Application of LIGER to integration of seqFISH and scRNA-seq

By Joshua Sodicoff and Dr. Joshua Welch

Department of Computational Medicine and Bioinformatics

University of Michigan - Ann Arbor

MEDICAL SCHOOL
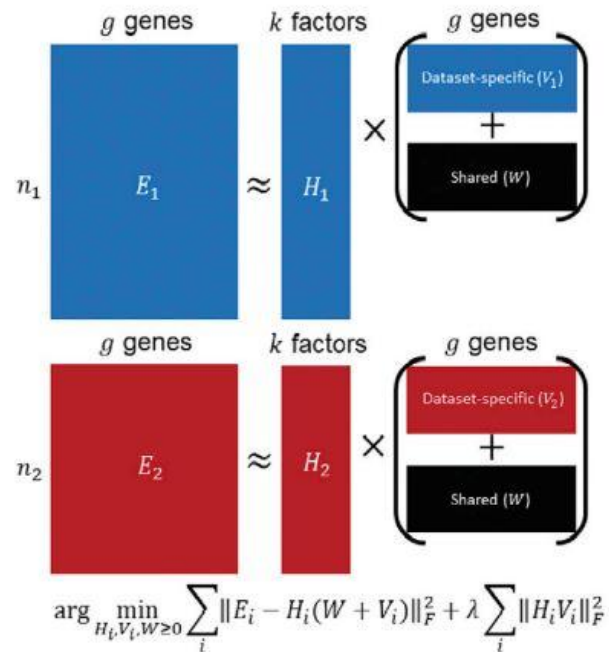UNIVERSITY OF MICHIGAN

Outline

- Overview of LIGER Approach
- Previous Spatial Integration Results
- Comparison of Integration from Different Spatial Transcriptomic Protocols

LIGER Integrates Diverse Single-Cell Datasets

Linked Inference of Genomic Experimental Relationships

- LIGER uses integrative NMF to jointly learns a low-dimensional space
  - Clustering with max factor assignment
  - Quantile normalization
  - Optional Louvain clustering on aligned space
- Allows for the integration of multiple datasets, from
  - Different samples
  - Different species
  - **Different modalities**

$$\arg\min_{H_i, V_i, W \geq 0} \sum_i \|E_i - H_i(W + V_i)\|_F^2 + \lambda \sum_i \|H_i V_i\|_F^2$$

Welch et al., *Cell*, 2019

Welch et al. (2019)

Outline

- Overview of LIGER Approach
- Previous Spatial Integration Results
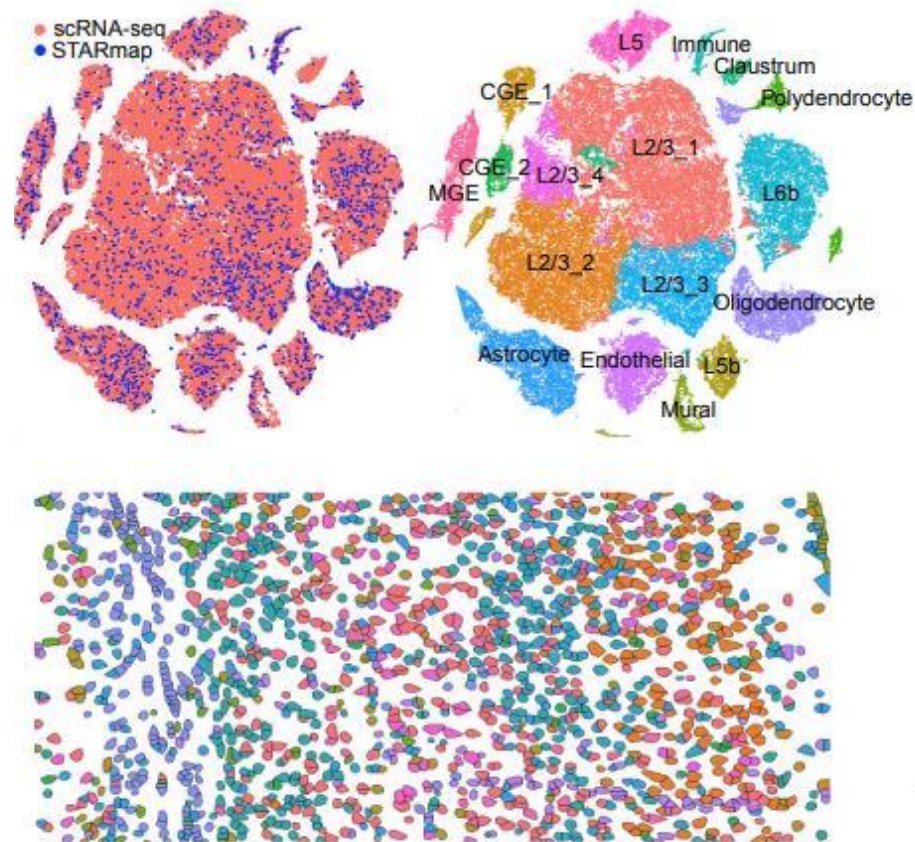- Comparison of Integration from Different Spatial Transcriptomic Protocols

Why Integrate scRNA-seq and Spatial Transcriptomic Data?

Spatial Transcriptomics	scRNA-seq
Usually few cells	Usually many cells
Only selected genes	All genes
In situ measurements	Tissue dissociation required
Spatial coordinates known	Spatial coordinates lost during dissociation
No dissociation bias	Dissociation may bias cell type proportions

- Integration with scRNA-seq better resolves cell subtypes
- Spatial transcriptomic data allows imputation of spatial trends and spatial cell type distributions

Previous Work: LIGER Integrates STARmap and scRNA-seq

- Integrated scRNA-seq (71000 cells) and STARmap (2000 cells) from mouse cortex
- Strong alignment between datasets
- Expression of known cell type markers confirmed accurate joint clustering
- Increased resolution for detecting clusters compared to STARmap alone

Welch et al. (2019)

Previous Work: LIGER Imputes Spatial Gene Expression

- Averaging of closest scRNA-seq samples imputes spatial distribution of genes not measured in STARmap
- Confirmed accuracy by holding out genes and comparing with Allen Brain Atlas
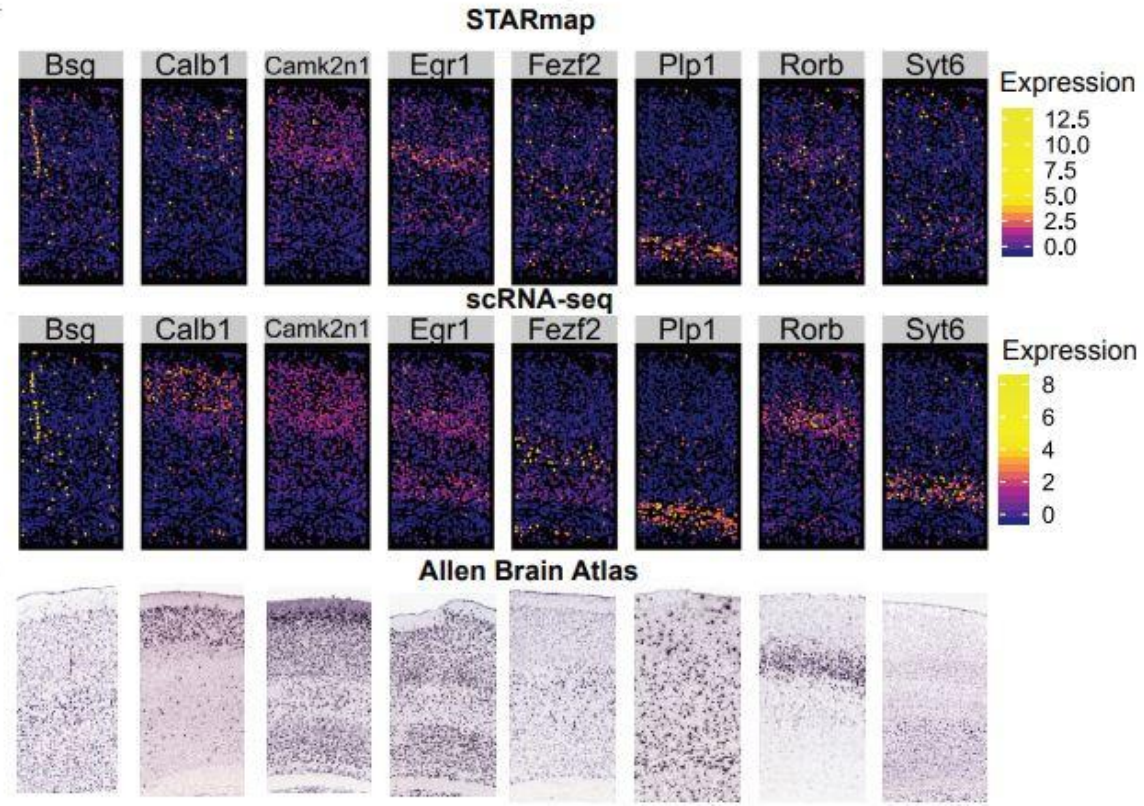
Welch et al. (2019)

Outline

- Overview of LIGER Approach
- Previous Spatial Integration Results
- Comparison of Integration Results from Different Spatial Transcriptomic Protocols

Guiding Questions

1) How do integration results differ by spatial transcriptomic protocol?
2) How does number of shared genes affect results?

Sequential FISH Protocol

- One of the earliest multiplexed FISH approaches
- Sequential rounds of barcoding hybridizations allow measurement of multiple genes using FISH
- Recent versions scale to thousands of genes

SeqFISH
Shah et al. (2016)
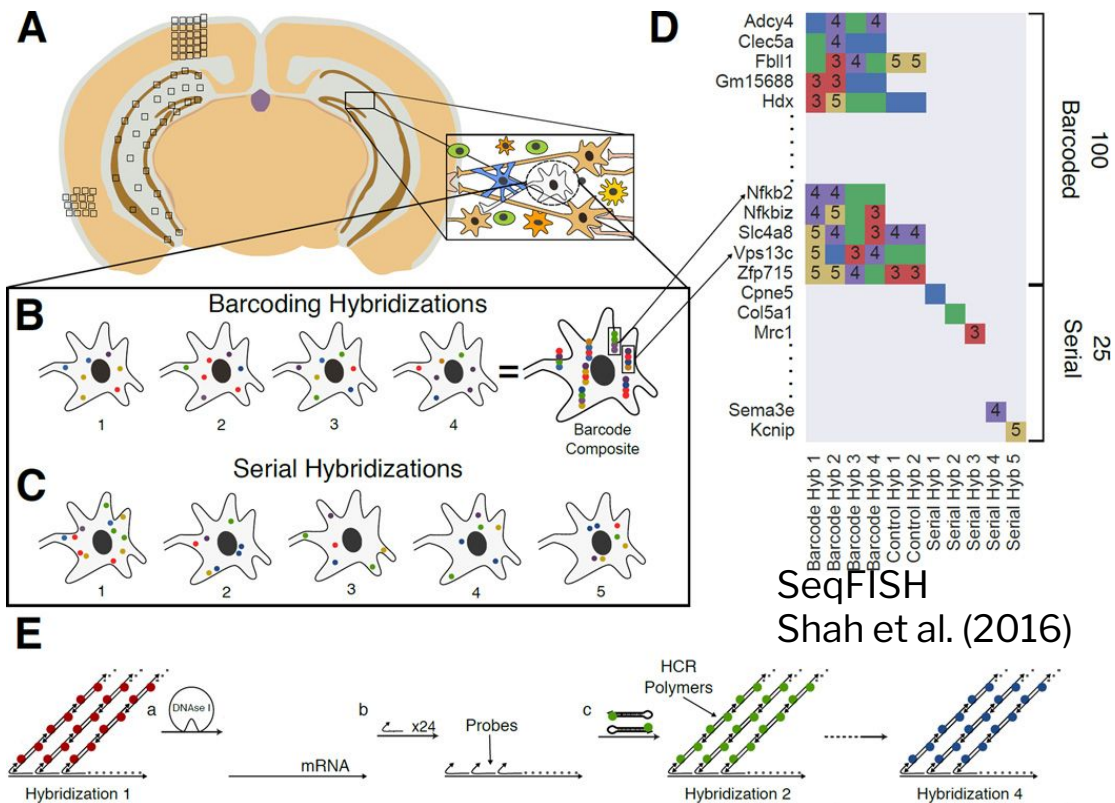
MERFISH Protocol

- Multiplexed error-robust FISH
- Uses error-correcting binary code to identify each gene
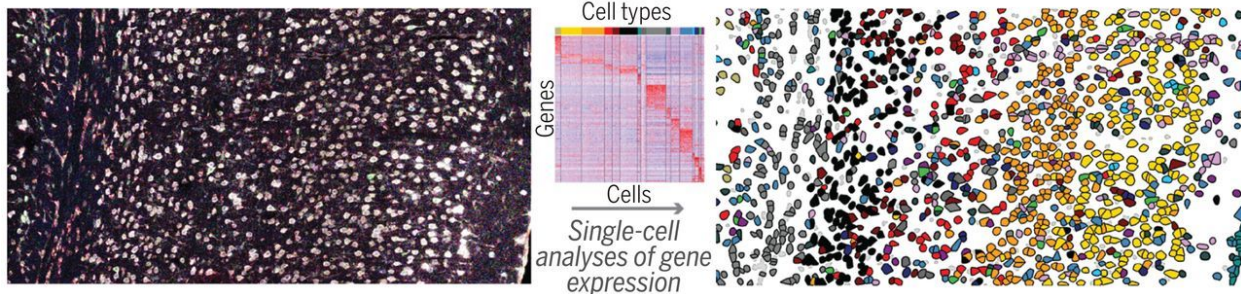- Recent versions scale to thousands of genes
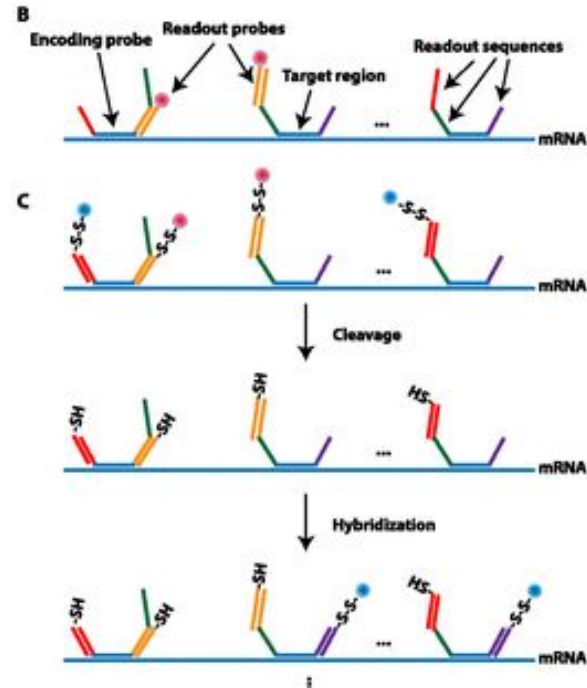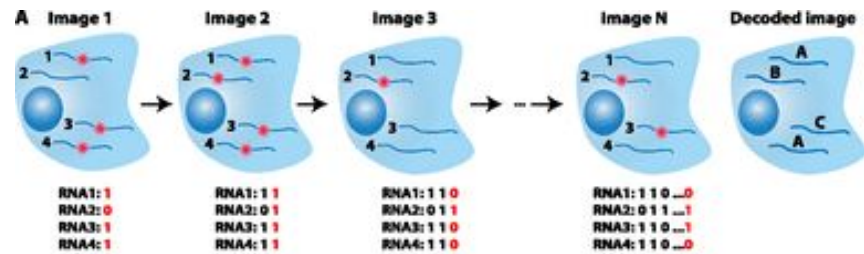
MERFISH Moffitt et al. (2016)

STARmap Protocol

- Converts tissue into hydrogel
- Rolling circle amplification using SNAIL probes
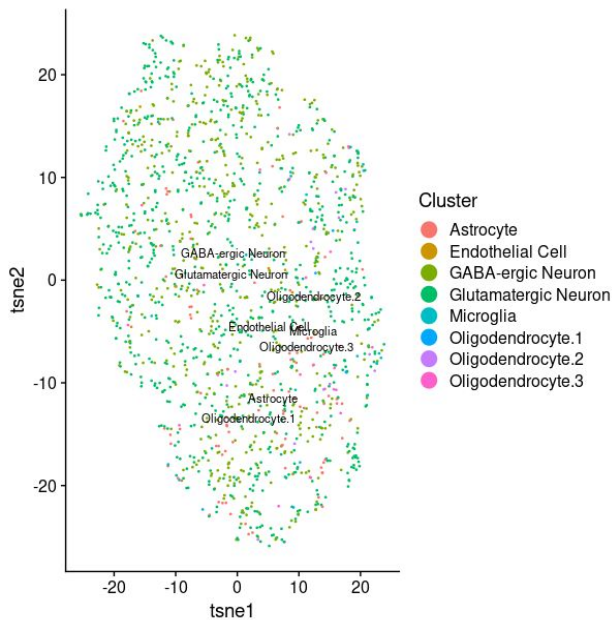- Allows measurement of 3D tissue volumes but limited to ~1000 genes
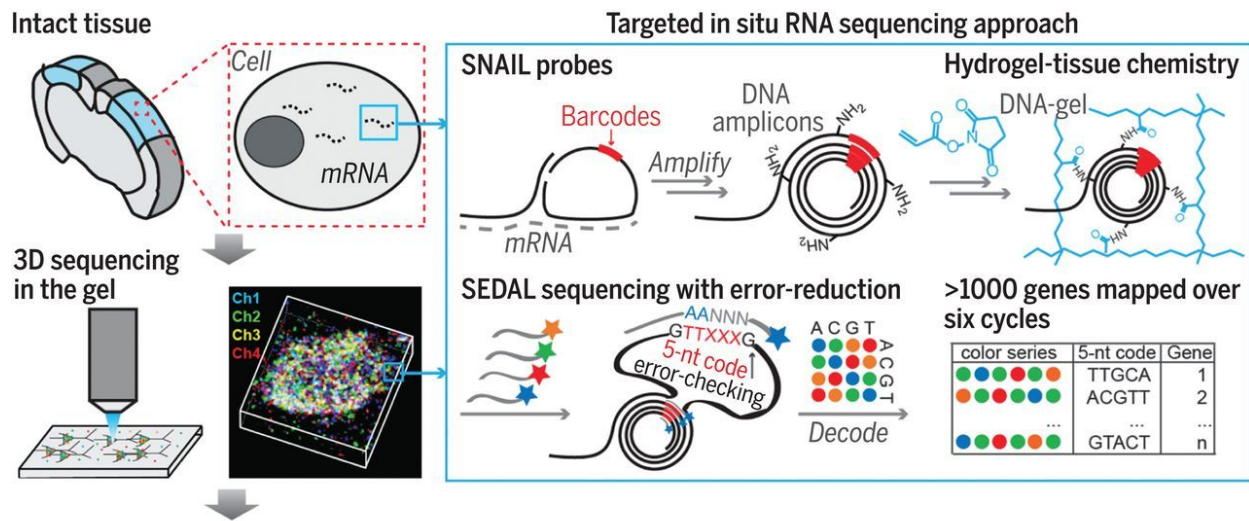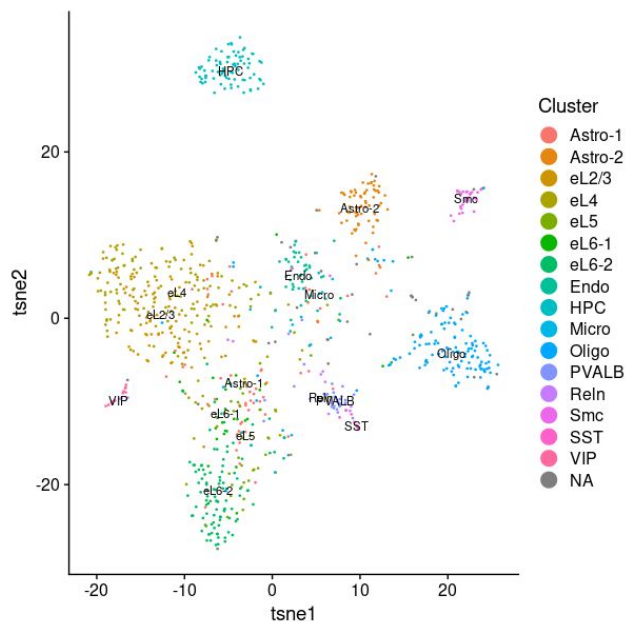

STARmap Wang and Allen et al. (2018)
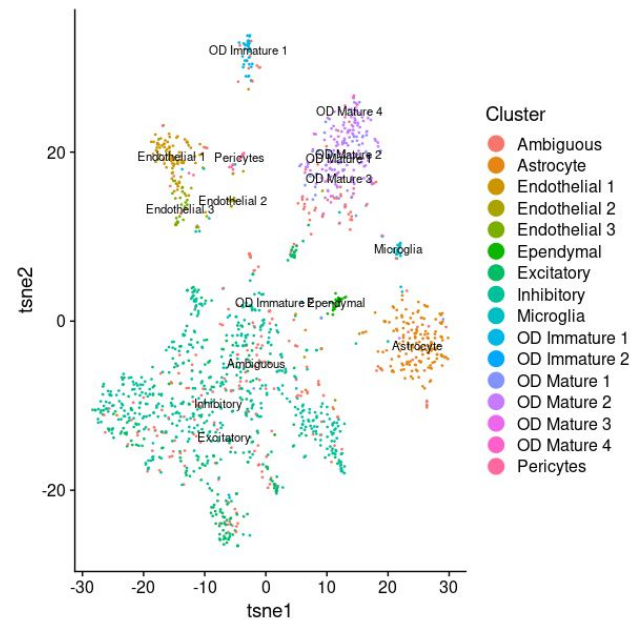
Comparing Spatial Transcriptomic Protocols

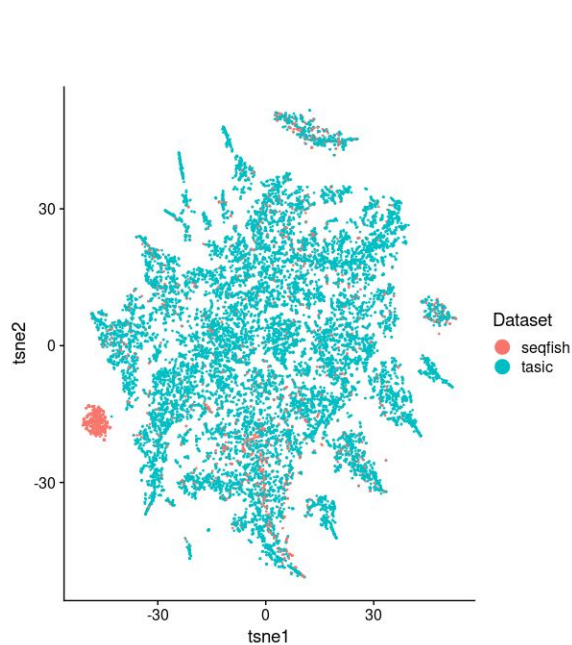SeqFISH - mouse cortex   STARmap - mouse cortex   MERFISH - mouse hypothalamus

Spatial Transcriptomic + scRNA-seq Integration Strategy

- Completed integrative analysis of given seqFISH data with ViSP subset from Tasic et al. (2018)
- Followed standard LIGER workflow
  - Preprocessing & variable gene selection %>% iNMF %>% quantile norm %>% clustering
  - 14662 scRNA-seq samples with ~43,000 genes and 1597 seqFISH samples with 113 genes
  - 111/113 genes used as features
- Guiding questions
1) How do integration results differ by spatial transcriptomic protocol?
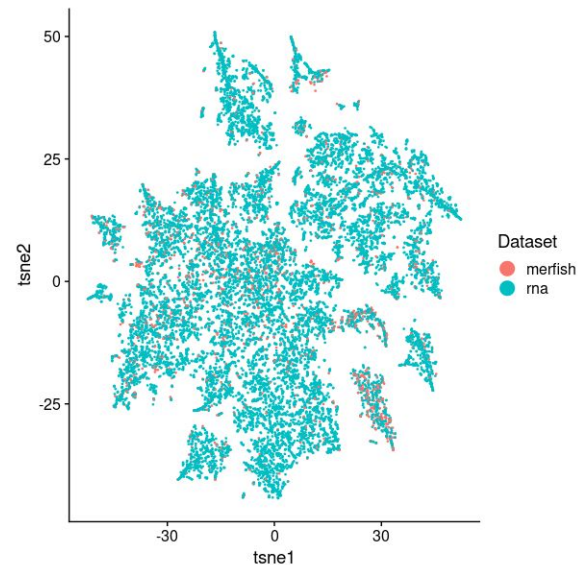2) How does number of genes affect results?

Comparing Integration Results Across Protocols
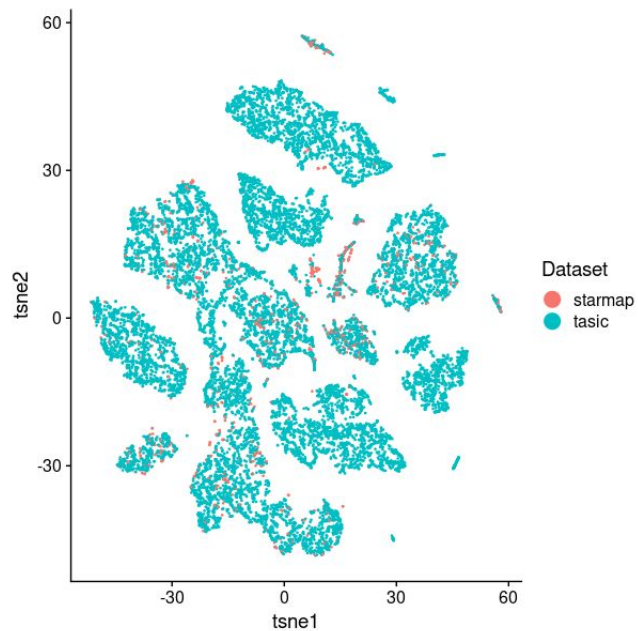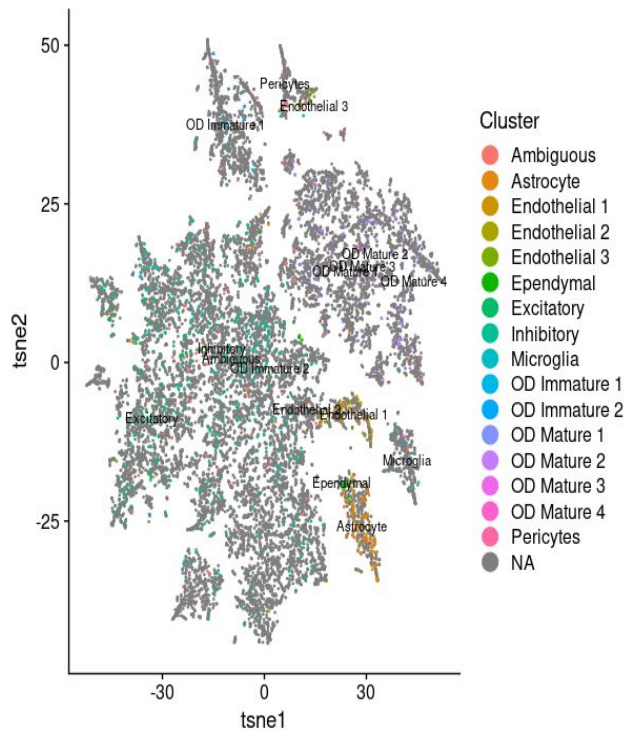

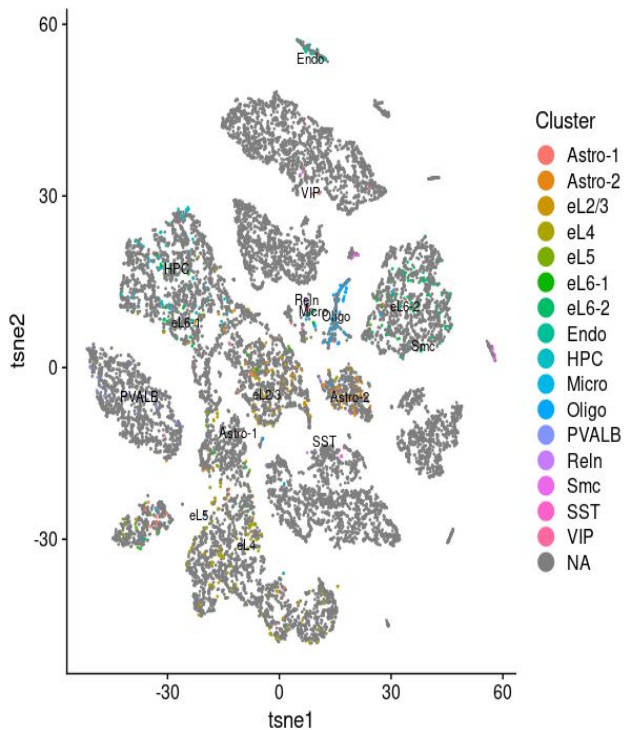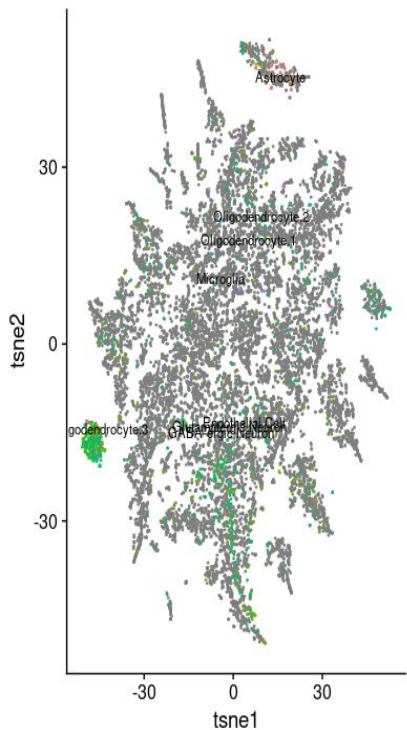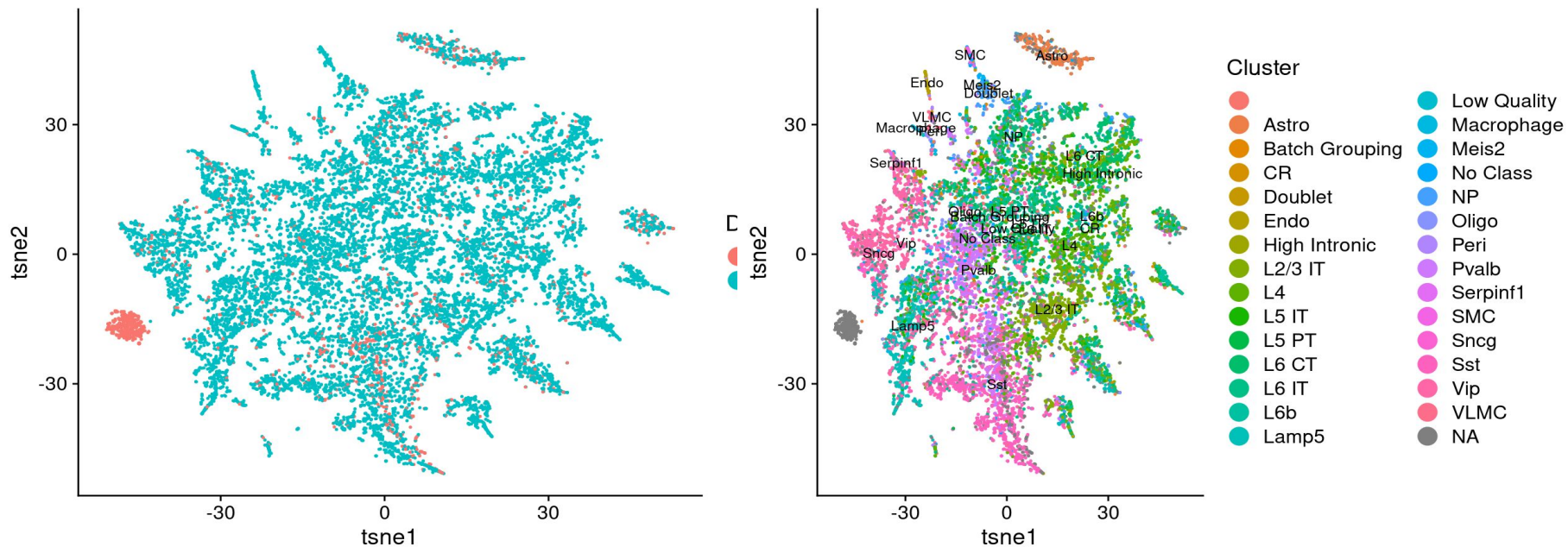SeqFISH - mouse cortex   STARmap - mouse cortex   MERFISH - mouse hypothalamus

Resolution was improved!
- More, better defined clusters
- Highly aligned integrations allow for imputation of spatial distribution

Suggests that more genes (1020 in STARmap) results in more informative integration

seqFISH   STARmap   MERFISH
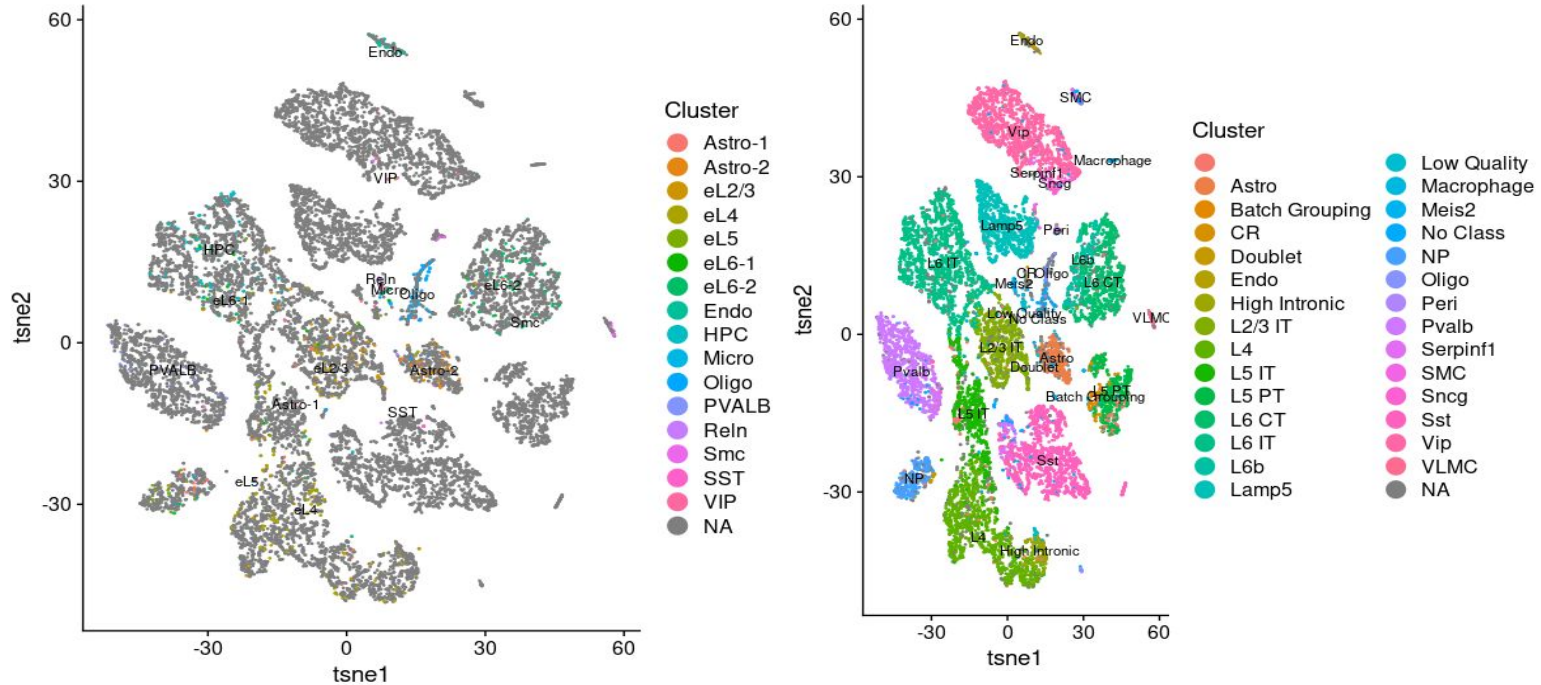
Comparing Integration Results - seqFISH

- Most seqFISH cells do not align at all with scRNA data
- Some loss of structure in scRNA data after integration

Comparing Integration Results - STARmap

- ~100 genes is on the lower end, so how many genes are needed?
- Tried repeatedly downsampling STARmap dataset with 1020 genes to determine at what number of genes recorded for the spatial data the integration breaks down
- Depicted here is an analysis with 1020 genes, for which the STARmap and scRNA-seq given clusterings agree, meaning the integration is informative
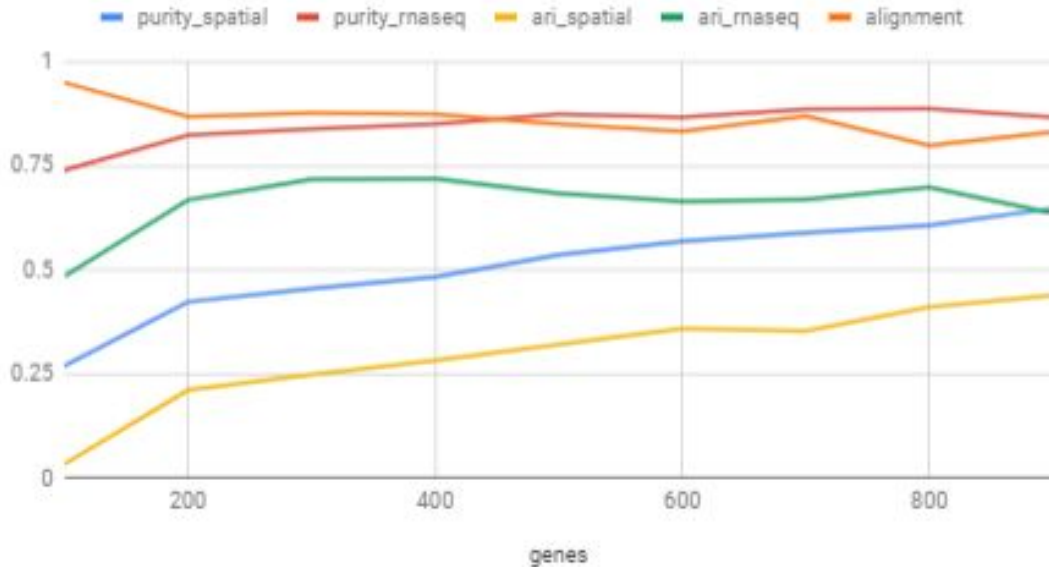
Comparing Integration Results - MERFISH
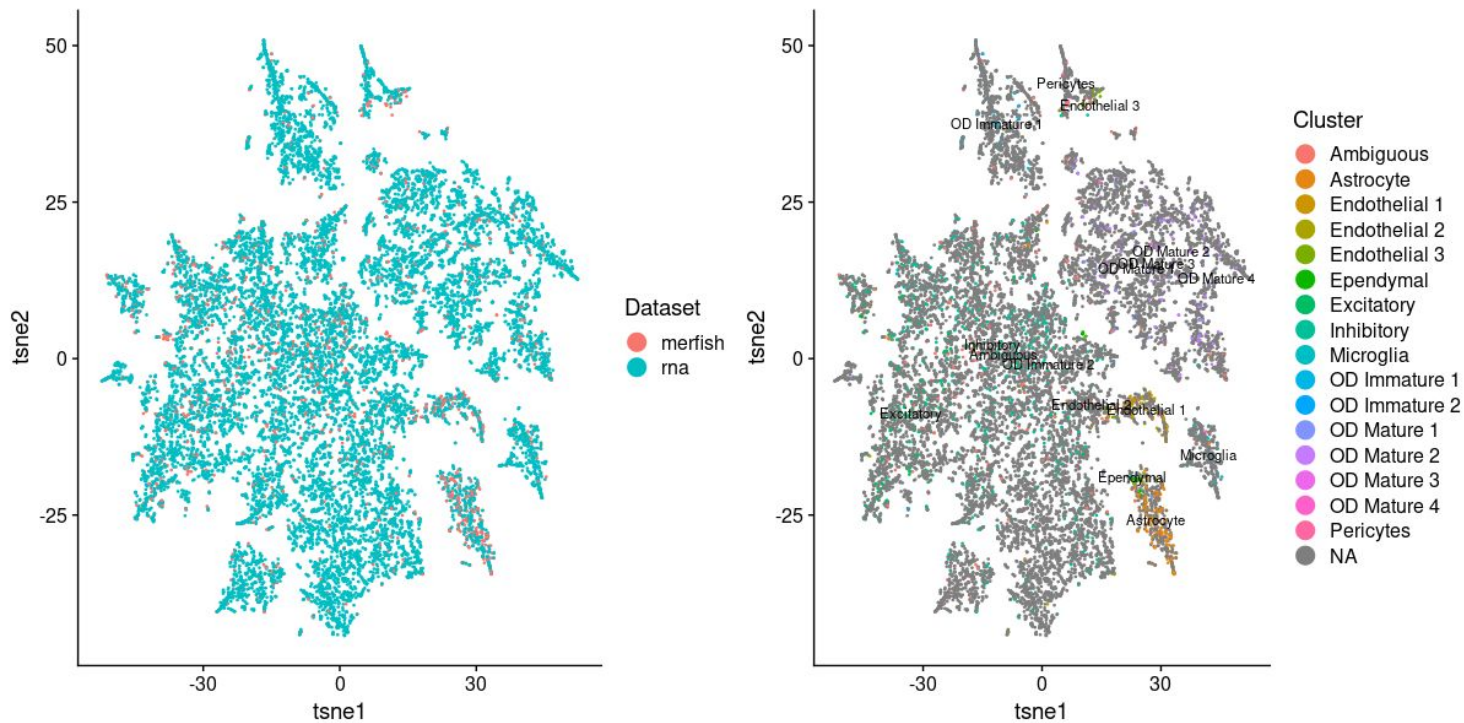
- Highly aligned after integration
- MERFISH enriched for astrocytes as compared to scRNA-seq

How Many Genes Are Needed for Integration?


summary statistics over downsampled genes

- Found that the increase in quality of spatial sample assignment is approximately linear
- Note that metrics for scRNA-seq samples level out quickly
    - Theorize that at the lowest level, STARmap data obfuscates real patterns in data, whereas with more genes the scRNA-seq is of high enough resolution and quantity to provide structure

Increasing Spatial Samples Improves Alignment

- MERFISH data used in analysis came from dataset with 1M samples
- Integration yielded alignment of 0.851 and many distinct clusters
- More informative demonstration of dissociation bias in relative numbers of cells found from each dataset per cluster

Kriebel (2020)
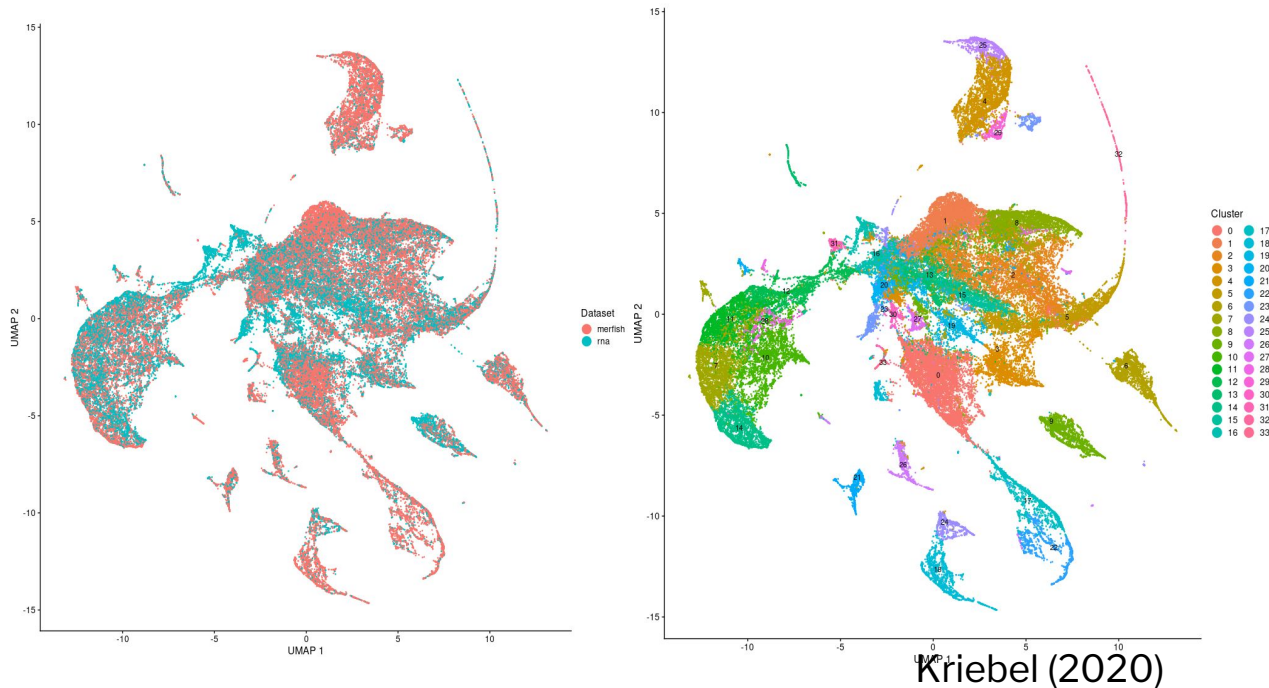
Comparing Integration Results Across Protocols
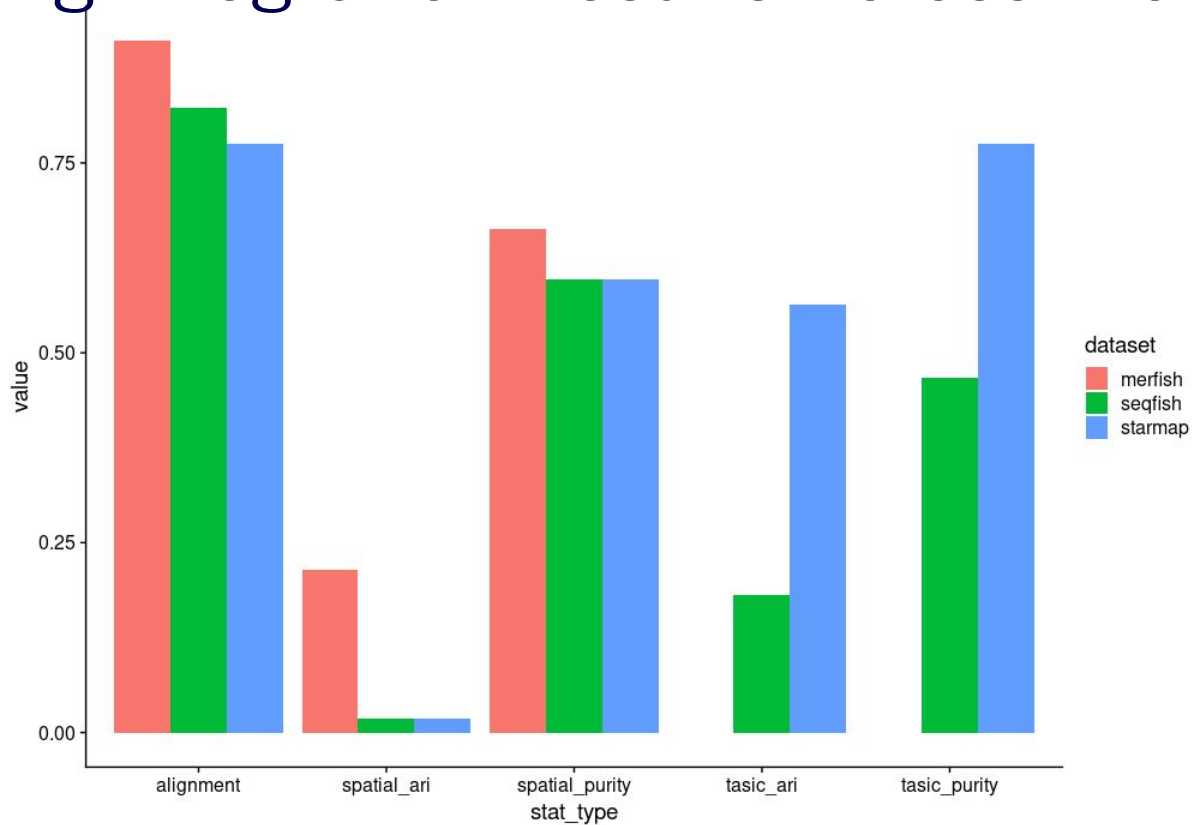
Acknowledgments

Welch Lab

- Jialin Liu
- April Kriebel

Thank you for your attention!

References

Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., & Macosko, E. Z. (2019). Single-Cell Multi-omic
Integration Compares and Contrasts Features of Brain Cell Identity. *Cell*, *177*(7), 1873–1887.e17.
https://doi.org/10.1016/j.cell.2019.05.006

Data:

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., ... & Penn, O. (2018). Shared and distinct
transcriptomic cell types across neocortical areas. *Nature*, *563*(7729), 72-78.

Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., ... & Nolan, G. P. (2018). Three-dimensional
intact-tissue sequencing of single-cell transcriptional states. *Science*, *361*(6400), eaat5691.

Xia, C., Fan, J., Emanuel, G., Hao, J., & Zhuang, X. (2019). Spatial transcriptome profiling by MERFISH reveals subcellular
RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of the National Academy of
Sciences*, *116*(39), 19490-19499.