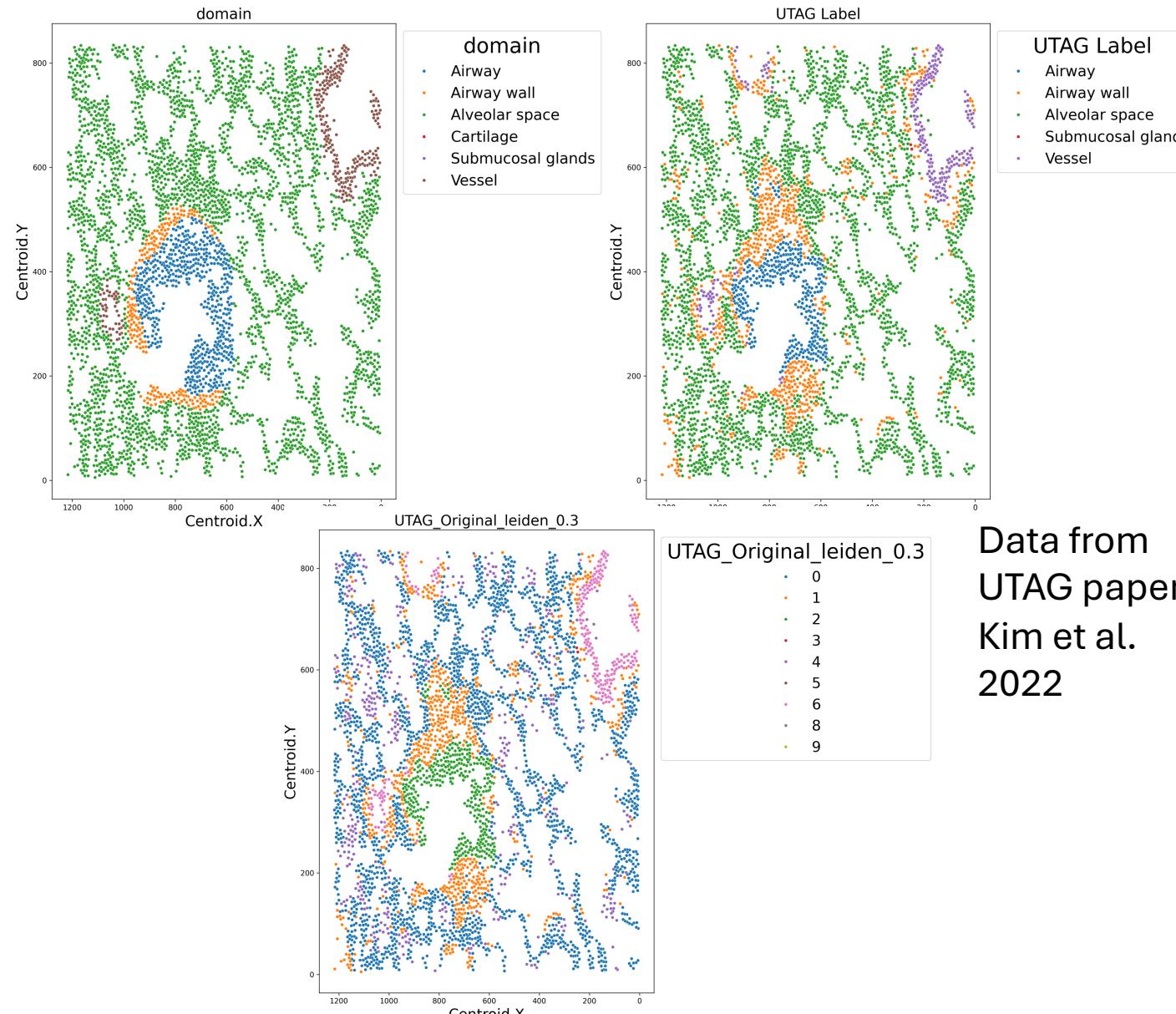


Unsupervised Clustering

Unsupervised Clustering Provides Multiple Advantages

- Multiplexed imaging and spatial transcriptomics allow direct observation of cellular phenotypes.
- Manual annotations are labor intensive and requires expertise of specialized pathologists.
- Unsupervised Clustering: A method used to group cells based on their markers expression/intensity profiles without any prior knowledge of cell types or labels.
- Advantages include scalability, reduced bias, and ability to discover new/rare cell types



Data from
UTAG paper,
Kim et al.
2022

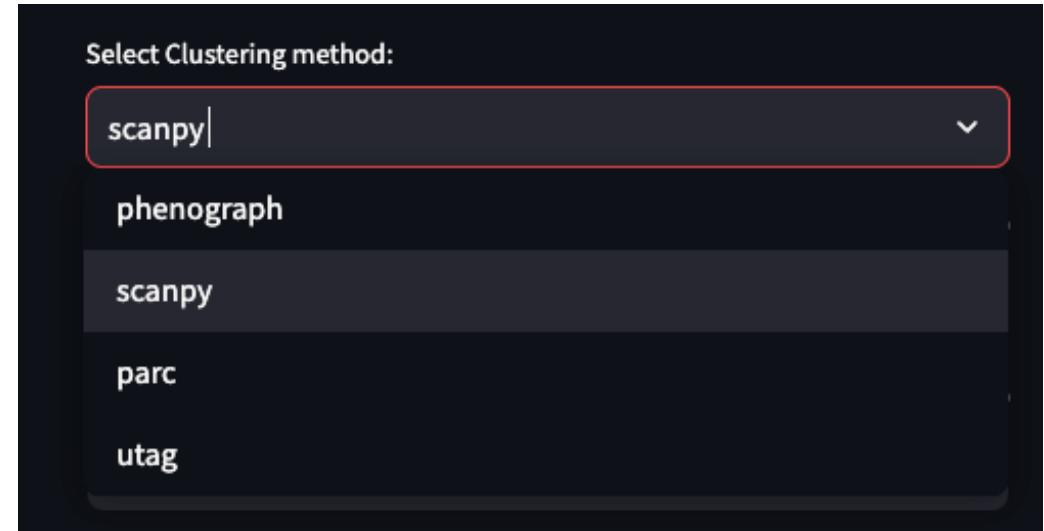
Main Steps in Unsupervised Clustering

- Data Preprocessing:
Normalization, PCA
- Build k-nearest neighbors' graph
- Use community detection
algorithm to identify clusters

The screenshot shows a user interface for unsupervised clustering. At the top, there are two rows of red buttons, each containing a combination of 'Ultivue' and a marker name (e.g., 'DAPI Nu...', 'CD8 Cyt...'). Below these are sections for 'Select columns for metadata:' and 'Select Clustering method:'. Under 'Select Clustering method:', 'scipy' is selected from a dropdown menu. The interface includes several input fields with numerical values and dropdown menus for parameters like 'Number of Principal Components' (set to 10), 'K Nearest Neighbors' (set to 10), 'Clustering resolution' (set to 0.1), 'n_jobs' (set to 7), 'n_iterations' (set to 5), and 'Distance metric:' (set to 'euclidean'). There are also checkboxes for 'Z-score normalize columns' (checked) and 'Fast' (unchecked). A large 'Run Clustering' button at the bottom is highlighted with a red border.

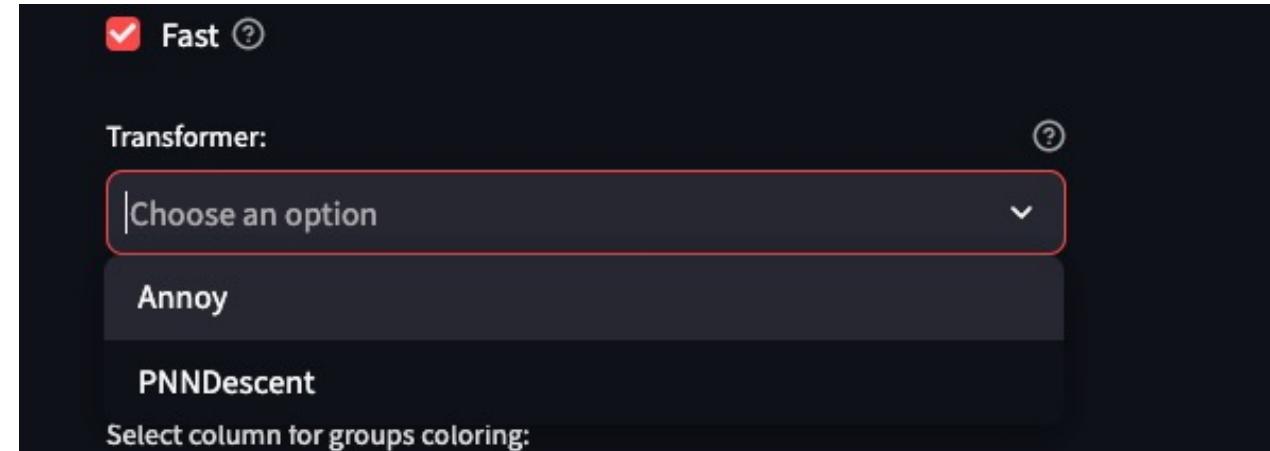
Four Unsupervised Clustering Methods

- Scanpy: simple, widely used in scRNA-seq, uses UMAP to assign weights to the graph
- Phenograph: well established, uses Jaccard similarity to assign weights, known for ability to identify rare communities
- PARC, fast knn graph building, performs local and global pruning of weak edges, tested for scRNA and imaging data
- UTAG, same as Scanpy but also accounts for spatial relationships between cells in the physical space



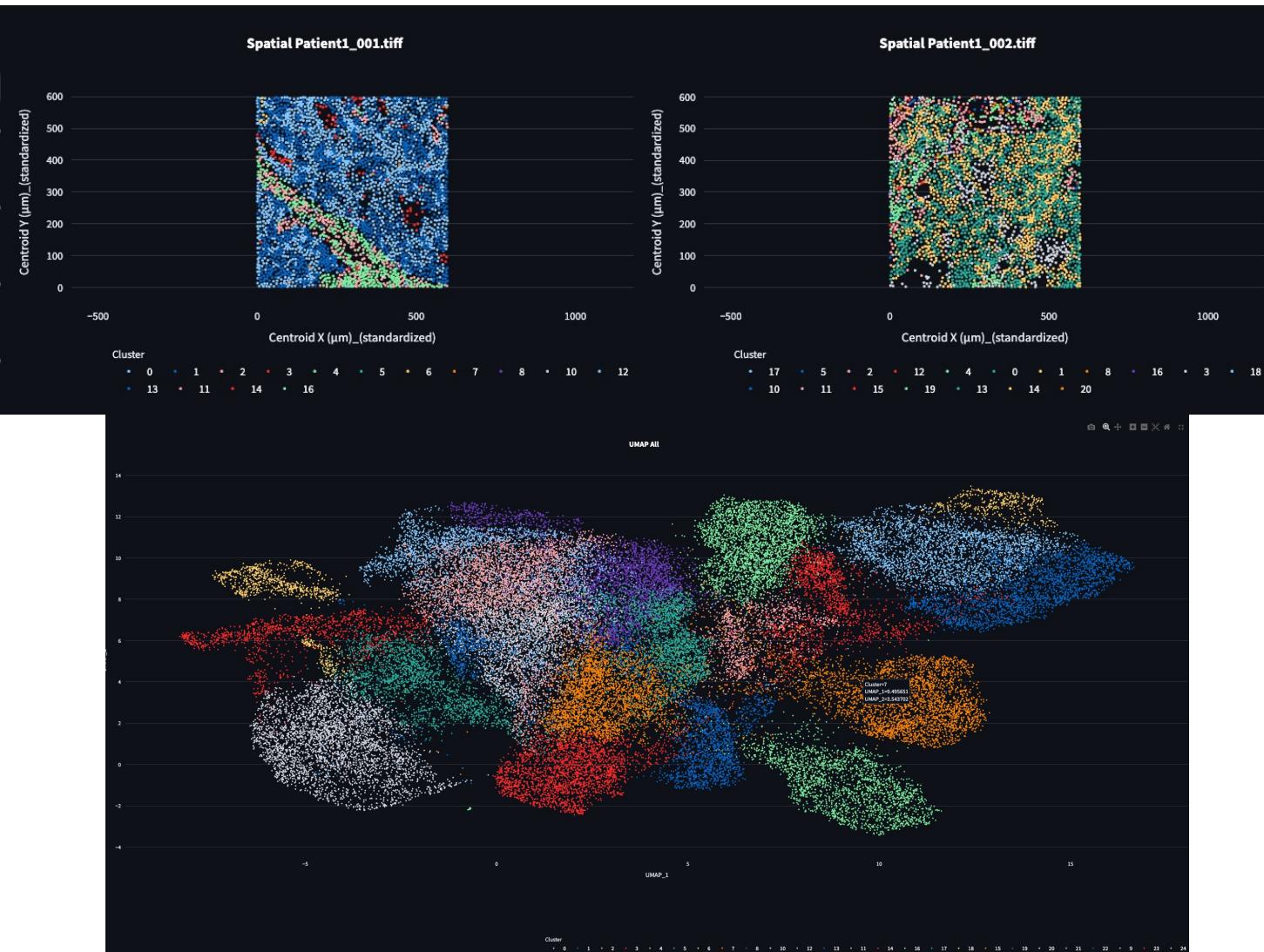
Implemented Fast Clustering With

- Each method has faster version
- Approximate nearest neighbor search
- iGraph's implementation of Leiden community detection method



Clustering Visualizations

- Spatial Plots: Visualize clusters in a physical space
- UMAP: Dimensionality reduction technique that aims to preserve the global structure of data while maintaining local relationships.



Differential Intensity

- Identifies markers that are significantly associated with each cluster
- Evaluates intensity/expression of each marker in the cluster vs the rest of the cell with Wilcoxon's sum rank test

Differential Intensity

Select column for differential intensity: Cluster

Select group for differential intensity table: 0 × 2 × 4 ×

Only Positive Markers

Run Differential Intensity

	group	names	scores	logfoldchanges	pvals	pvals_adj
12	0	CarbonicAnhydrase	35.8829	0.2828	5.7e-282	1.5e-281
13	0	HLADR	31.6765	6.9122	3.3e-220	7.3e-220
14	0	CD40	30.3752	0.3085	1.2e-202	2.5e-202
15	0	CD7	29.7734	2.3042	8.6e-195	1.7e-194
16	0	CD4	28.219	1.1389	3.4e-175	6.5e-175
17	0	GrzB	25.6736	0.5542	2.3e-145	4.2e-145
18	0	LAG3 / LAG33	24.3169	0.2845	1.3e-130	2.3e-130
19	0	ICOS	21.3963	0.6782	1.4e-101	2.4e-101
20	0	B2M	21.0393	0.7523	2.9e-98	4.6e-98
21	0	FOXP3	12.7645	0.6019	2.6e-37	3.5e-37
22	0	DNCF04	20.6449	0.6026	4.7e-172	6.2e-172

Differential Intensity/Expression Visualizations

- Rank Plots
- Heatmap
- Intensity UMAP

