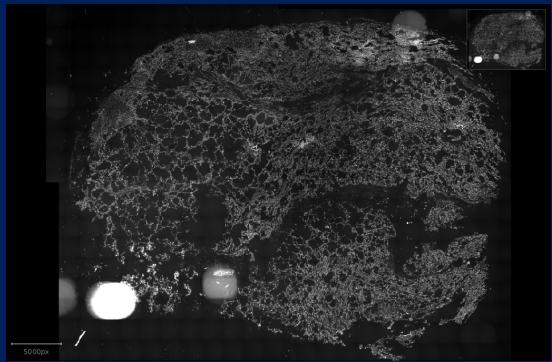
- Define a measure to quantify the similarity between two images.
- Apply affine transformations to morph the moving image.
- Optimize the similarity metric by tuning the affine transformations.

Image Registration

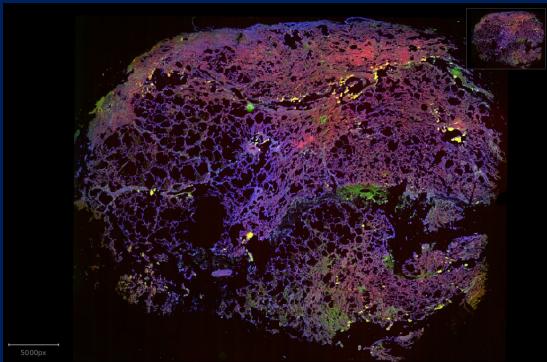


<https://valis.readthedocs.io/en/latest/>

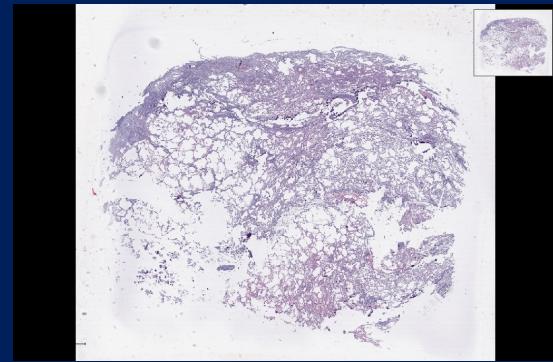
DAPI



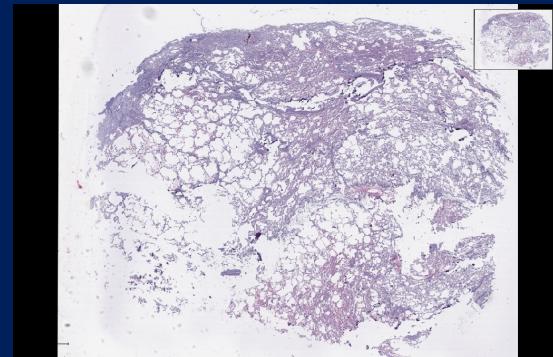
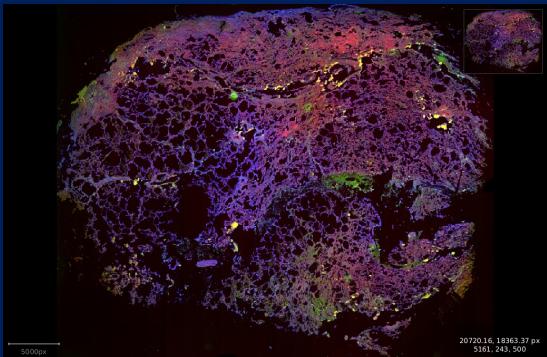
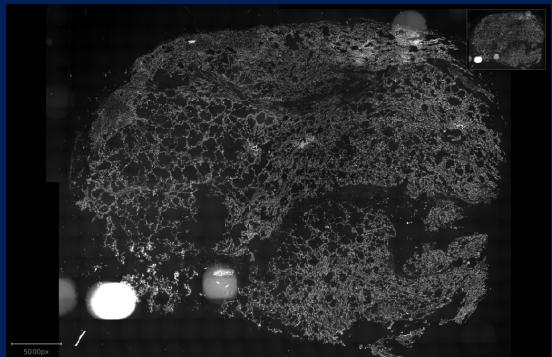
Membrane Staining



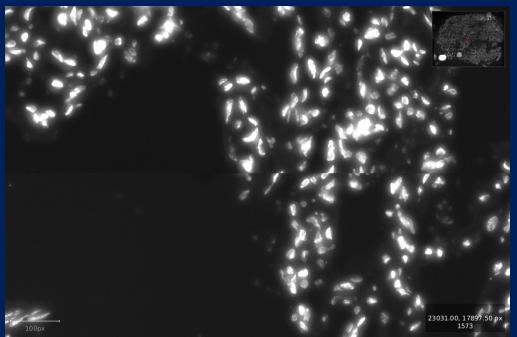
H&E Imaging



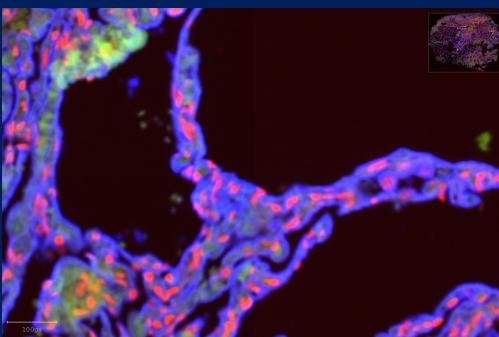
spot any differences?



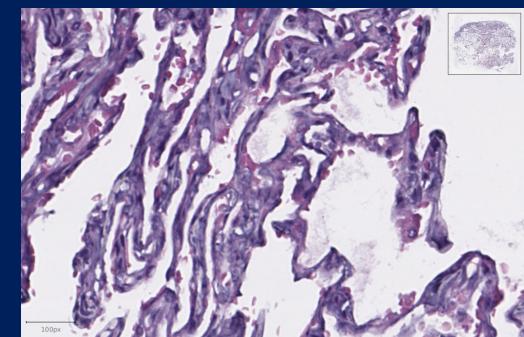
DAPI



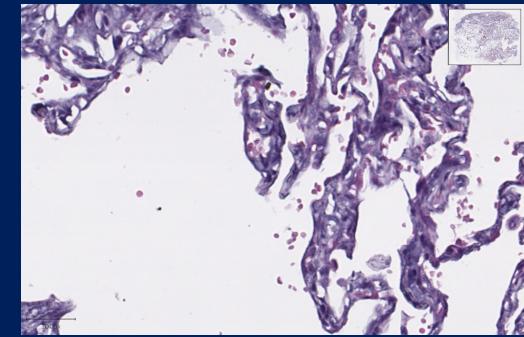
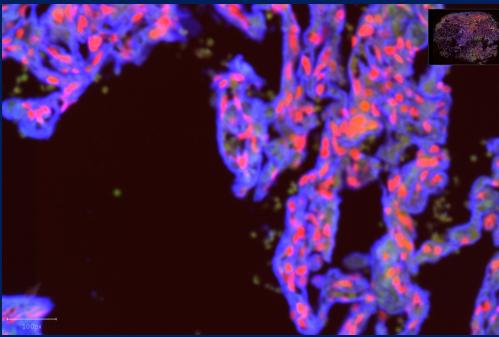
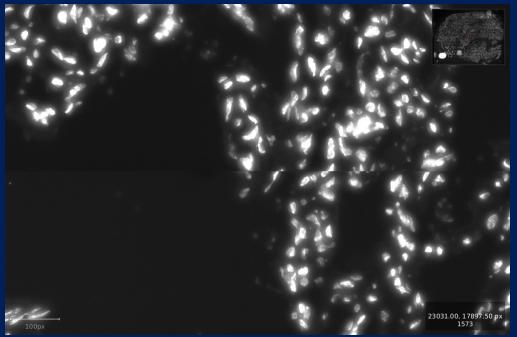
Membrane Staining



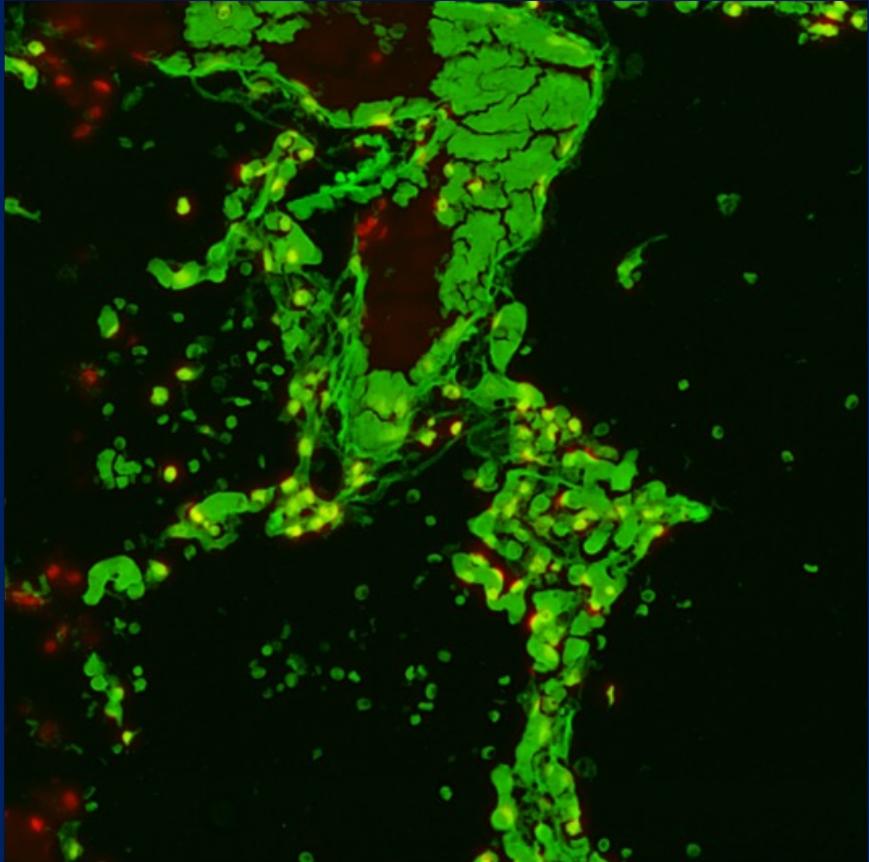
H&E Imaging



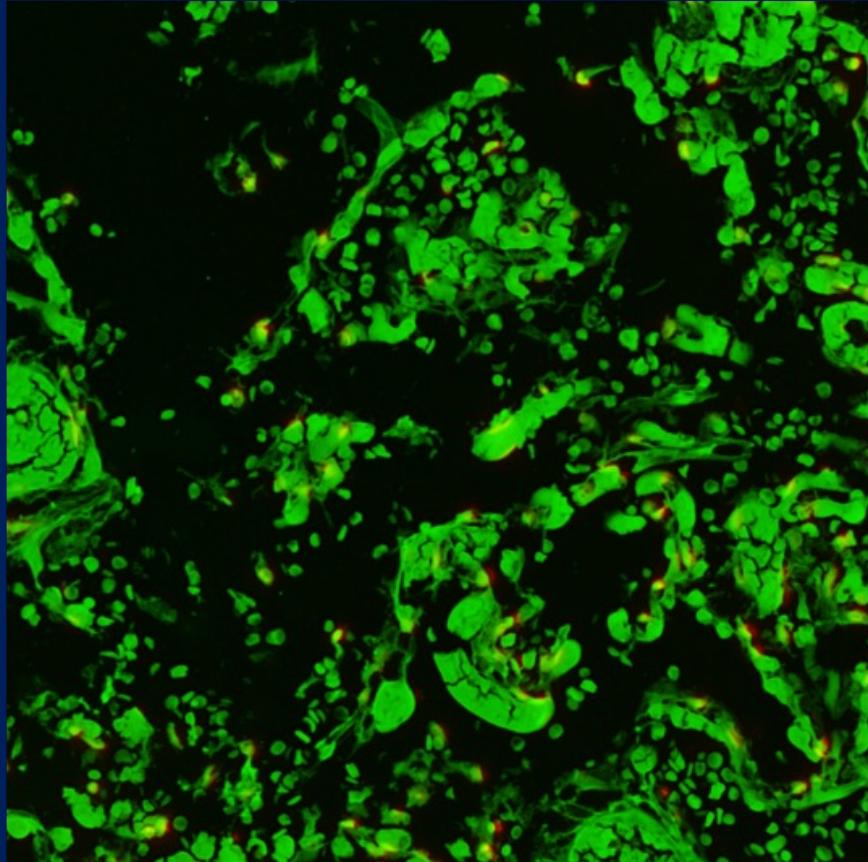
Spot any differences now?

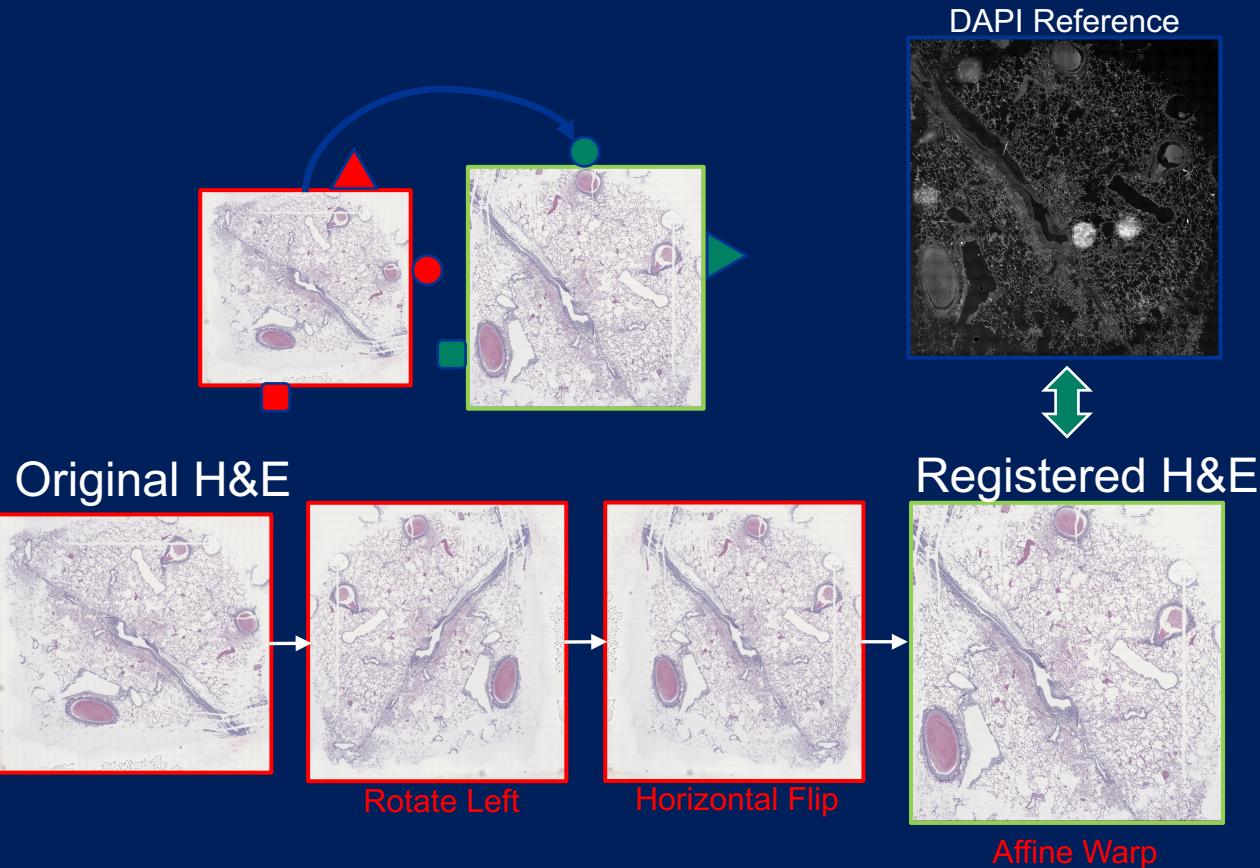


1-2C MATLAB_VALIS Registration H&E
Red Channel: DAPI, Blue Channel: Zero
Green Channel: gray image of H&E

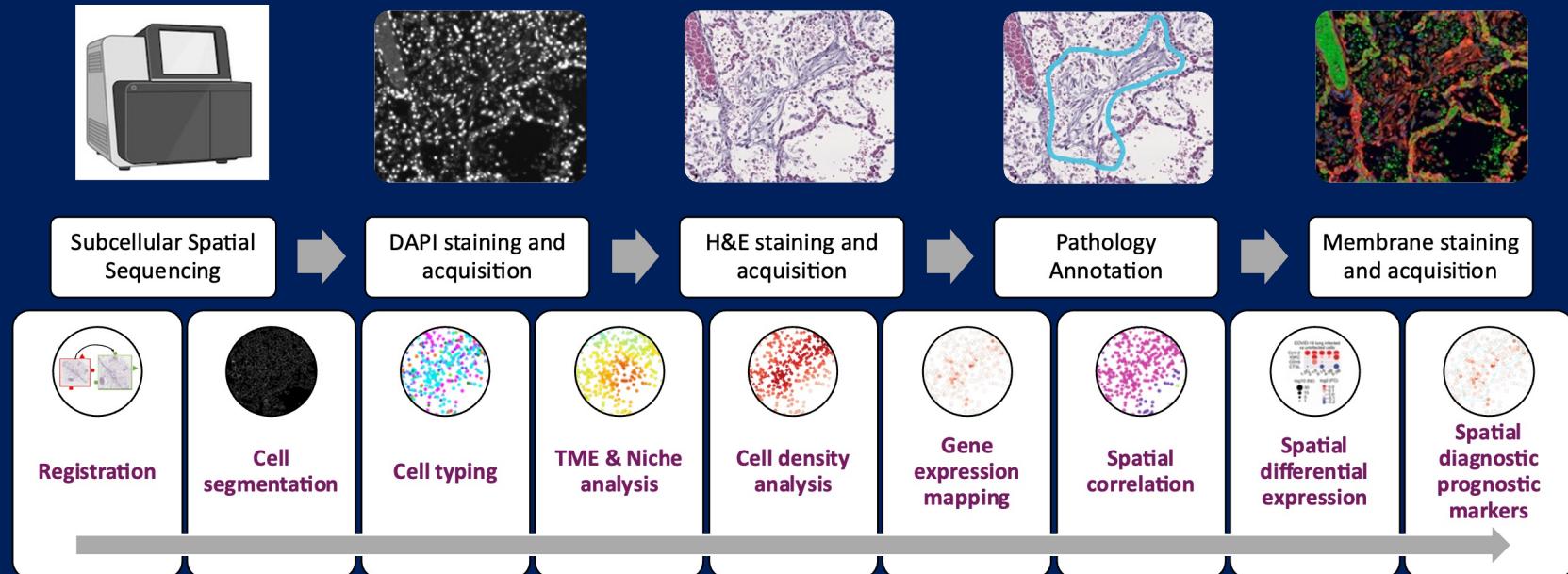


4-3B MATLAB_VALIS Registration H&E
Red Channel: DAPI, Blue Channel: Zero
Green Channel: gray image of H&E





Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



Cell Segmentation

- Image-based – CellPose
- Transcriptomics-based – Baysor

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>

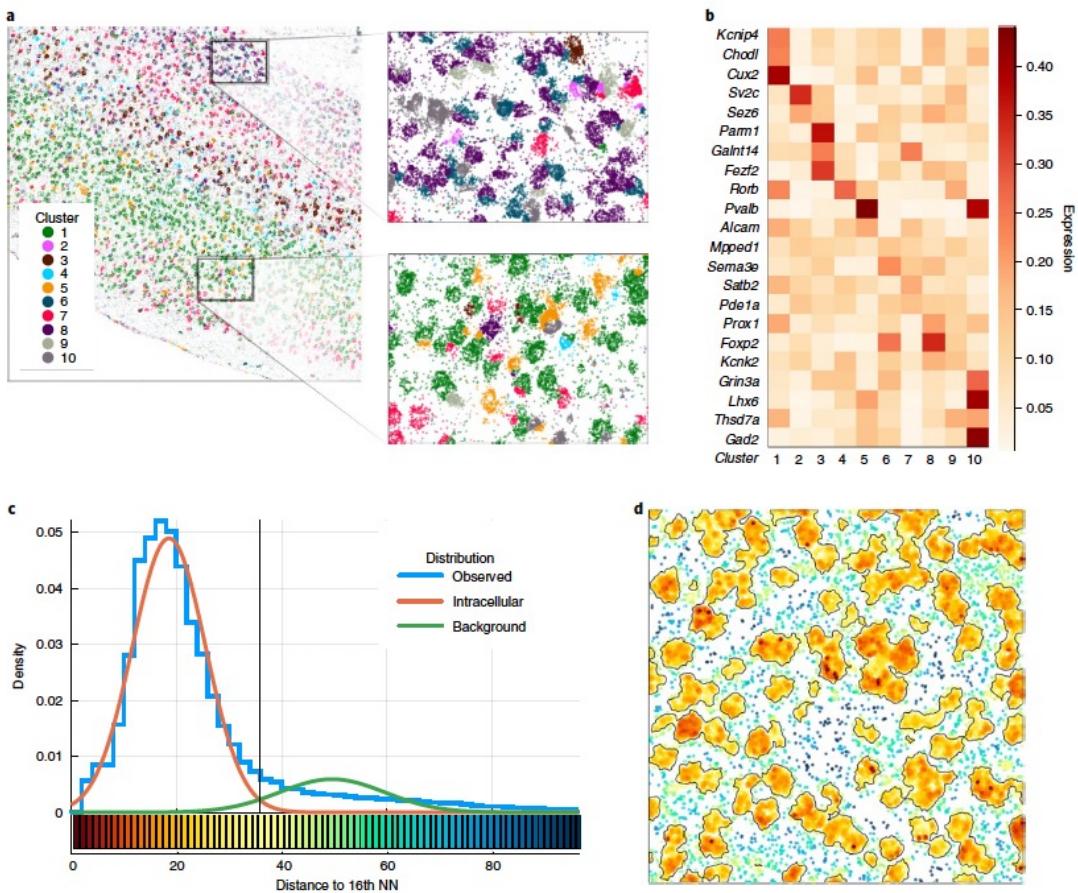


Fig. 2 | Application of an MRF framework for segmentation-free cell-type inference and background filtration. **a,b**, Individual molecules were clustered into major cell types by modeling the tissue as a mixture of multinomial distributions with an MRF prior. Cluster labels per molecule are shown in **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>

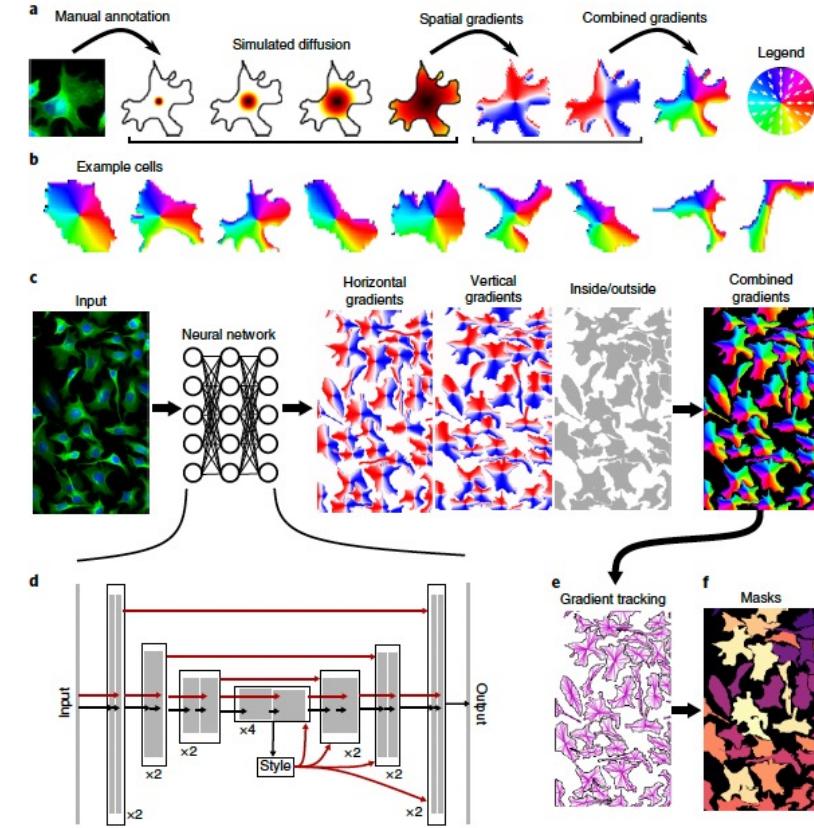
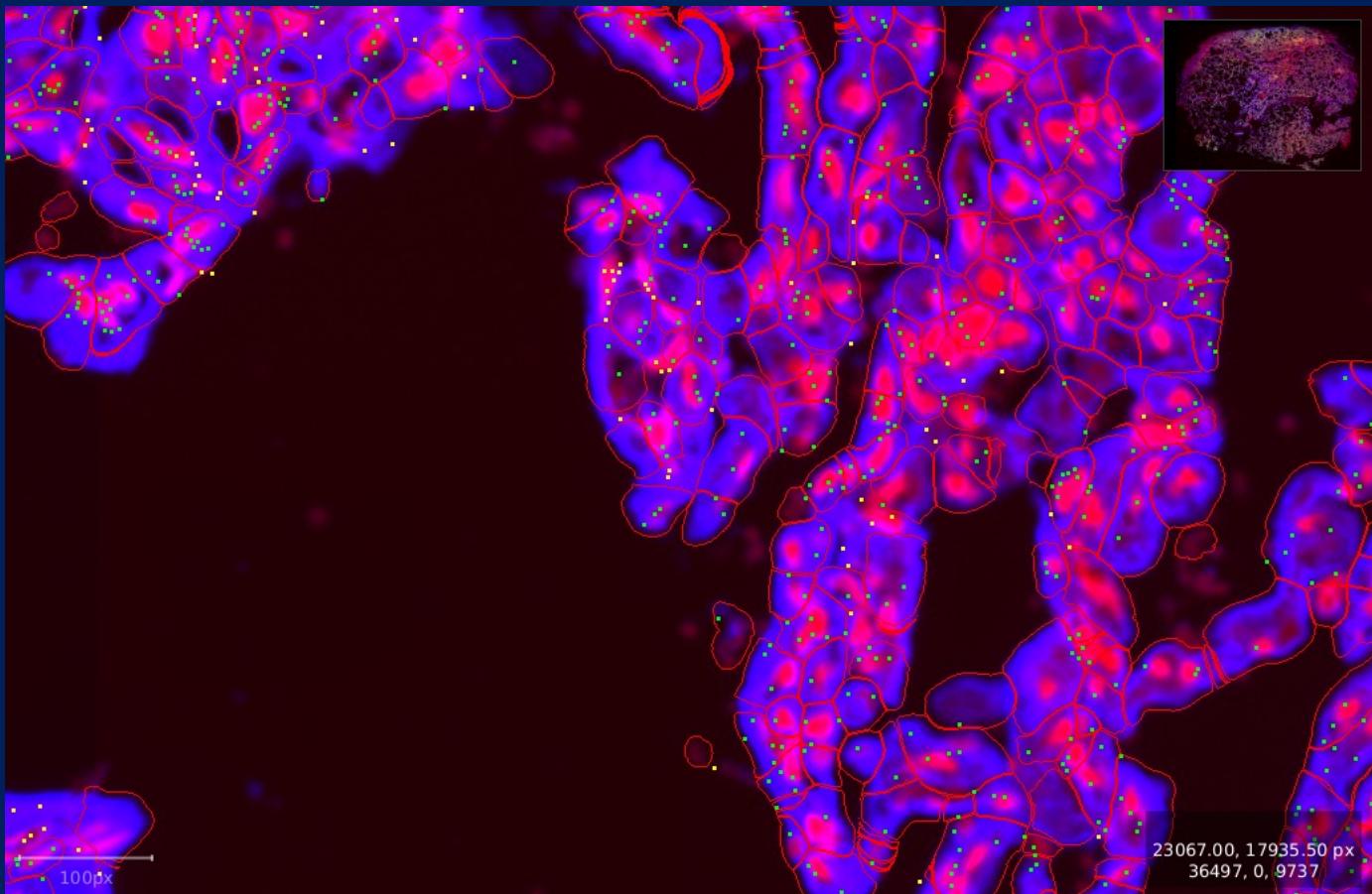
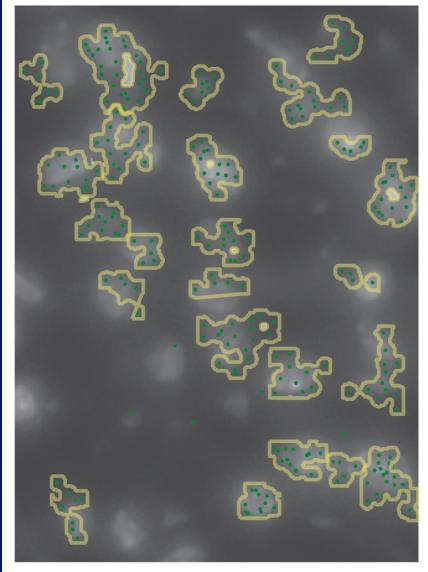


Fig. 1 | Model architecture. **a**, Procedure for transforming manually annotated masks into a vector flow representation that can be predicted by a neural network. A simulated diffusion process starting from the center of the mask is used to derive spatial gradients that point toward the center of the cell, potentially indirectly around corners. The x and y gradients are combined in a single direction from 0° to 360° . **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-Net³ 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

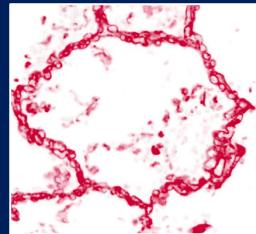
Baysor

Figure S4

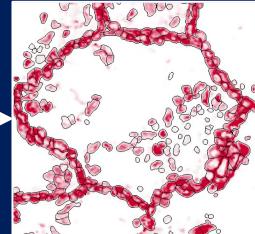


Registered
Membrane Staining

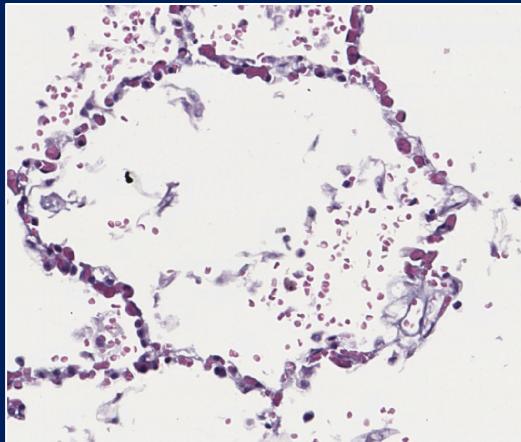
Cell Segmentation



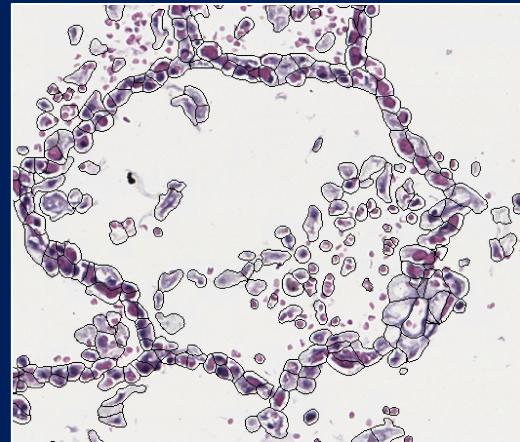
→ CellPose →



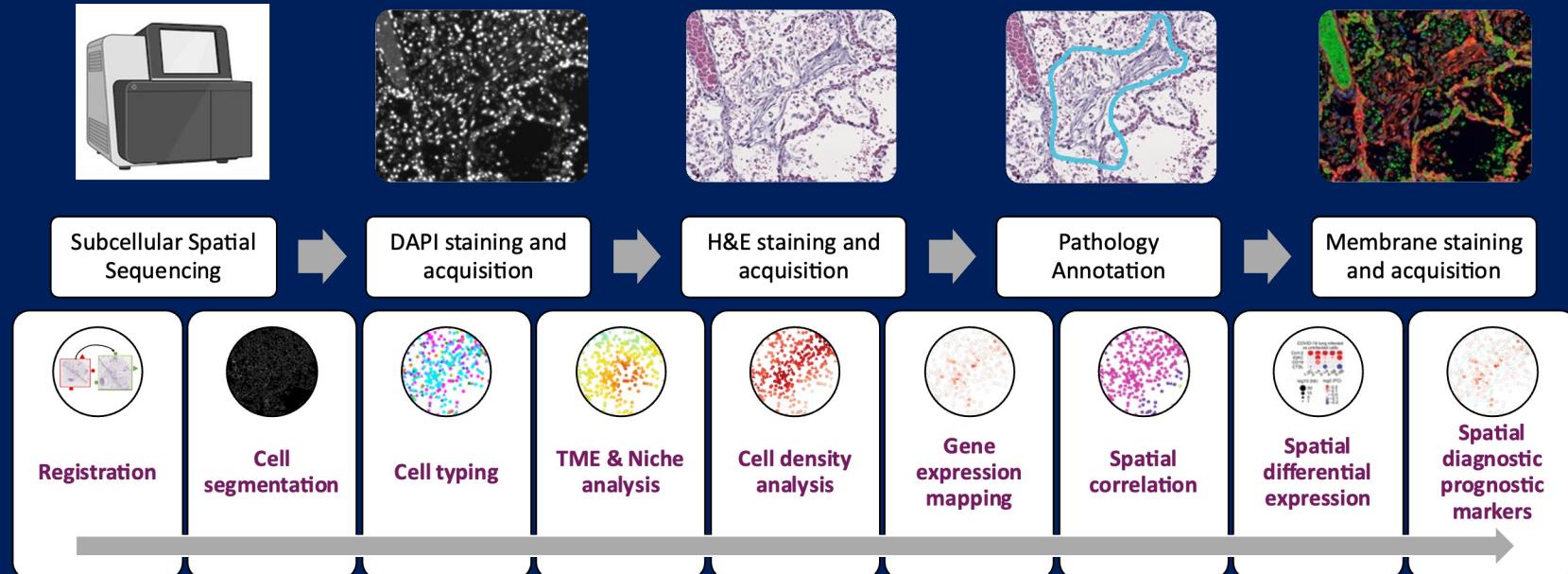
Registered H&E image



Cell Segmentation Overlay



Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



 University of
Pittsburgh®

UPMC
LIFE CHANGING MEDICINE

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 Zubair¹, Richard H. Chapple¹, Sivaraman Natarajan¹, William C. Wright¹, Min Pan¹, Hyoeng-Min Lee¹, Heather Tillman², John Easton^{3,4} and Paul Geeleher^{3,1*}

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

²Department 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 Janevski^{1,2}, Anil K. Giri^{1,3} & Tero Aittokallio^{1,2,3,4,5*}

Bioinformatics Advances, 2023, vbad030
<https://doi.org/10.1093/bioadv/vbad030>
Advance Access Publication Date: 13 March 2023
Original Paper

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

Spatial-ID: a cell typing method for spatially resolved transcriptomics via transfer learning and spatial embedding

Received: 17 June 2022

Accepted: 25 November 2022

Published online: 10 December 2022

Rongbo Shen^{1,2}, Lin Liu^{2,3}, Jizhan Wu^{1,7}, Ying Zhang^{2,7}, Zhiyuan Yuan^{1,3}, Junfu Guo^{2,7}, Fan Yang¹, Chao Zhang², Bichao Chen^{1,2}, Wanwan Feng^{1,4}, Chao Liu², Jing Quo², Guochen Fan², Yong Zhang^{2,5}, Yuxiang Li^{2,6}, Xun Xu^{1,2*} & Jianhua Yao^{1,2}

COMPUTATIONAL
AND STRUCTURAL
BIOTECHNOLOGY
JOURNAL

journal homepage: www.elsevier.com/locate/csbj

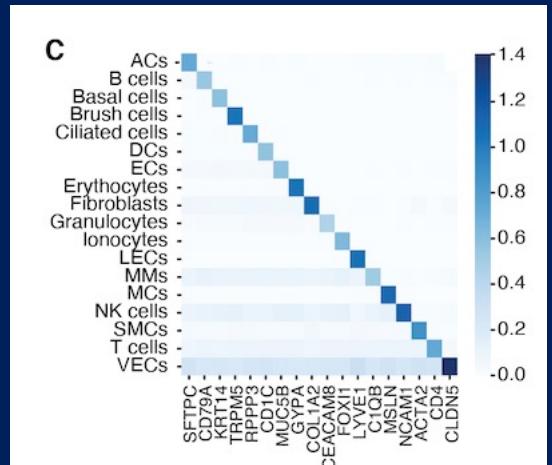
Automated methods for cell type annotation on scRNA-seq data

Giovanni Pasquini^{3,4,*}, Jesus Eduardo Rojo Arias⁵, Patrick Schäfer³, Volker Busskamp^{3,4,C,*}

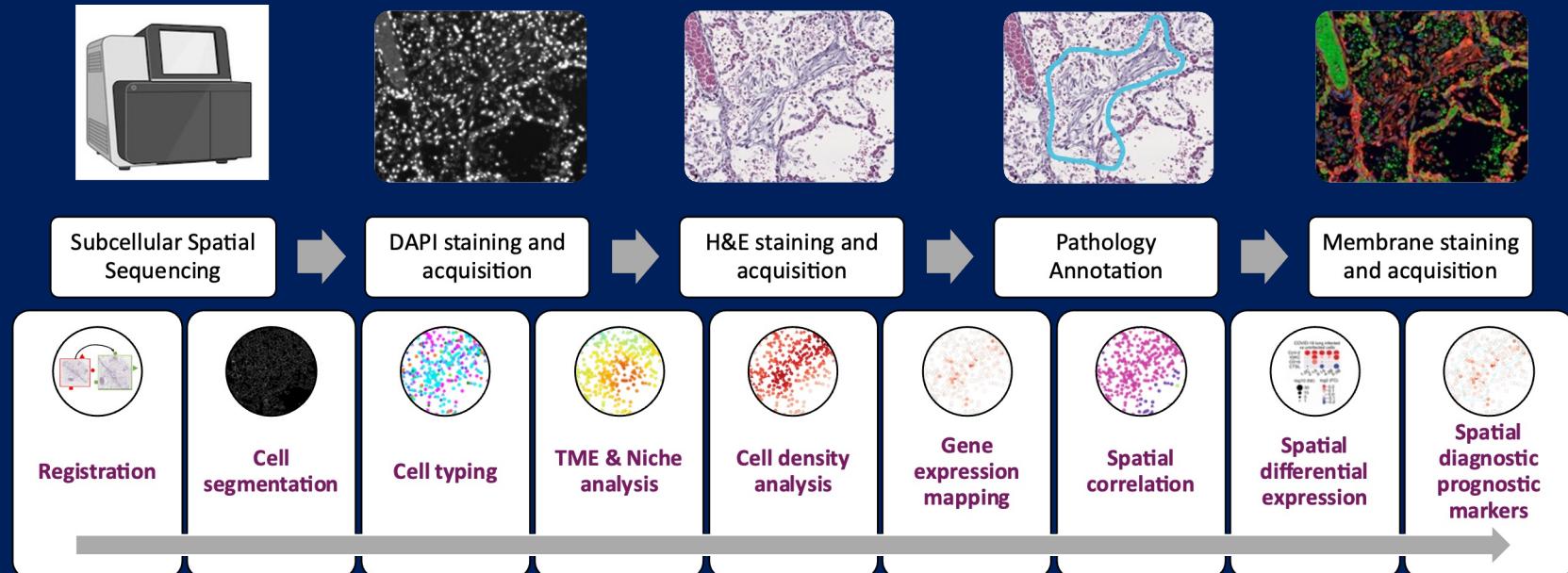
³ Technische Universität Dresden, Center for Molecular and Cellular Bioengineering (CMCB), Center for Regenerative Therapies Dresden (CRTD), Dresden 01307, Germany

⁵ Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheung Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK

⁴ Universitäts-Augenklinik Bonn, University of Bonn, Department of Ophthalmology, Bonn 53127, Germany



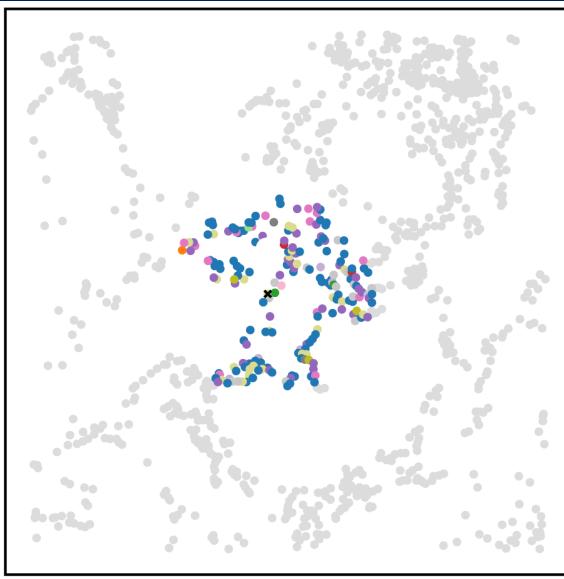
Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



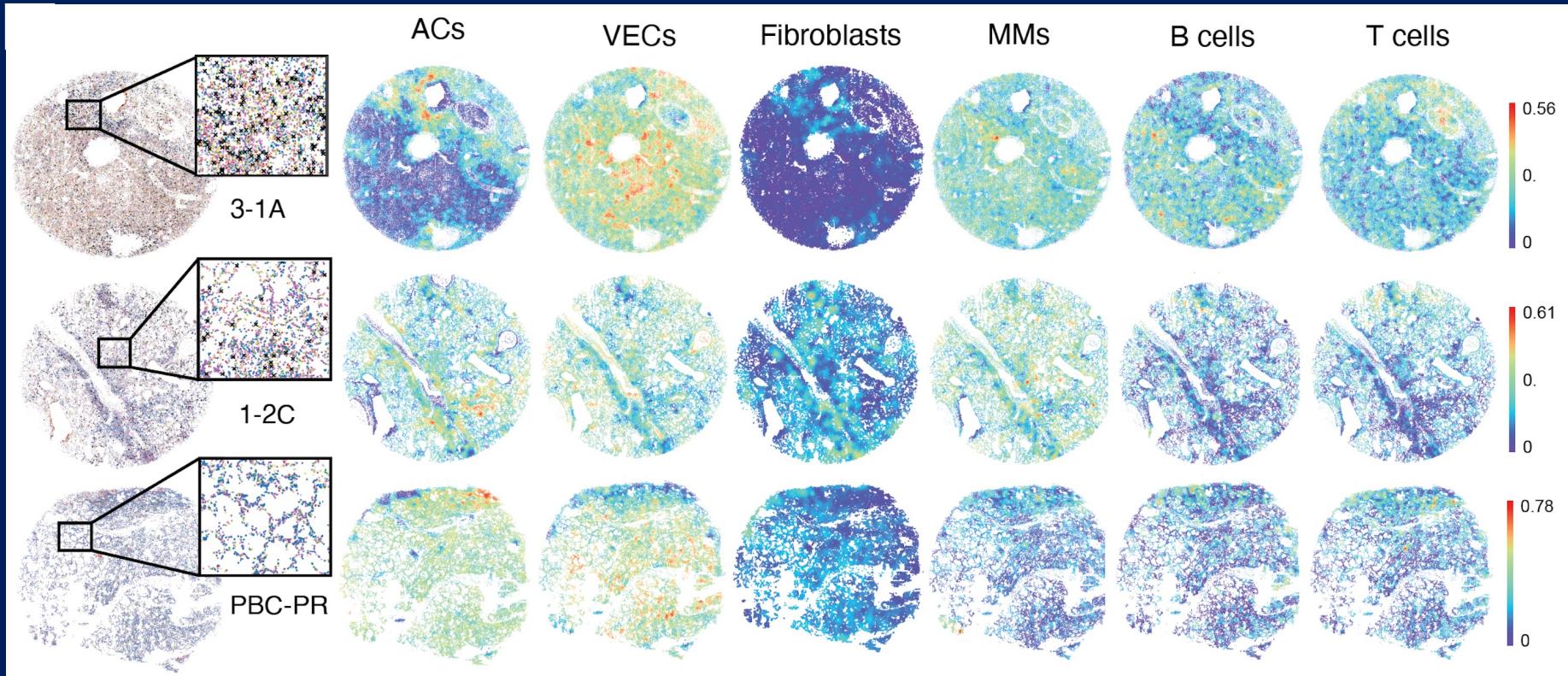
Niche Analysis

Algorithm:

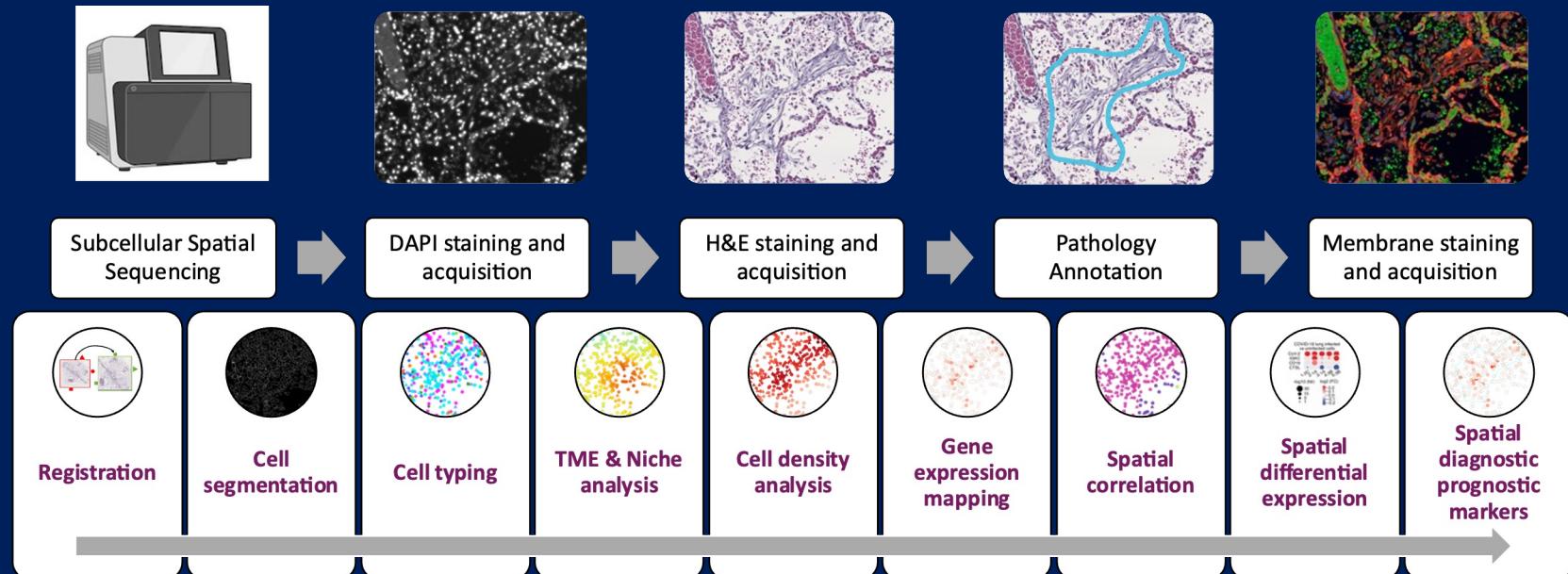
1. Pick one cell.
2. Count the number of each cell type in the neighborhood.
 - We now have a cell type count vector for which is the local ***Neighborhood Cell-type Composition (NCTC)***.
3. Repeat steps 1-2.



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

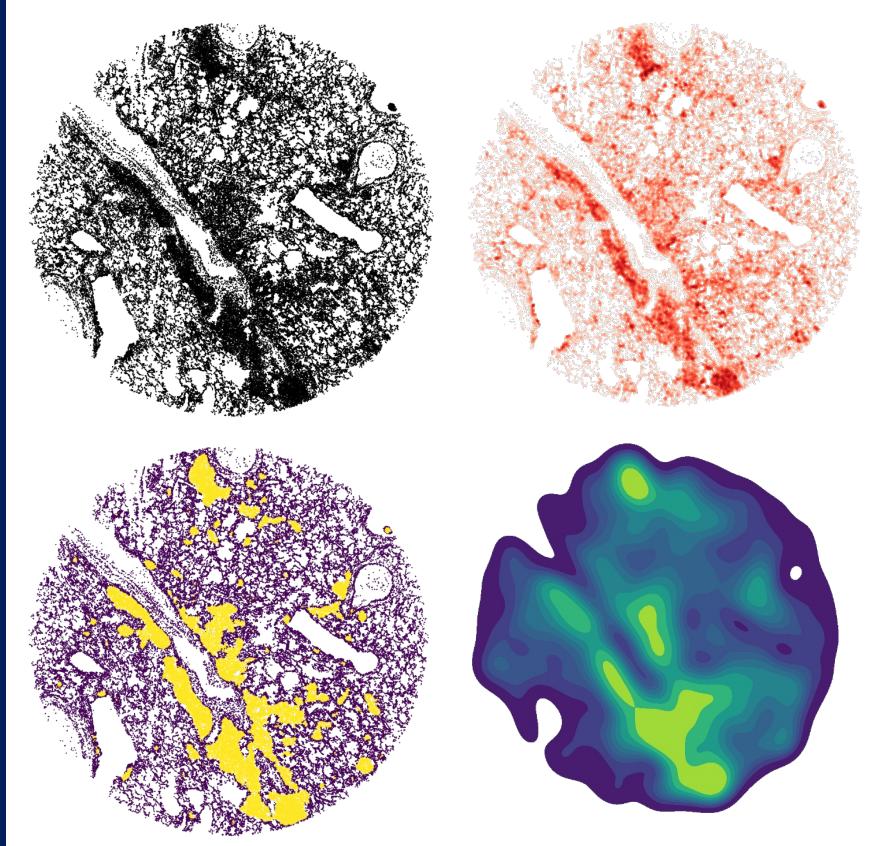


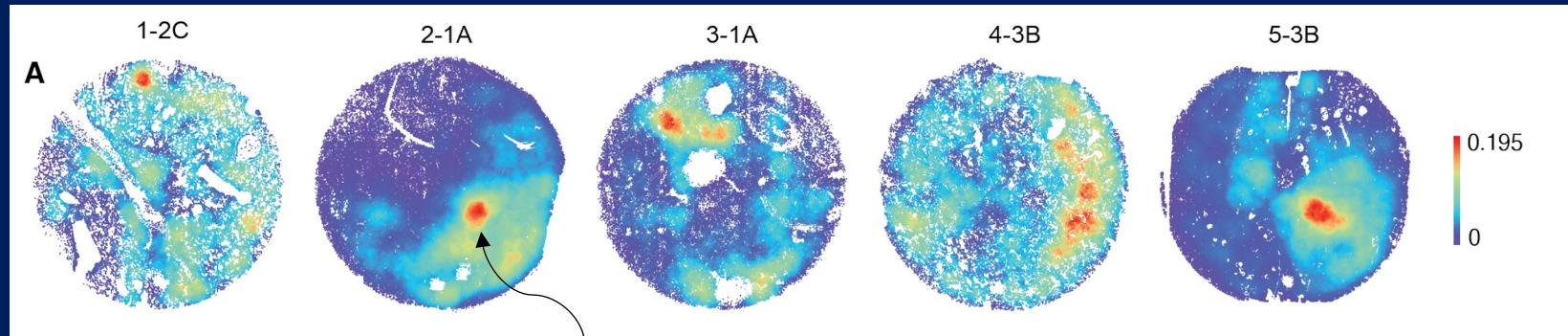
Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



Cell Density Analysis

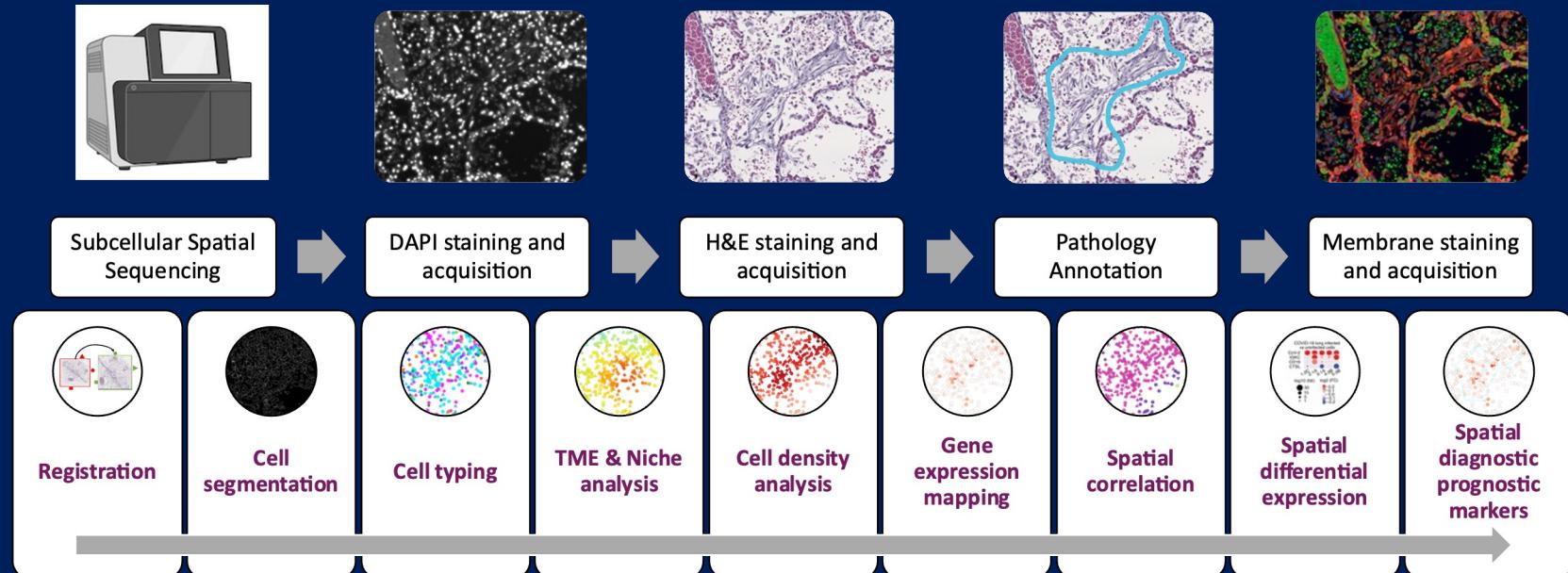
Instead of fixing the number of cells like we did for NCTC, take a count all the cells inside a circular region of fixed radius.





High infection regions

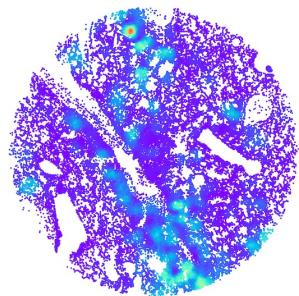
Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



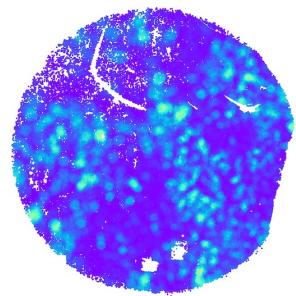
Gene Expression Mapping

COL1A1

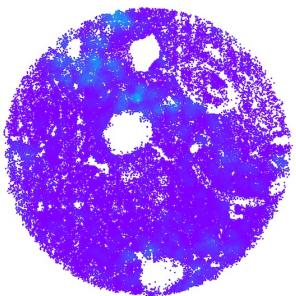
1-2C



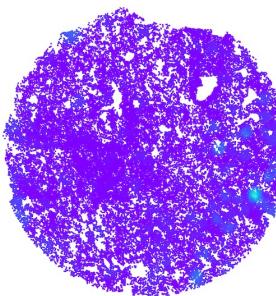
2-1A



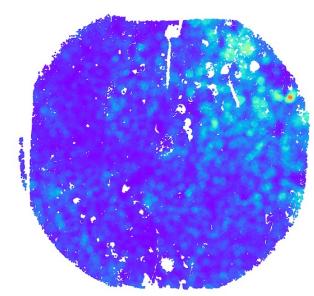
3-1A



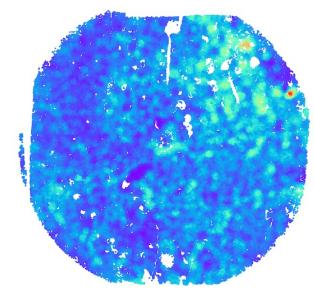
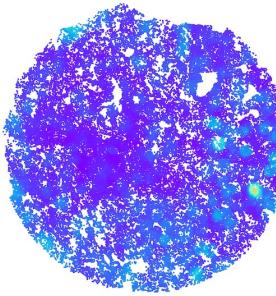
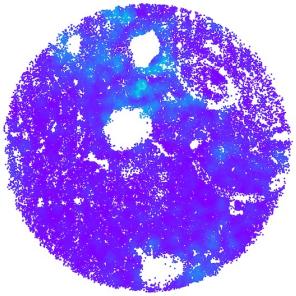
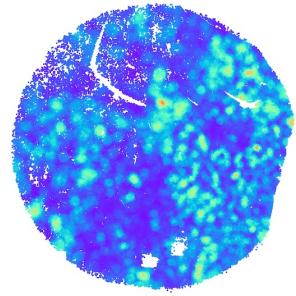
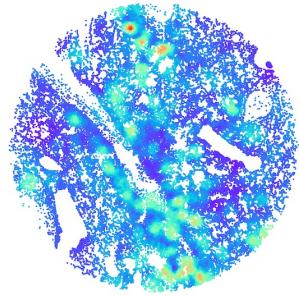
4-3B



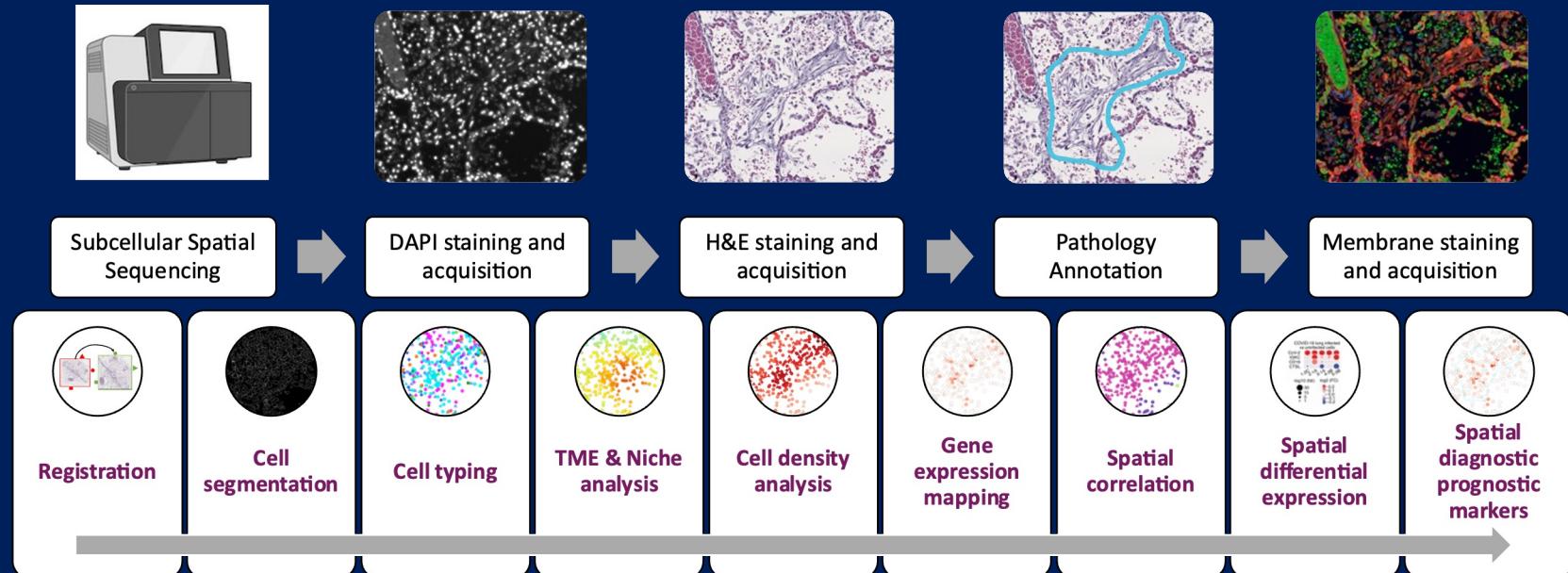
5-3B



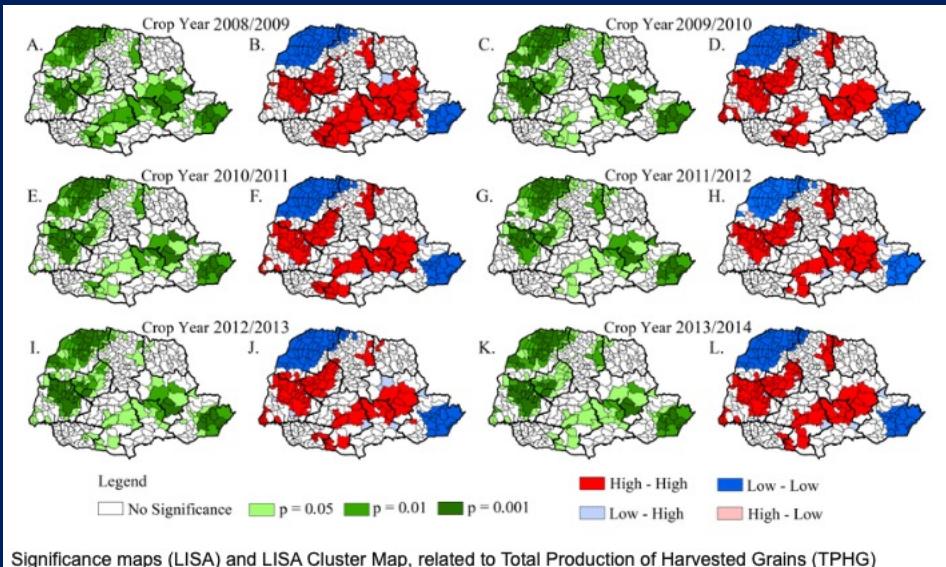
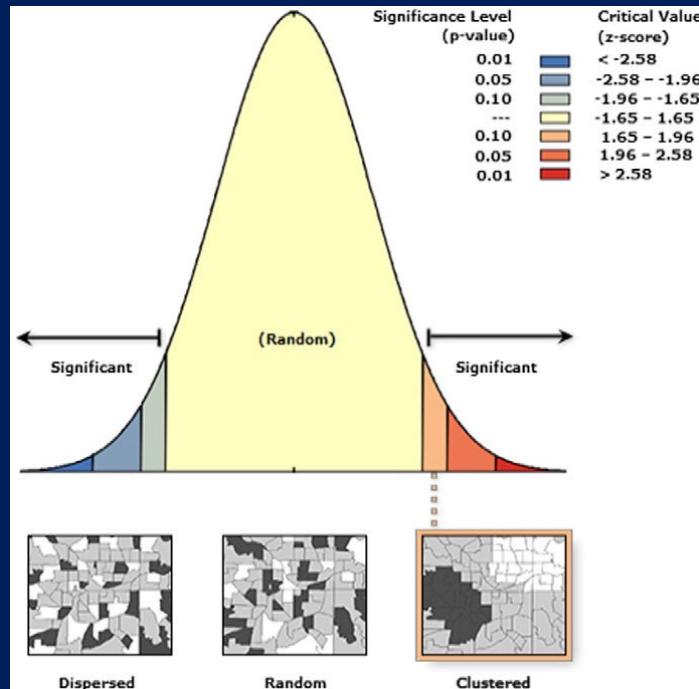
COL1A2



Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



Spatial Correlation Analysis



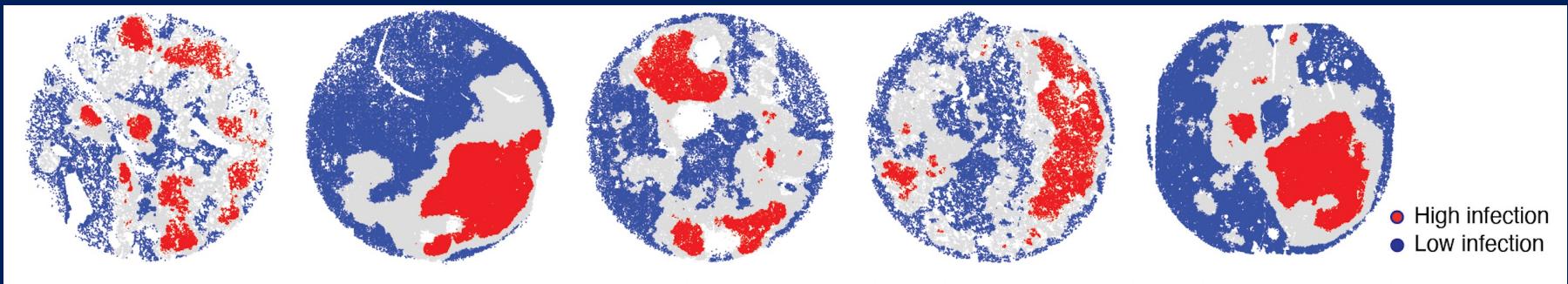
Significance maps (LISA) and LISA Cluster Map, related to Total Production of Harvested Grains (TPHG) (soybean, corn 1st and 2nd harvests and wheat) for the 2008/2009 harvest years (A and B); 2009/2010 (C and D); 2010/2011 (E and F); 2011/2012 (G and H); 2012/2013 (I and J) and 2013/2014 (K and L).

Spatial Correlation Analysis

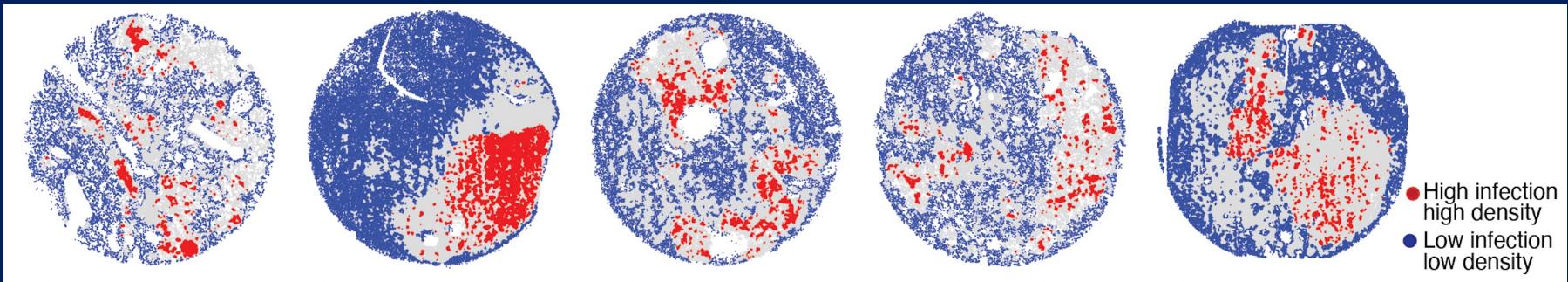
- What kind of questions can we ask?
- Can we identify high-density spatial clusters?
- What about spatial hotspots of infection?

Spatial Correlation Analysis

Moran's I Autocorrelation

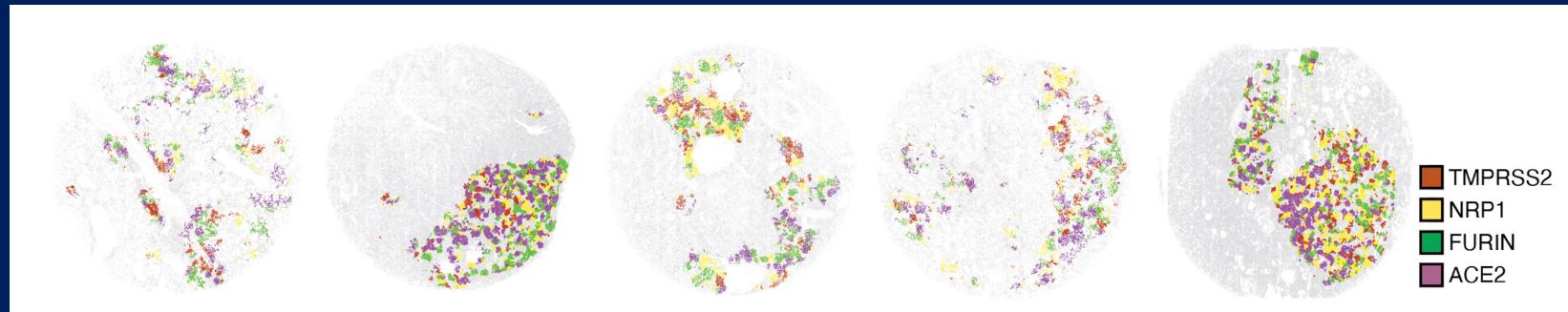


Moran's I Cross-correlation

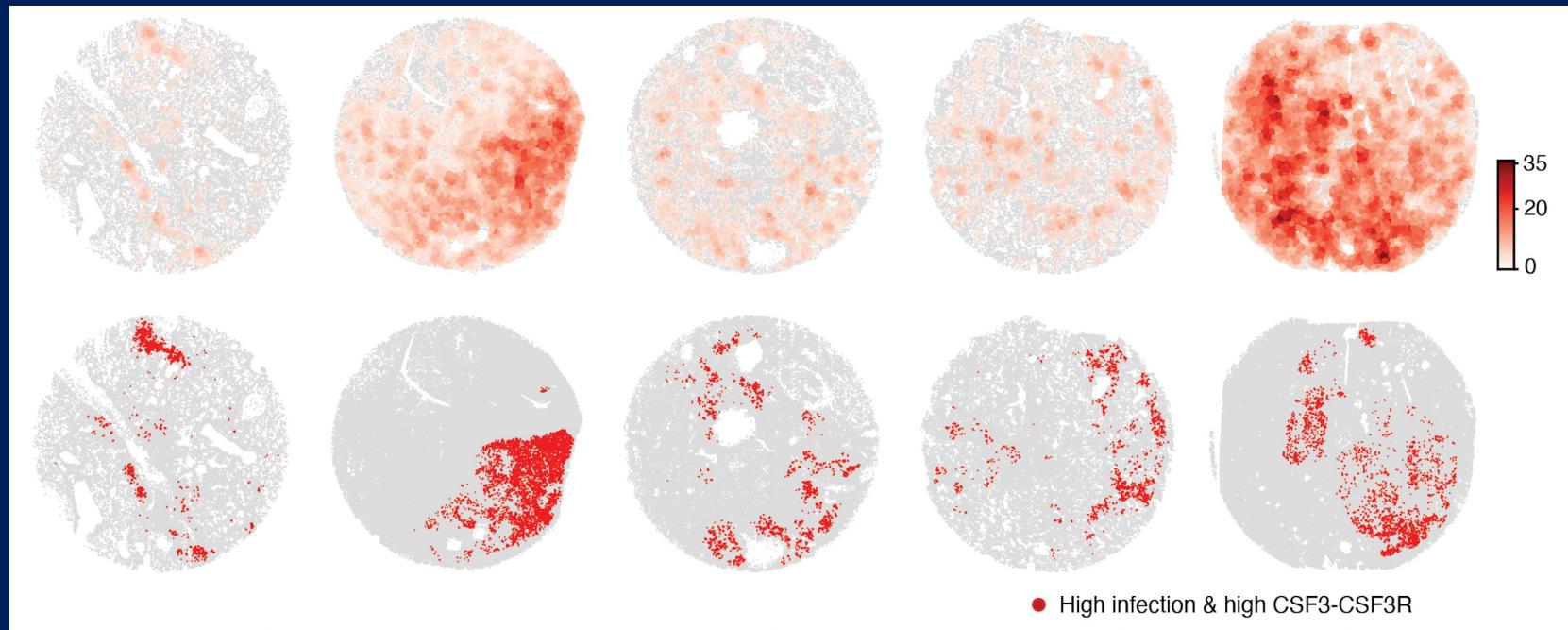


Spatial Correlation Analysis

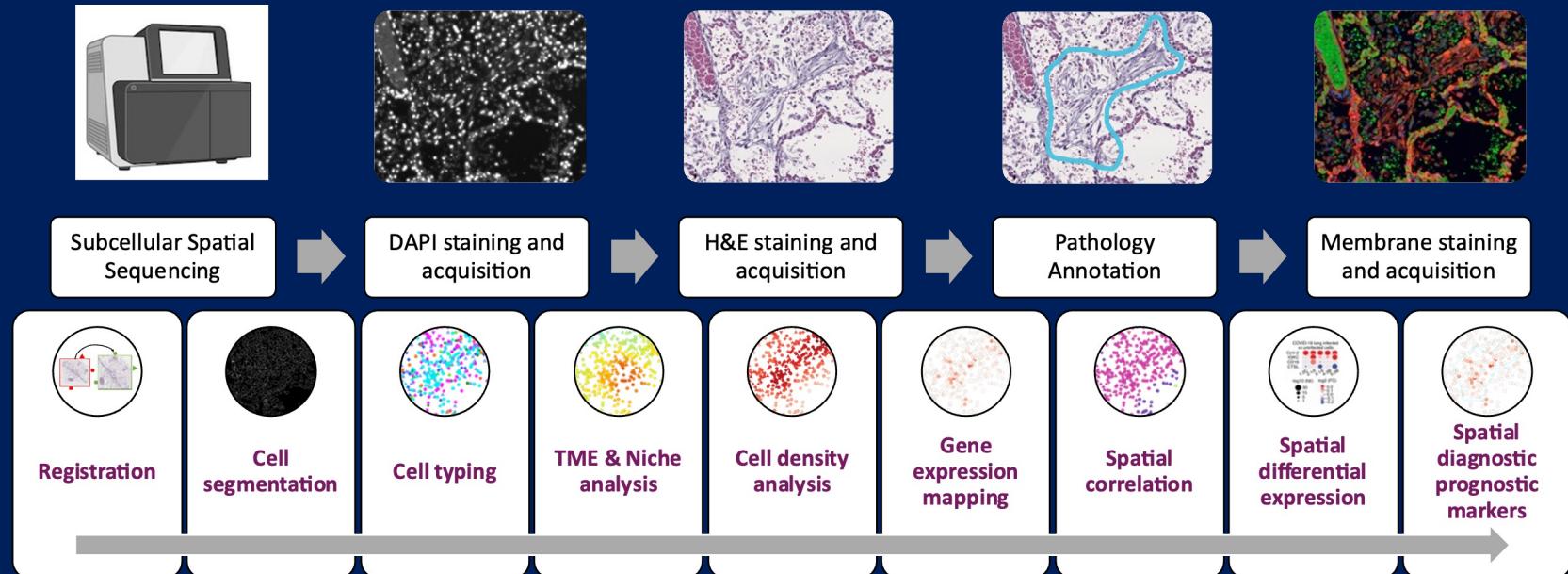
Bivariate Moran's I analysis to summarize high-infection regions with high entry protein expression in each tissue sample.



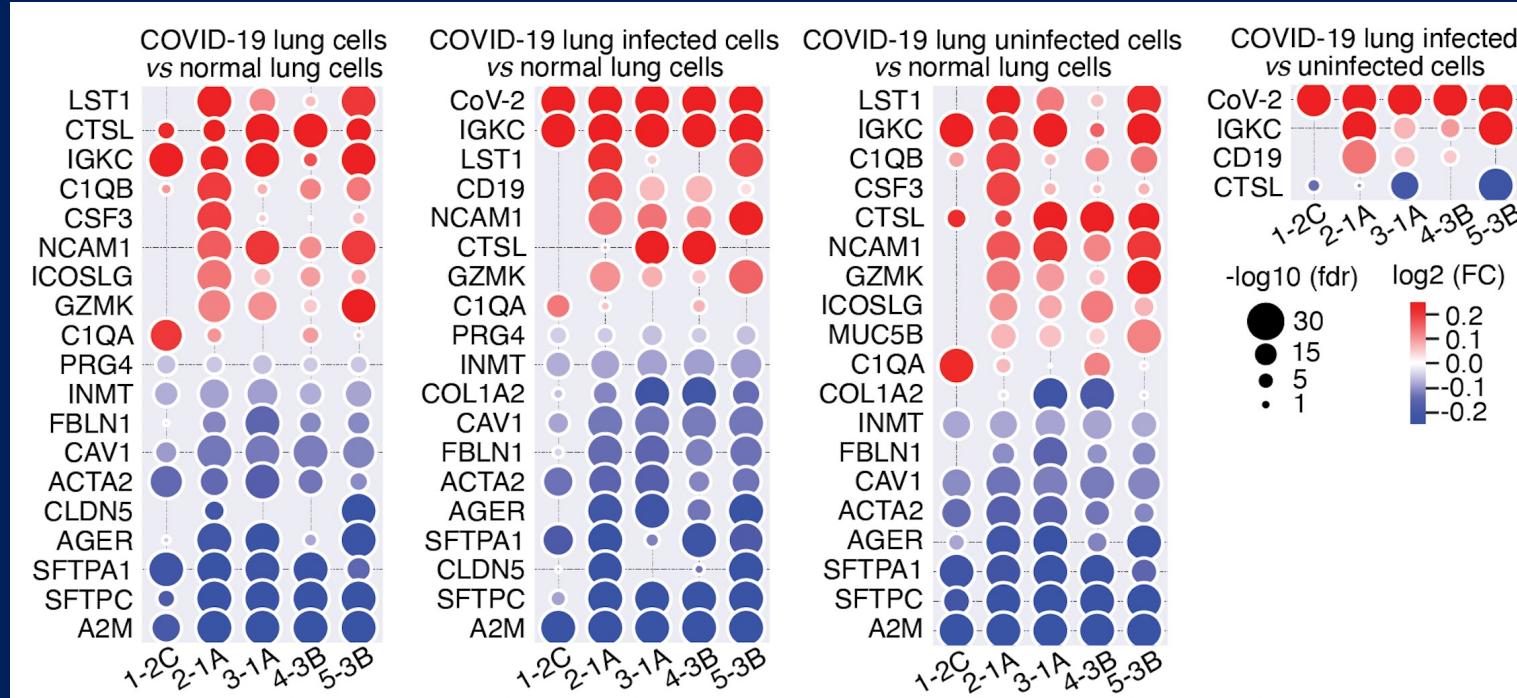
Spatial Correlation Analysis



Single-Cell Spatial Transcriptomics Analysis Pipeline (ScSTA)



Differential Expression Analysis



Let's Code:

https://github.com/arundasan91/ScSTA_tutorial

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