

# SPATIAL OMICS DATA ANALYSIS WITH MAWA 4: NEIGHBORHOOD ANALYSIS USING SPATIAL UMAP

DANTE J SMITH, PHD (NCI)

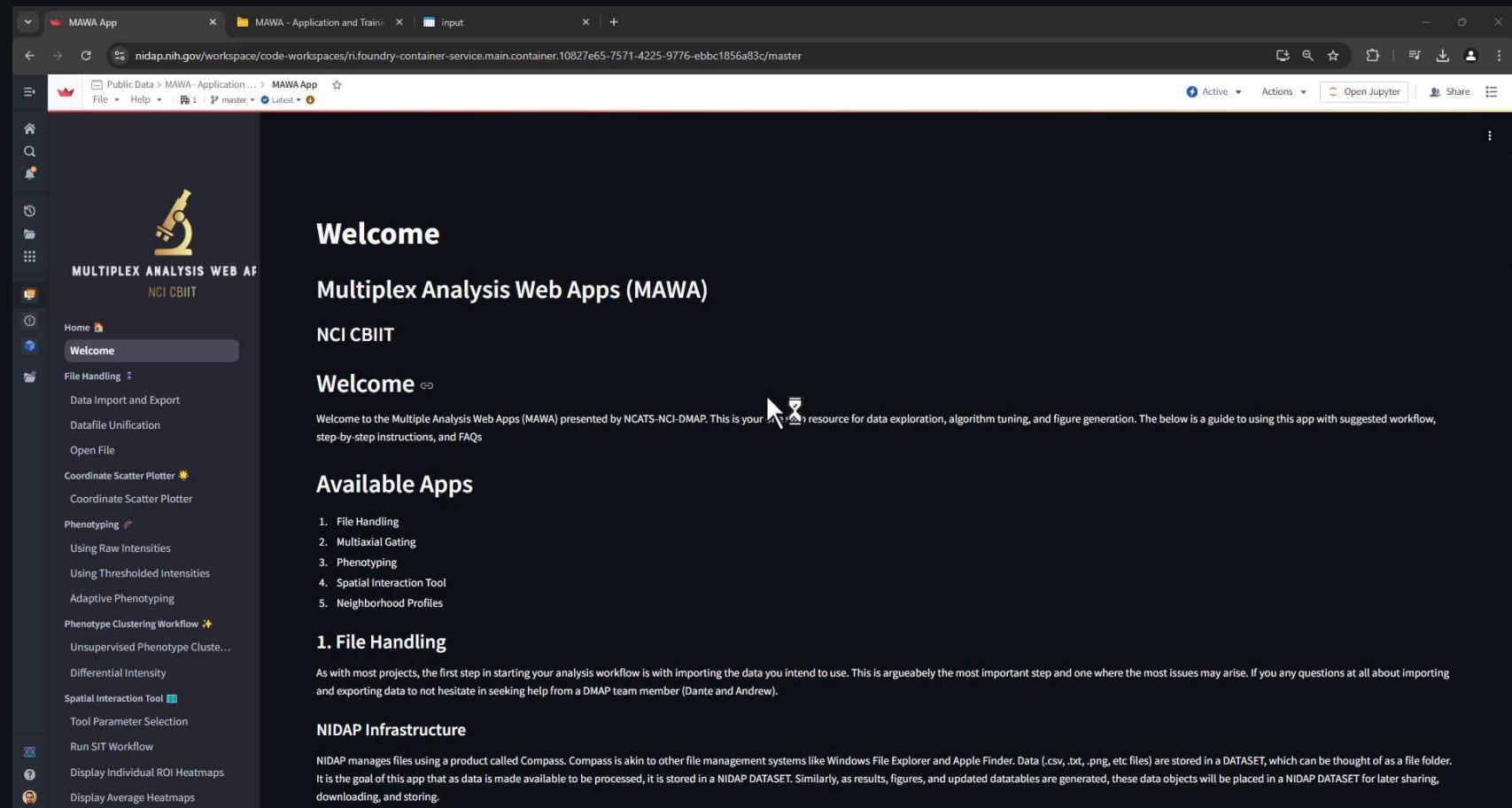
WEDNESDAY 11/6/2024

# Agenda

- Review Data Ingestion and Phenotyping
- Introduction to Neighborhood Profiles and UMAP
- Using Neighborhood Profiles on MAWA

# Datafile Import and Export

- Select *Data Import and Export*
- Table on left displays all data available on NIDAP for MAWA
- Verify your data is present by checking the **INPUT** folder in the top-level NIDAP folder



# Datafile Import and Export

- Select *Data Import and Export*
- Table on left displays all data available on NIDAP for MAWA
- Verify your data is present by checking the **INPUT** folder in the top-level NIDAP folder
- Select which files you want to load into MAWA
- Click the Load Button
- Verify all expected files appear in the table on the right

The screenshot shows the MAWA App interface with the following components:

- Left Sidebar:** Displays the "MULTIPLEX ANALYSIS WEB APP NCI CBIIT" logo and a navigation menu with options like Home, Welcome, File Handling (selected), Data Import and Export (highlighted in blue), Datafile Unification, Open File, Coordinate Scatter Plotter, Coordinate Scatter Plotter, Phenotyping, Using Raw Intensities, Using Thresholded Intensities, Adaptive Phenotyping, Phenotype Clustering Workflow, Unsupervised Phenotype Cluste..., Differential Intensity, Spatial Interaction Tool, Tool Parameter Selection, Run SIT Workflow, Display Individual ROI Heatmaps, and Display Average Heatmaps.
- Main Content Area:**
  - Data Import and Export:** A title section.
  - Input:** A table titled "Available input data on NIDAP" listing several CSV files:
    - 11431\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - 12123\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - 12801\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - 13780\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - 16171\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - Multiplexed\_Image\_Sample.csv
    - all\_five\_images\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - mawa-unified\_datafile-1\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - mawa-unified\_datafile-Demo3Images\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
    - mawa-unified\_datafile-Fused\_images\_test-20241017\_112907\_EDT.csv
  - Load input data into MAWA:** A button with a green icon and the text "Load selected NIDAP input data".
  - Save MAWA-unified datafile to NIDAP:** A button with a green icon and the text "Save selected (above) MAWA-unified datafile to NIDAP".
  - Results:** A title section.
  - Available results archives (i.e., saved results) on:** A table listing results:
    - Load results
    - Results in MAWA

# Datafile Unification

- Select *Datafile Unification*
- Select the files to combine
- Follow Steps 2-6 to preprocess your dataset to conform to MAWA's format
- Save a copy of the mawa\_unified\_datafile
- Select *Open File*
- Ensure toggle is on to load dataset from Datafile Unifier
- Load and review your data

The screenshot shows the MAWA App interface on a web browser. The left sidebar contains a navigation menu with options like Home, Welcome, File Handling (selected), Data Import and Export (highlighted in red), Datafile Unification, Open File, Coordinate Scatter Plotter, Phenotyping, Using Raw Intensities, Using Thresholded Intensities, Adaptive Phenotyping, Phenotype Clustering Workflow, Unsupervised Phenotype Clustering, Differential Intensity, Spatial Interaction Tool, Tool Parameter Selection, Run SIT Workflow, Display Individual ROI Heatmaps, and Display Average Heatmaps.

The main content area is titled "Data Import and Export". It has two sections: "Available input data on NIDAP" and "Input data in MAWA".

**Available input data on NIDAP:** A list of CSV files:

- 11431\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- 12123\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- 12801\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv (checkbox checked)
- 13780\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- 16171\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- Multiplexed\_Image\_Sample.csv
- all\_five\_images\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- mawa-unified\_datafile-1\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- mawa-unified\_datafile-Demo3Images\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- mawa-unified\_datafile-Fused\_images\_test-20241017\_112907\_EDT.csv

**Input data in MAWA:** A list of CSV files:

- 13780\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- 12801\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv
- 11431\_NOS2\_and\_COX2\_and\_CD8\_Fused\_ALW\_05212021\_entire\_image.csv

Buttons in the center include "Load selected NIDAP input data", "Save MAWA-unified datafile to NIDAP", "Reset file selections", and "Delete selected (above) loaded input files".

**Results:** A section titled "Available results archives (i.e., saved results) on NIDAP" with a "Load results" button. Below it is a "Results in MAWA" section.

# Phenotyping

- Select *Using Thresholded Intensities*
- Load your data
- Review your images
- Select a phenotyping method
- Review the phenotyped scatter plot

The screenshot shows the MAWA App interface running in a web browser. The left sidebar contains a navigation menu with options like Home, Welcome, File Handling, Data Import and Export, Datafile Unification, Open File (which is selected), Coordinate Scatter Plotter, Phenotyping, Using Raw Intensities, Using Thresholded Intensities, Adaptive Phenotyping, Phenotype Clustering Workflow, Unsupervised Phenotype Clustering, Differential Intensity, Spatial Interaction Tool, Tool Parameter Selection, Run SIT Workflow, Display Individual ROI Heatmaps, and Display Average Heatmaps. The main content area has a dark background with white text. It displays an "Open File" header, "Input options" section with a "Load the selected input dataset" button, and a "Loaded dataset" section with a "Properties" table. The "Properties" table lists the following details:

- Datafile: loaded from Datafile Unifier
- Coordinate units: 1.0 microns/coord
- Dataset format: <class 'dataset\_formats.Standardized'>
- Number of rows: 289317
- Number of columns: 66
- Minimum coordinate spacing: 0.2000 microns
- Loaded memory usage: 43.31 MB

Below the properties table is a "Dataframe sample" section with a "Refresh dataframe sample" button and a table showing a sample of the loaded data. The table has columns: Slide ID, tag, Cell X Position, Cell Y Position, Phenotype NOS2, Phenotype COX2, Phenotype CD8, Image ID\_(standardized), Centroid X (μm)\_ (standardized), and Centroid Y (μm)\_ (standardized). The first few rows of data are:

Slide ID	tag	Cell X Position	Cell Y Position	Phenotype NOS2	Phenotype COX2	Phenotype CD8	Image ID_(standardized)	Centroid X (μm)_ (standardized)	Centroid Y (μm)_ (standardized)
2,019	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	2,841	928.4	-	+	Ultivue_NOS2COX2_11431.tif	2,841	
7,960	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	3,661	1,260	-	-	Ultivue_NOS2COX2_11431.tif	3,661	
11,170	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	2,025.6	1,332.8	+	-	Ultivue_NOS2COX2_11431.tif	2,025.6	
11,539	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	2,252	1,316.6	-	-	Ultivue_NOS2COX2_11431.tif	2,252	
12,200	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	2,886.3999	1,379	-	+	Ultivue_NOS2COX2_11431.tif	2,886.3999	
16,430	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	6,162.2002	1,323.8	-	-	Ultivue_NOS2COX2_11431.tif	6,162.2002	
17,328	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	1,371	1,601.8	-	-	Ultivue_NOS2COX2_11431.tif	1,371	
19,902	1A-Ultivue_NOS2COX2_11431.tif	1A-Ultivue_NOS2COX2_11431.tif_roi_[4360,4054]	4,152.6001	1,689.6	-	-	Ultivue_NOS2COX2_11431.tif	4,152.6001	

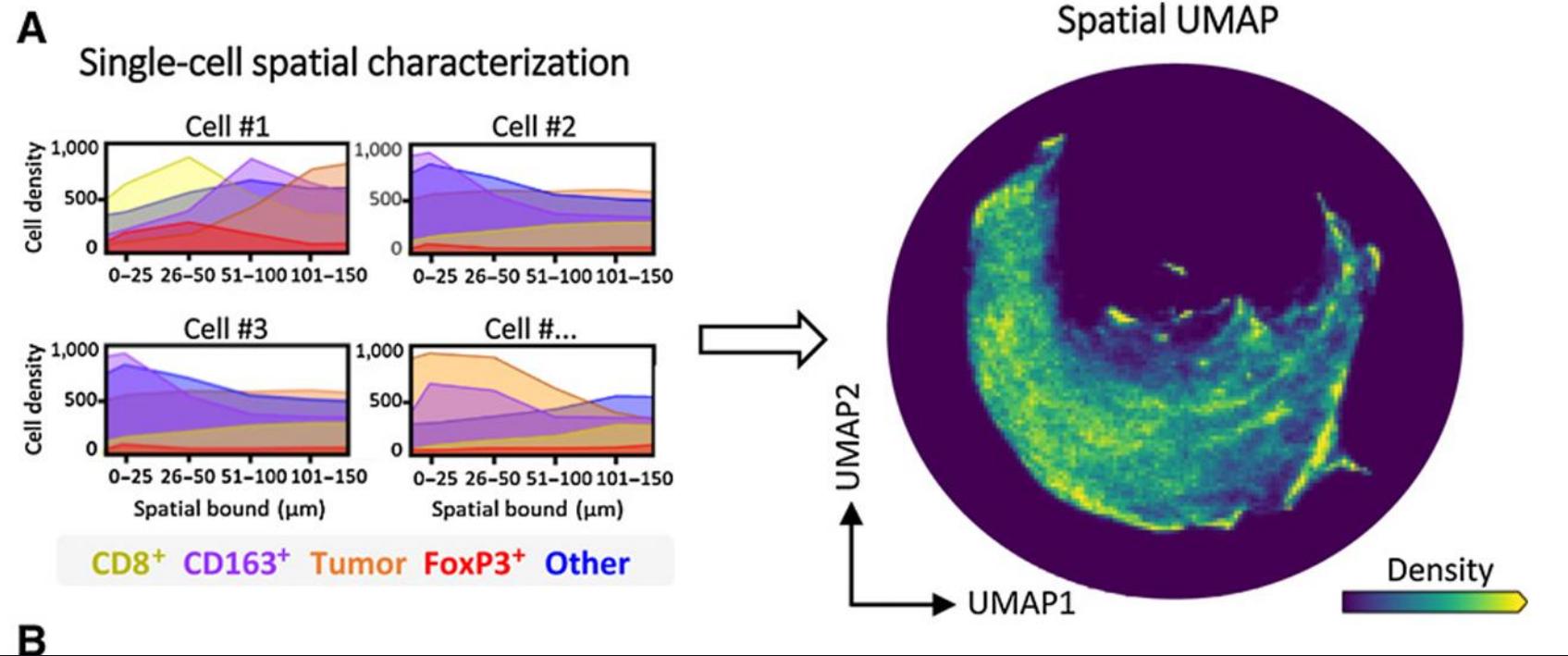
# Neighborhood Profiles - Giraldo et al. 2021

CANCER IMMUNOLOGY RESEARCH | PRIORITY BRIEF

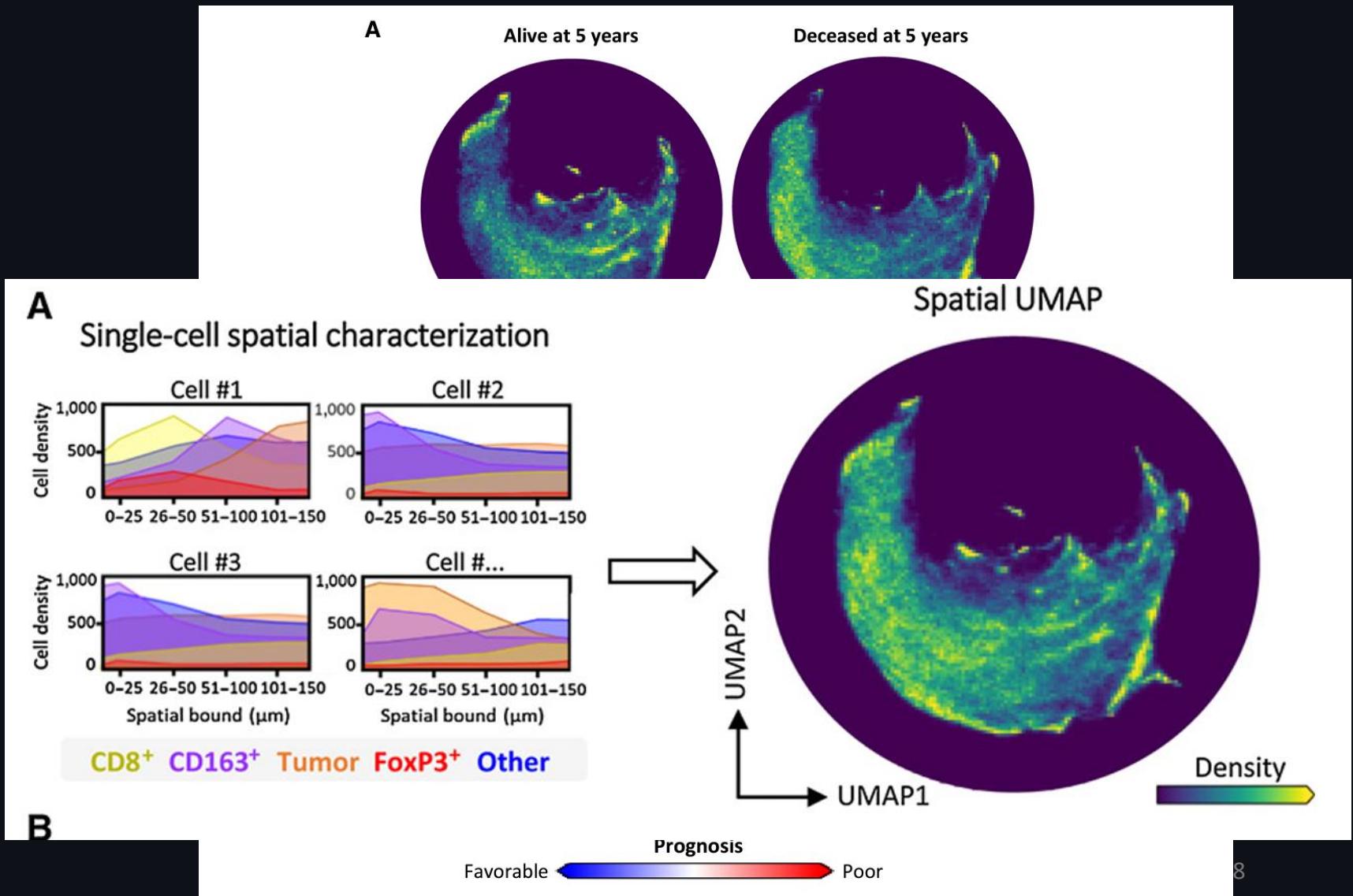
## Spatial UMAP and Image Cytometry for Topographic Immuno-oncology Biomarker Discovery



Nicolas A. Giraldo<sup>1</sup>, Sneha Berry<sup>2</sup>, Etienne Becht<sup>3</sup>, Deniz Ates<sup>4</sup>, Kara M. Schenk<sup>2</sup>, Elizabeth L. Engle<sup>5</sup>, Benjamin Green<sup>2</sup>, Peter Nguyen<sup>5</sup>, Abha Soni<sup>5</sup>, Julie E. Stein<sup>5</sup>, Farah Succaria<sup>5</sup>, Aleksandra Ogurtsova<sup>5</sup>, Haiying Xu<sup>5</sup>, Raphael Gottardo<sup>3</sup>, Robert A. Anders<sup>1</sup>, Evan J. Lipson<sup>2</sup>, Ludmila Danilova<sup>2</sup>, Alexander S. Baras<sup>1</sup>, and Janis M. Taube<sup>1,2,5</sup>



# Neighborhood Profiles - Giraldo et al. 2021



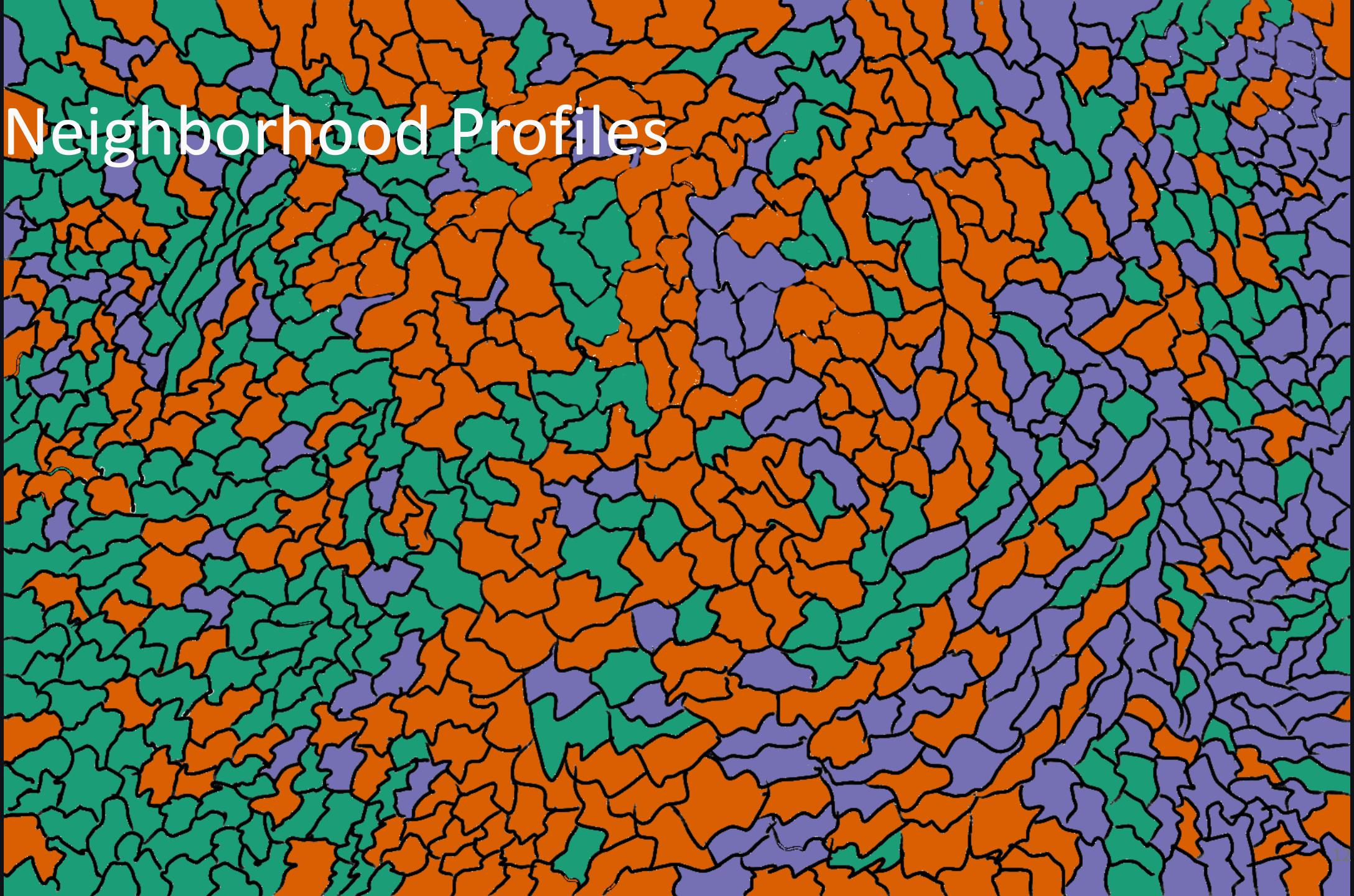
# Neighborhood Profiles

- Step 1: Perform Density measures of each cell in the sample
- Step 2: Perform UMAP to reduce the number of variables
- Step 3: Apply clustering algorithms to identify areas of the UMAP that are most similar and most disparate.

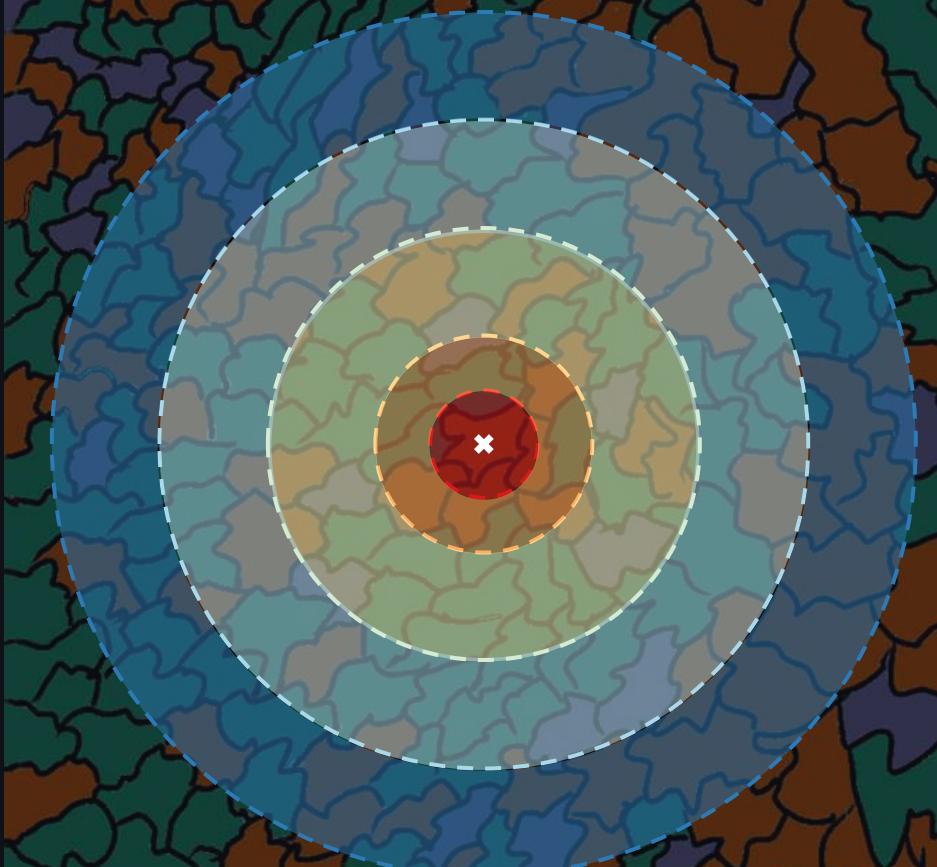
# Neighborhood Profiles

# Neighborhood Profiles

# Neighborhood Profiles



# Neighborhood Profiles

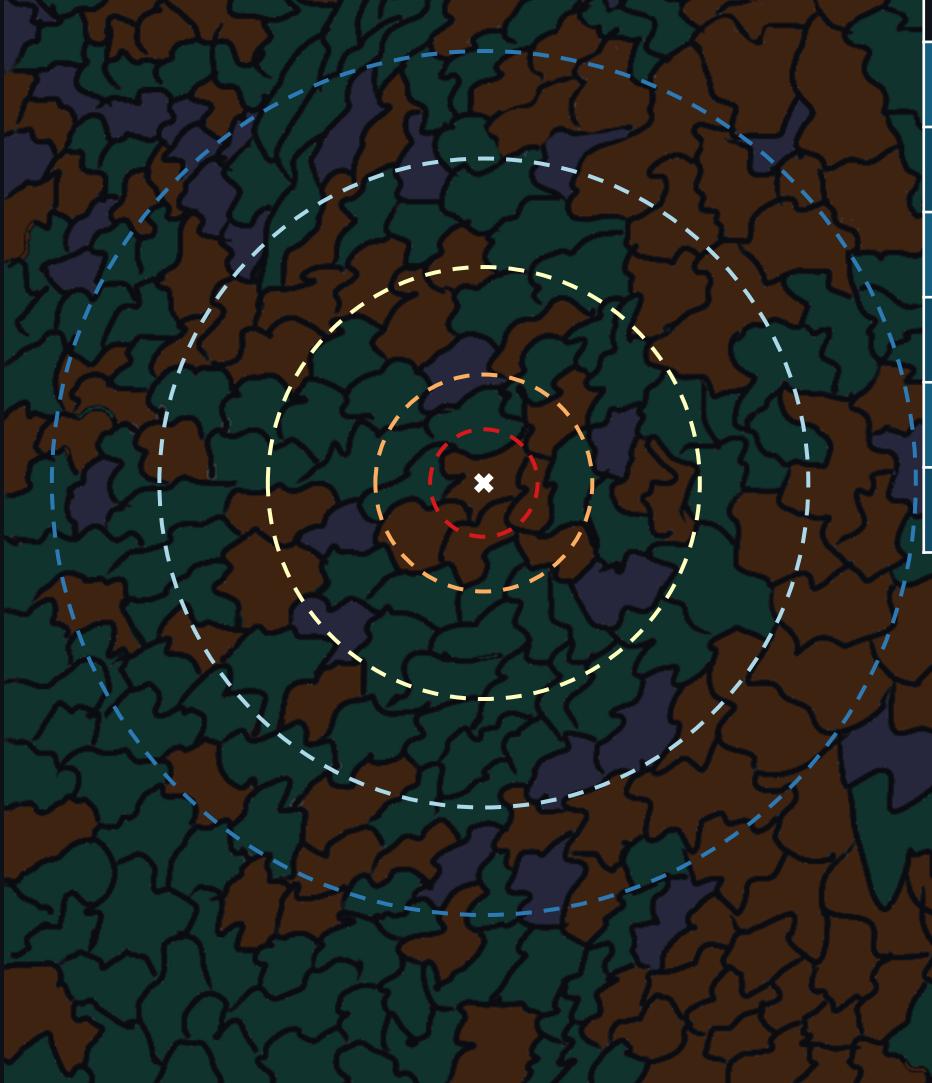


Radii:  $25\mu\text{m}$ ,  $50\mu\text{m}$ ,  $100\mu\text{m}$ ,  $150\mu\text{m}$ ,  $200\mu\text{m}$

Areas:  $A_{25\mu}$ ,  $A_{50\mu}$ ,  $A_{100\mu}$ ,  $A_{150\mu}$ ,  $A_{200\mu}$

Densities:  $D_{25\mu}$ ,  $D_{50\mu}$ ,  $D_{100\mu}$ ,  $D_{150\mu}$ ,  $D_{200\mu}$

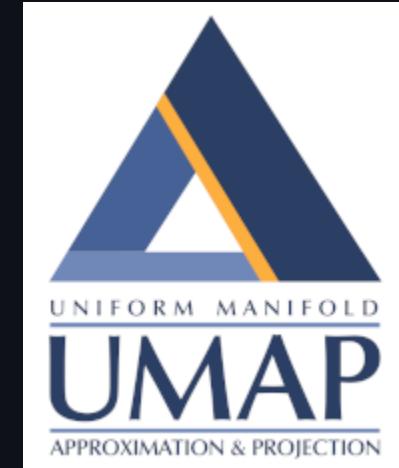
# Neighborhood Profiles



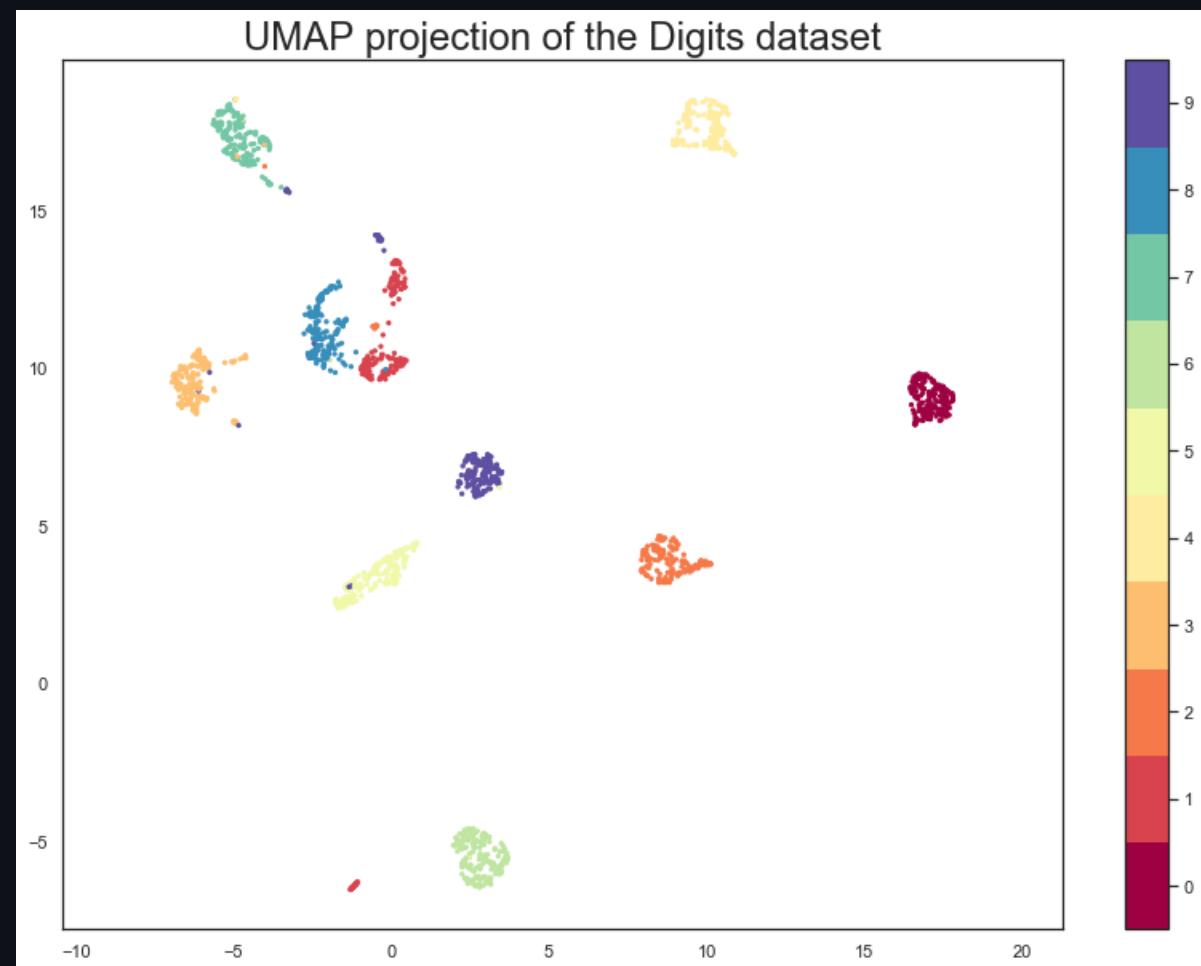
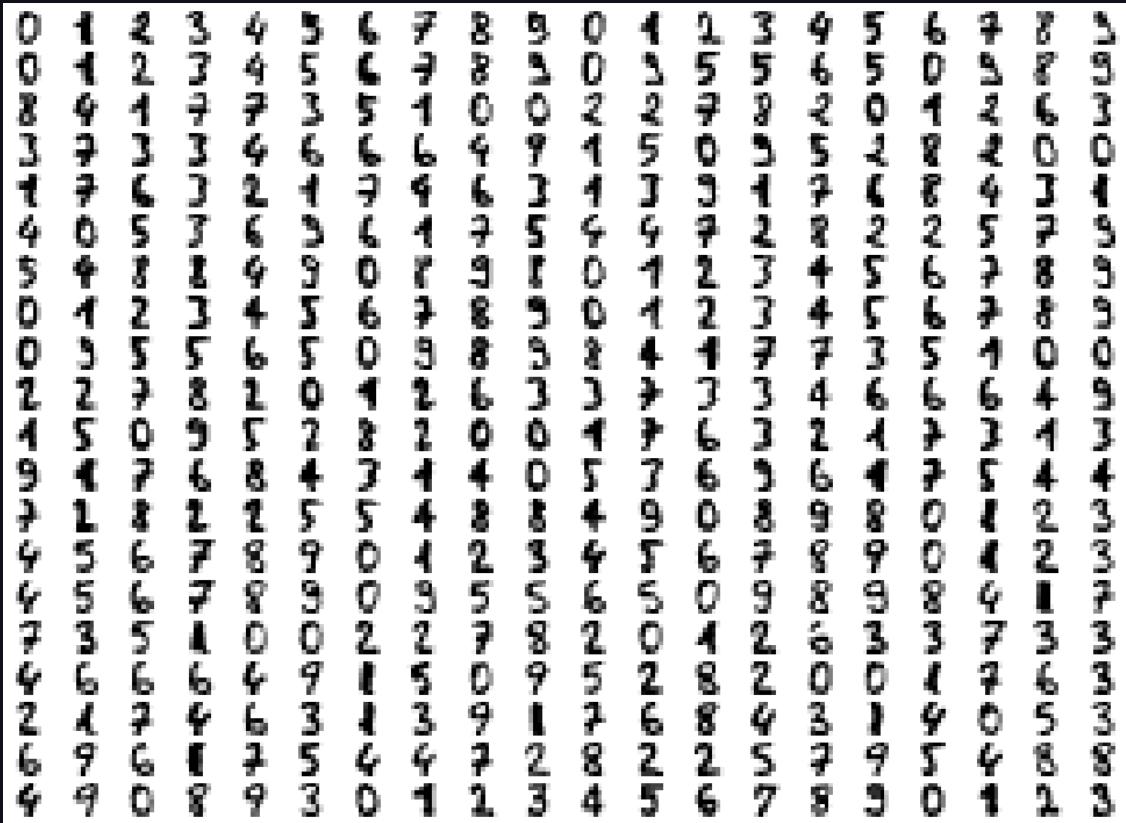
	Ph1 $D_{25\mu}$	Ph1 $D_{50\mu}$	Ph1 $D_{100\mu}$	Ph1 $D_{150\mu}$	Ph1 $D_{200\mu}$
Cell 1					
Cell 2					
Cell 3					
Cell 4					
....					
Cell N					

# UMAP

- Uniform Manifold Approximation & Projection
- Developed by Leland McInnes
- Dimensionality reduction technique
  - Like tSNE and PCA
- Reduces the number of features from  $n$  to 2
  - Accounting for the variance and error between observations

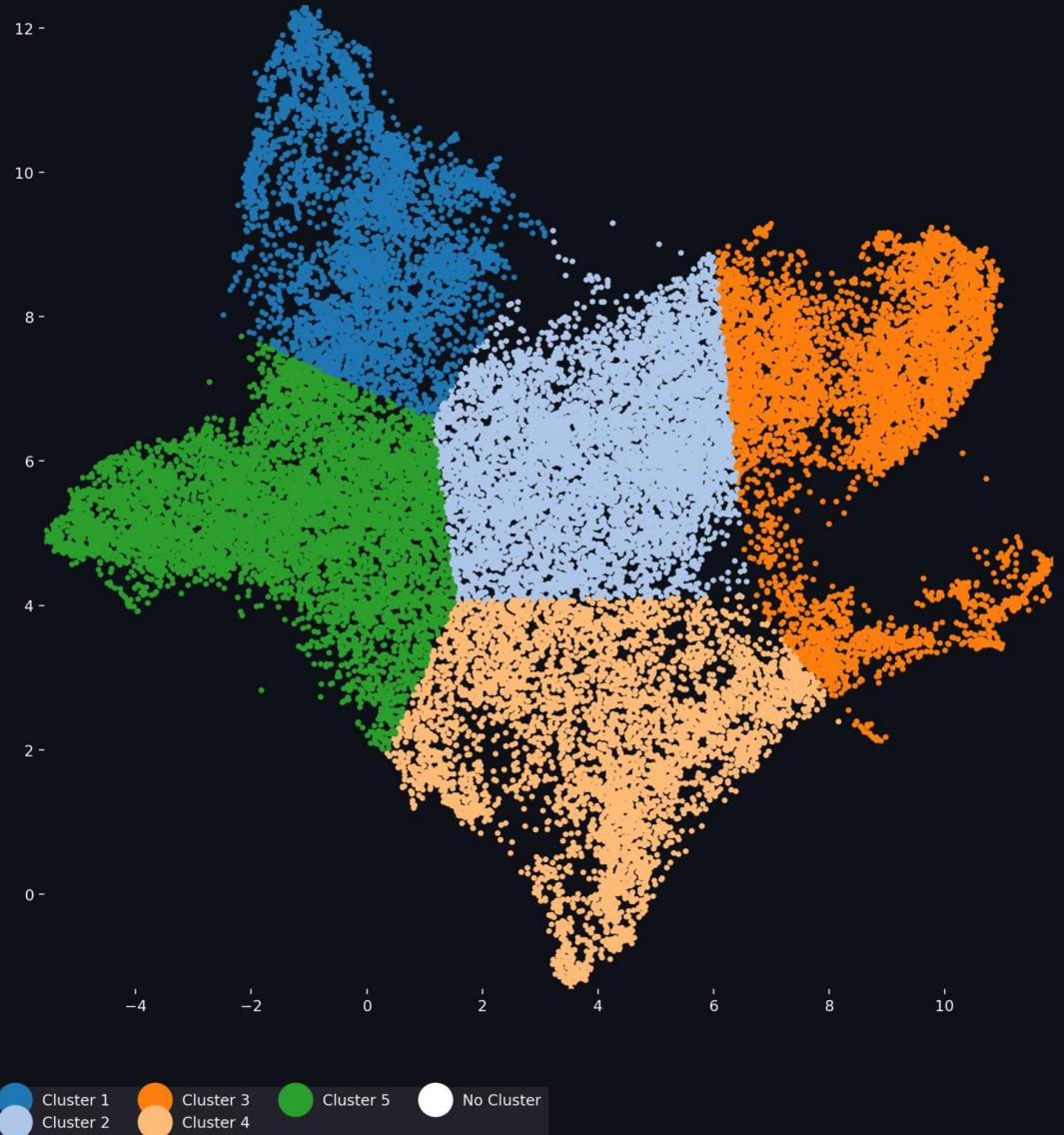
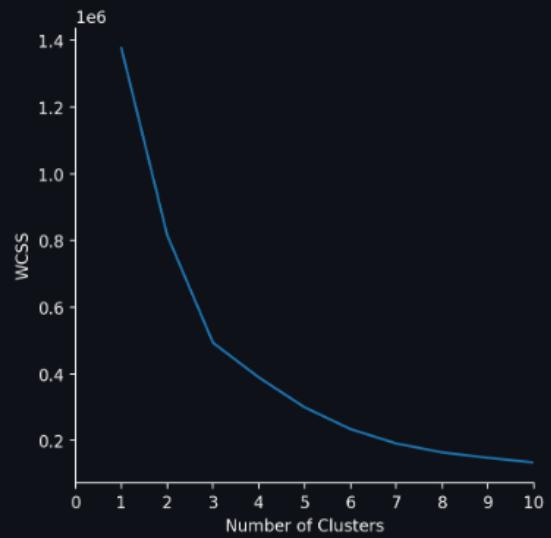


# UMAP



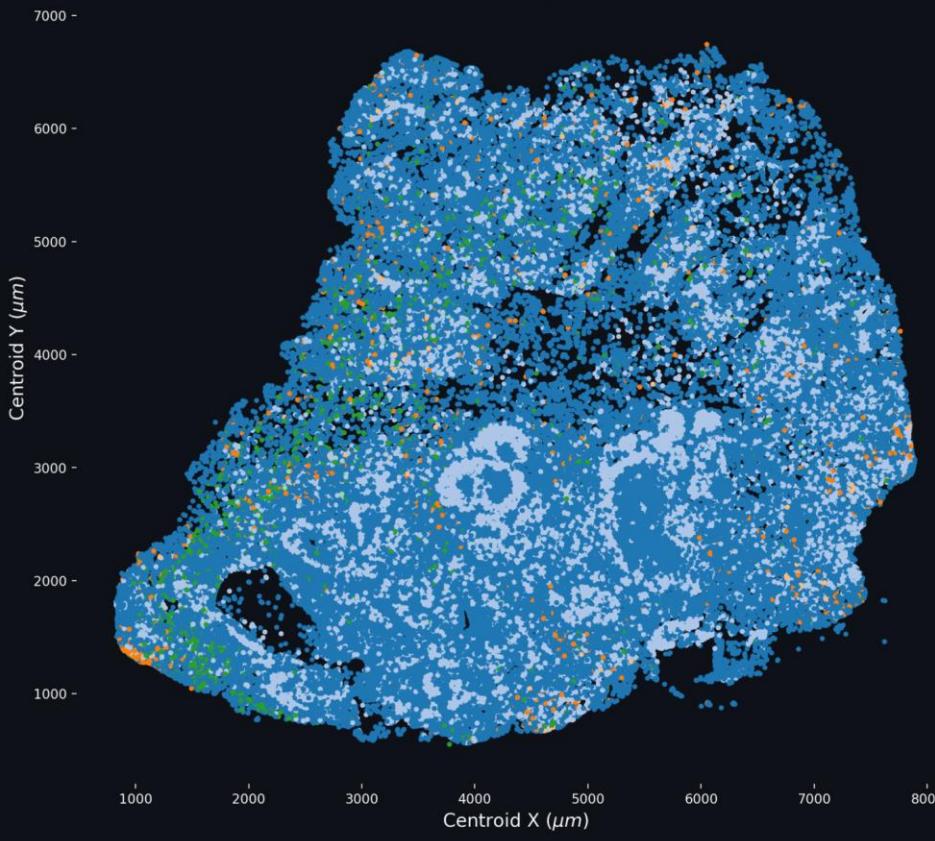
# Clustering

- k-nearest neighbors (kNN)
- Performed on 2D UMAP
- Choose a number of clusters
- Iterate to reduce error



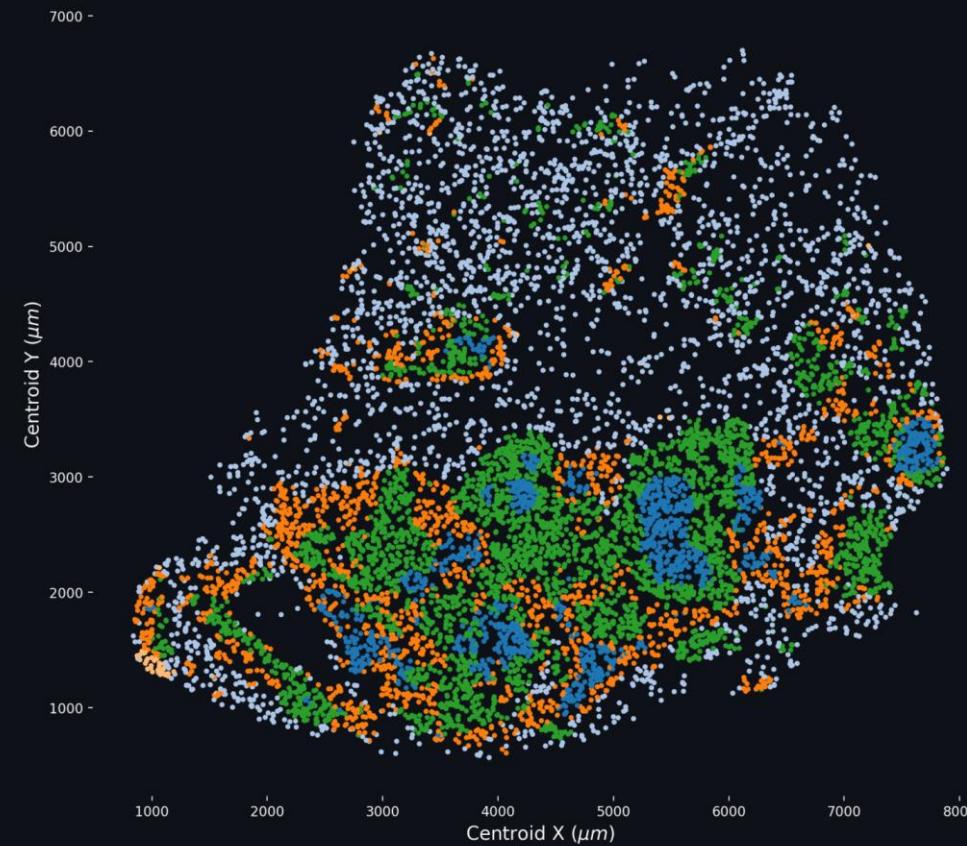
# Average Neighborhood Profiles

DATASET: from\_memory  
PHENO METHOD: Species  
SLIDE ID: 1A-Ultivue\_NOS2COX2\_11431.tif



● Other COX2+   ● CD8+ COX2+   ● NOS2+   ● NOS2+ CD8+

DATASET: from\_memory  
PHENO METHOD: Species  
SLIDE ID: 1A-Ultivue\_NOS2COX2\_11431.tif

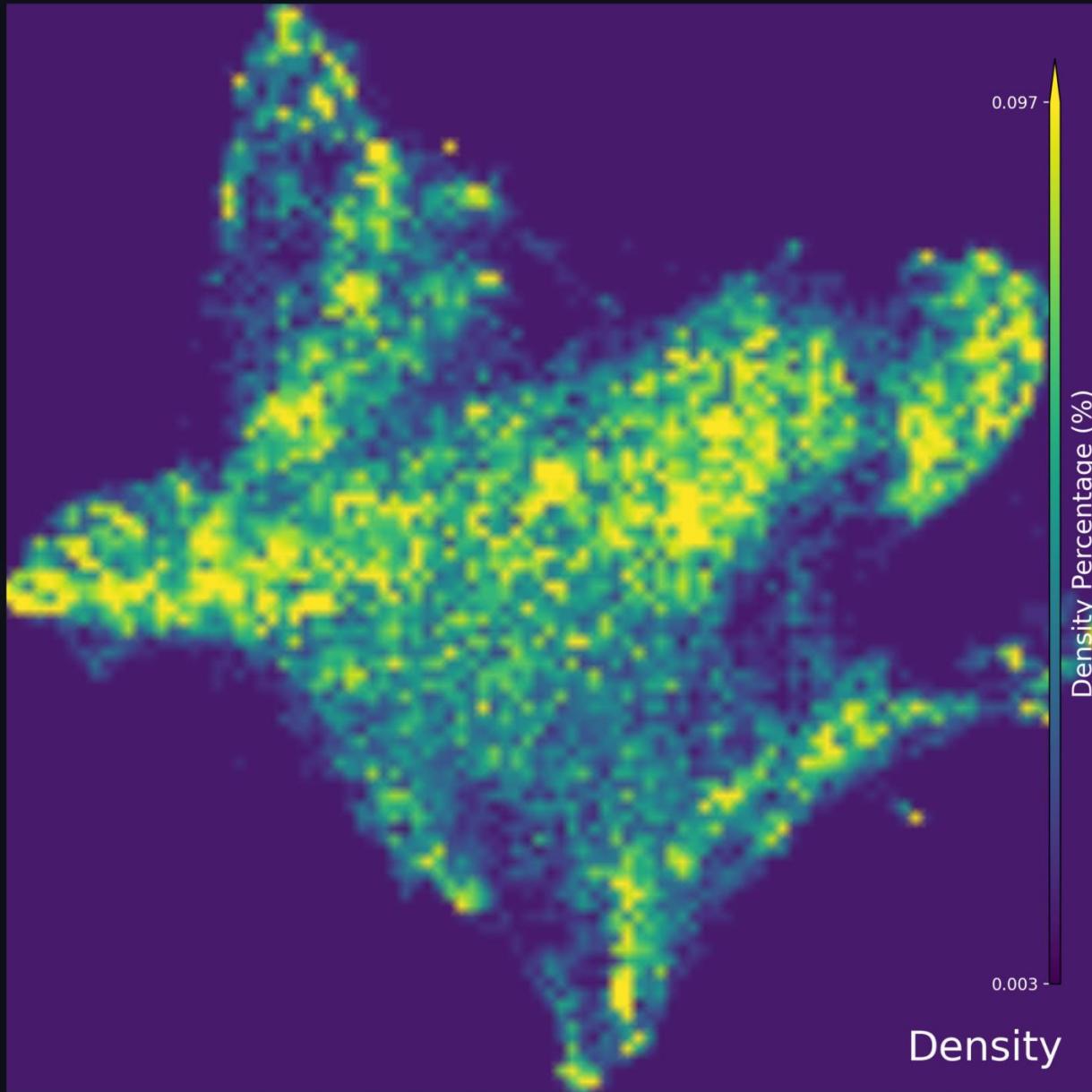


● Cluster 1 Cluster 2   ● Cluster 3 Cluster 4   ● Cluster 5 Cluster 4   ● No Cluster

# Neighborhood Profiles

- Can we identify regions of cell phenotype combinations that are similar across tissue samples? Can we start to use them as a biomarker?
- Do these average neighborhood profiles appear (or not appear) more frequently with desirable (or undesirable) outcomes?

# Demo



# Acknowledgements

MAWA Team

- Andrew Weisman
- Andrei Bombin

NCI

- Janelle Cortner

Axle Informatics

OMAL

- Will Heinz
- David Wink

Palantir and NIDAP Team

# Questions?