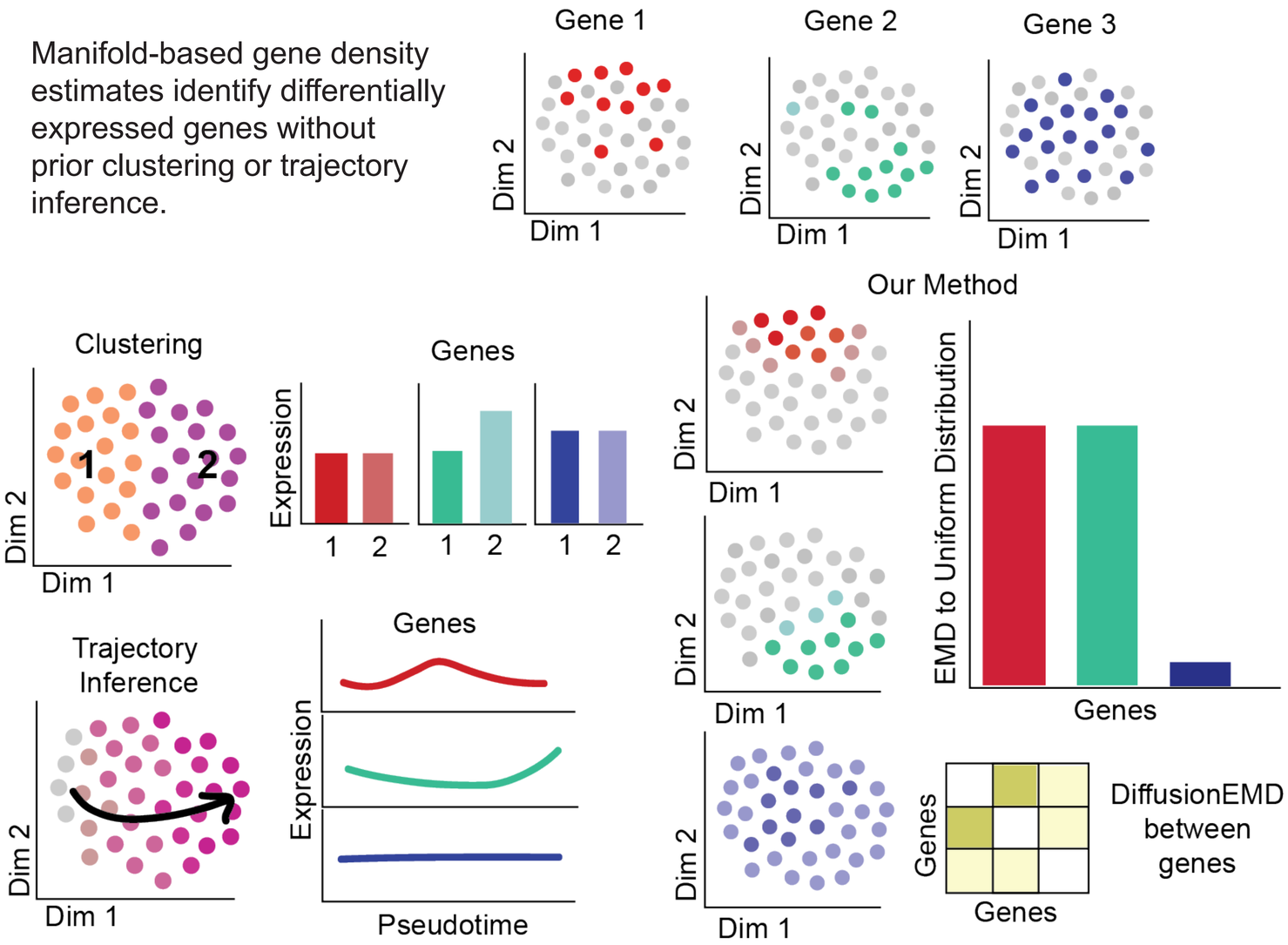
Manifold-based gene density estimates reveal immune signaling in meningioma

Aarthi Venkat[1], Danielle Miyagishima[2,3], Alexander Tong[4], Murat Günel[1,2,3], Smita Krishnaswamy[1,3,4]

[1]Computational Biology and Bioinformatics Program, Yale University [2]Department of Neurosurgery, Yale School of Medicine [3]Department of Genetics, Yale School of Medicine
[4]Department of Computer Science, Yale University

Manifold-based gene density estimates identify differentially expressed genes without prior clustering or trajectory inference.



Algorithm details

1. A cell similarity graph is built over data, where each node is a cell and graph edges connect cells with similar gene expression values.

2. Gene expression *x* for each cell is used to create the gene *x* expression probability distribution signal.

3. Each signal is then smoothed over the graph using a generalized form of kernel density estimation to manifolds[1].

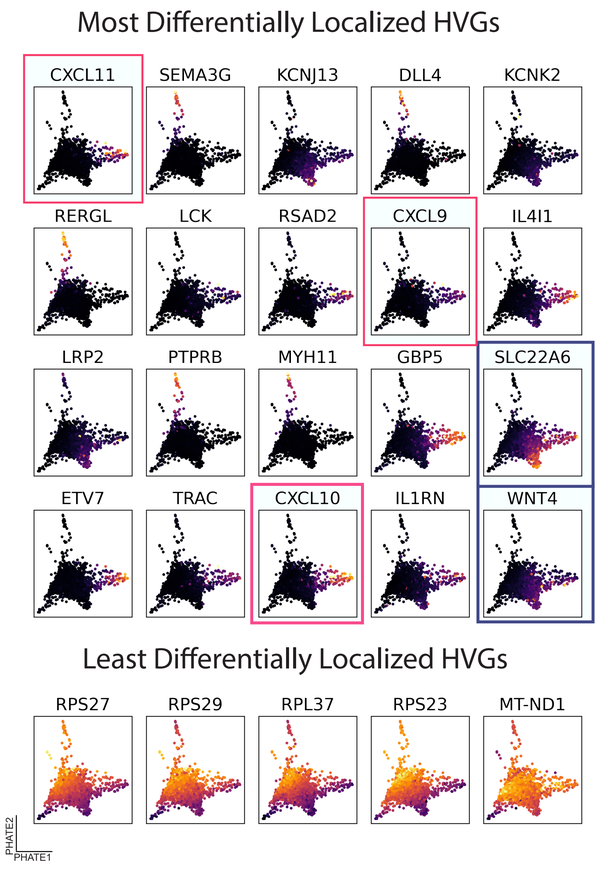
4. EMD (Earth Mover's Distance) is computed between each gene density distribution and the uniform distribution.

5. Genes with the highest EMD to the uniform distribution are considered least uniformly expressed, or most differentailly localized along the manifold.

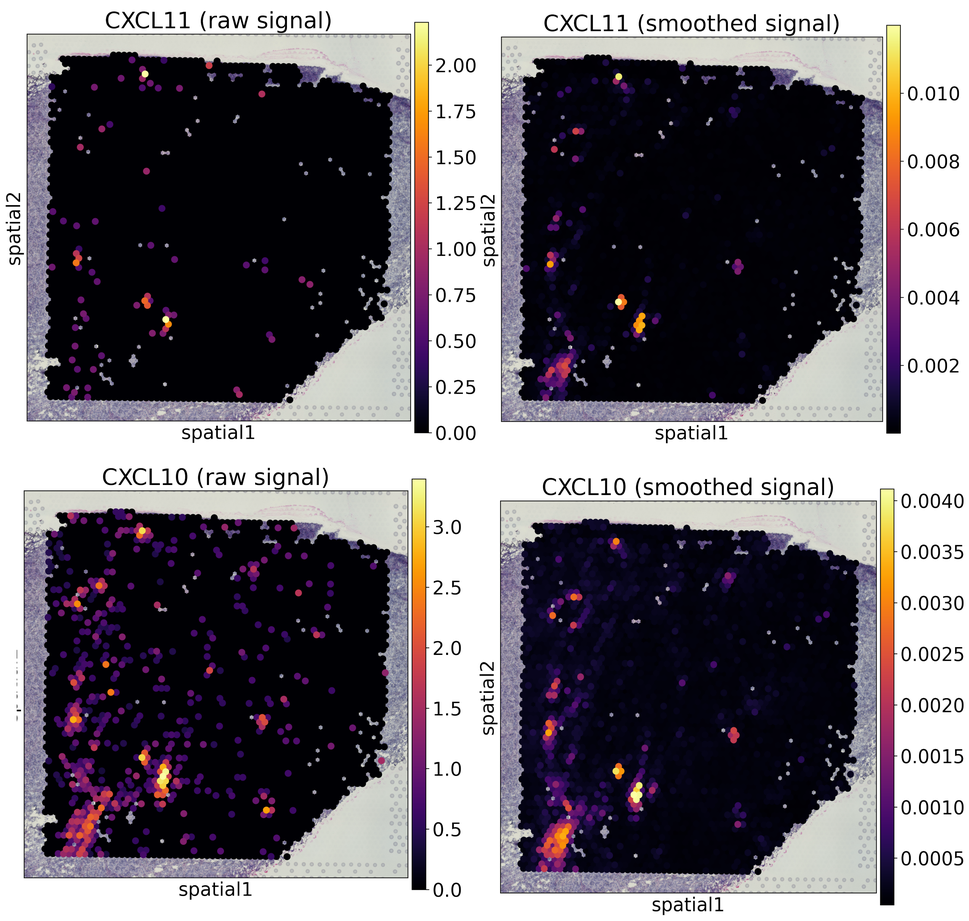
6. DiffusionEMD[2] between genes that are localized to same regions of the manifold is low, enabling gene set analysis and cell state labeling independent of clustering or trajectory inference.

7. Smoothed density estimates enable improved visualization and generalized biological discovery.

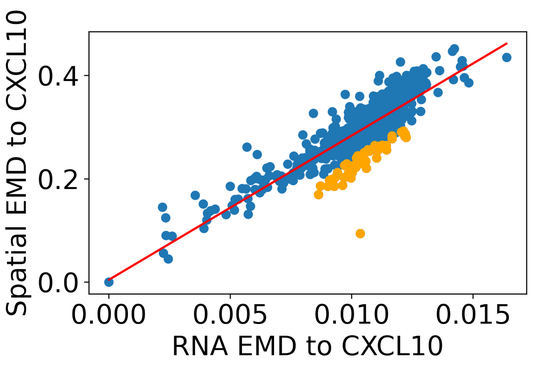
I. Method identifies immune signature in NF2-mut grade 1 meningioma tumor, corroborating link between NF2 loss and immune infiltration.



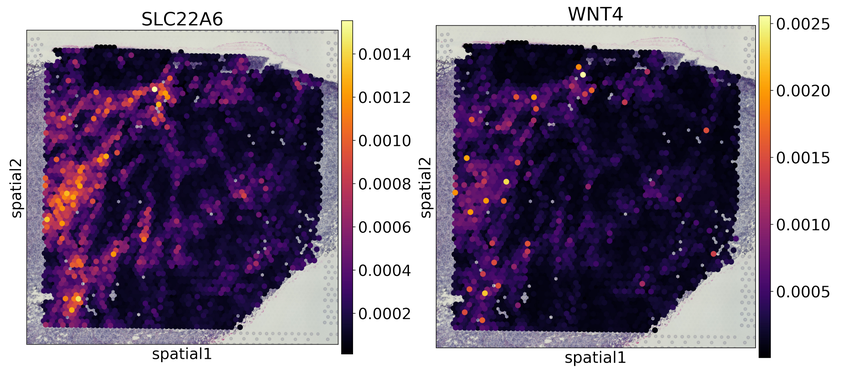
II. Density estimates recover and smooth immune signal in spatial data visualization.



III. DiffusionEMD between CXCL10 distribution and all genes' over scRNA graph vs. spatial graph identifies SLC-high cells as closer to the immune population in spatial graph than in RNA manifold.



This discovers these two cell states as putative communicating populations in a cluster-independent fashion.



REFERENCES

1. MELD: Burkhardt, D.B., Stanley, J.S., Tong, A. et al. Quantifying the effect of experimental perturbations at single-cell resolution. Nat Biotechnol (2021). https://doi.org/10.1038/s41587-020-00803-5

2. DiffusionEMD: Tong, A. et al. Diffusion Earth Mover's Distance and Distribution Embeddings. https://arxiv.org/abs/2102.12833

3. PHATE: Moon, K. R. et al. Visualizing Transitions and Structure for Biological Data Exploration. bioRxiv 120378 (2018). doi:10.1101/120378

github.com/KrishnaswamyLab

This work is supported in part by the Gruber Foundation