

*This pdf presentation contains screenshots instead of the originally presented videos. For interest in the video material please reach out to the speaker.*

# Visual Analytics for Imaging-Based Spatial Profiling

Robert Krueger, Ph.D.

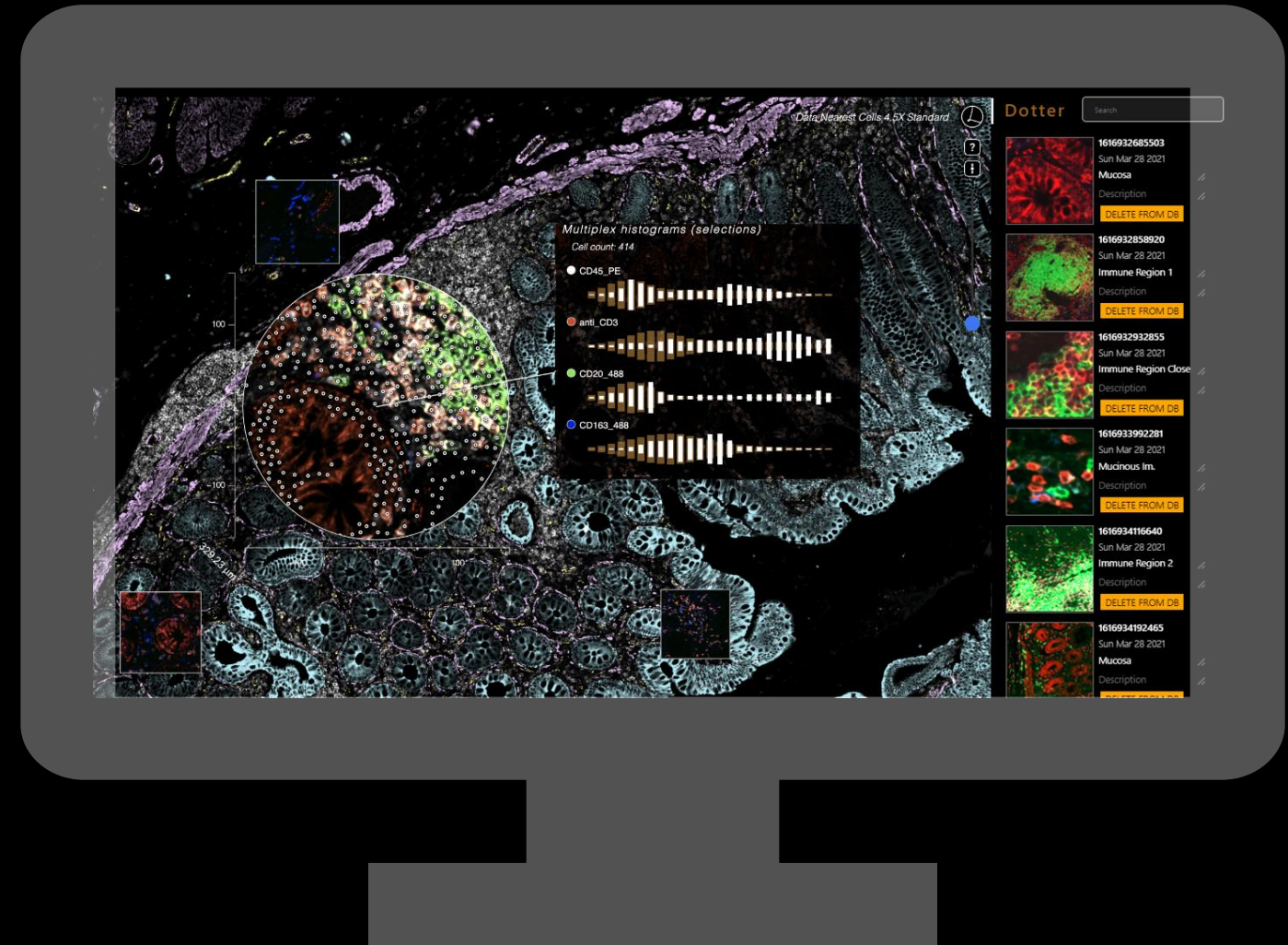


Visual Computing Group @  
Harvard School of  
Engineering and Applied  
Sciences



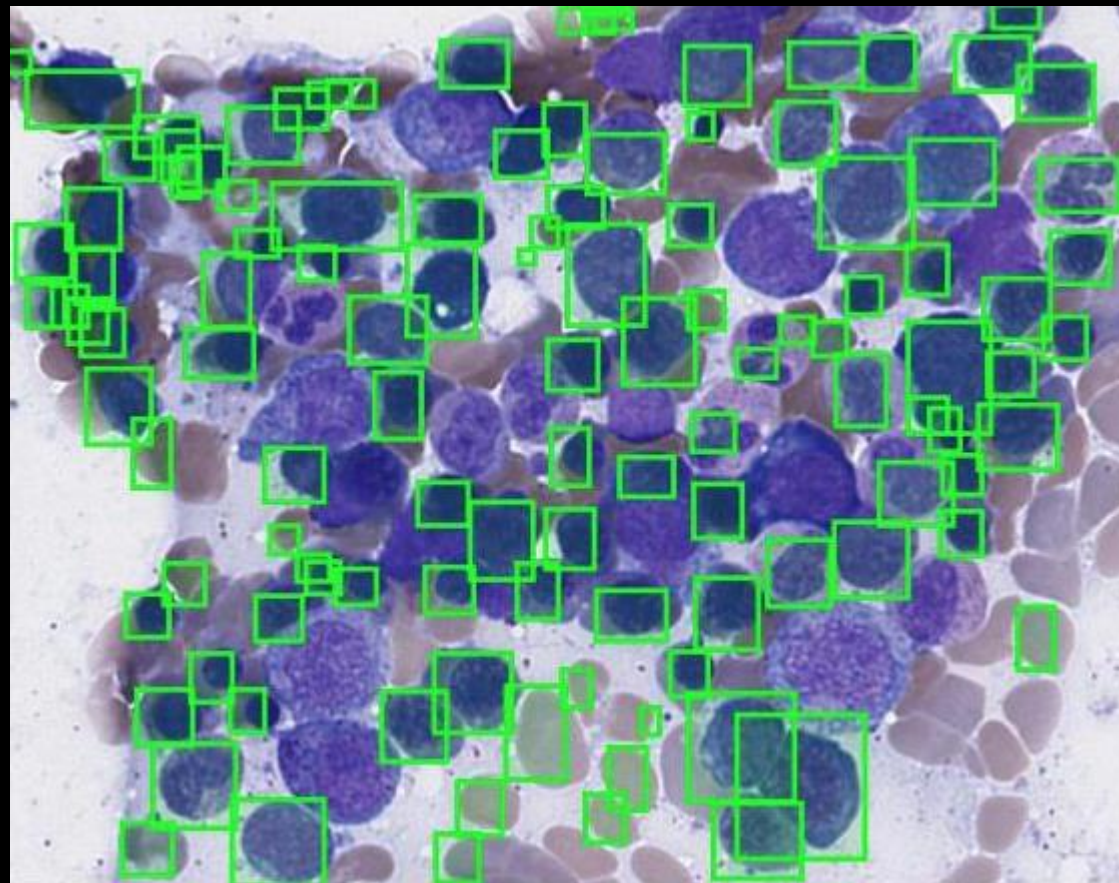
Laboratory of Systems  
Pharmacology  
@Harvard Medical School

# Digital Pathology

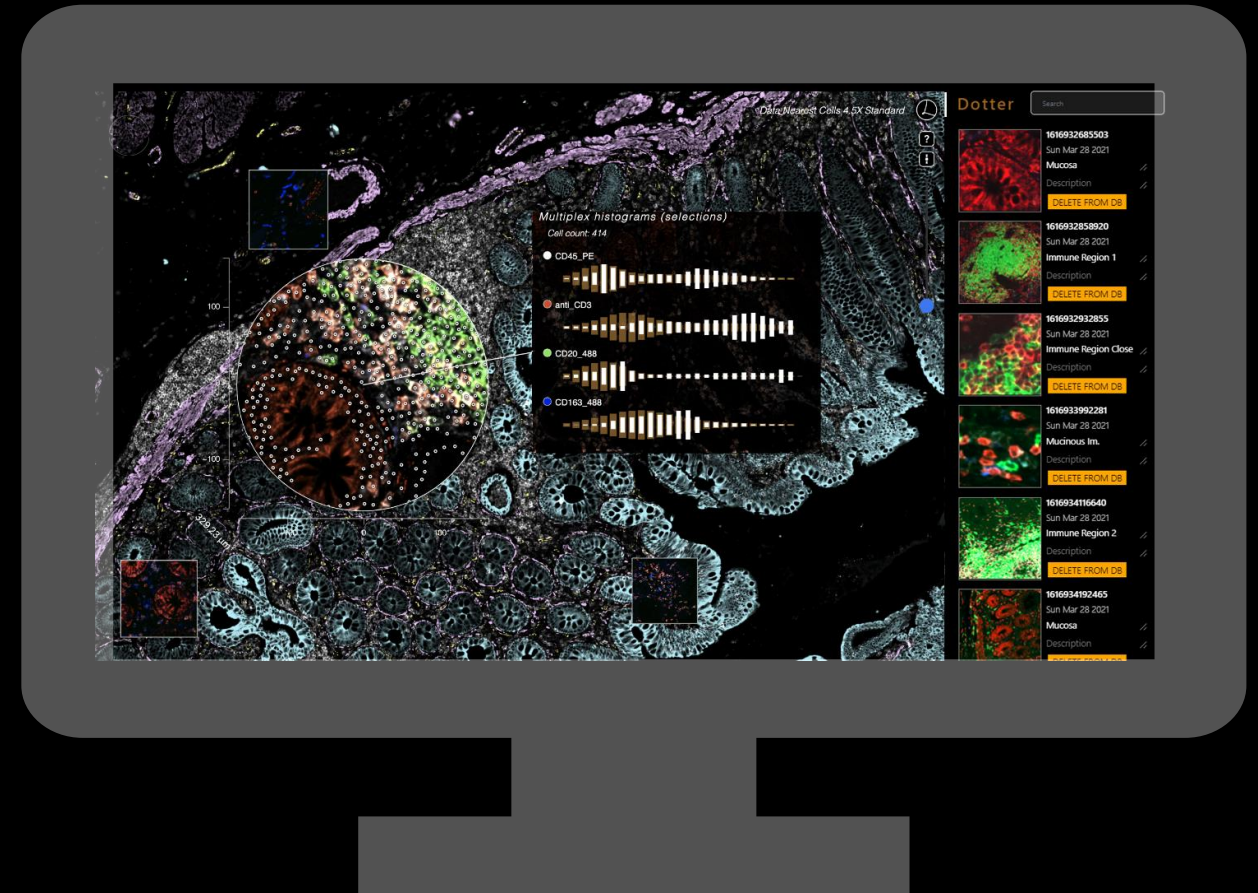


# Semi-Automated Workflows

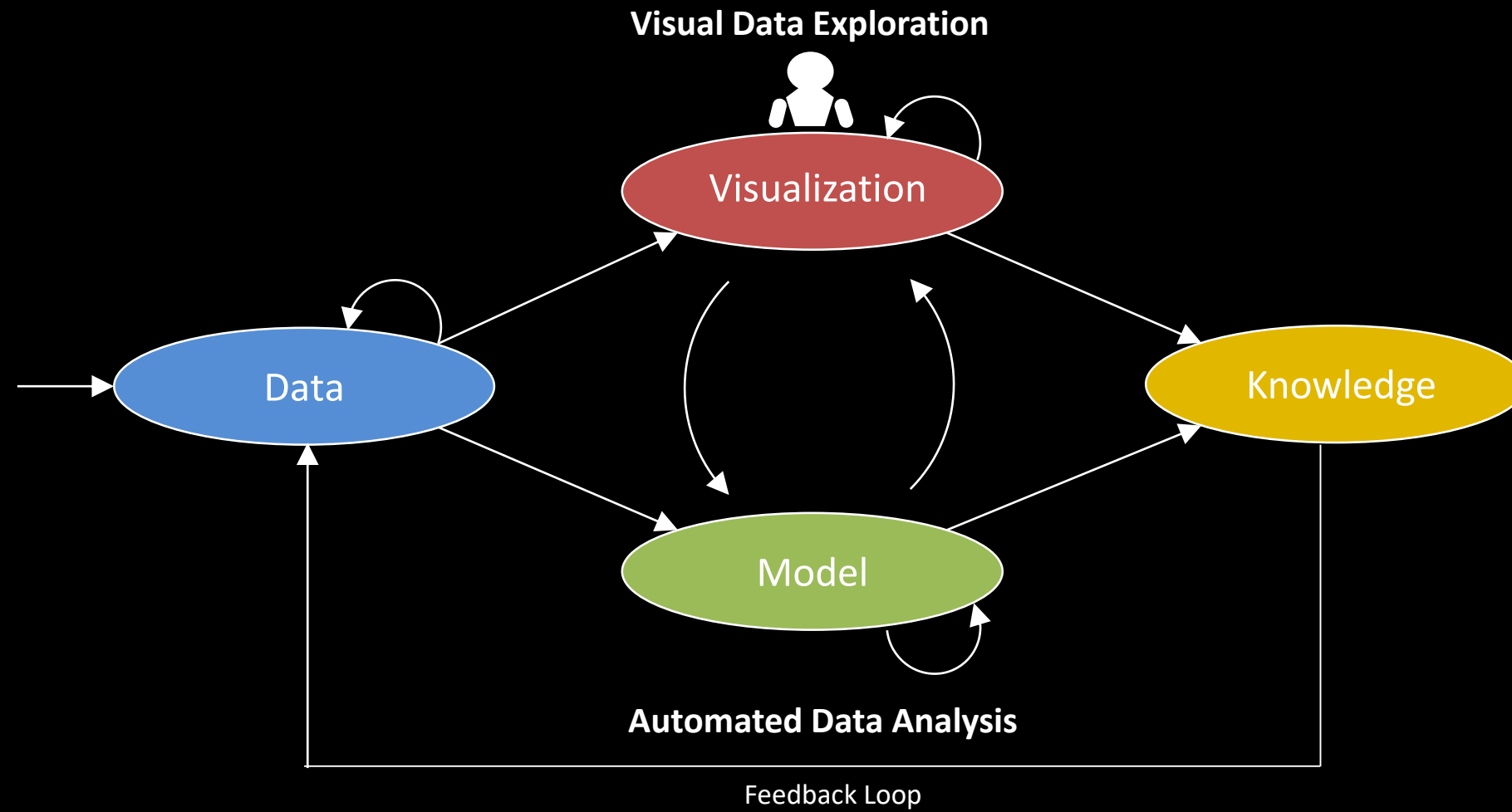
## Machine Learning



## Expert-in-the-Loop

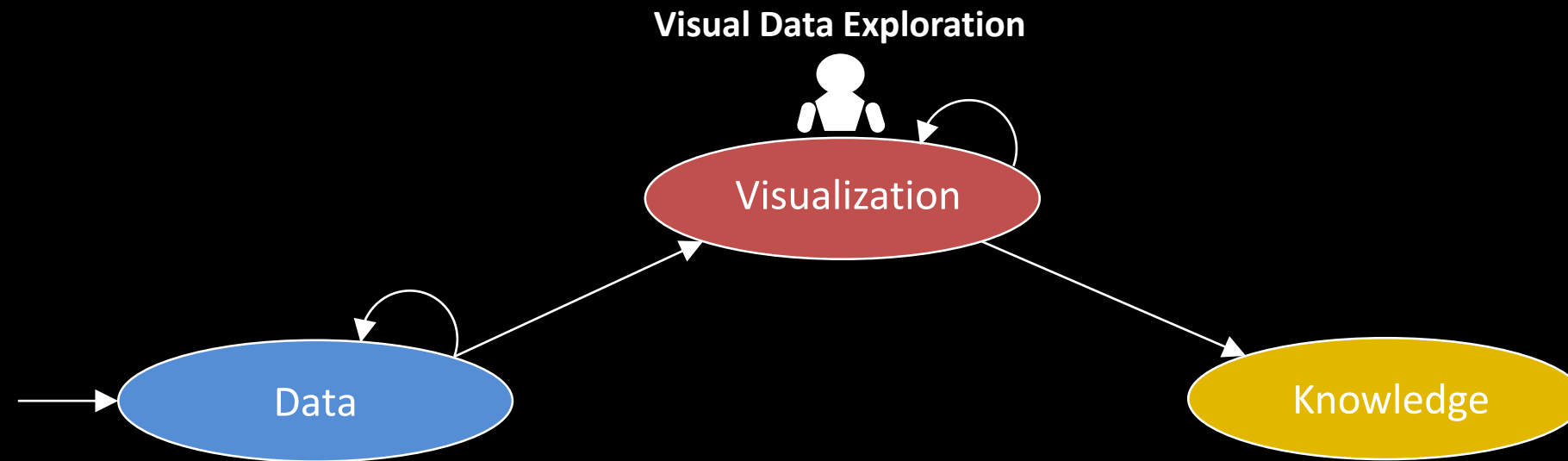


# Visual Analytics



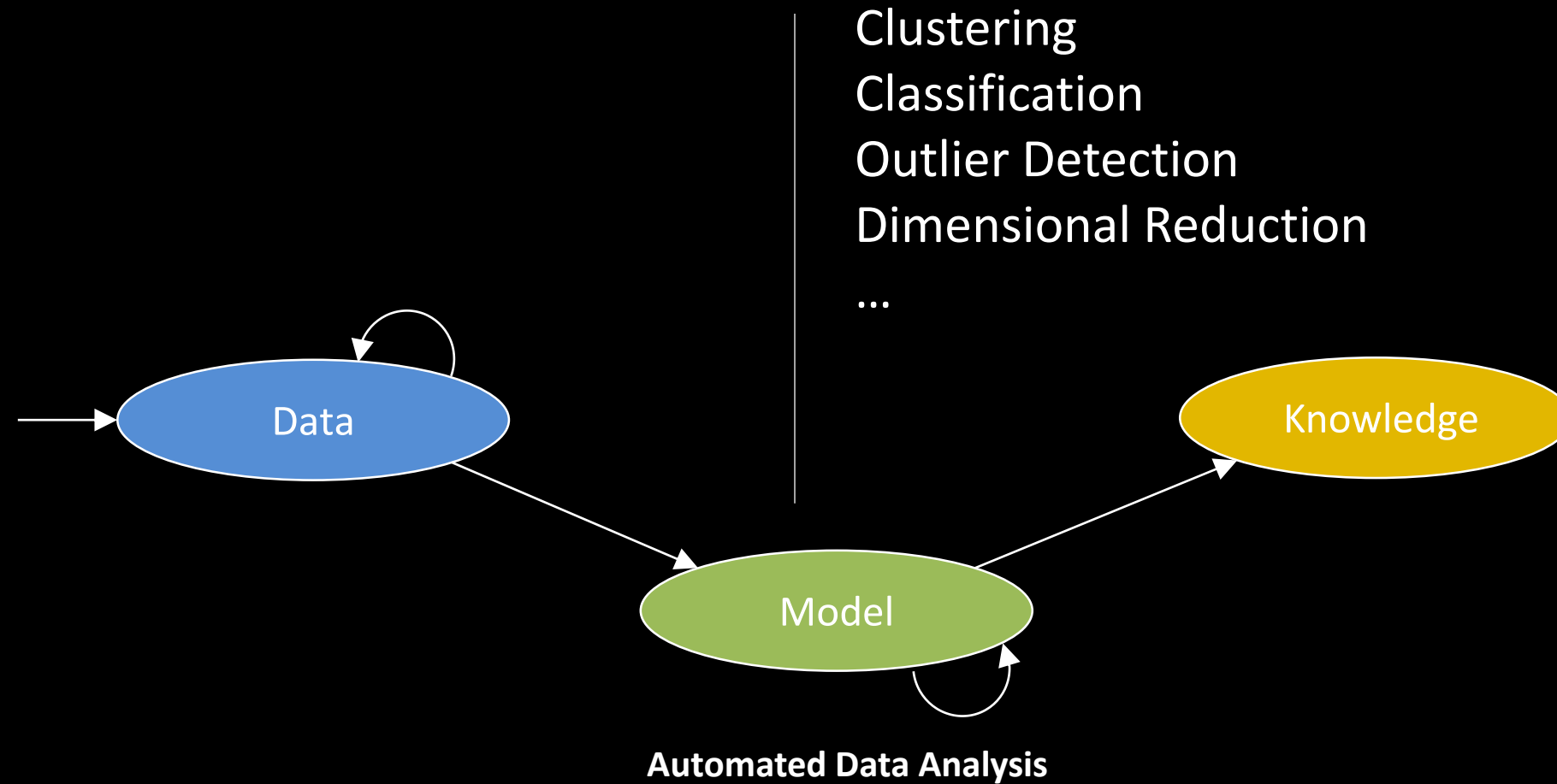
D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008.

# Visual Analytics



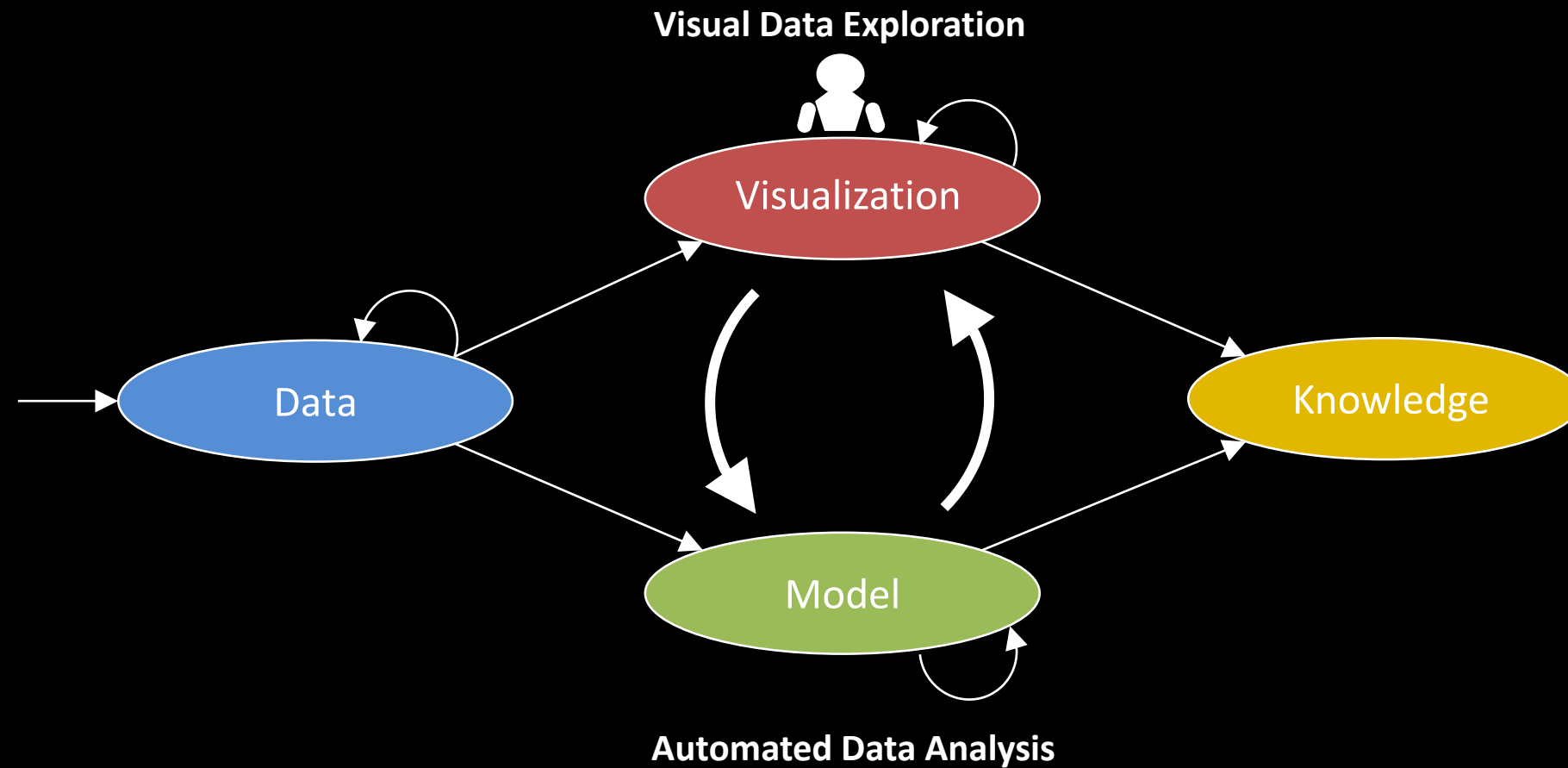
D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008.

# Visual Analytics



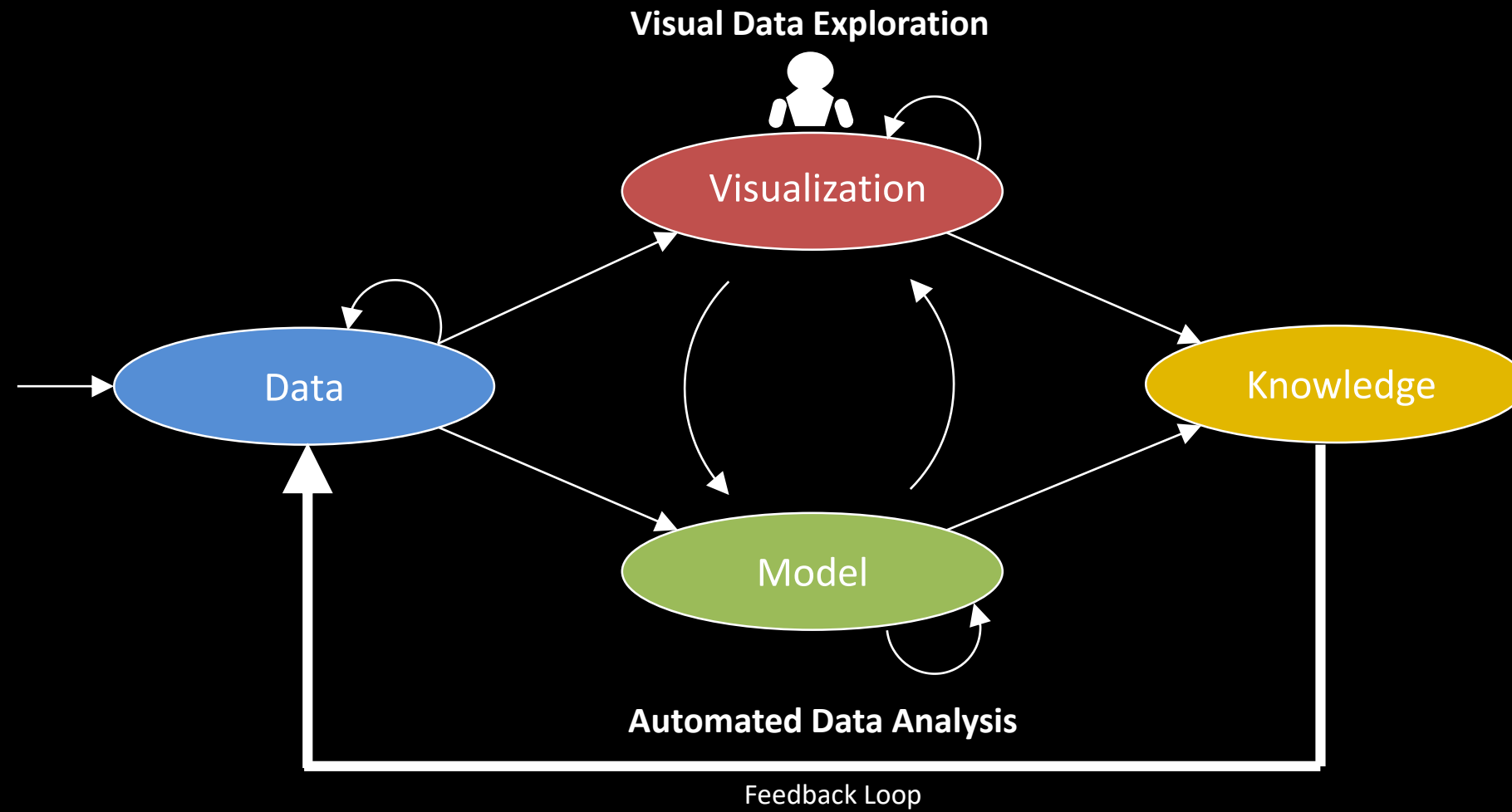
D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008.

# Visual Analytics



D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008.

# Visual Analytics



D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008.



# About Me



Assistant Professor of Computer Science  
@ New York University | Tandon School of Engineering

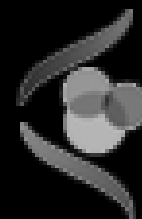
>> Research: Data Visualization and Visual Analytics

>> Application Field: Biomedical Data Analysis

>> Previously at Harvard – LSP / VCG



Visual Computing Group @  
Harvard School of  
Engineering and Applied  
Sciences



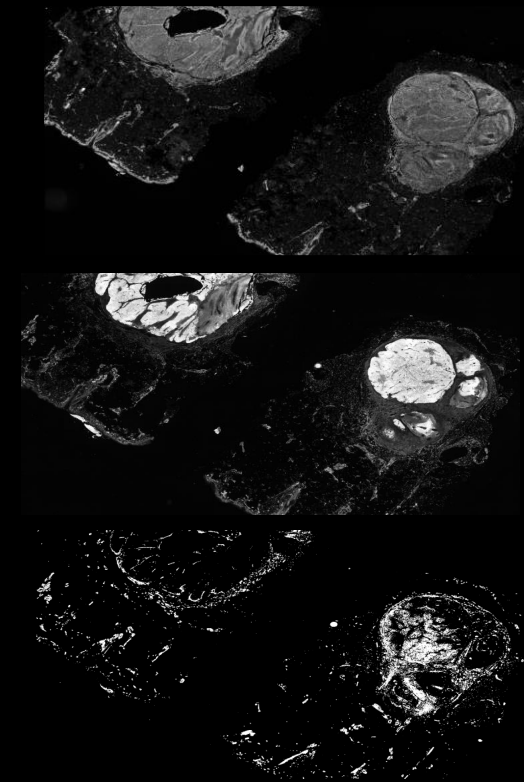
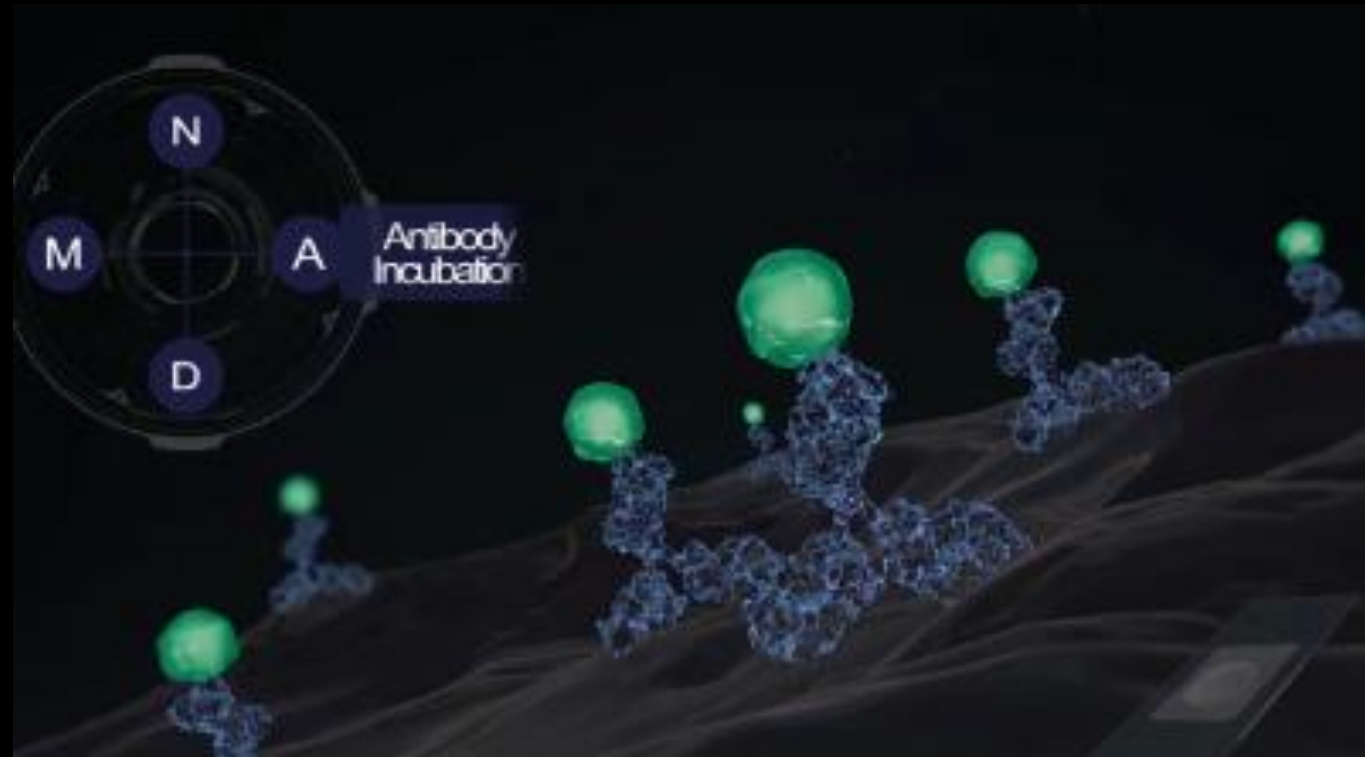
Laboratory of Systems  
Pharmacology  
@Harvard Medical School

# Multiplexed Tissue Imaging

- Making proteins visible (give cells their identity)
- Understand and target tumor microenvironment
- High throughput imaging methods  
e. g., CyCIF - Cyclic Immunofluorescence Imaging

# CyCIF - Cyclic Immunofluorescence Imaging

Lin, J.R., et al., 2018. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife*, 7.



## Dataset

60 image channels

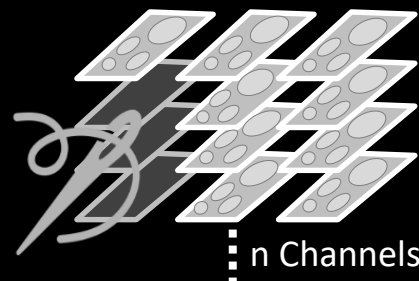
Res: up to 60k x 60k px,  
200nm per pixel

Size: often > 1TB  
> 1 million cells

[www.cycif.org]



Imaging



Registration & Stitching

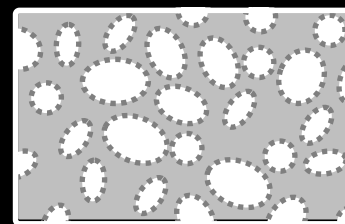
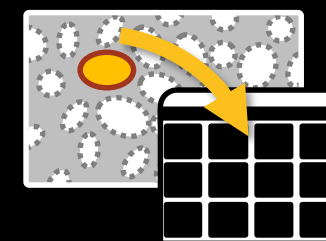


Image Segmentation

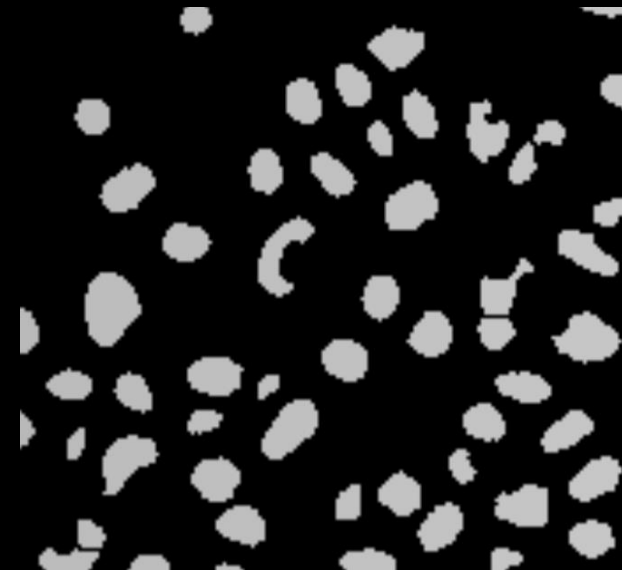
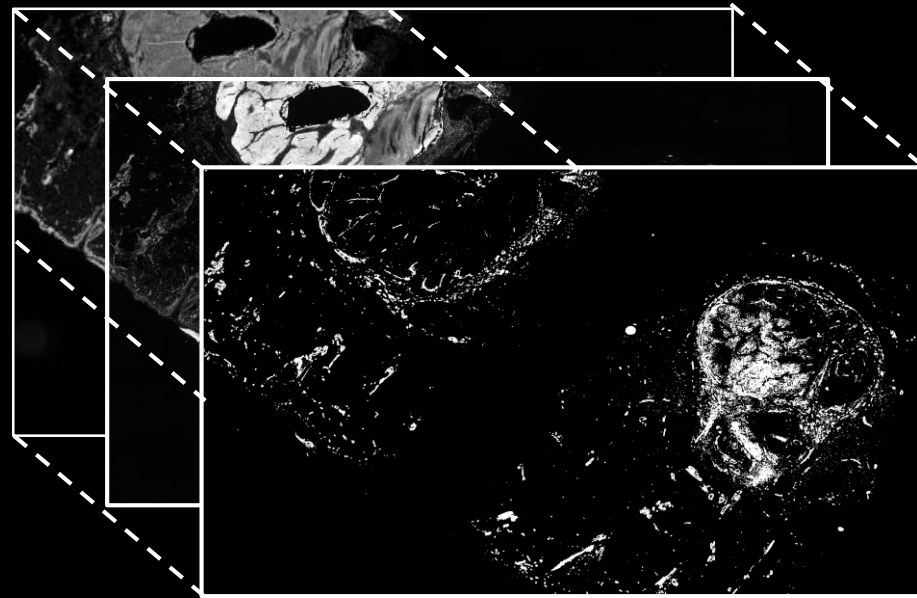


Feature Extraction

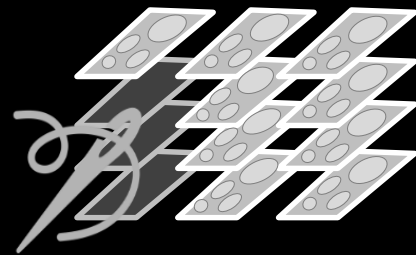


Analysis

# Registration, Segmentation



Imaging



Registration & Stitching

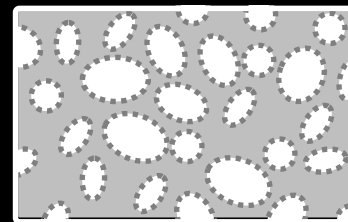
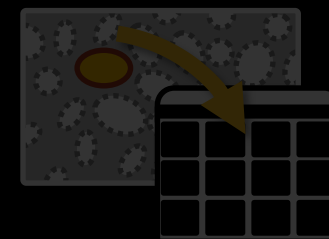
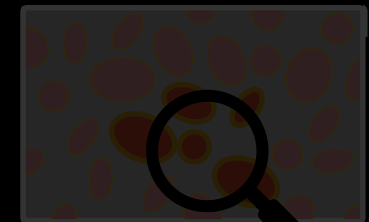


Image Segmentation

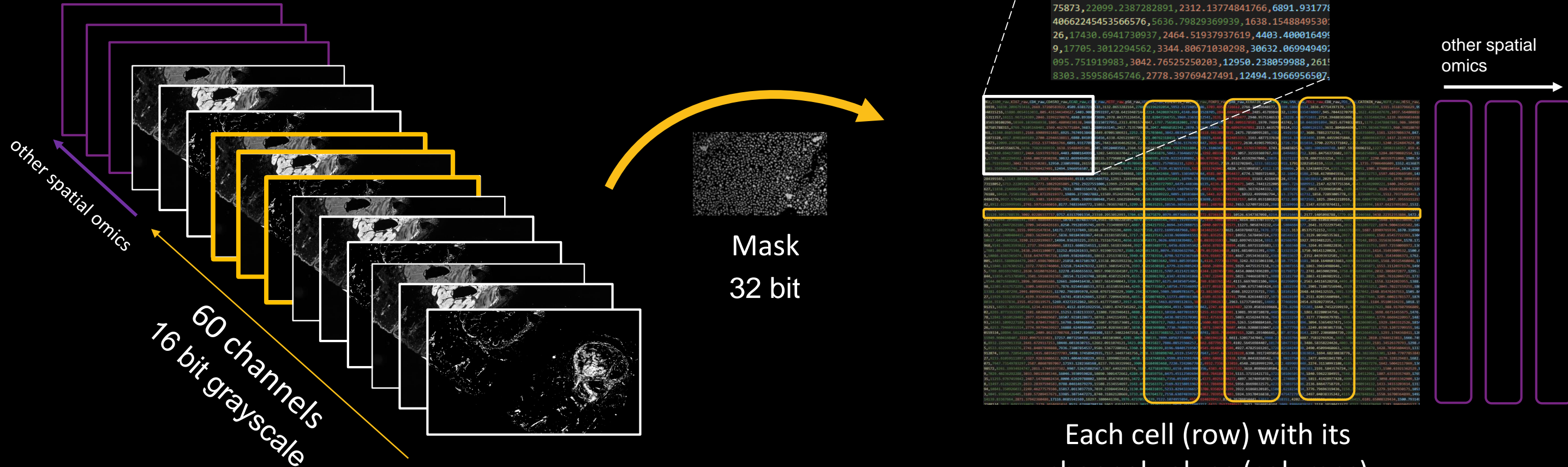


Feature Extraction

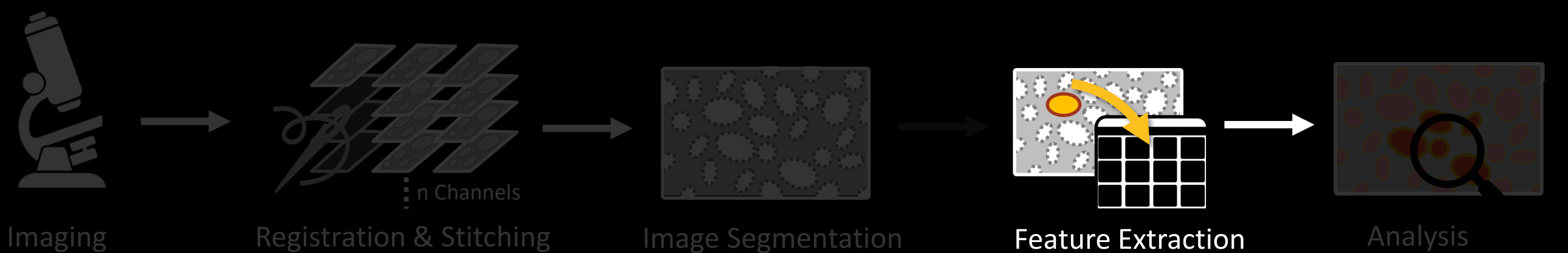


Analysis

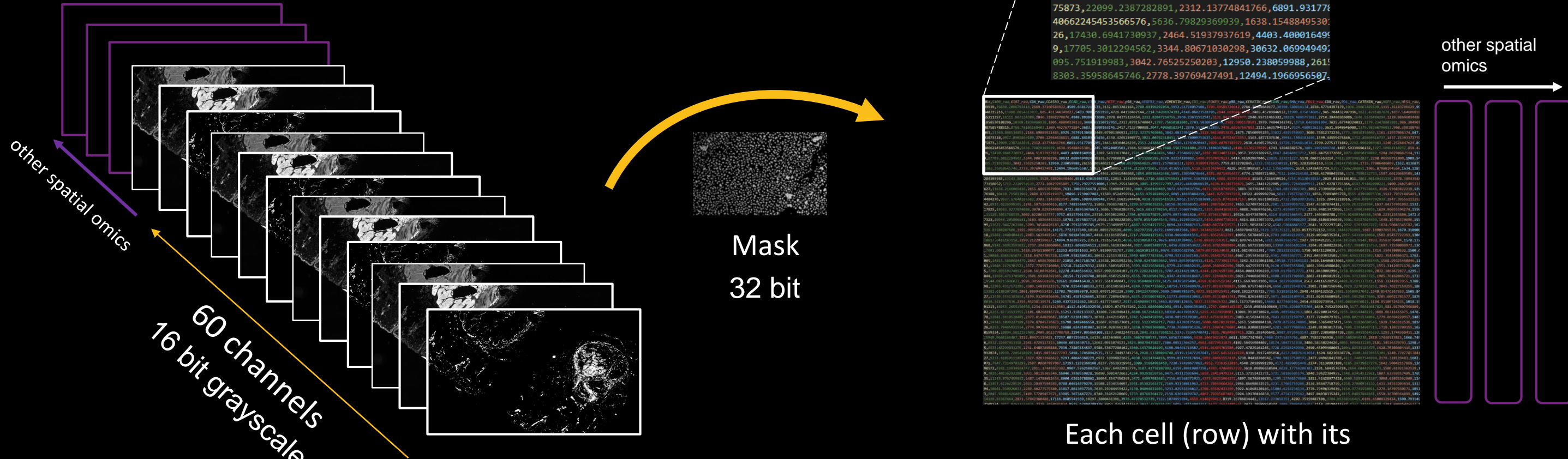
# Feature Extraction



Each cell (row) with its channel values (columns)



# Feature Extraction

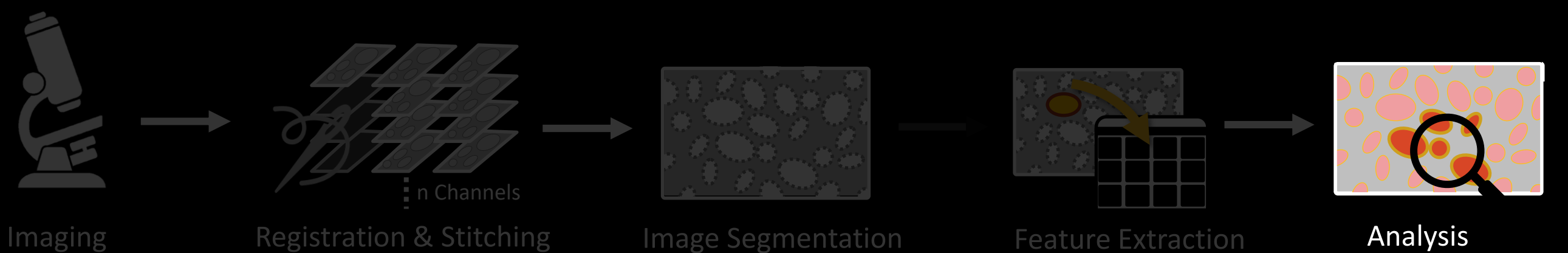


```

RKz_S100_raw,KI67_raw,CD4_raw,CD45RO_raw,ECAD_raw,
49939,36830,2096793416,2669,37260583922,4509,63817,
340315216,15886,0654513031,805,431344349627,5403,96
55311357,16111,967124389,2046,31992270876,4848,893
1856530108296,10369,1039466938,1805,48090230138,34
087585788315,8769,76105168401,1569,46276771884,368
061,21340,868534893,2166,69089921485,6925,767491384
01873328,6917,8901869189,2700,22946538811,6888,8416
75873,22099,2387282891,2312,13774841766,6891,93177
40662245453566576,5636,79829369939,1638,1548849530
26,17430,6941730937,2464,51937937619,4403,400016499
9,17705,3012294562,3344,80671030298,30632,06994949
095,751919983,3042,76525250203,12950,238059988,261
8303,35958645746,2778,39769427491,12494,1966956507

```

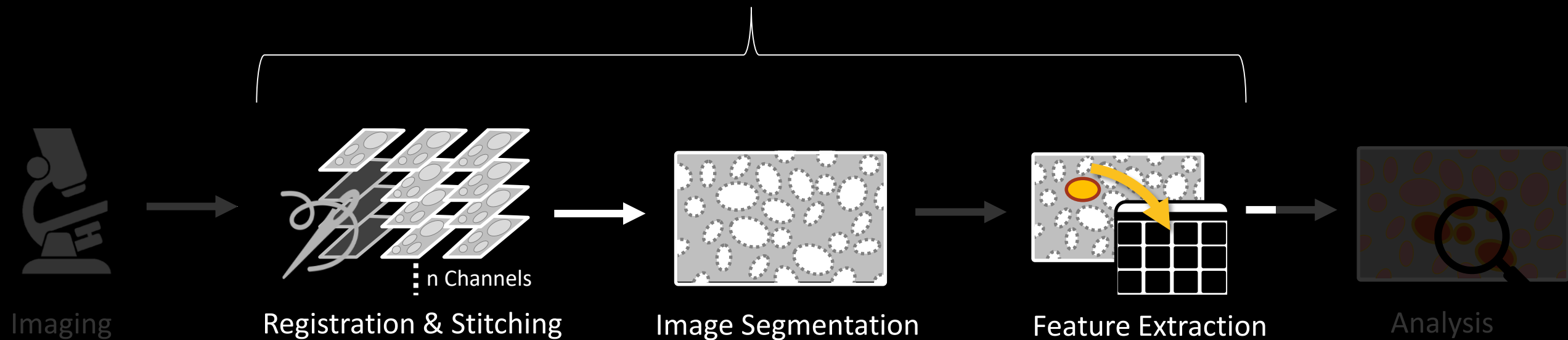
Each cell (row) with its channel values (columns)

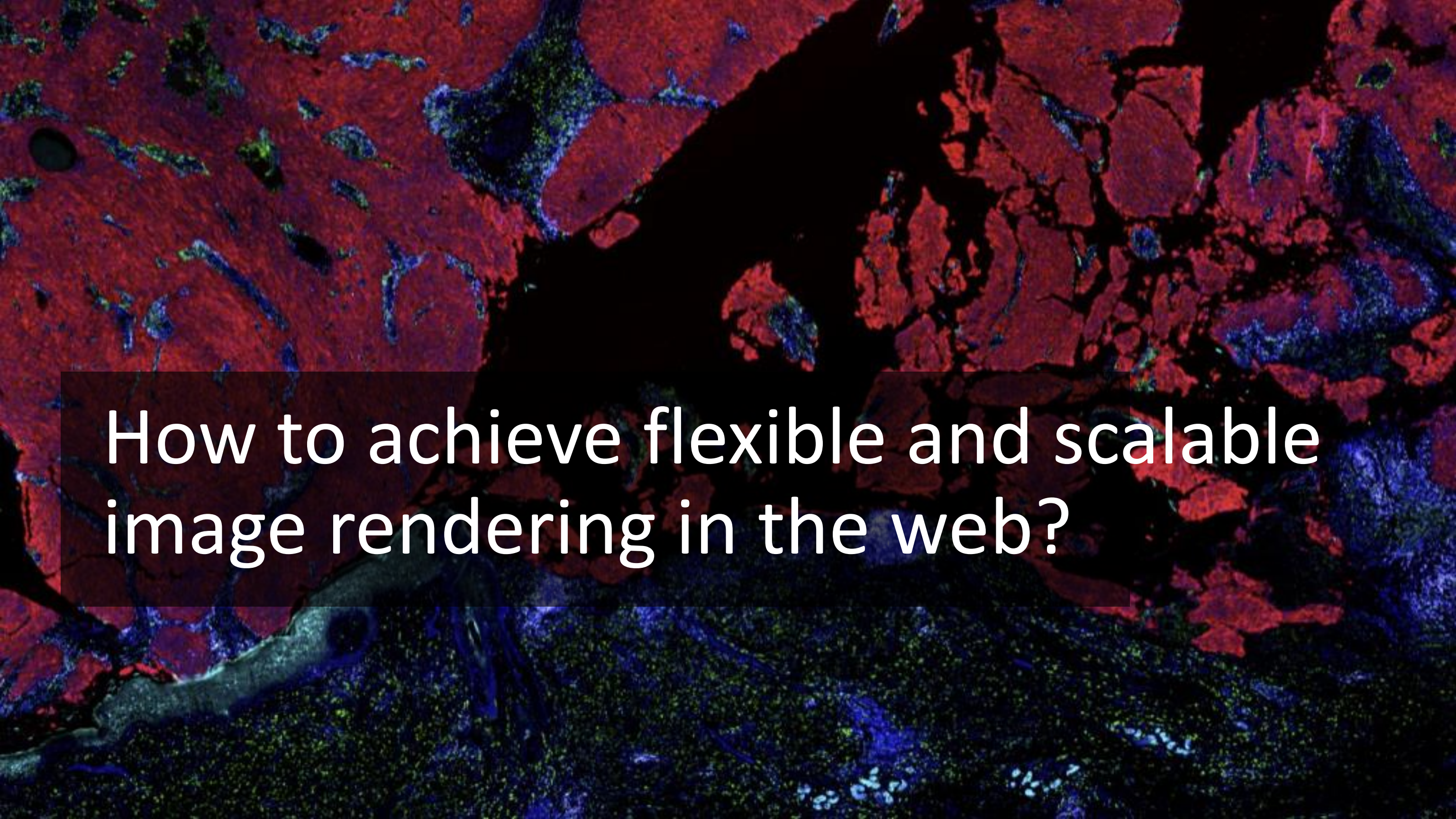


# MCMICRO

<https://mcmicro.org>

Schapiro et al. 2022. MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nature methods*



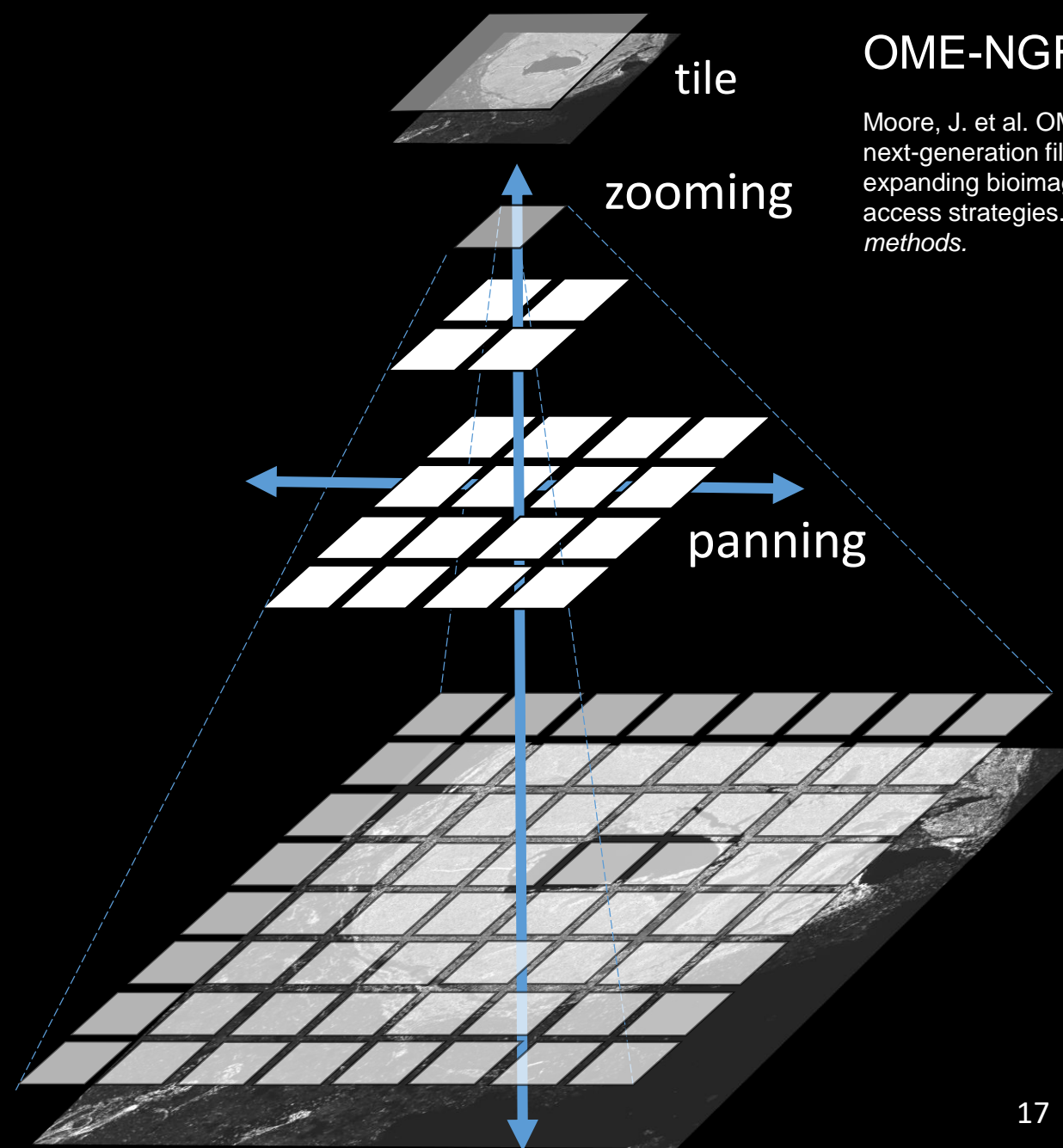
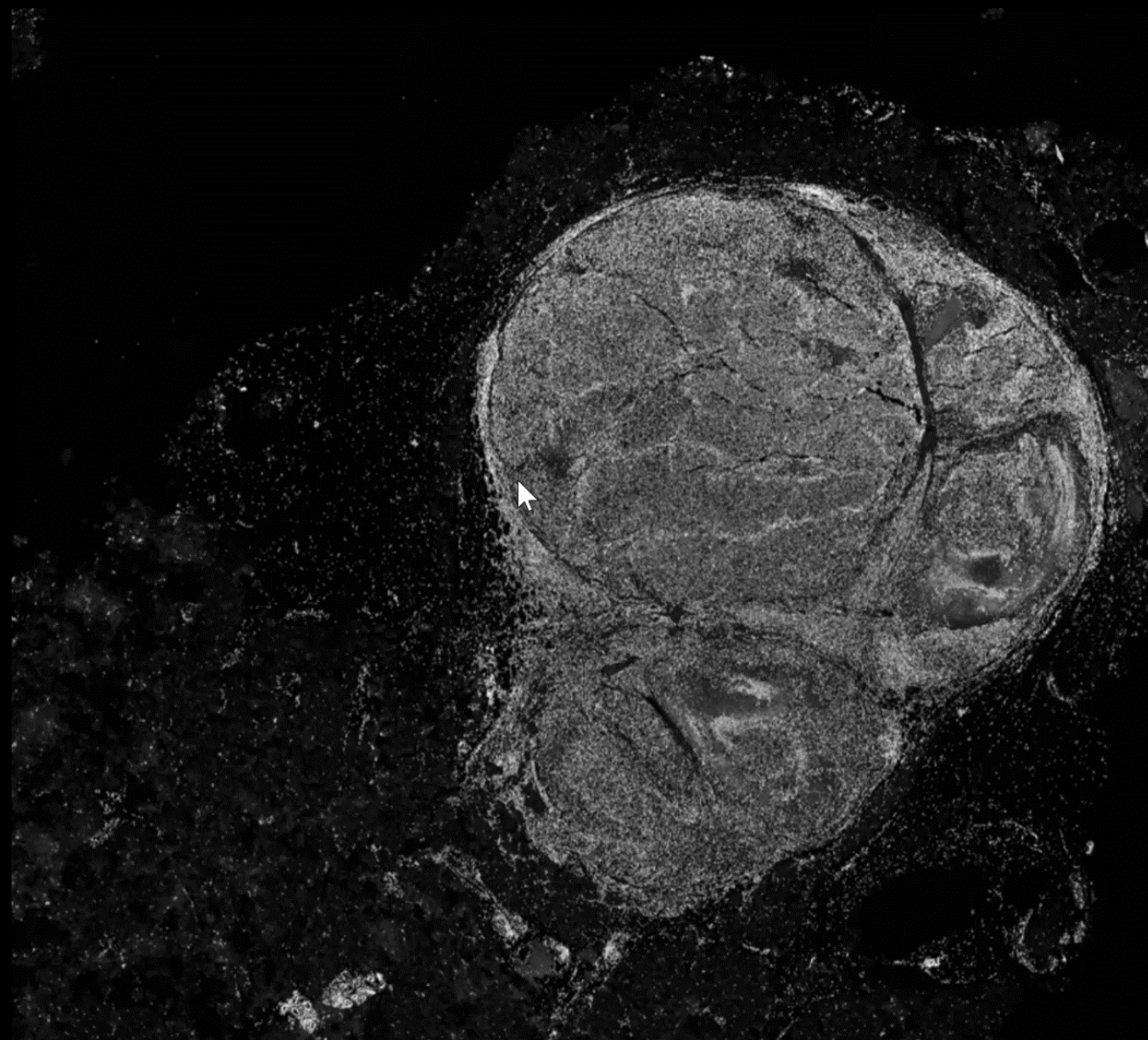


How to achieve flexible and scalable  
image rendering in the web?



# Scalable Multiplex Image Rendering for the Web

- Big data: up to 1TB image

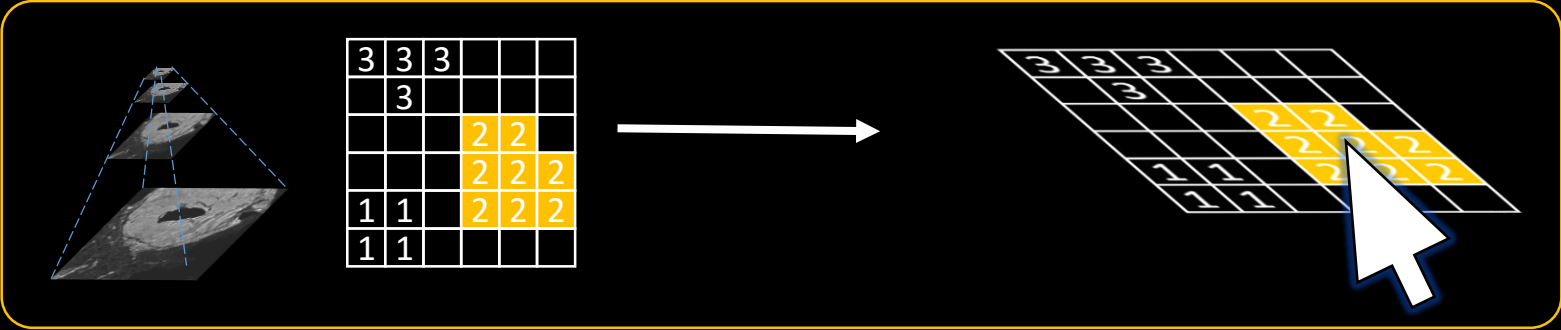
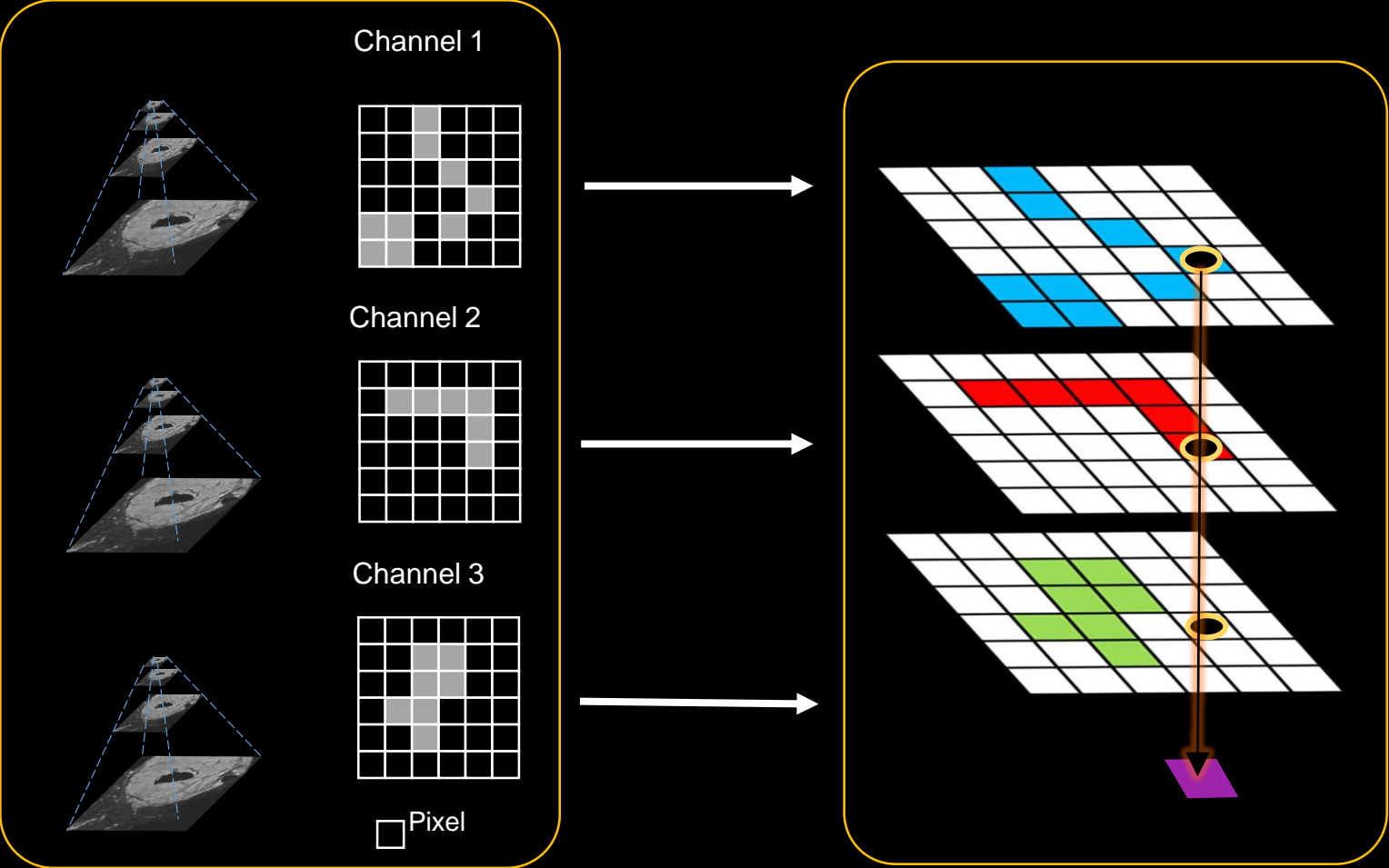
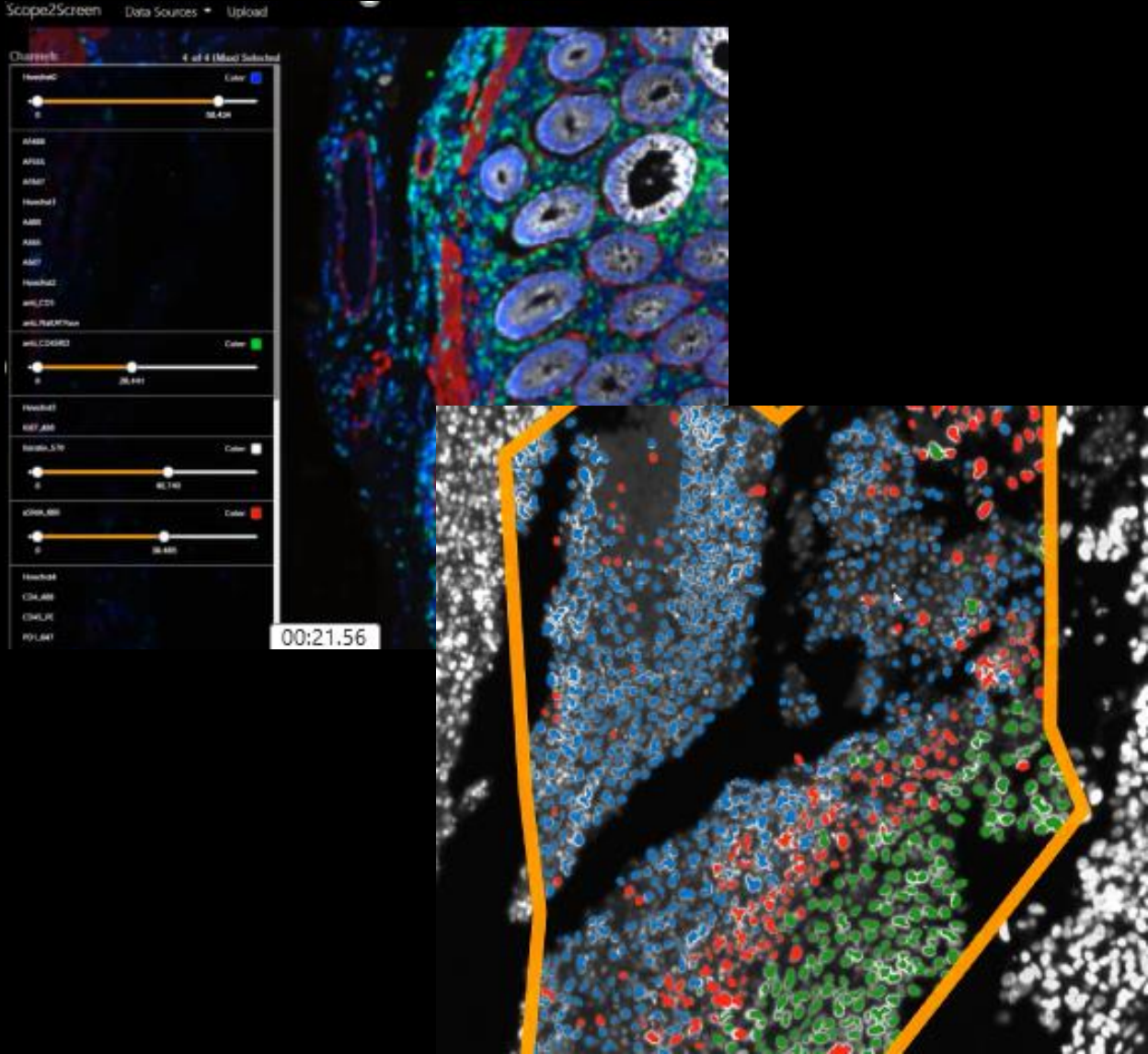


OME-TIFF

OME-NGFF

Moore, J. et al. OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies. *Nature methods*.

# Scalable Multiplex Image Rendering for the Web



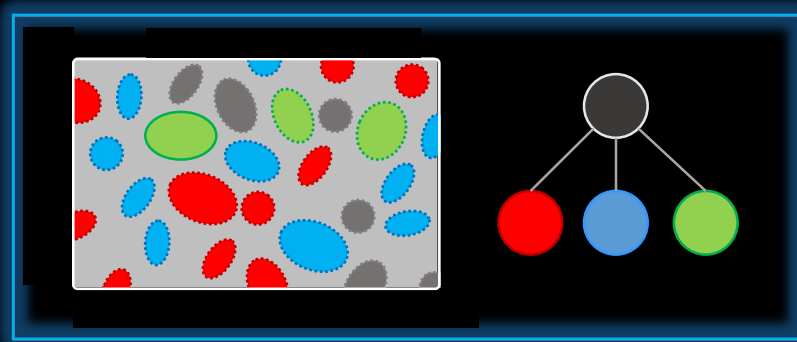


# Visual Analytics Tasks

# Visual Analysis of Multiplexed Immunofluorescence Data

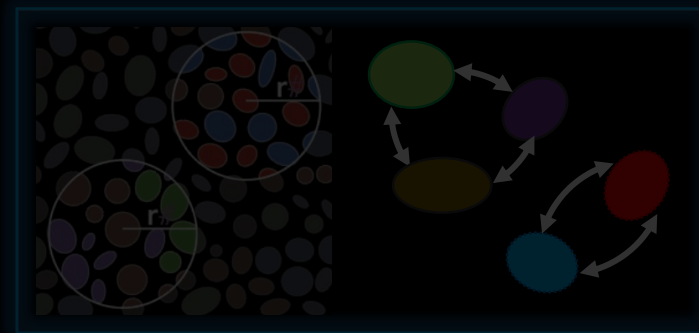
## Cell Classification

Data Facetting with Machine Learning



## Interaction Analysis

Cellular Neighborhoods



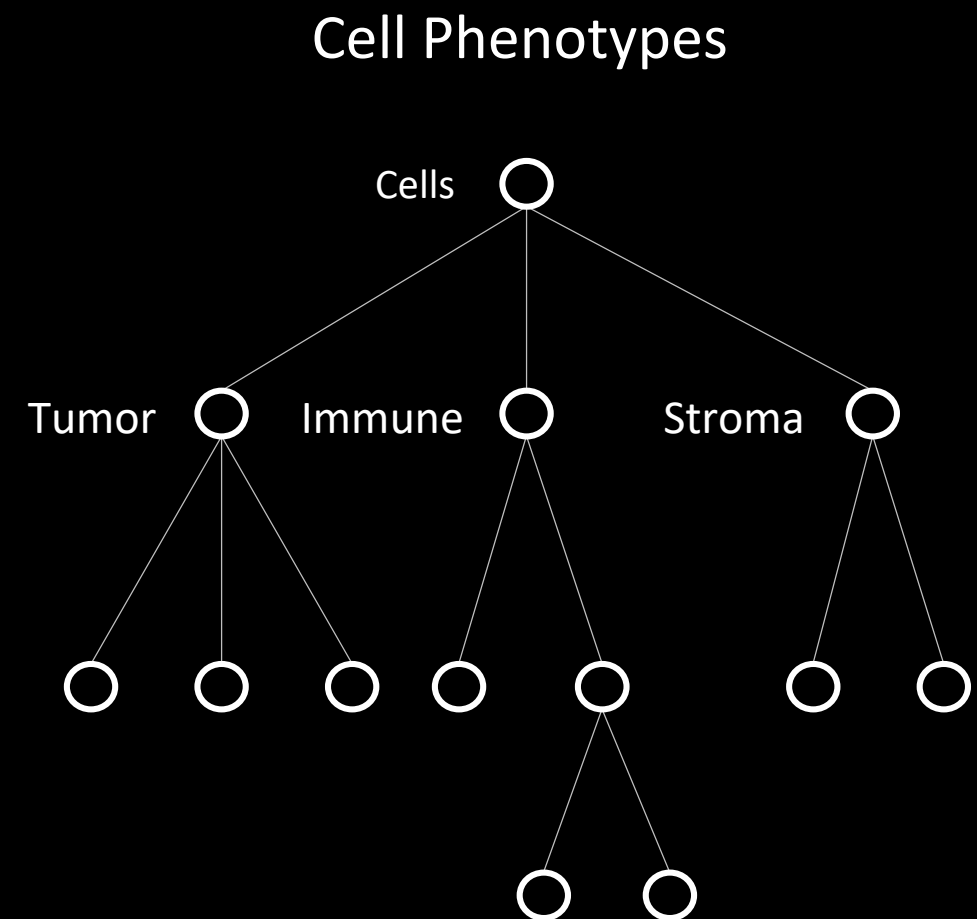
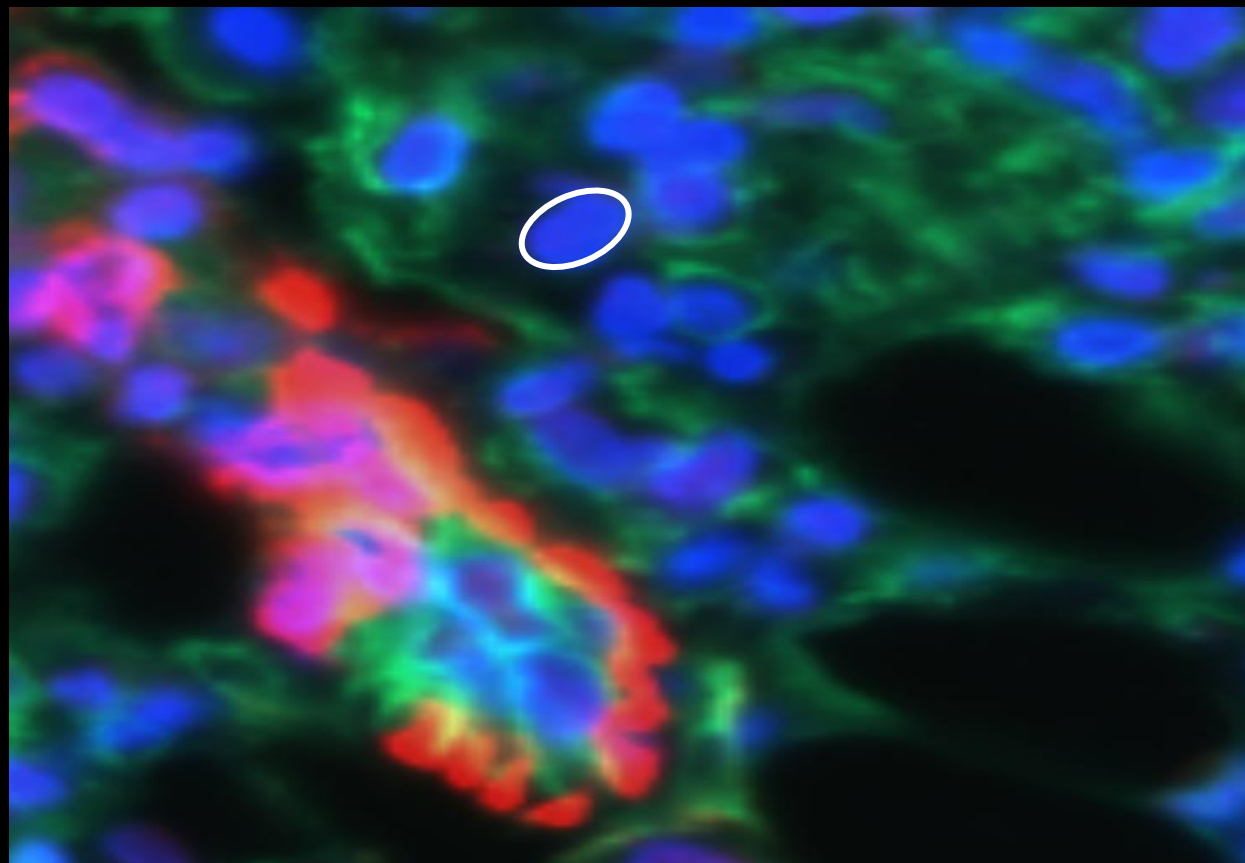
## Annotation

Focus+Context Pathology Assessment



# Facet Data into Cell Types

- Identify cell types based on protein expressions in image channels

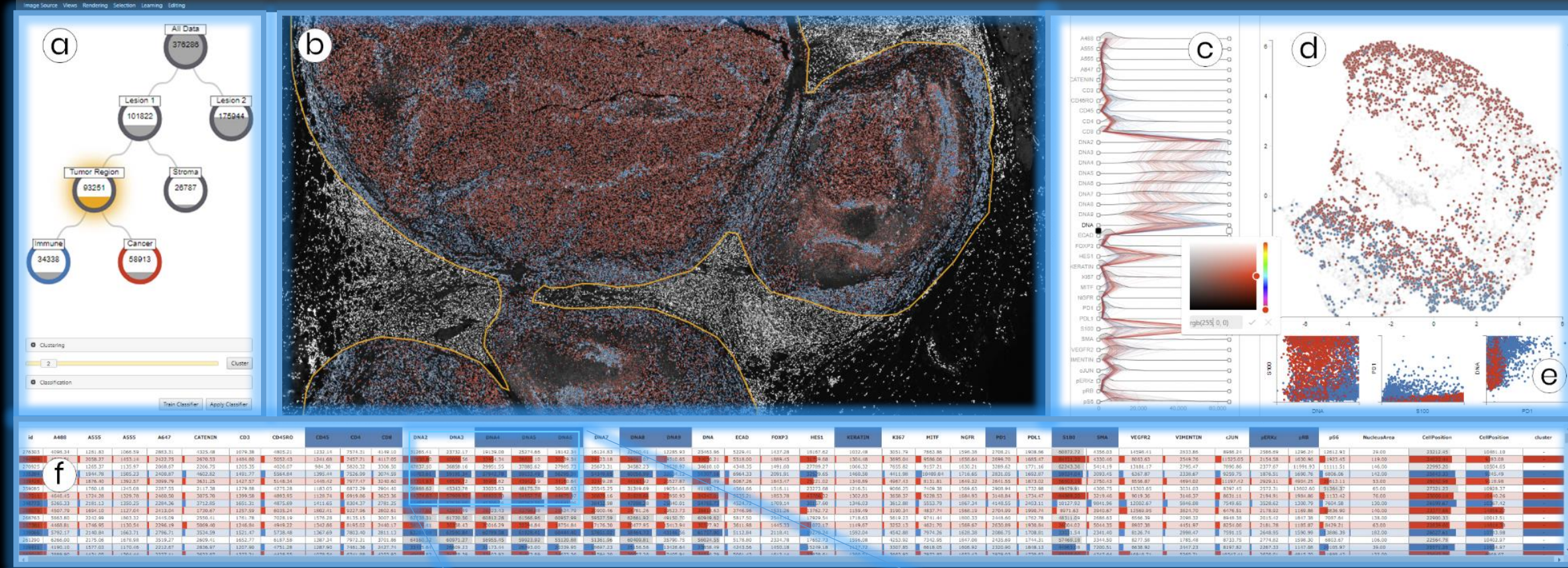


# Facetto

## Hierarchical Faceting

## Image Exploration

## Feature Exploration

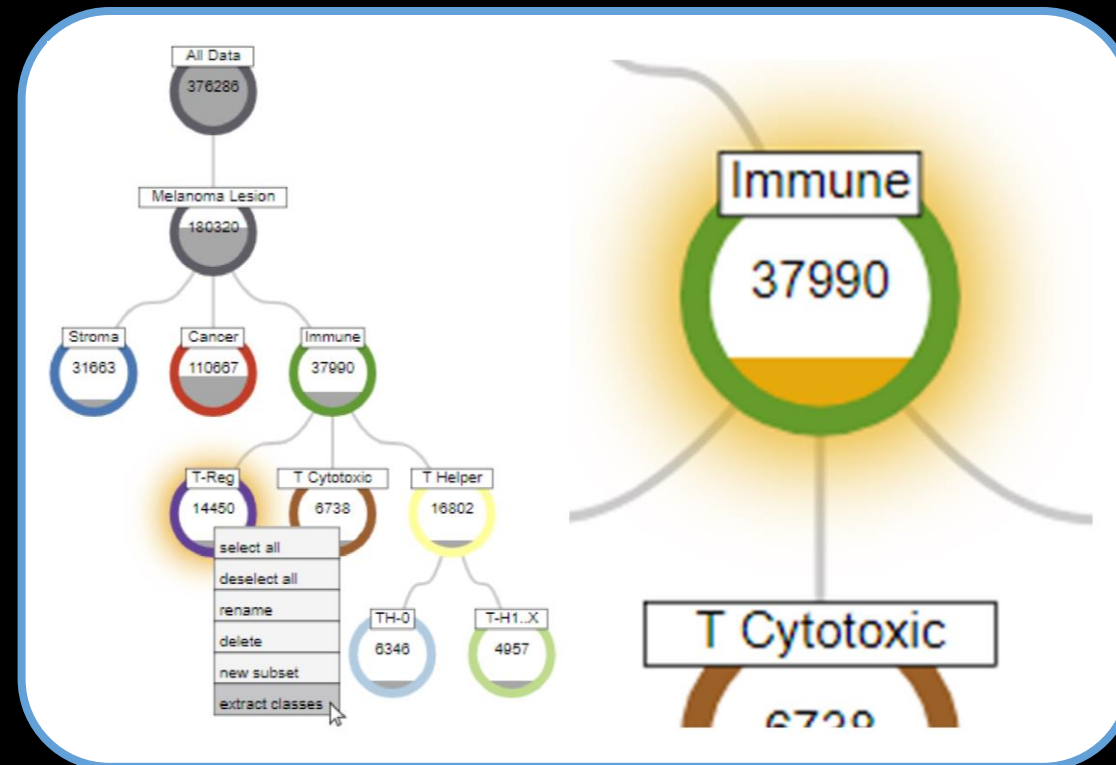
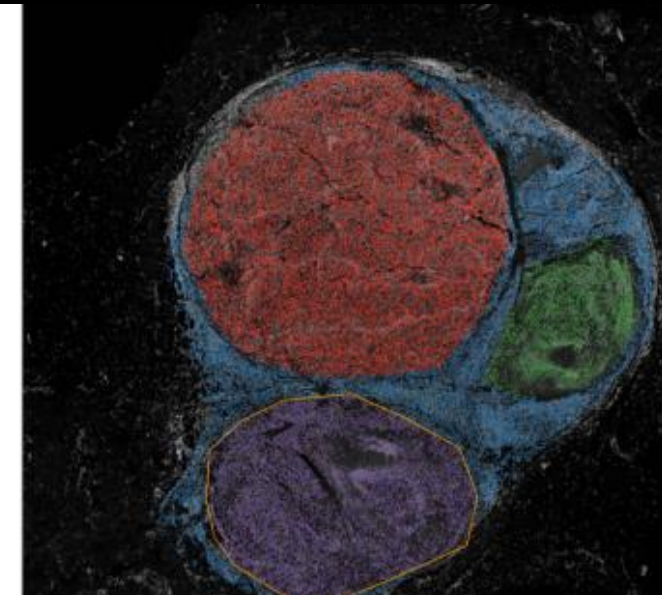
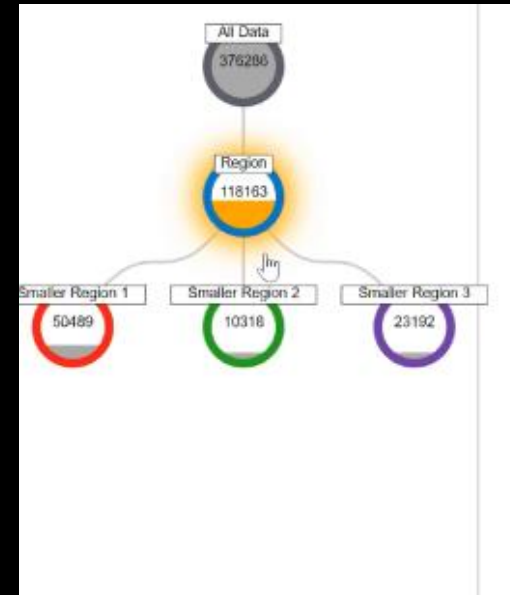


Krueger, R., et al., 2019. Facetto: Combining unsupervised and supervised learning for hierarchical phenotype analysis in multi-channel image data. *IEEE TVCG*

DNA2	DNA3	DNA4	DNA5	DNA6
31265.41	23732.17	19139.00	25274.66	18142.31
47930.85	40080.56	37994.34	76026.10	30279.34
47837.10	36858.16	29951.55	37885.62	27985.73
60703.61	59195.34	57942.78	59033.49	56295.39
57313.83	44539.32	39905.62	43842.19	34180.64
56488.63	45343.78	33035.63	48175.78	30458.63
64374.67	57969.92	48820.70	54857.74	44675.97
63454.06	57538.09	45300.16	60267.35	41641.96

# Manual Faceting

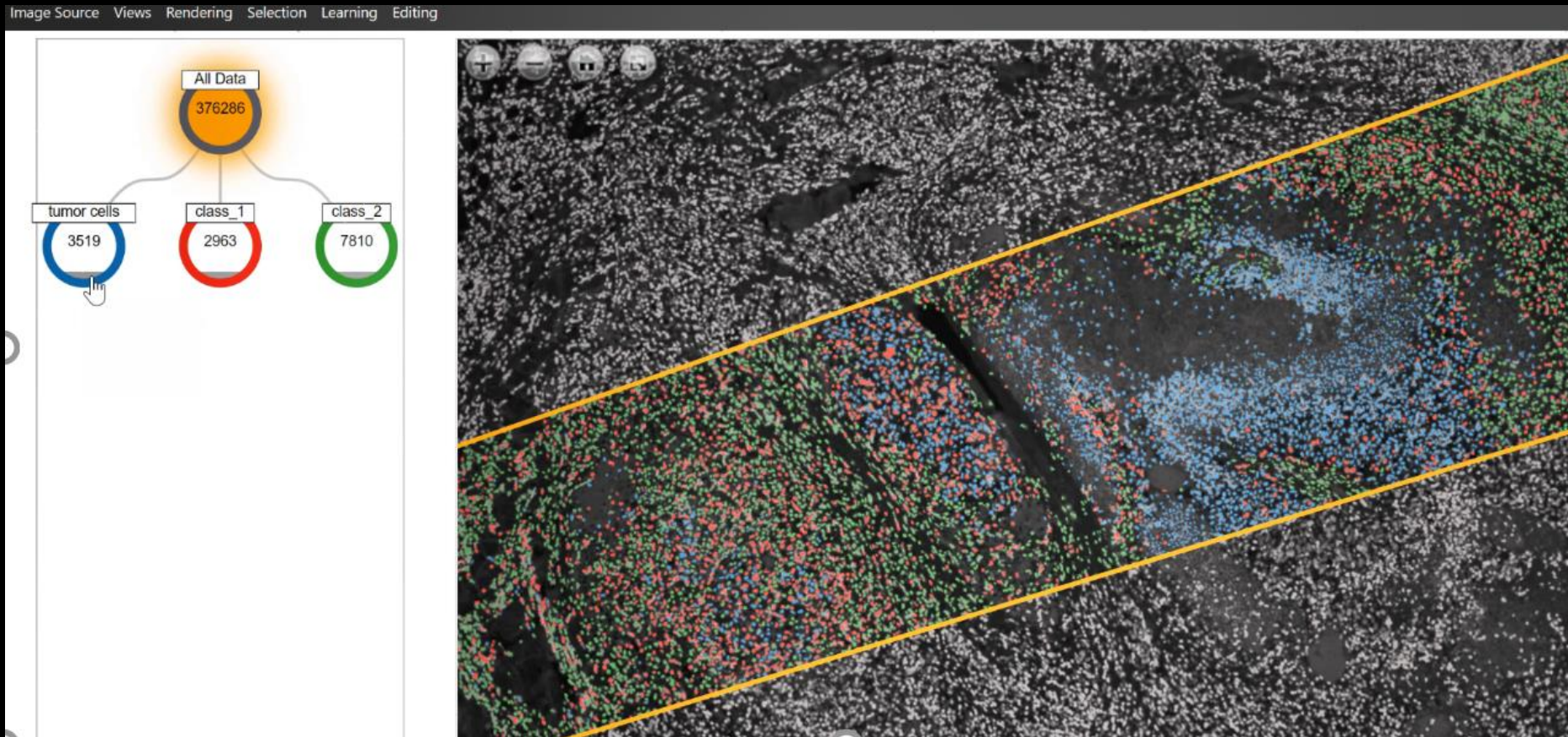
- Facet data into meaningful subsets (cell types, regions)
- Keep track of consecutive filtering actions



# How can ML Support Analysis & Faceting?

- **Bottom-Up Analysis**

- Unsupervised learning supports exploration and faceting
- EM-Clustering is fast, detects clusters of different size, density, shape

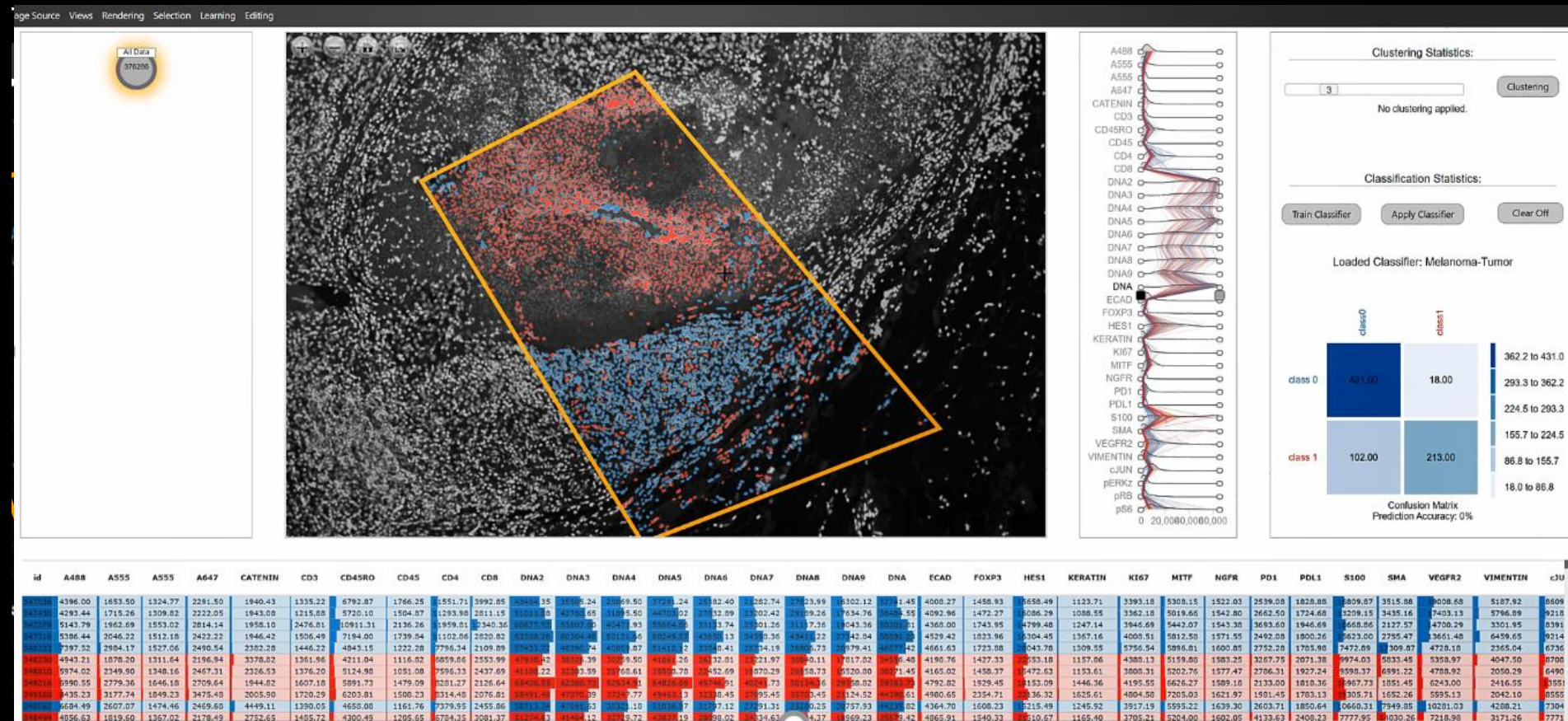




# How can ML Support Analysis & Faceting?

- Top-Down Analysis

- Supervised learning applies gained knowledge to new data
- Train classifier on manually refined clustering results
- CNN Classification uses image tiles as input to classify cell phenotypes



# Visual Analysis of Multiplexed Immunofluorescence Data

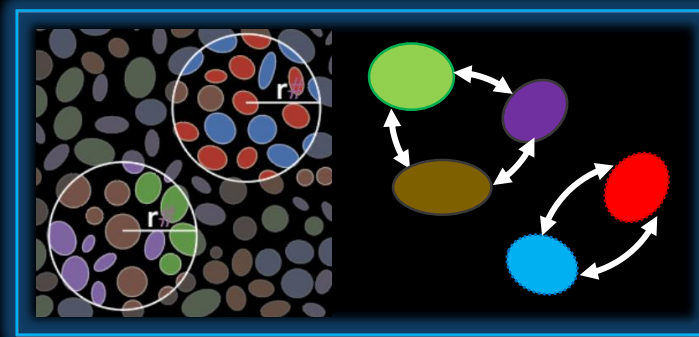
## Cell Classification

Data Facetting with Machine Learning



## Interaction Analysis

Cellular Neighborhoods

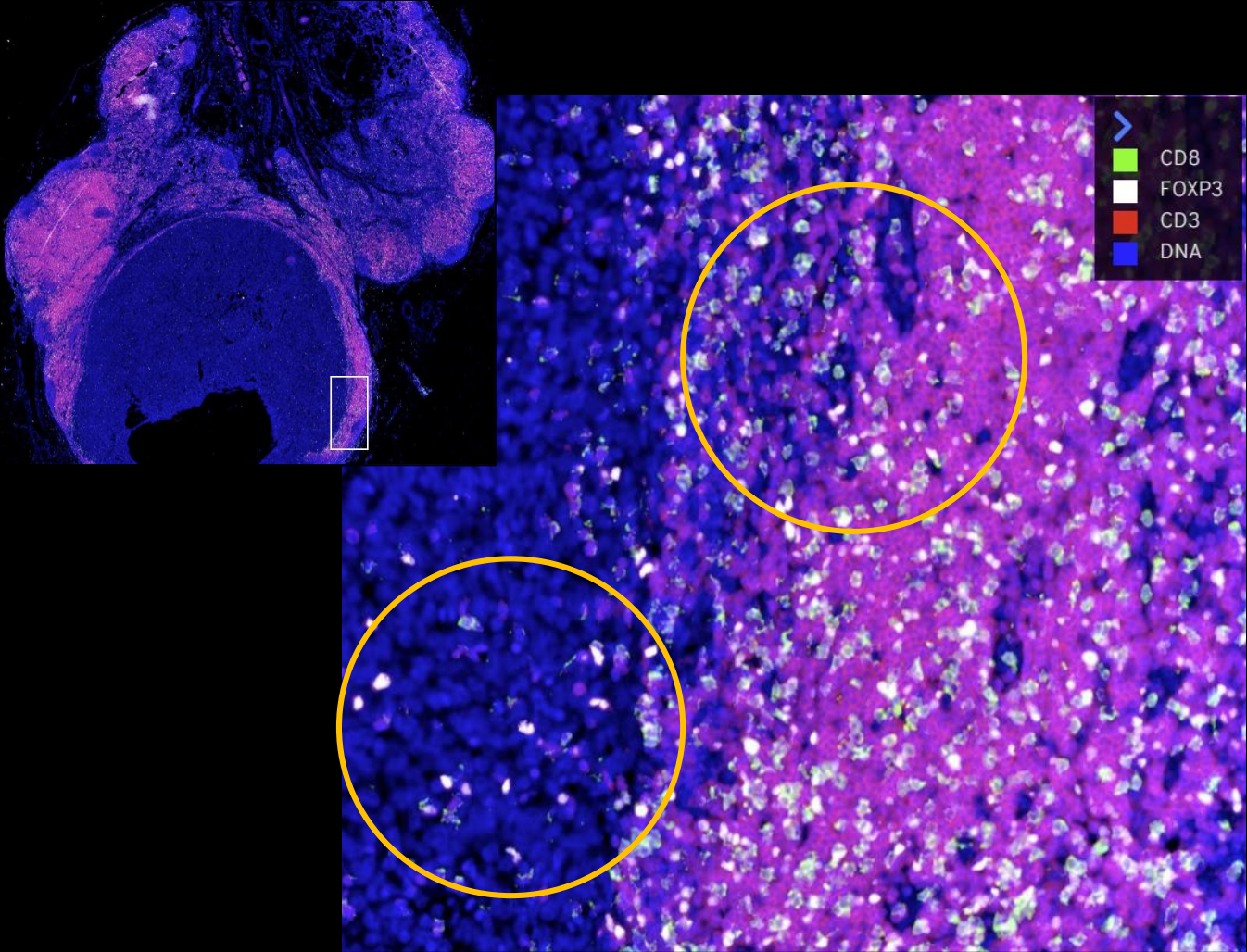


## Annotation

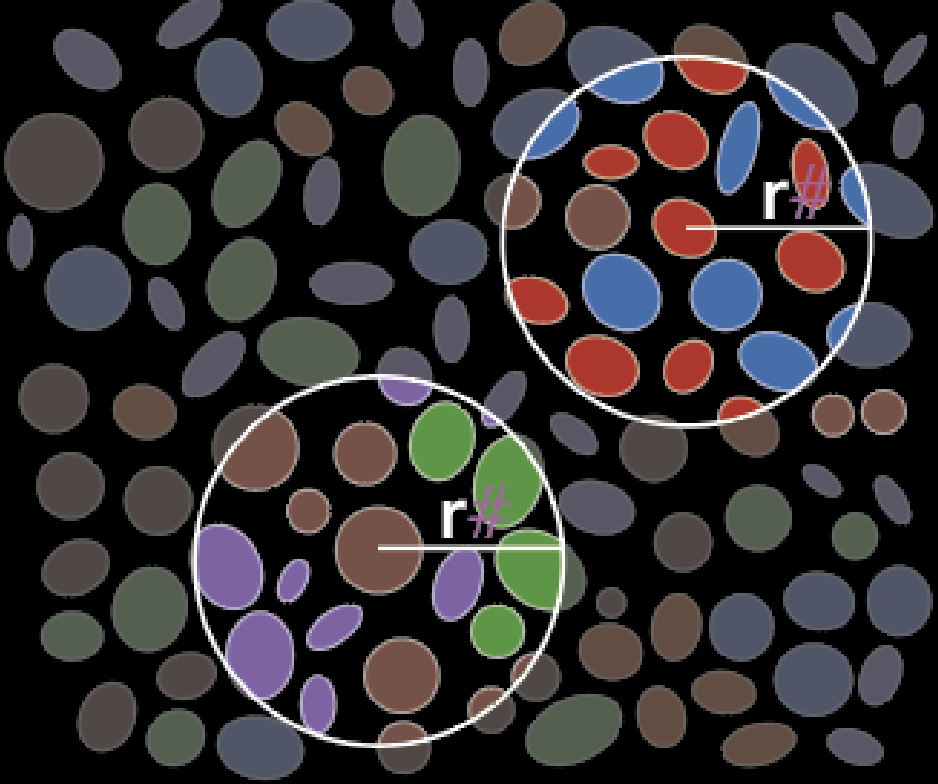
Focus+Context Pathology Assessment



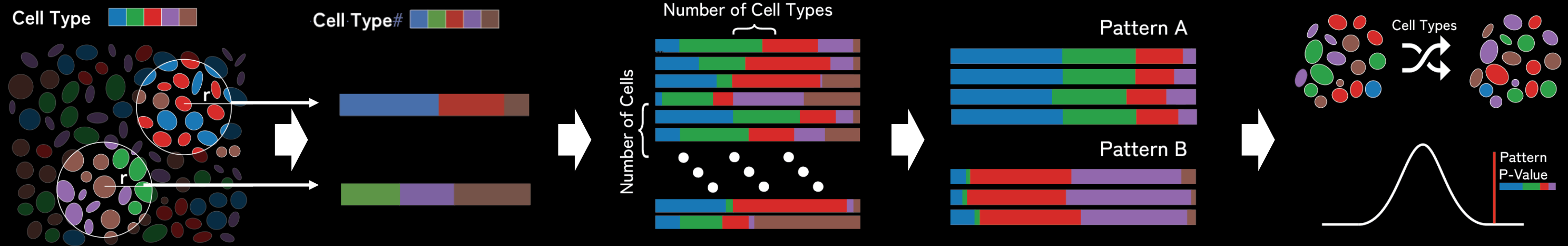
# From Single-Cell Analysis to Cellular Neighborhoods



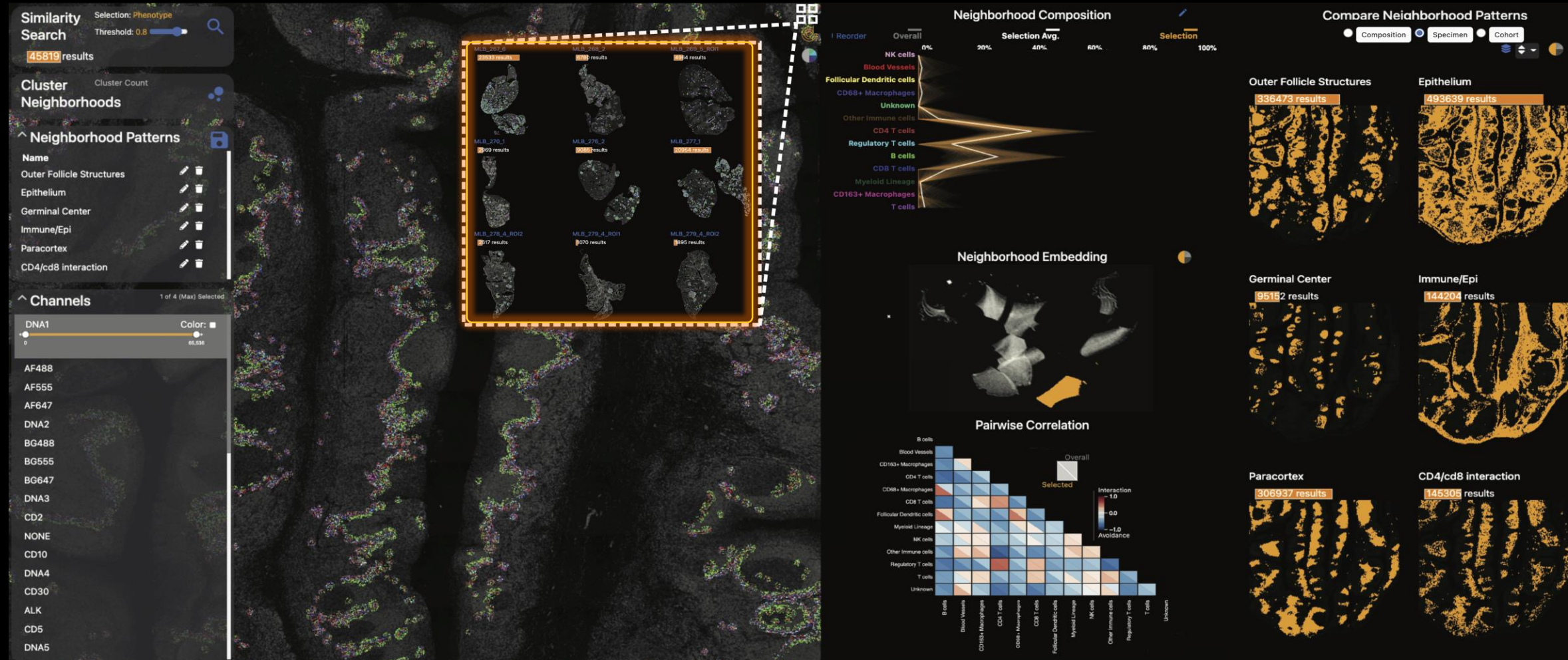
Cell Type#



# Neighborhood Quantification



# Visinity: Visual Spatial Neighborhood Analysis



Warchol, S., Krueger, R. et al., 2022.

**Visinity: Visual Spatial Neighborhood Analysis for Multiplexed Tissue Imaging Data.** *IEEE TVCG.*

Gaglia, Giorgio, et al., 2023. **Lymphocyte networks are dynamic cellular communities in the immunoregulatory landscape of lung adenocarcinoma.** *Cancer Cell.*

**Similarity Search** Selection: Cluster  
 Threshold: 0.83

**Cluster Neighborhoods** 10

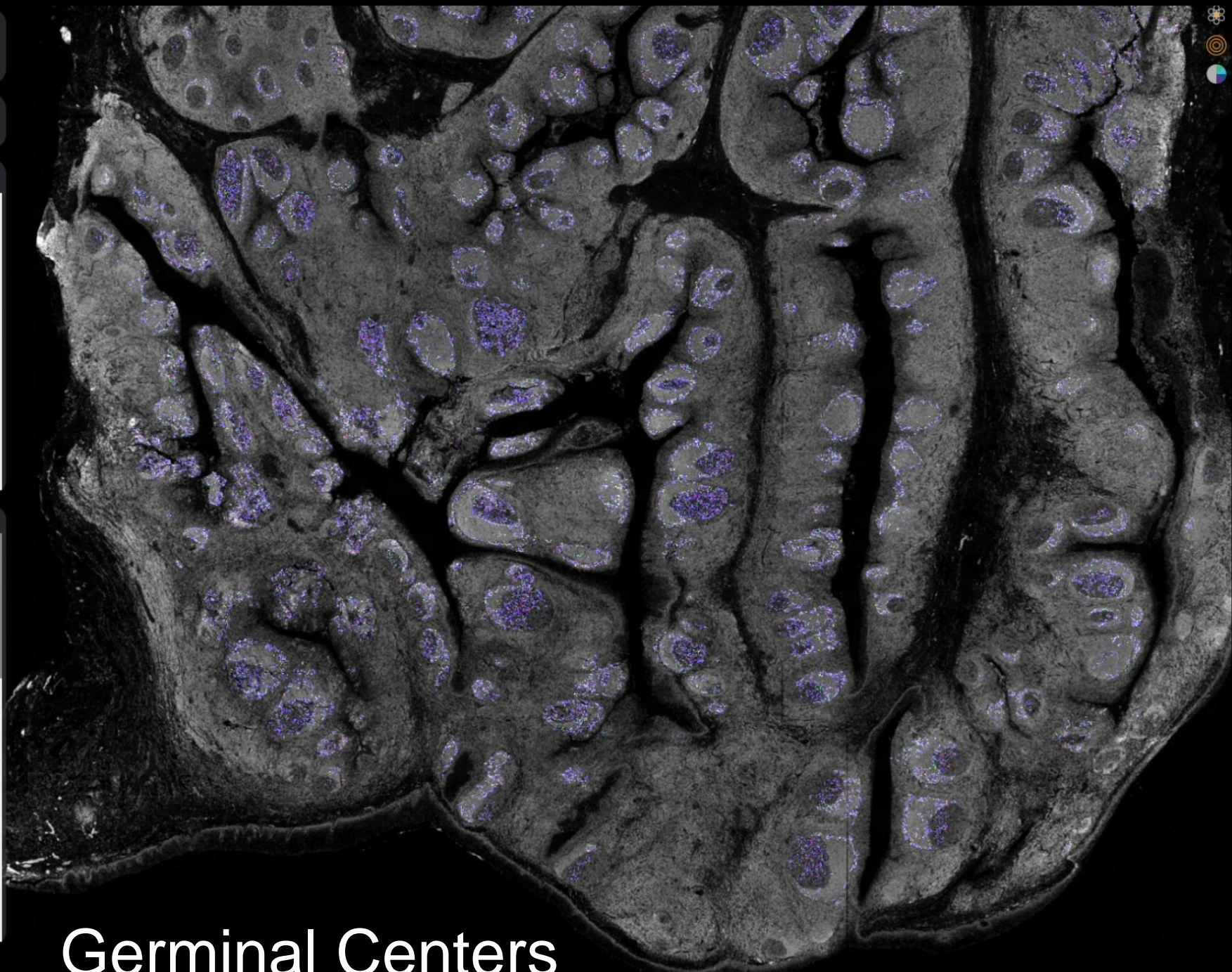
**Neighborhood Patterns**

Name	Actions
Germinal Center	[Edit] [Delete]
Custom Cluster 0 / 10	[Edit] [Delete]
Custom Cluster 1 / 10	[Edit] [Delete]
Custom Cluster 2 / 10	[Edit] [Delete]
Custom Cluster 3 / 10	[Edit] [Delete]
Custom Cluster 4 / 10	[Edit] [Delete]
Custom Cluster 5 / 10	[Edit] [Delete]
Custom Cluster 6 / 10	[Edit] [Delete]
Custom Cluster 7 / 10	[Edit] [Delete]
Custom Cluster 8 / 10	[Edit] [Delete]
Custom Cluster 9 / 10	[Edit] [Delete]

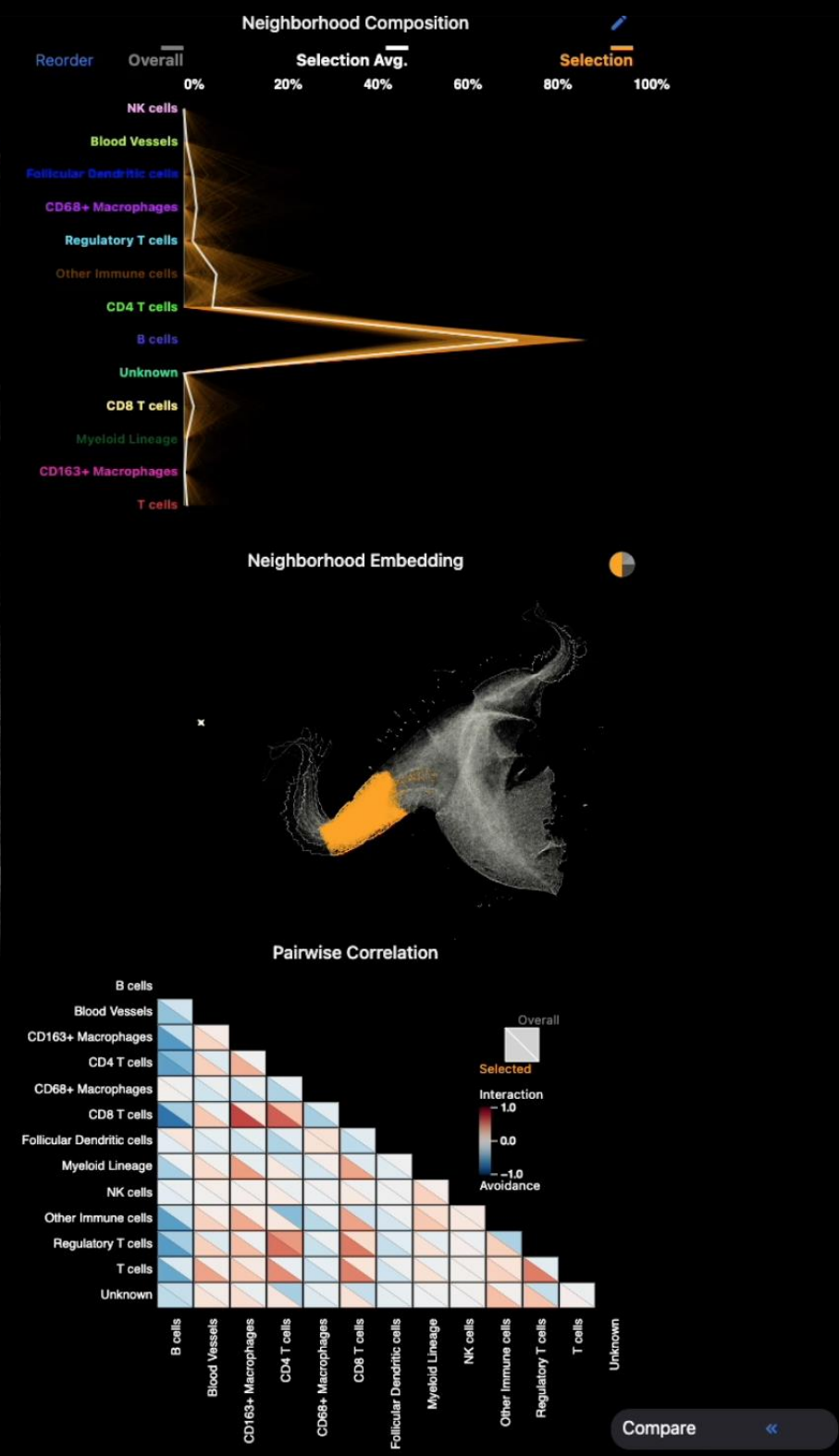
**Channels** 1 of 4 (Max) Selected

DNA1 Color: [Color Picker]

- AF488
- AF555
- AF647
- DNA2
- BG488
- BG555
- BG647
- DNA3
- CD2
- NONE
- CD10
- DNA4
- CD30
- ALK



**Germinal Centers**  
 B-cells mature



**Similarity Search** Selection: Cluster  
 Threshold: 0.83

**Cluster Neighborhoods** 10

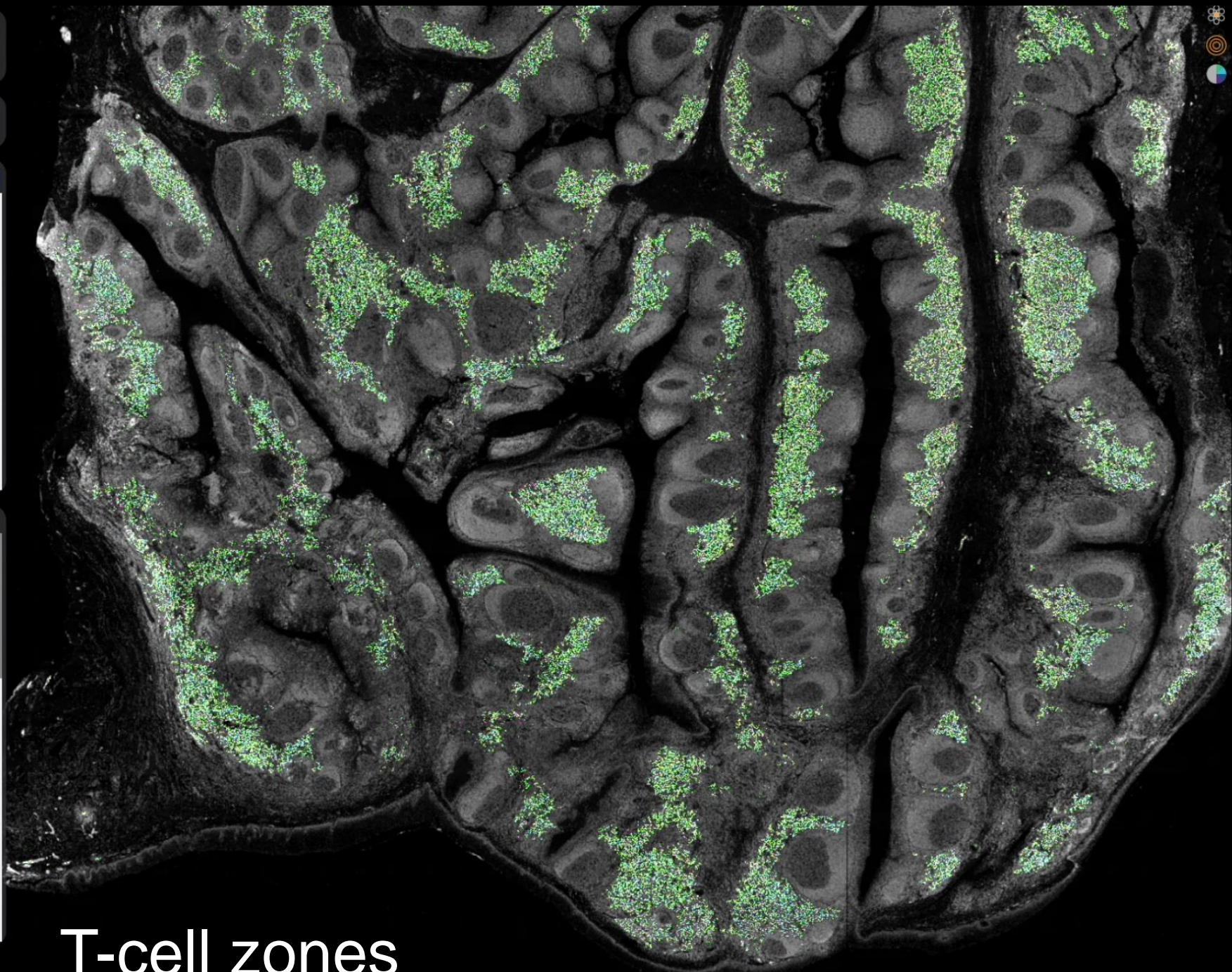
**Neighborhood Patterns**

Name	Actions
Germinal Center	✎ ✕
Custom Cluster 0 / 10	✎ ✕
Custom Cluster 1 / 10	✎ ✕
Custom Cluster 2 / 10	✎ ✕
Custom Cluster 3 / 10	✎ ✕
Custom Cluster 4 / 10	✎ ✕
Custom Cluster 5 / 10	✎ ✕
Custom Cluster 6 / 10	✎ ✕
Custom Cluster 7 / 10	✎ ✕
Custom Cluster 8 / 10	✎ ✕
Custom Cluster 9 / 10	✎ ✕

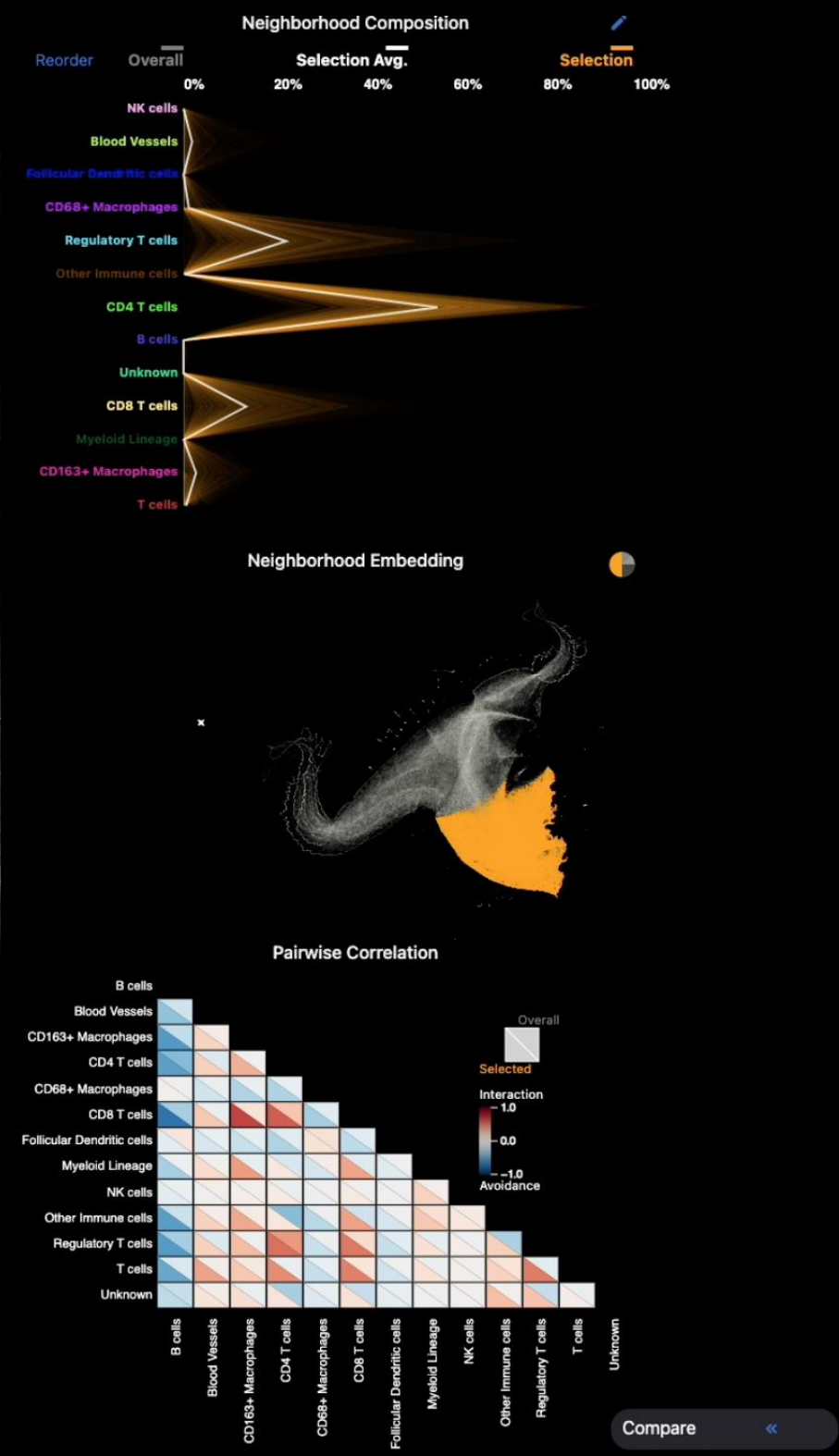
**Channels** 1 of 4 (Max) Selected

DNA1 Color: ■

- AF488
- AF555
- AF647
- DNA2
- BG488
- BG555
- BG647
- DNA3
- CD2
- NONE
- CD10
- DNA4
- CD30
- ALK



**T-cell zones**  
 T-cells activate B-cells



**Similarity Search** Selection: Cluster  
 Threshold: 0.83

**Cluster Neighborhoods** 10

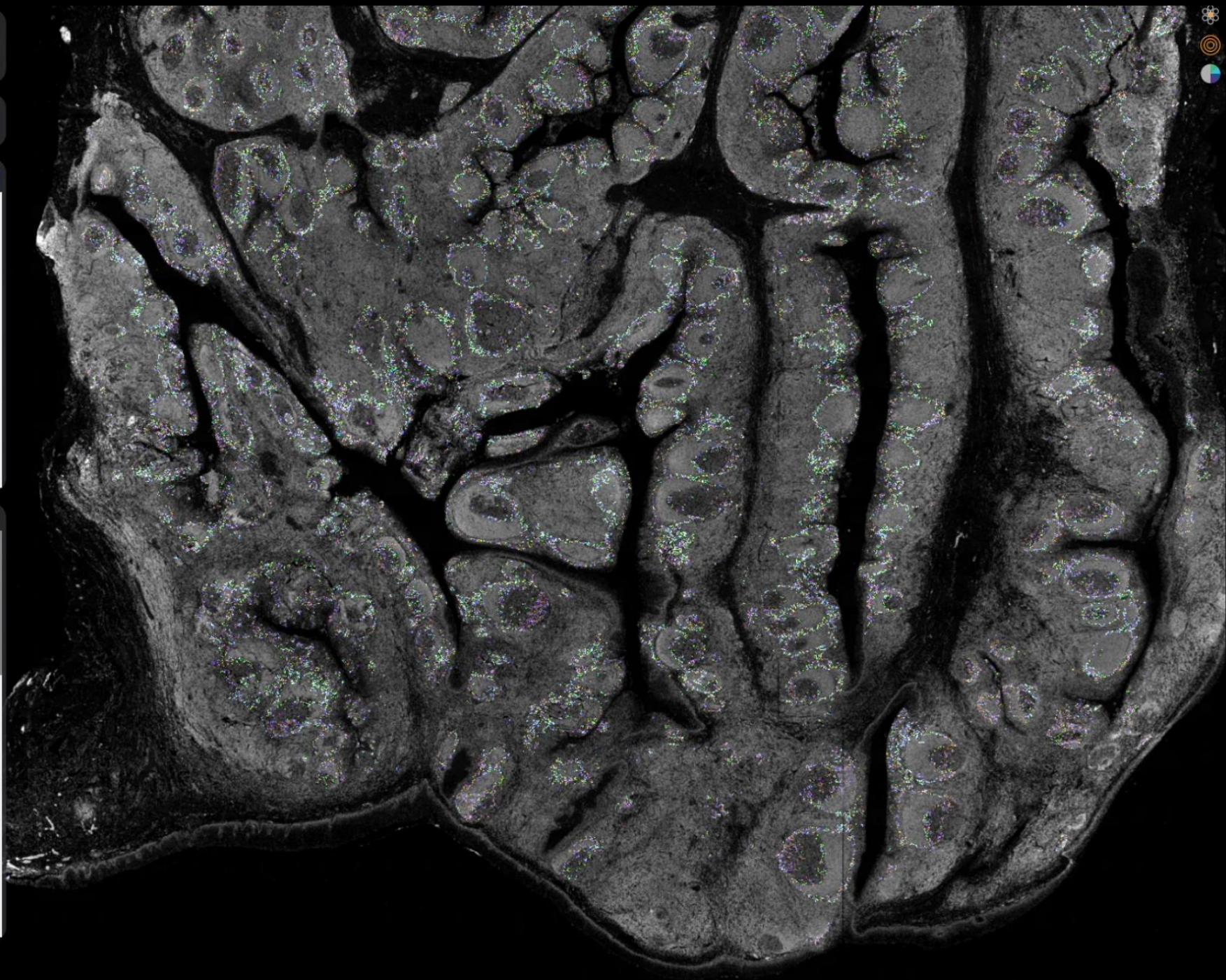
**Neighborhood Patterns**

Name	Actions
Germinal Center	[Edit] [Delete]
Custom Cluster 0 / 10	[Edit] [Delete]
Custom Cluster 1 / 10	[Edit] [Delete]
Custom Cluster 2 / 10	[Edit] [Delete]
Custom Cluster 3 / 10	[Edit] [Delete]
Custom Cluster 4 / 10	[Edit] [Delete]
Custom Cluster 5 / 10	[Edit] [Delete]
Custom Cluster 6 / 10	[Edit] [Delete]
Custom Cluster 7 / 10	[Edit] [Delete]
Custom Cluster 8 / 10	[Edit] [Delete]
Custom Cluster 9 / 10	[Edit] [Delete]

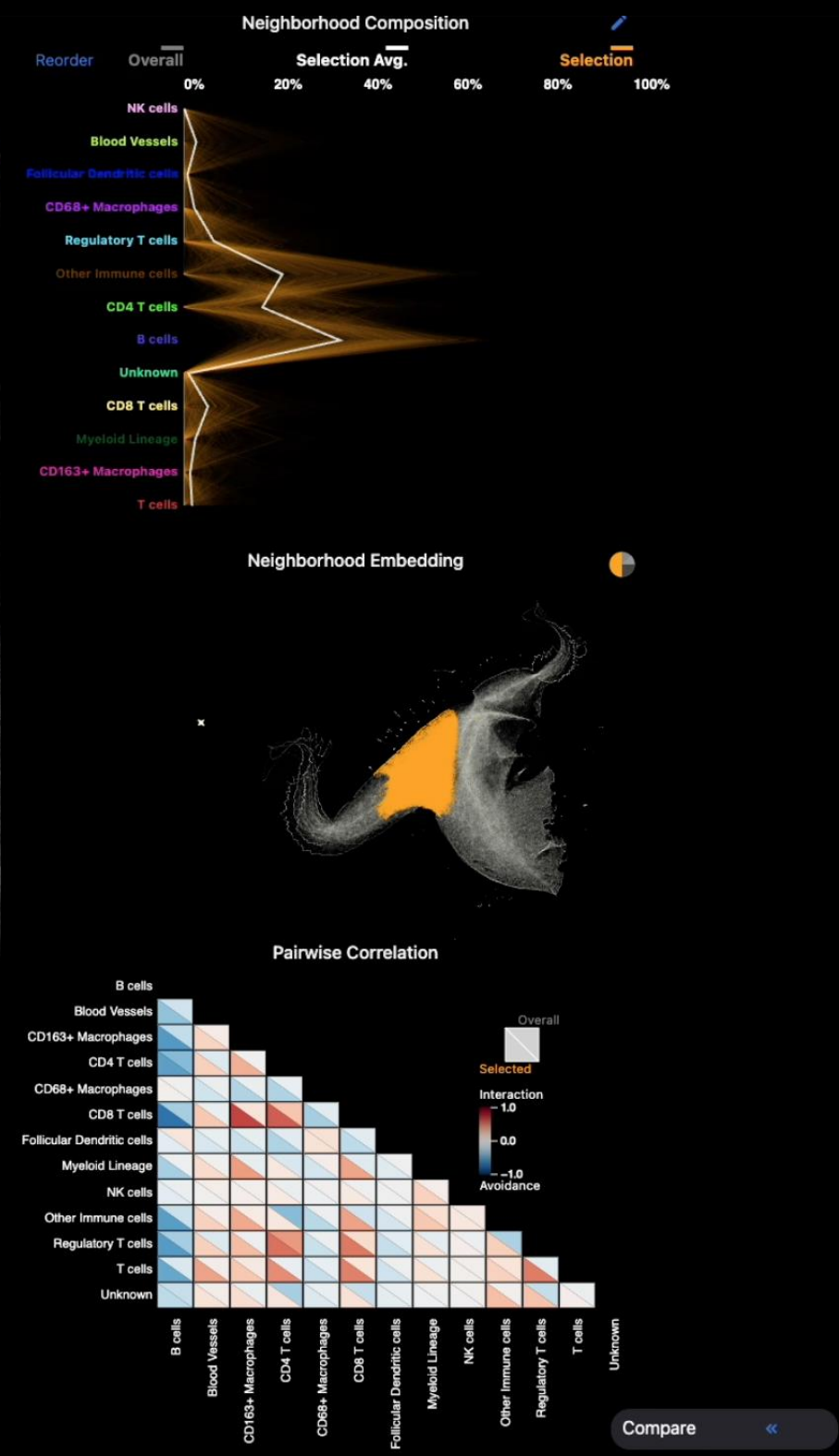
**Channels** 1 of 4 (Max) Selected

DNA1 Color: [Color Picker]

- AF488
- AF555
- AF647
- DNA2
- BG488
- BG555
- BG647
- DNA3
- CD2
- NONE
- CD10
- DNA4
- CD30
- ALK



T-cell – B-cell margin



Compare



**Similarity Search**

Selection: Cluster

Threshold: 0.8

**Cluster Neighborhoods**

5

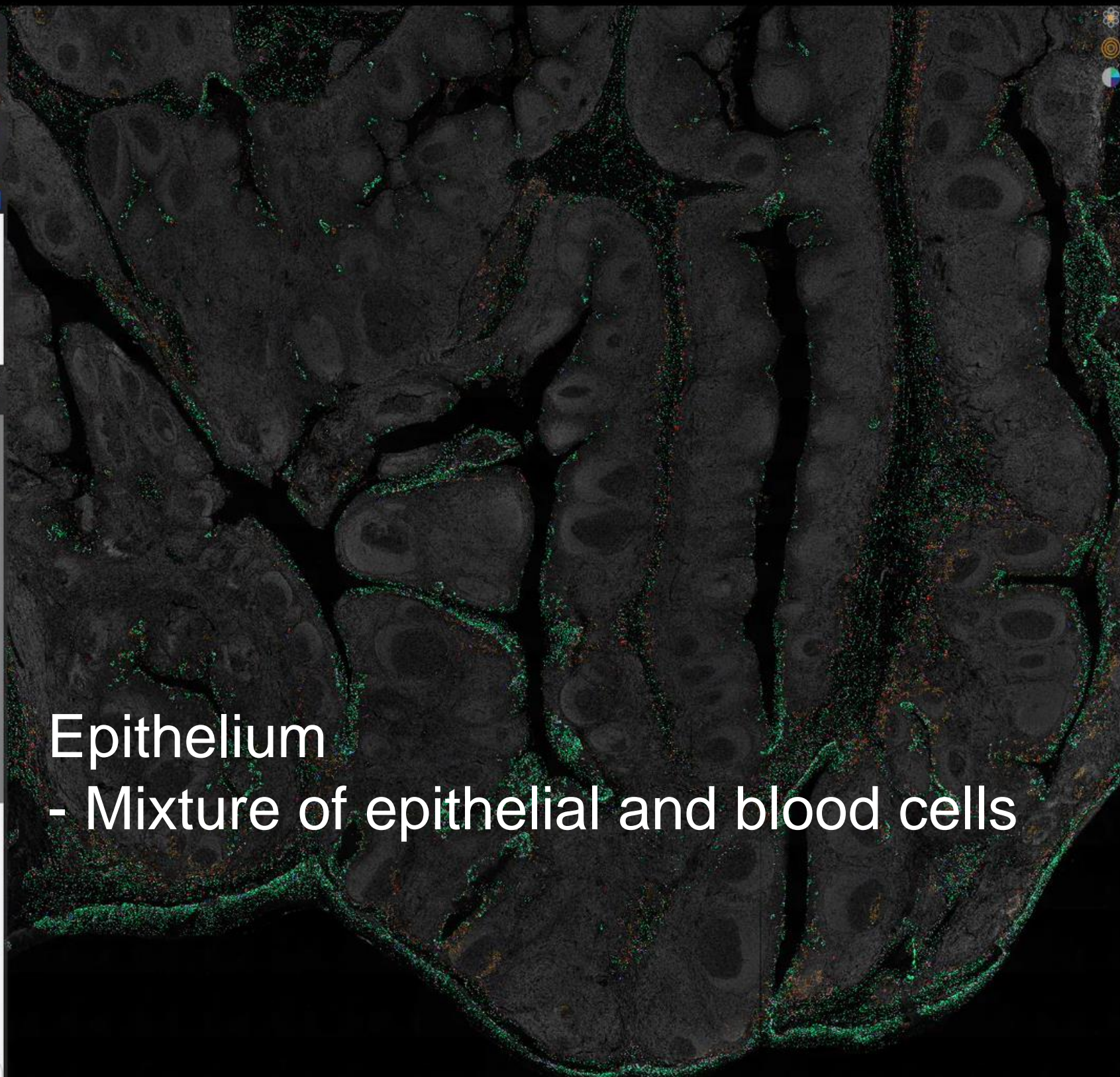
**Neighborhood Patterns**

Name	Actions
Cluster 1 / 5	
Cluster 2 / 5	
Cluster 3 / 5	
Cluster 4 / 5	
Cluster 5 / 5	

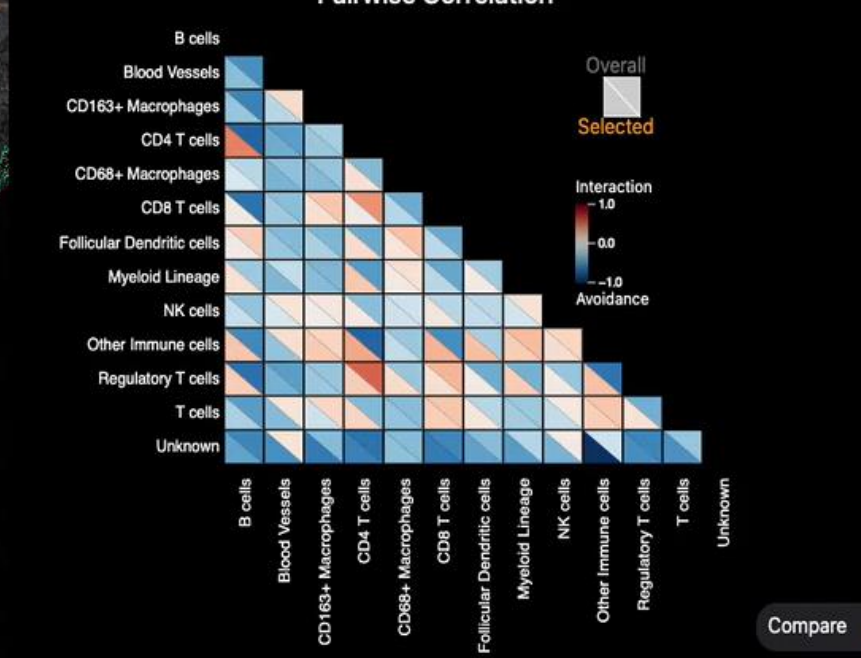
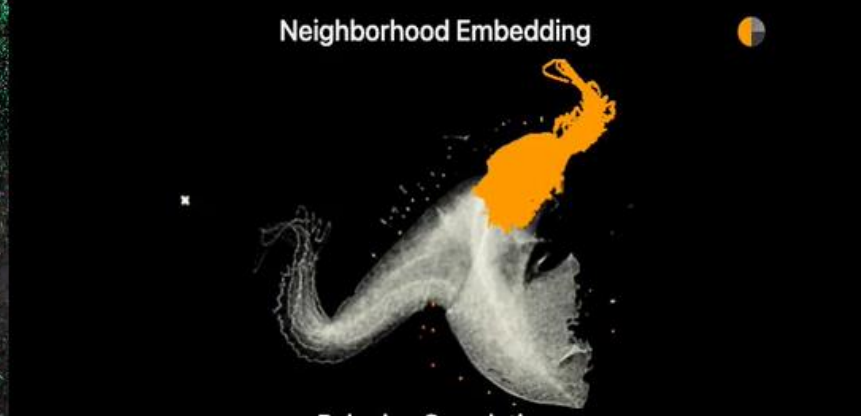
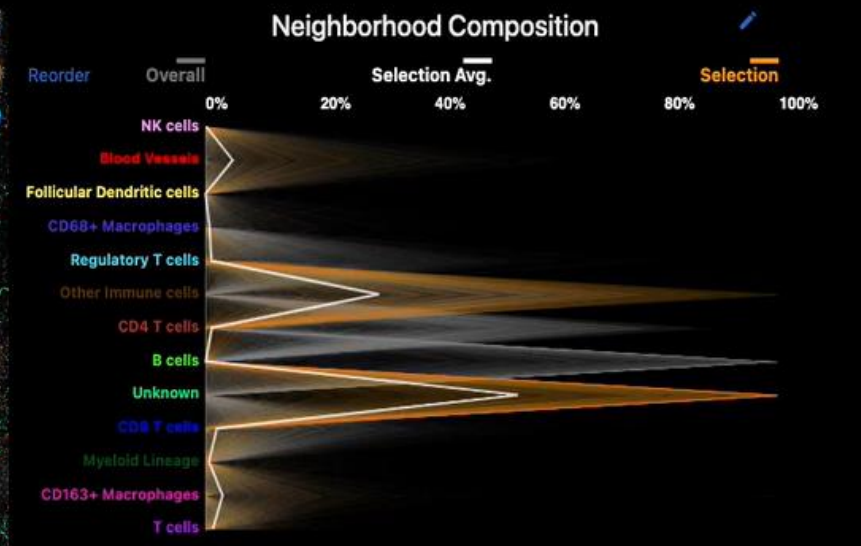
**Channels** 1 of 4 (Max) Selected

DNA1 Color:

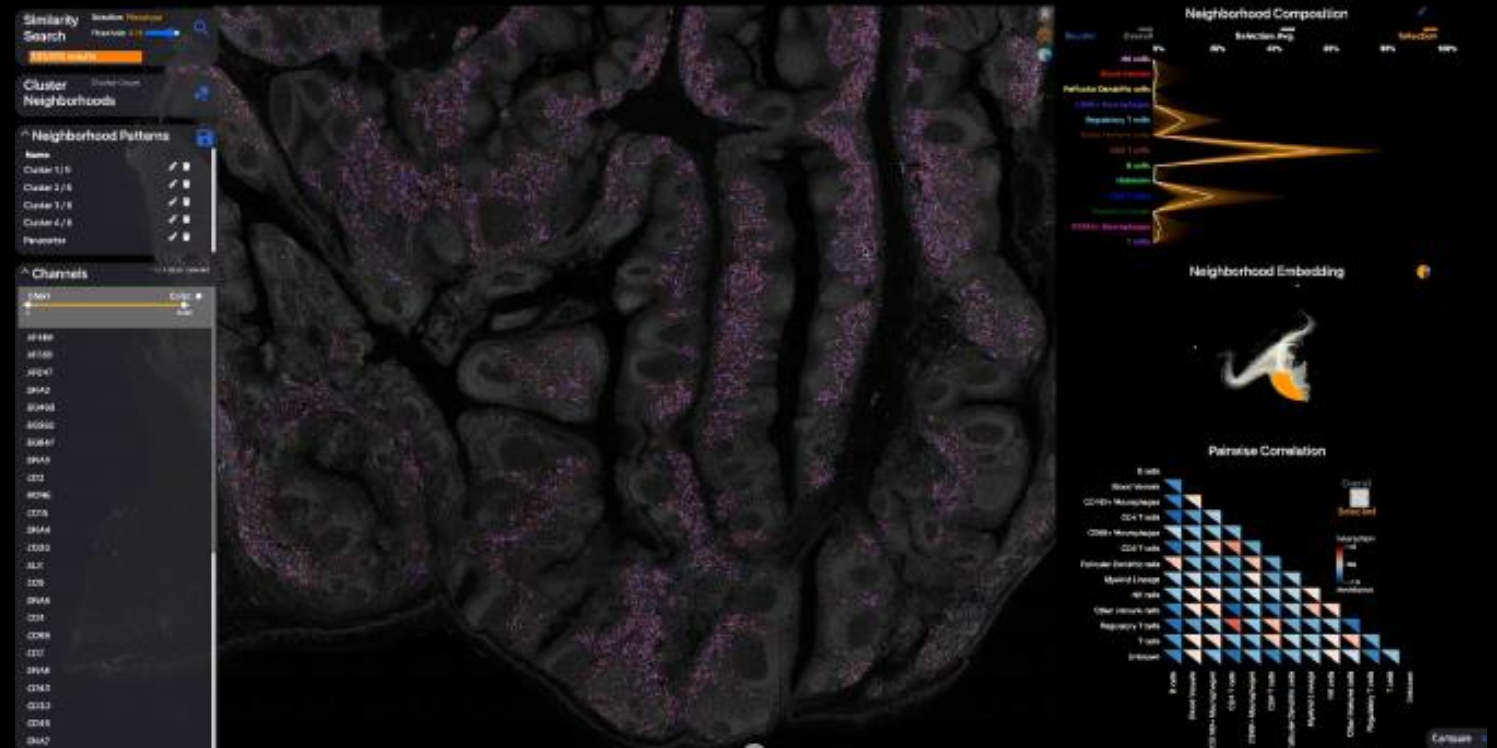
- AF488
- AF555
- AF647
- DNA2
- BG488
- BG555
- BG647
- DNA3
- CD2
- NONE
- CD10
- DNA4
- CD30
- ALK
- CD5
- DNA5
- CD4
- CD68
- CD7
- DNA6
- CD43
- CD3D
- CD45
- DNA7



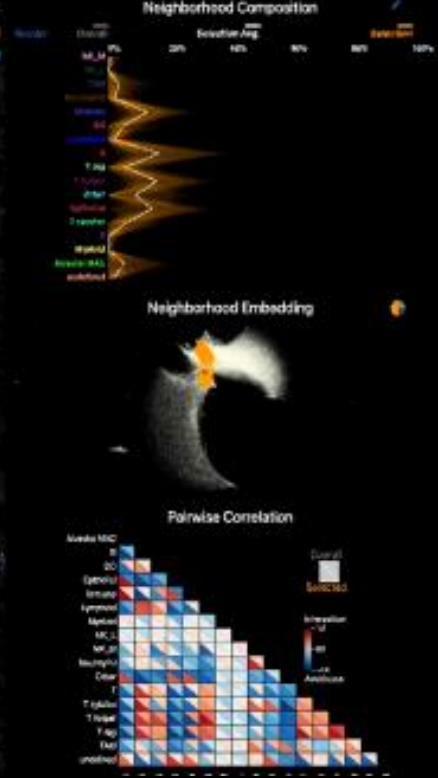
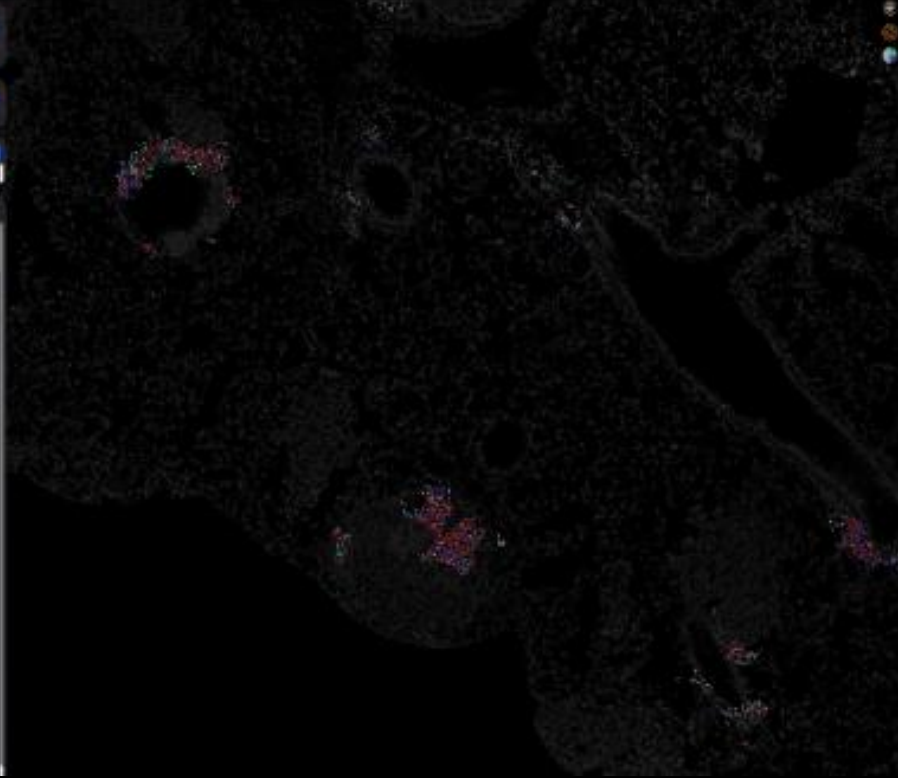
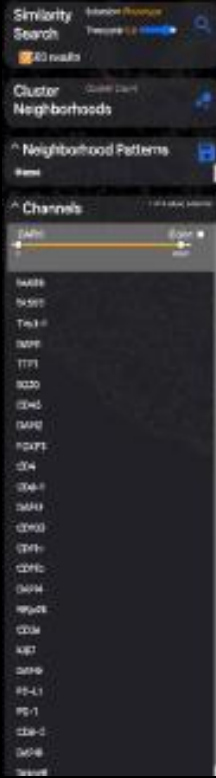
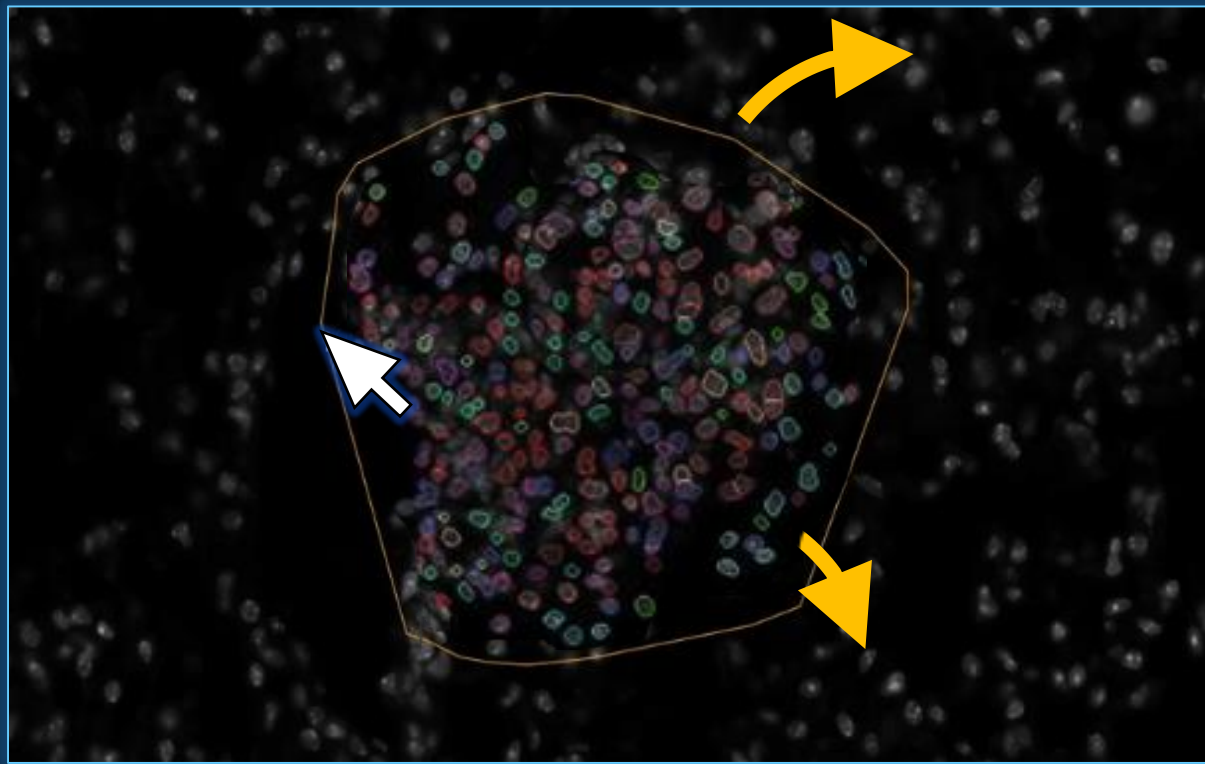
Epithelium  
- Mixture of epithelial and blood cells



# Top-Down: Hypothesis Formulation & Query



# Top-Down: Search/Find by Example



# Visual Analysis of Multiplexed Immunofluorescence Data

## Cell Classification

Data Facetting with Machine Learning



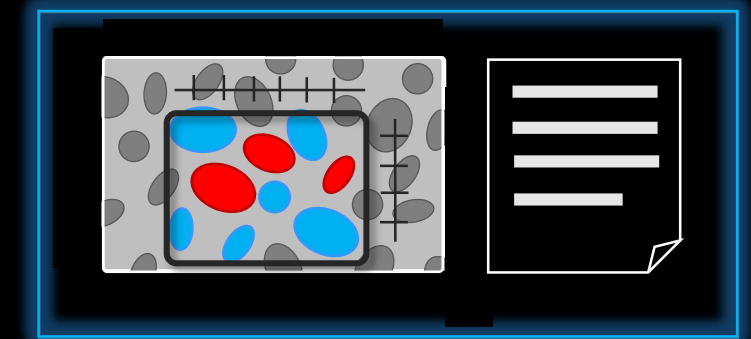
## Interaction Analysis

Cellular Neighborhoods

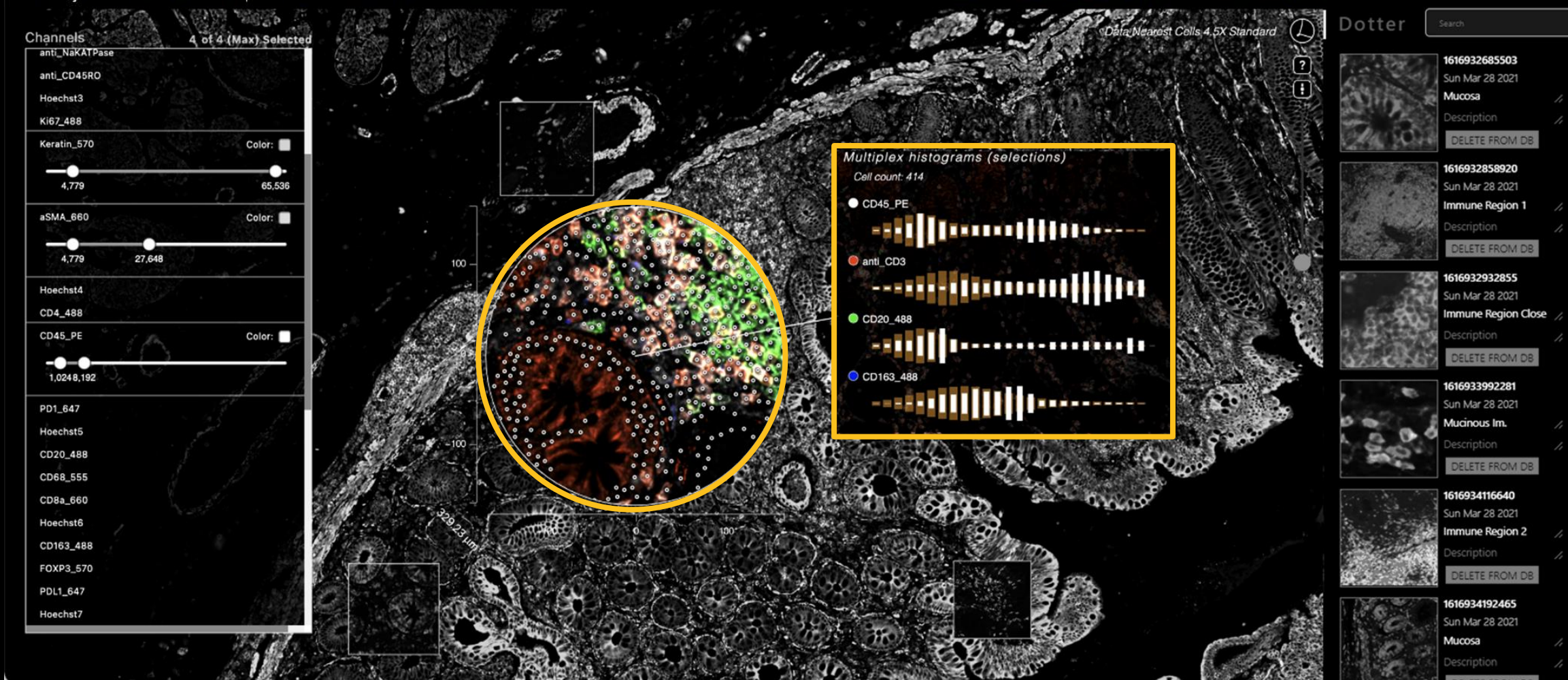


## Annotation

Focus+Context Pathology Assessment

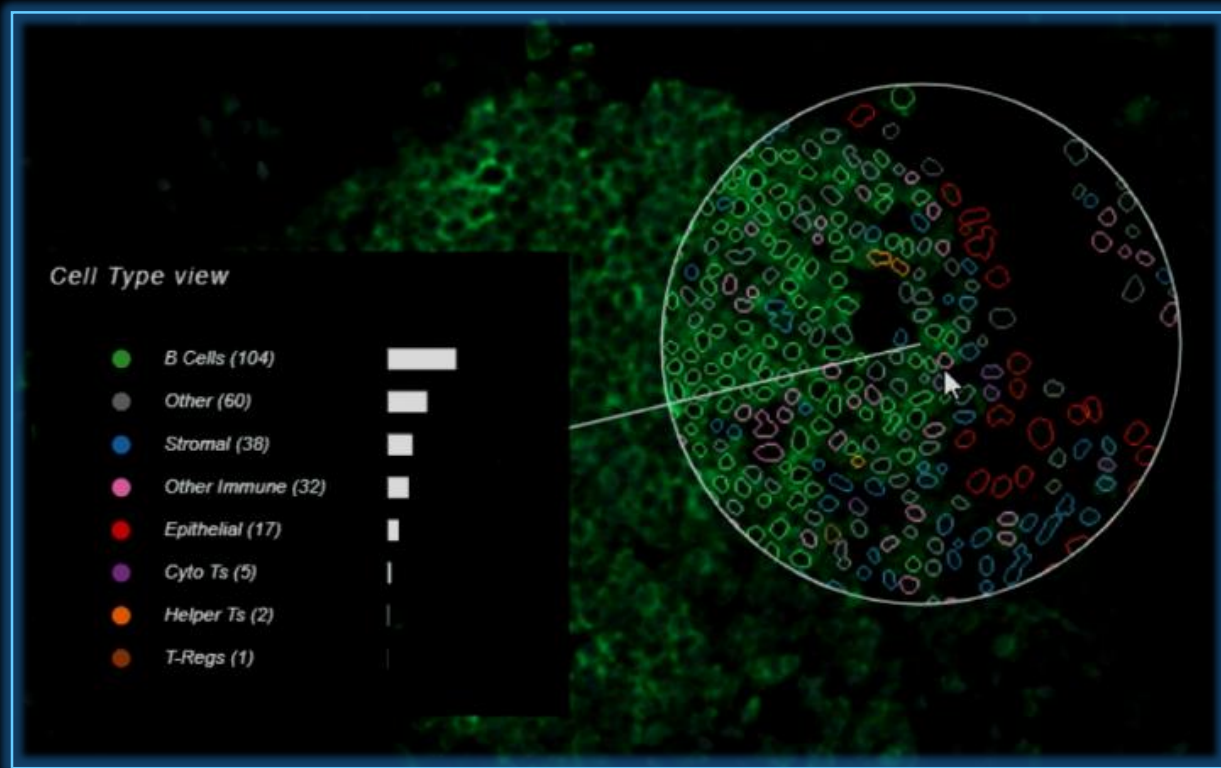


# From Scope 2 Screen



Jessup, J.& Krueger, R., et al., 2021. Scope2Screen: Focus+Context Techniques for Pathology Tumor Assessment in Multivariate Image Data. *IEEE TVCG*.

# Filters



Feature-based

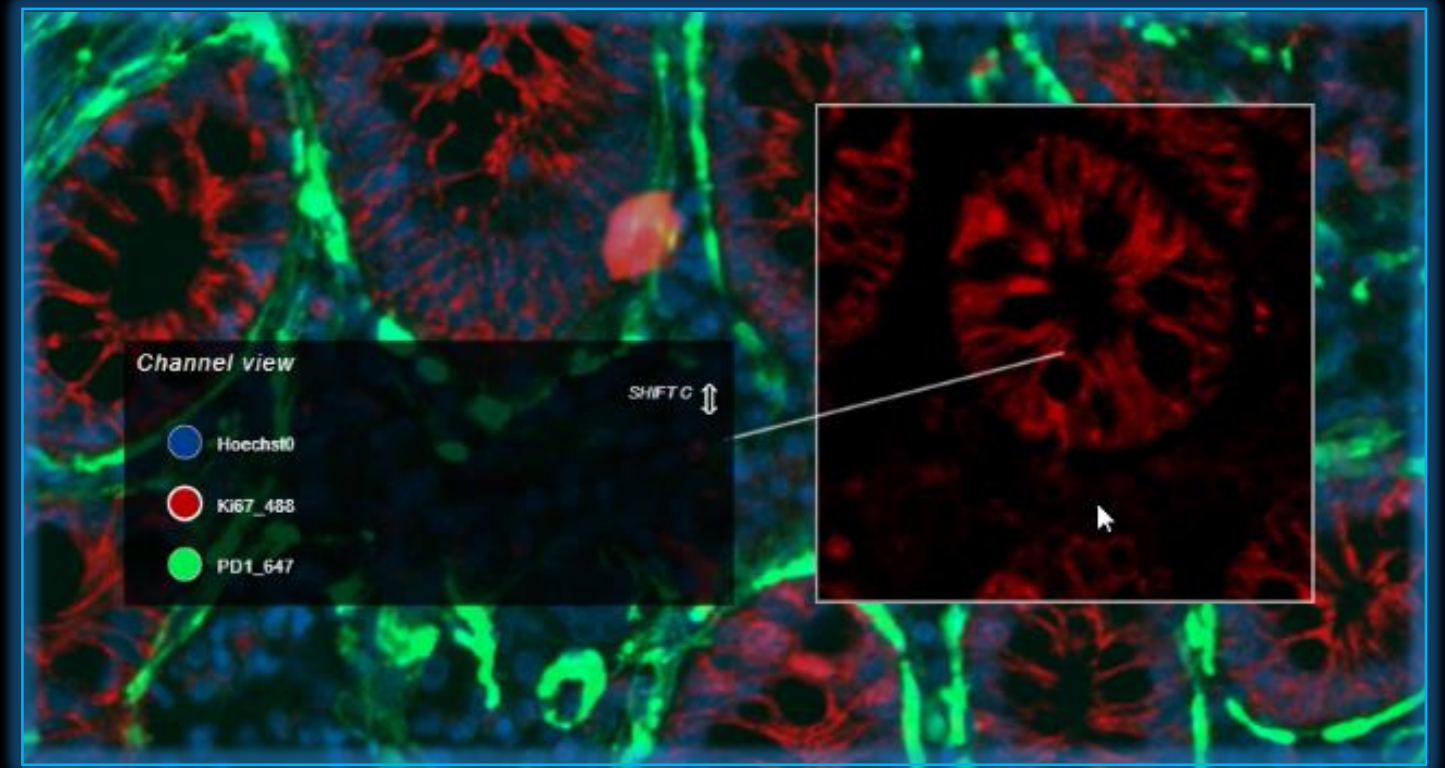
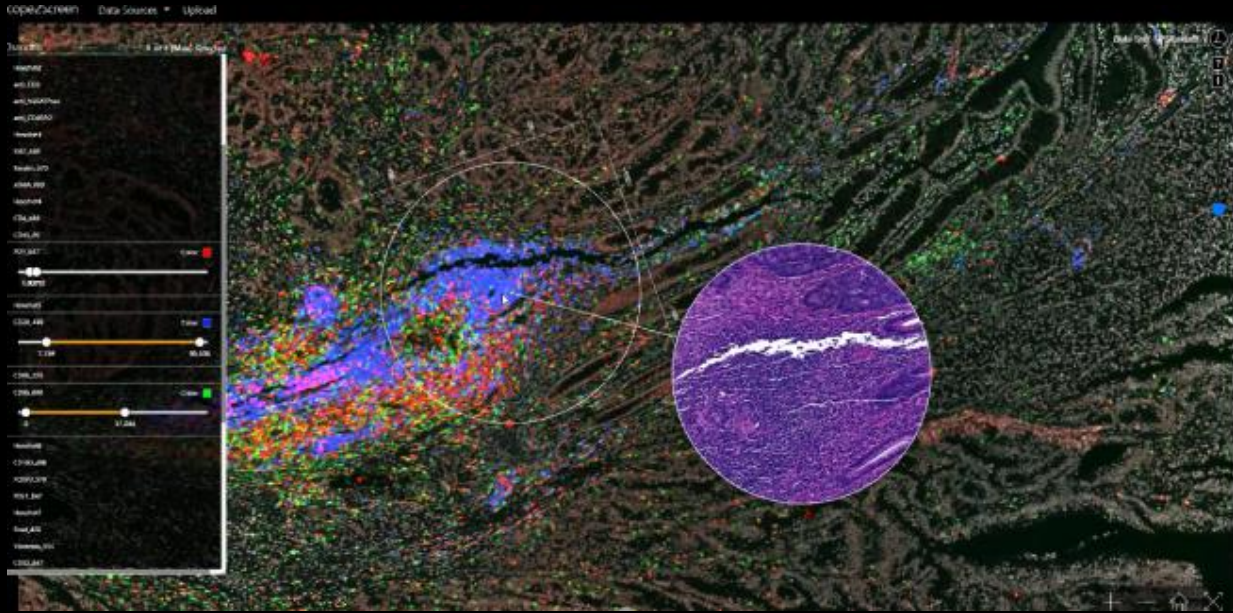
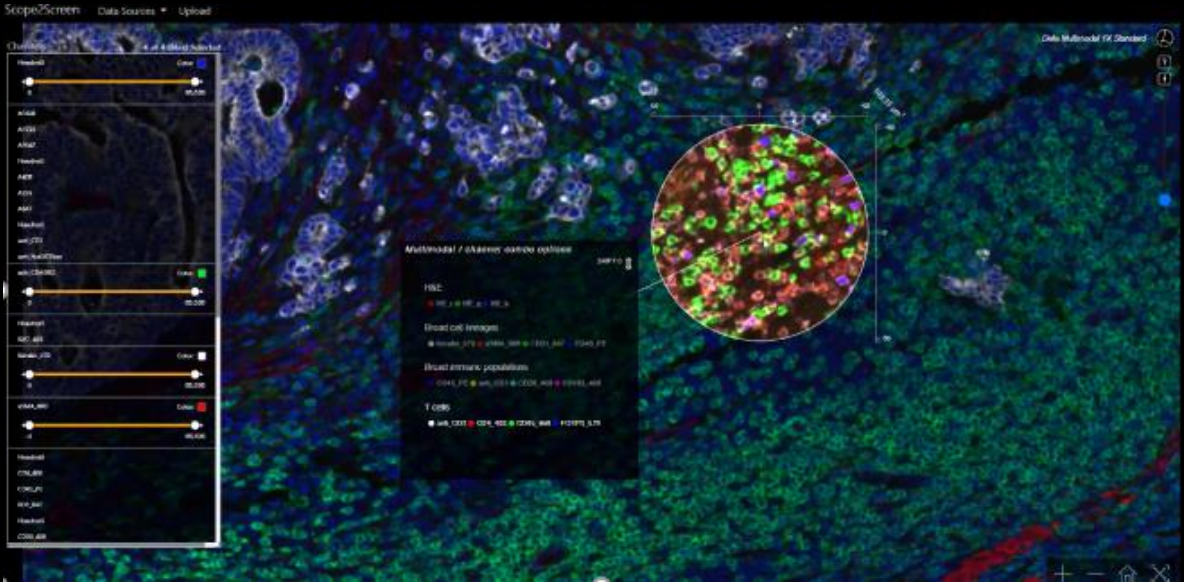
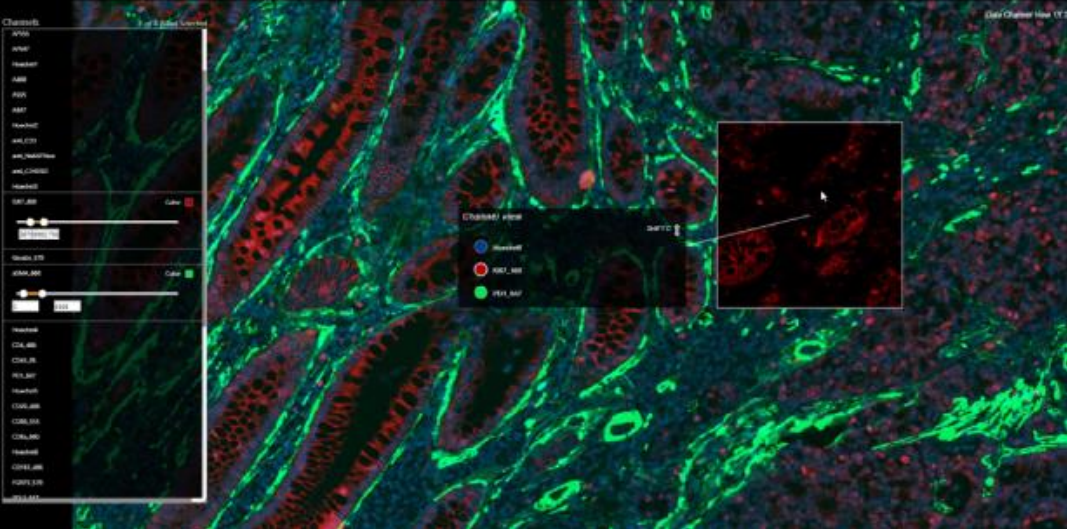


Image-based

# Image-Based Filters



# Annotation

The screenshot displays a histology image analysis software interface. The main window shows a tissue section with a color scale on the left and a list of annotated regions on the right. The color scale includes channels for Hoechst, Ki67, and Ki67, with sliders for intensity and color selection. The list of annotated regions includes details such as ID, date, name, description, and a 'DELETE FROM DB' button for each entry.

**Channels:** 1 of 4 (Max) Selected

- Hoechst
- Ki67
- Ki67
- Ki67

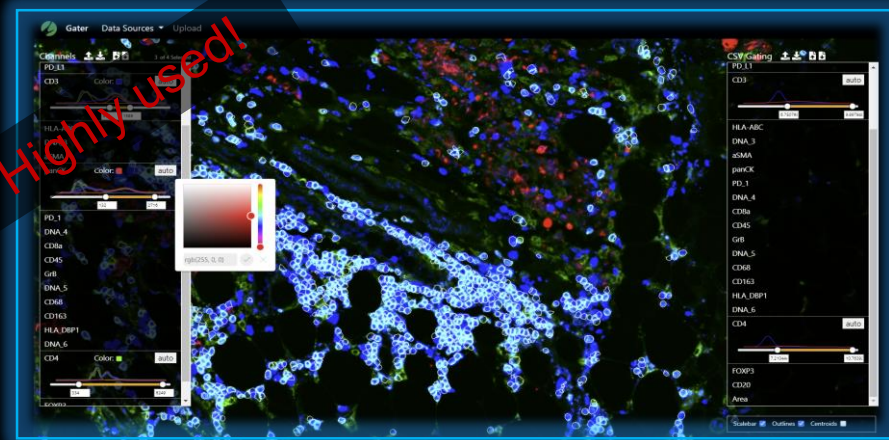
**Annotations:**

ID	Date	Name	Description	Action
M0715701805	Wed Mar 31 2021	Lymphocyte Region	This region shows...	DELETE FROM DB
M0715701860	Wed Mar 31 2021	Name	Description	DELETE FROM DB
M0715701234	Wed Mar 31 2021	Name	Description	DELETE FROM DB
M0715701806	Wed Mar 31 2021	Name	Description	DELETE FROM DB
M0715701848	Wed Mar 31 2021	Name	Description	DELETE FROM DB



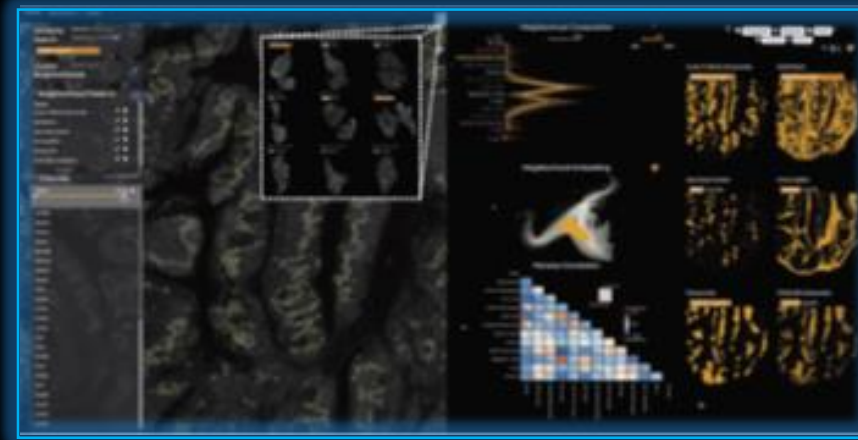


# Analysis Suite: 3 Tools



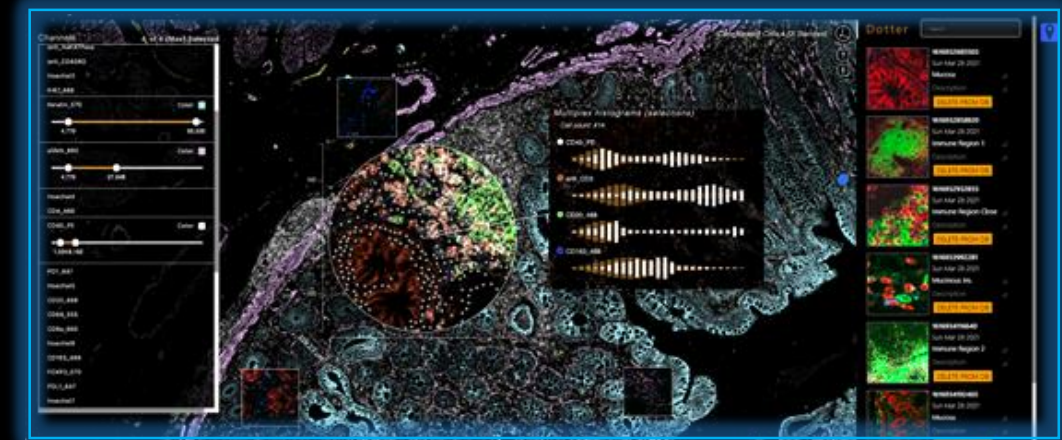
Gater: Tabular Data Cleaning with Visual Feedback

<https://github.com/labsyspharm/gater>



Visinity (Facetto): Cell Typing and Neighborhood Analysis

<https://github.com/labsyspharm/visinity>



Scope2Screen: Focus+Context ROI Exploration and Annotation

<https://github.com/labsyspharm/scope2screen>



Visual Computing Group @  
Harvard School of  
Engineering and Applied  
Sciences

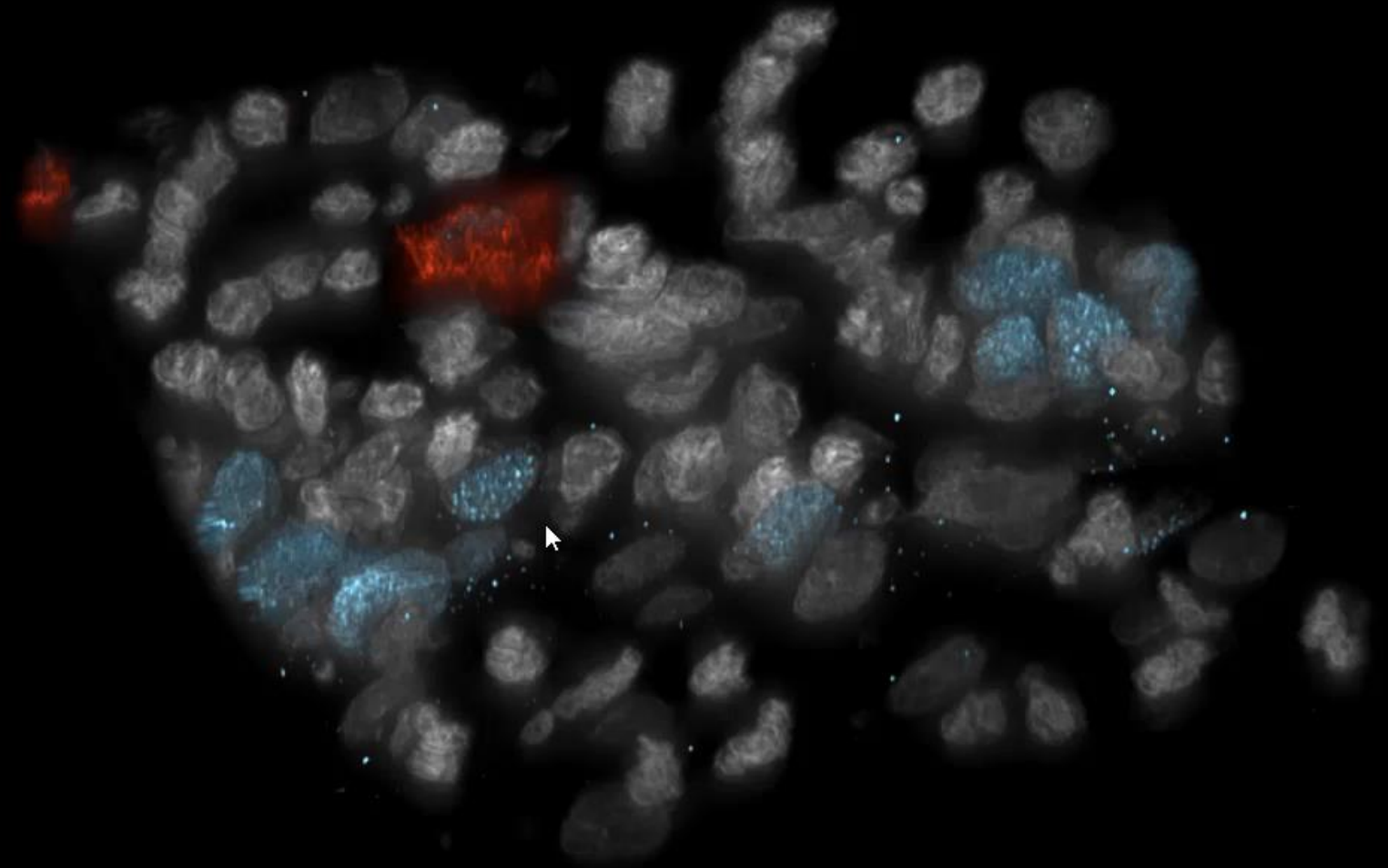


Laboratory of Systems  
Pharmacology  
@Harvard Medical School

# Future Research Areas

## 2D to 3D

- TB-sized multi-volumetric data
- Scalable multi-resolution rendering
- Image-based cell and interaction analysis



Herzberger, Lukas et al. 2023. Residency Octree: A Hybrid Approach for Scalable Web-Based Multi-Volume Rendering. IEEE TVCG.

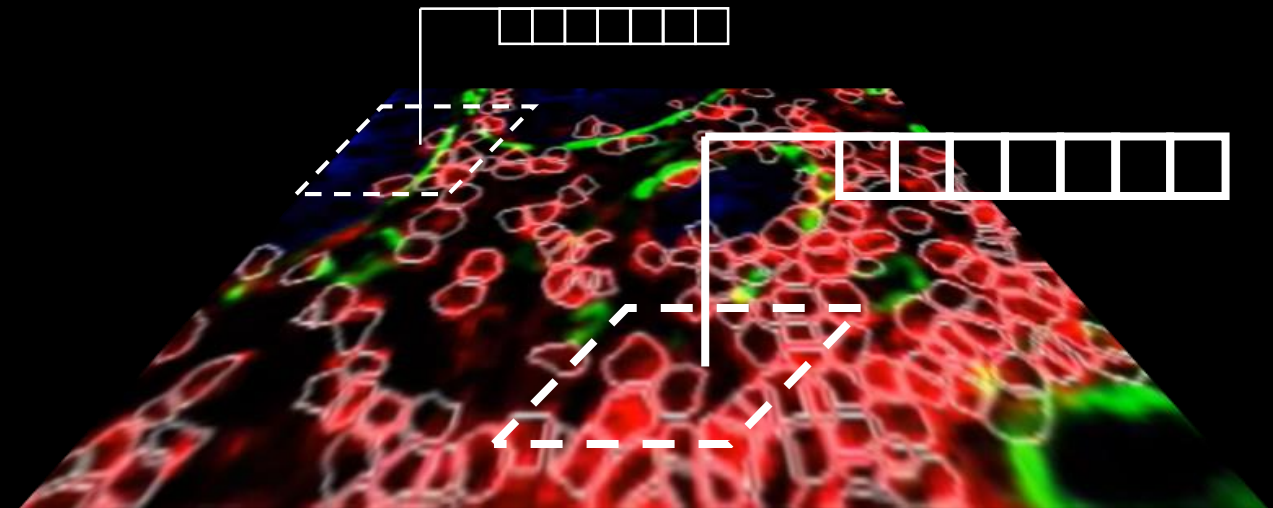
# Future Research Areas

- 2D to 3D

- TB-sized multi-volumetric data
- Scalable multi-resolution rendering
- Image-based cell and interaction analysis

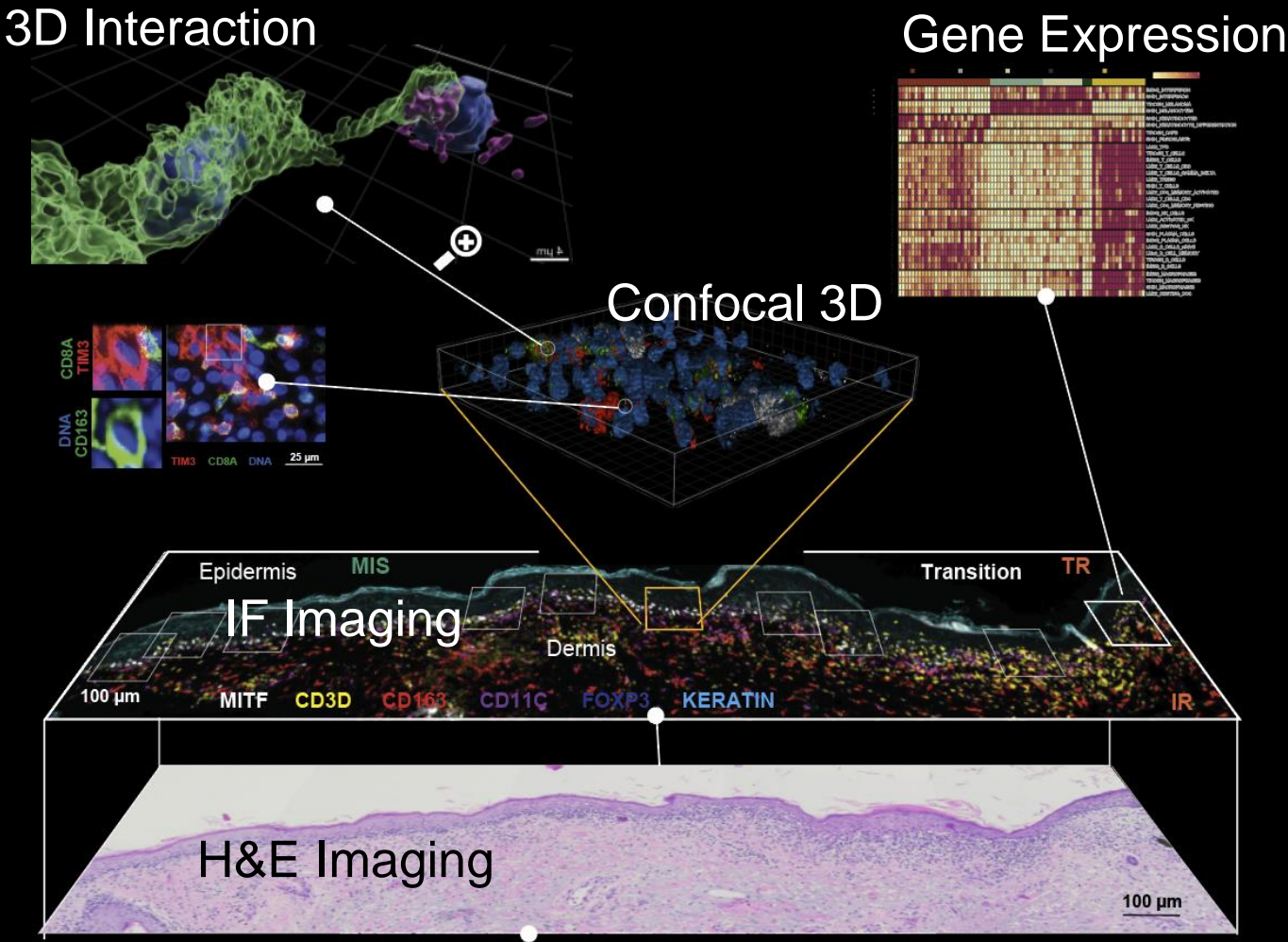
- Holistic Orchestration and Analysis

- Multiple modalities
- Spatially referenced gene expressions
- Textual data (annotations, meta data)

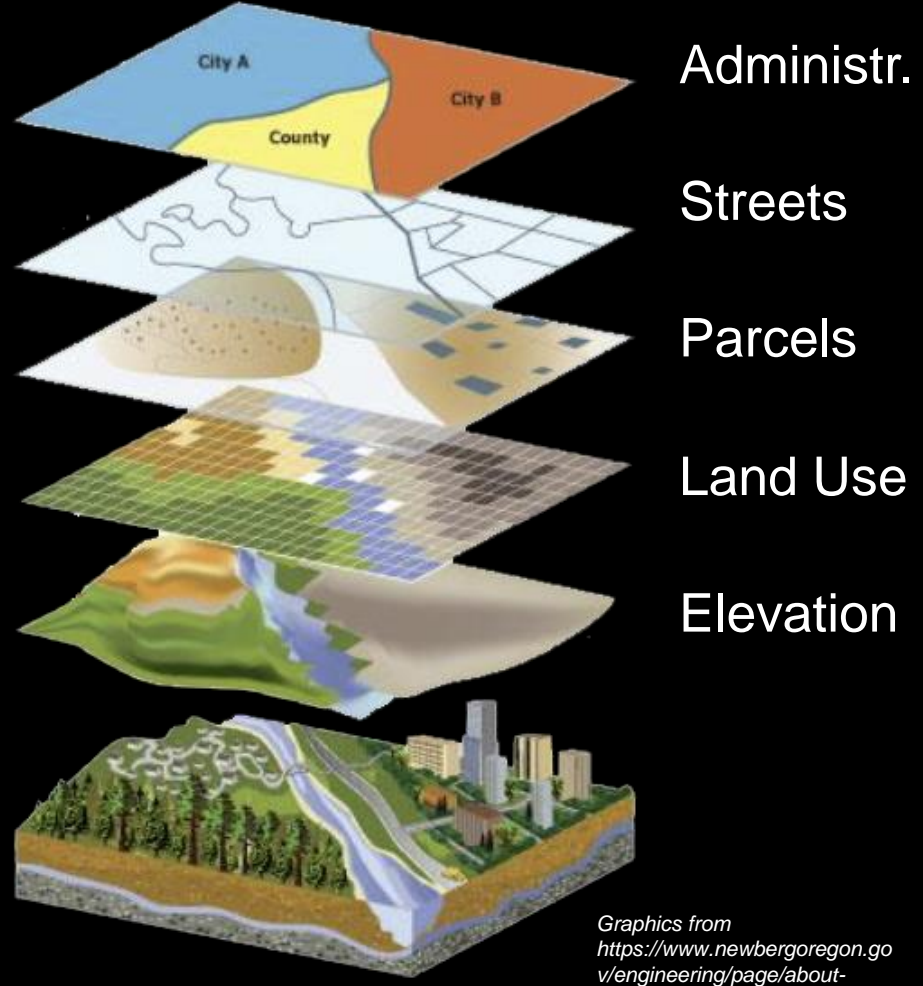


# Spatial Analysis in Other Domains

Nirmal et al. The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution, Cancer Discovery, 2022.



Biomedical Spatial Information Systems



Geographical Information Systems

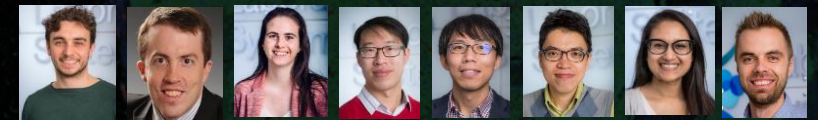
Graphics from <https://www.newbergoregon.gov/engineering/page/about-newberg-gis>

# Thank You!

## Questions?

rk4815@nyu.edu

### Acknowledgments:



This work is supported by the Ludwig Center at Harvard Medical School, and by NIH/NCI grants NCI U2C-CA233262, NCI U2C-CA233280 and NCI U54-CA225088.



Visual Computing Group  
@School of Engineering and  
Applied Sciences

Hi+S



Laboratory of Systems Pharmacology,  
@Harvard Medical School