



Debrief of Brainstorming Session
Integrating Spatial Information in Single-Cell Transcriptomics Analysis

Ruben Dries
Guo-Cheng Yuan

BIRSBioIntegration, June 16th, 2020

Speakers

Alexis Coullomb

Hang Xu

Dario Righelli

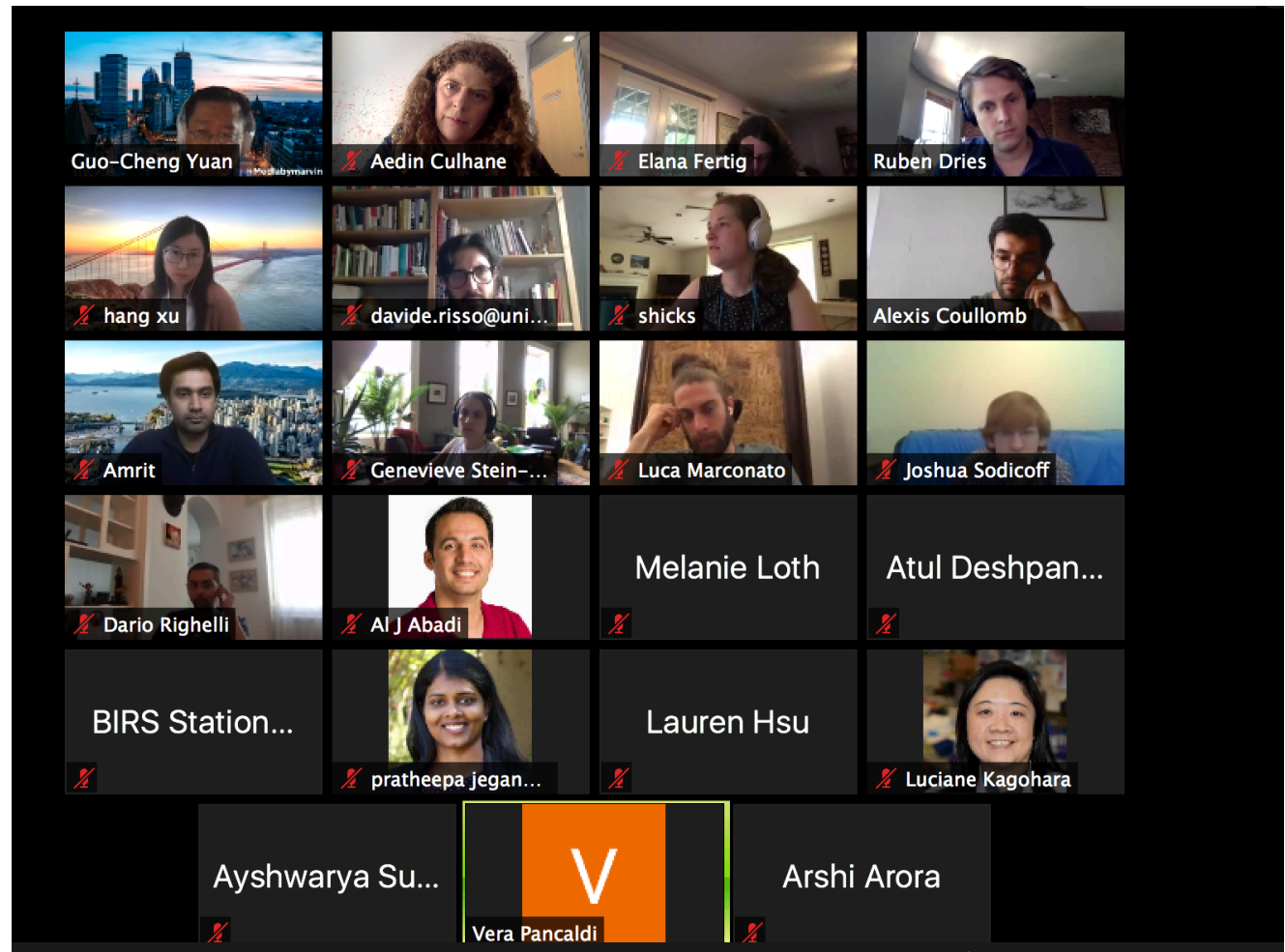
Amrit Singh

Joshua Sodicoff

Notes

Aedin Culhane

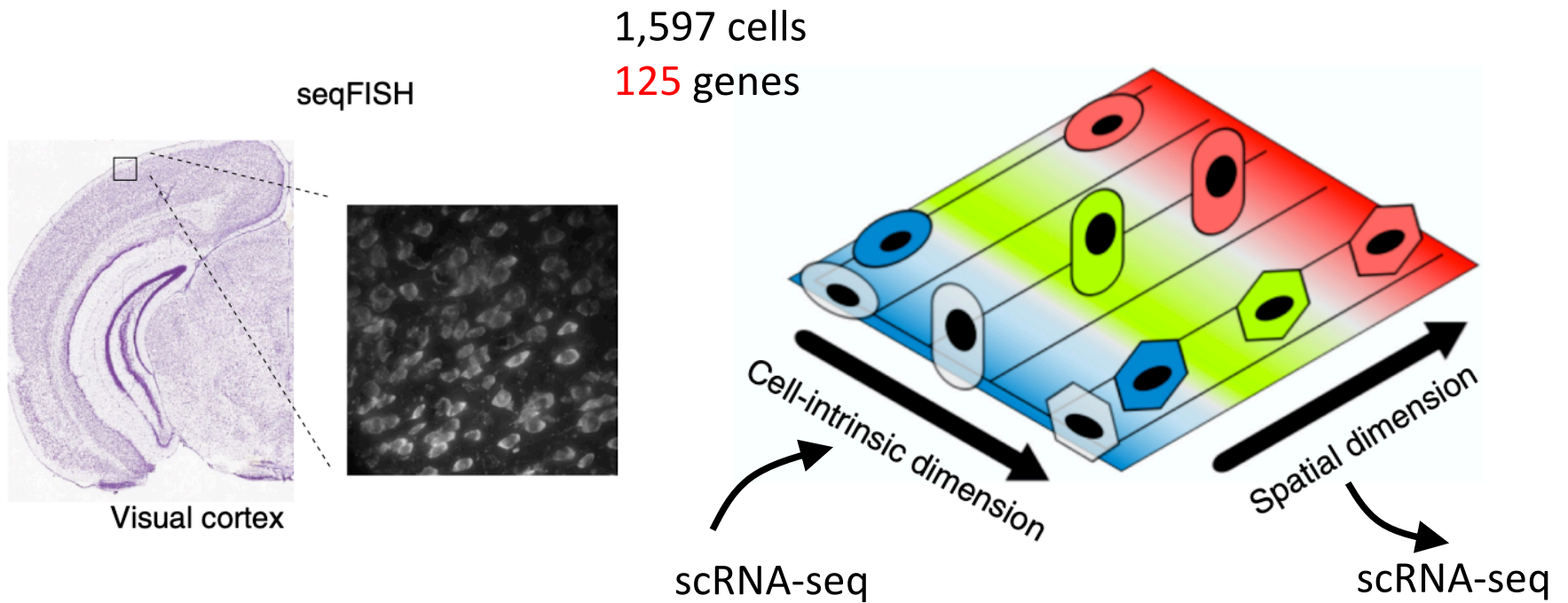
Elana Fertig



<https://docs.google.com/document/d/1UrJ4wjLe7XGHXTU6UwqbsliLqaOZAIR6-9Rq3Nt-Roo/edit?pli=1>

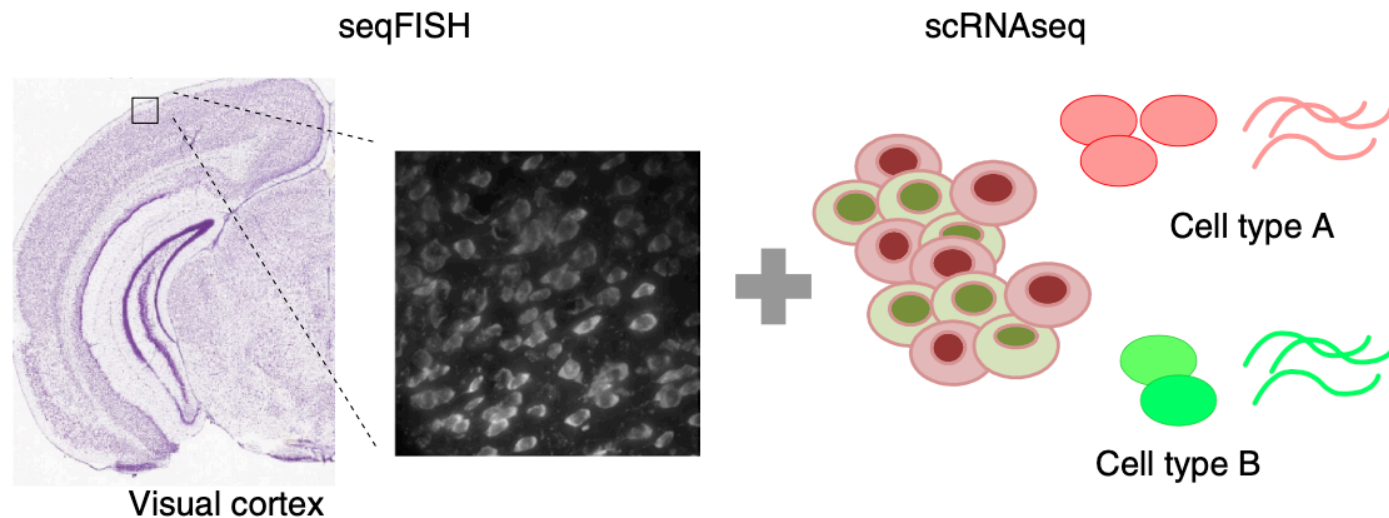
Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence *in situ* hybridization data

Qian Zhu¹, Sheel Shah^{2,3}, Ruben Dries¹, Long Cai² & Guo-Cheng Yuan¹



Adult mouse cortical cell taxonomy revealed by single cell transcriptomics

Bosiljka Tasic^{1,2}, Vilas Menon^{1,2}, Thuc Nghi Nguyen¹, Tae Kyung Kim¹, Tim Jarsky¹, Zizhen Yao¹, Boaz Levi¹, Lucas T Gray¹, Staci A Sorensen¹, Tim Dolbeare¹, Darren Bertagnolli¹, Jeff Goldy¹, Nadiya Shapovalova¹, Sheana Parry¹, Changkyu Lee¹, Kimberly Smith¹, Amy Bernard¹, Linda Madisen¹, Susan M Sunkin¹, Michael Hawrylycz¹, Christof Koch¹ & Hongkui Zeng¹

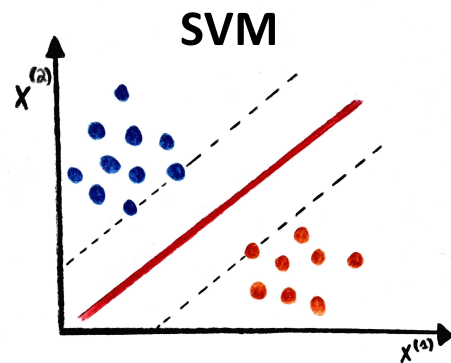
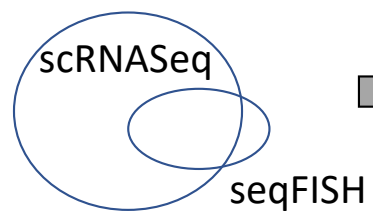


- Can scRNA-seq data be overlaid onto seqFISH for resolution enhancement?

- What is the minimal number of genes needed for data integration?

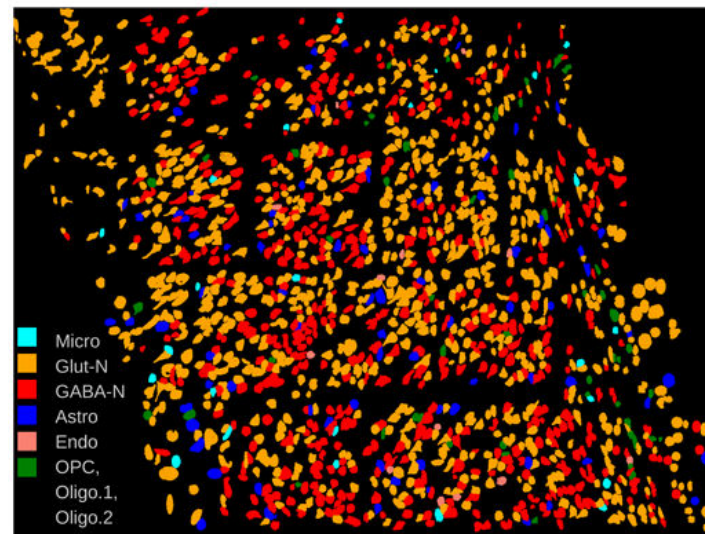
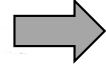
- Are there signatures of cellular co-localization or spatial coordinates in the non-spatial scRNA-seq data?

gene features

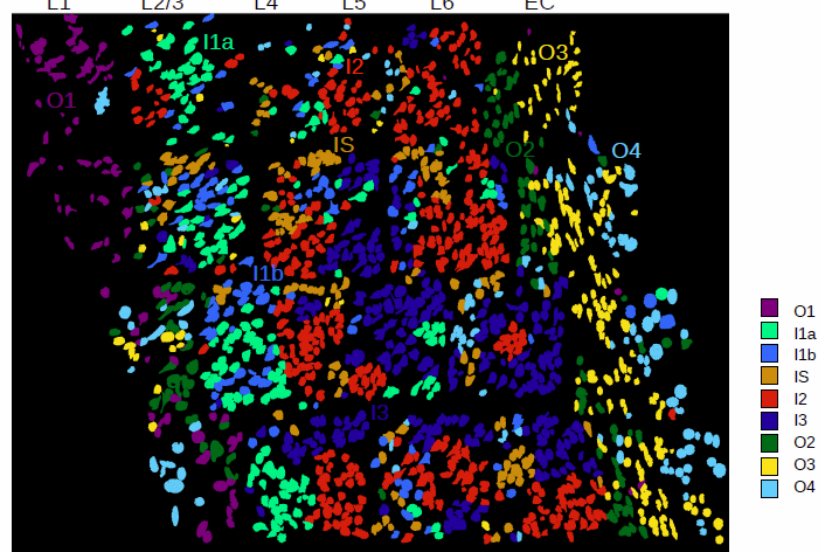
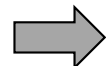
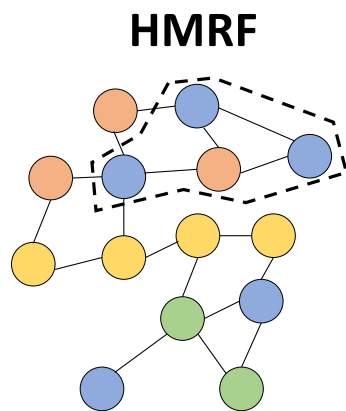
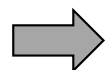


Training set: scRNA-seq

Test set: seqFish

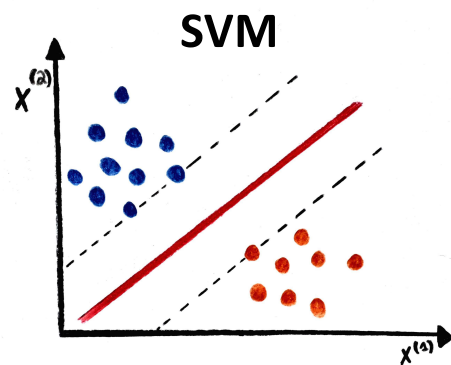
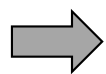
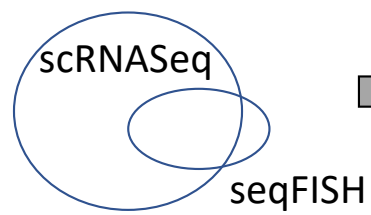


spatial coherent genes



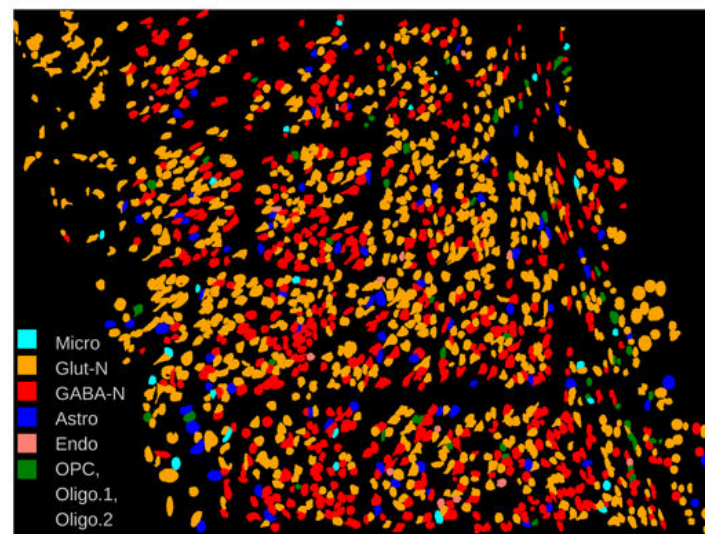
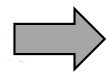
Experimental design

gene features

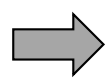


Training set: scRNA-seq

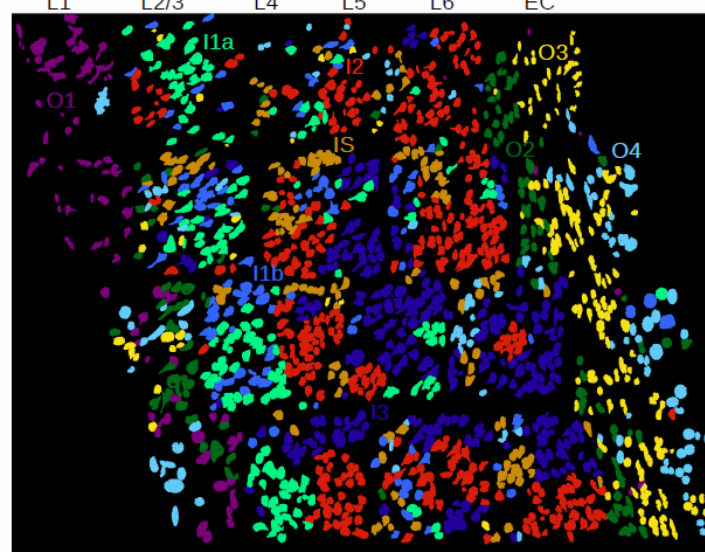
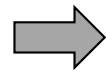
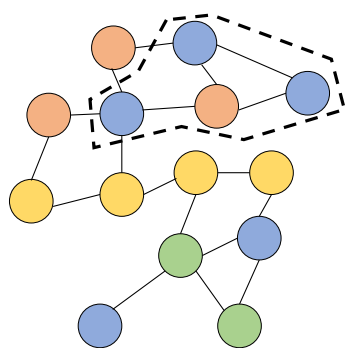
Test set: seqFish



spatial coherent genes



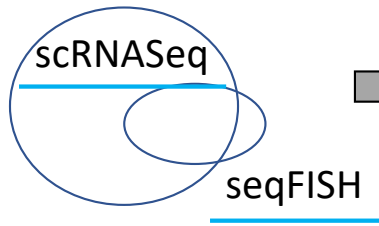
HMRF



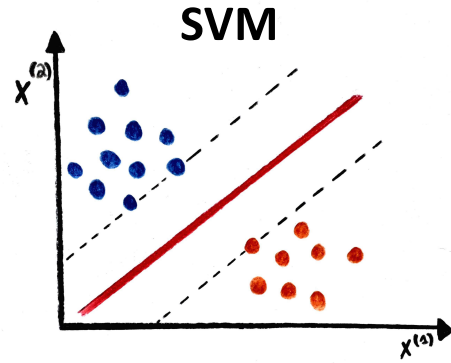
Experimental design

- What is the goal of the project? Spatial cell type composition or spatial patterns?
 - merFISH vs seqFISH approach
- Can we use literature or single-cell datasets to create an informative gene set?
 - minimum number of genes?
 - genes with similar distributions between different technologies?

gene features

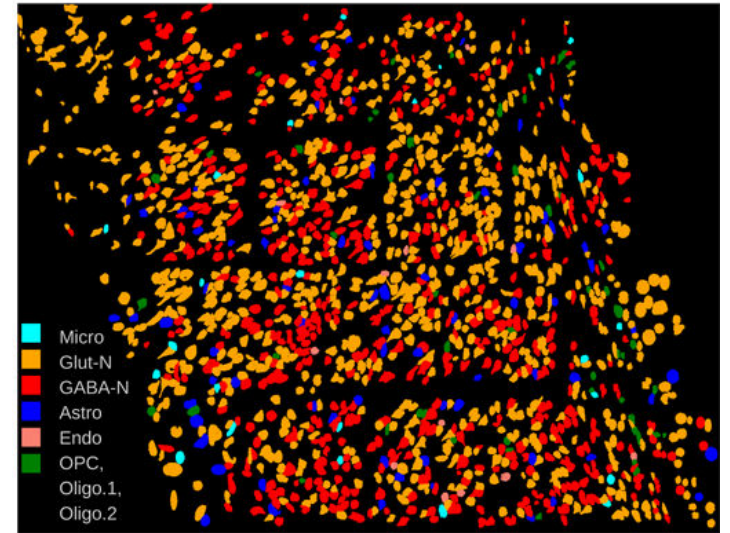


Platform specific biases

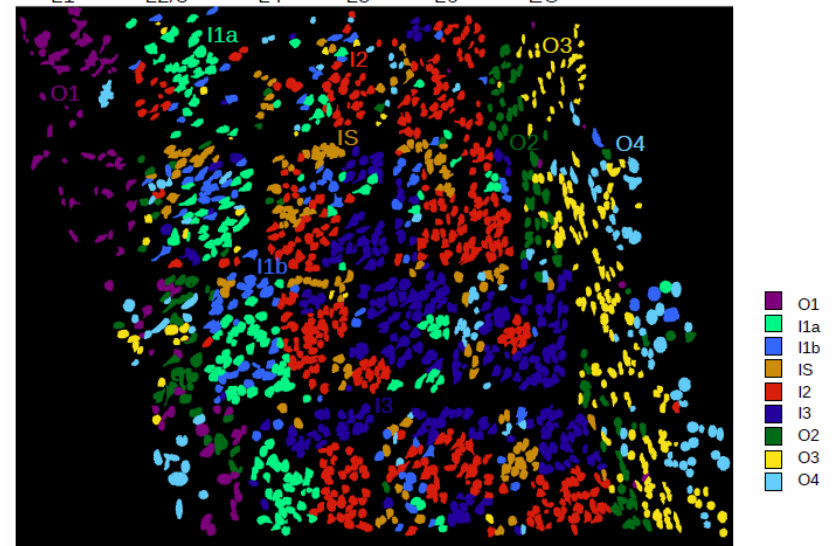
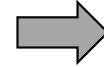
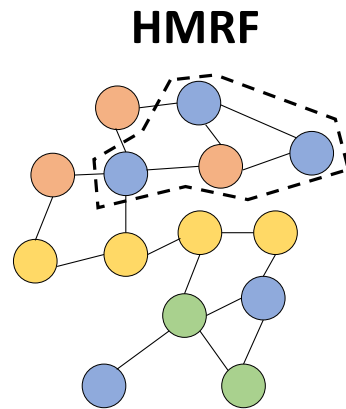
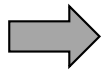


Training set: scRNA-seq

Test set: seqFISH

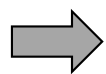
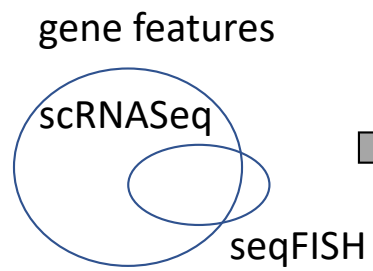


spatial coherent genes

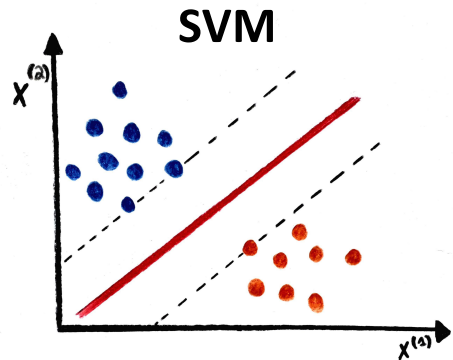


Platform specific biases

- bias in gene detection / coverage.
 - low vs high expressed genes
 - ISH vs poly(A)-tail enrichment
- bias in cell composition
 - dissociation methods
 - global expression differences between scRNA-seq and seqFISH due to stress response?
- (incomplete) subcellular information
 - transcripts that are not assigned?
 - segmentation challenge
 - essential for cell type identification?
 - Is there information in morphology and subcellular distribution?

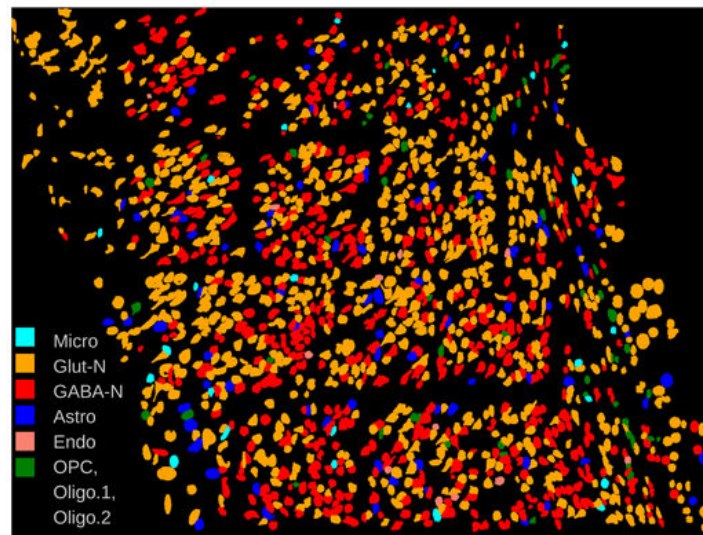
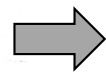


Objective assessment



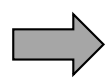
Training set: scRNA-seq

Test set: seqFish

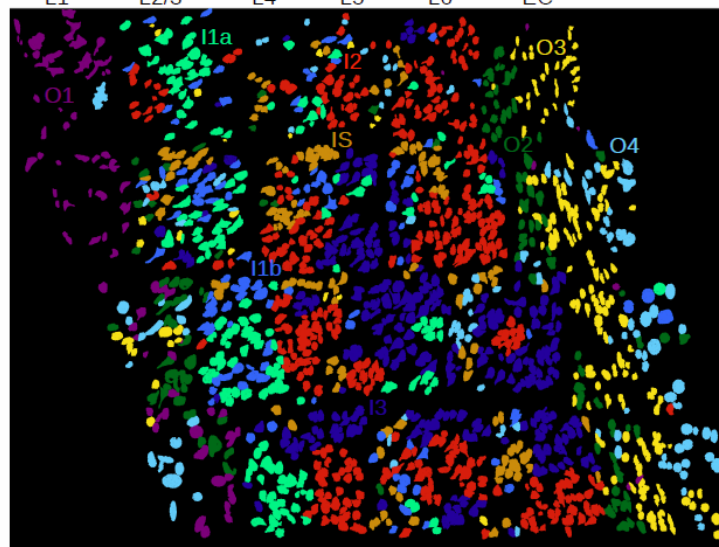
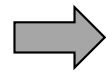
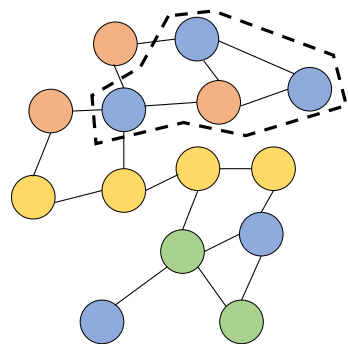


L1 L2/3 L4 L5 L6 EC

spatial coherent genes



HMRF

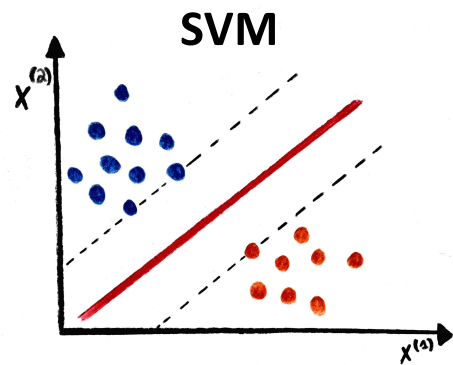
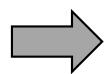
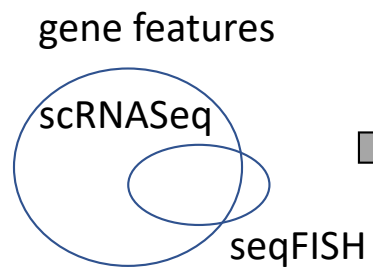


Objective assessment

- How do we know if our results are good or which method is better?
- Benchmark:
 - Need for manually curated data
 - Bench marks (Matthew Ritchie)
 - Cluster robustness - prediction strength (Tibsharani)
 - Organoid cultures, artificial reconstructed datasets
 - Gold standard to detect problems and issues with different platforms

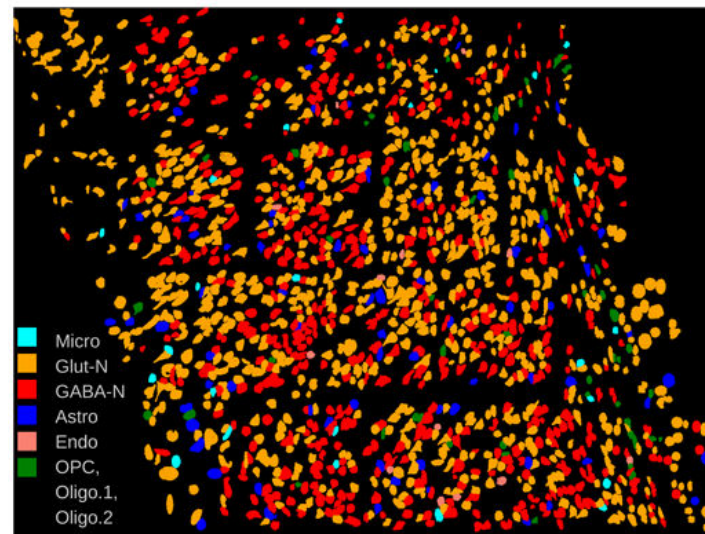
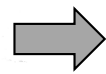
Important: do you gain new biological insight?

- Human cortex with 10x visium with manually curated spots into cortex layers from the DLPFC (**Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex**). The benefit is the DLPFC is an incredibly well studied area of the brain, so lots of biology already known (Stephanie Hicks)



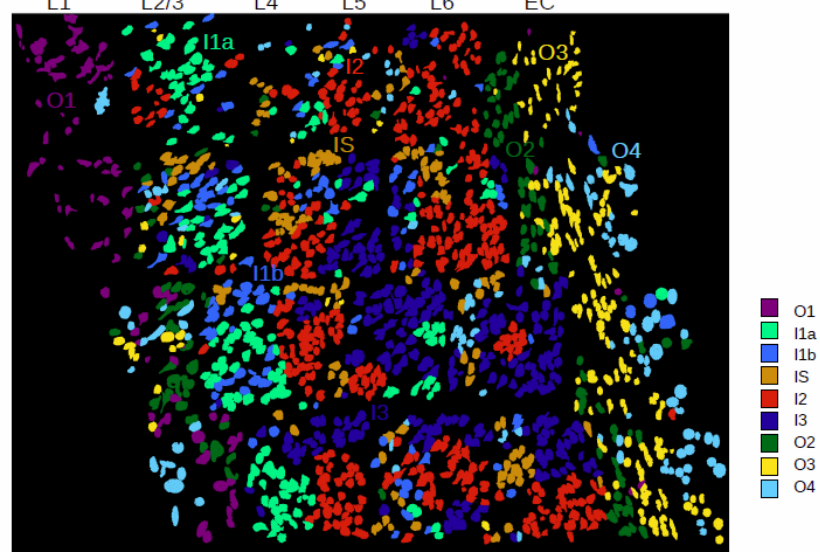
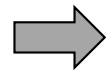
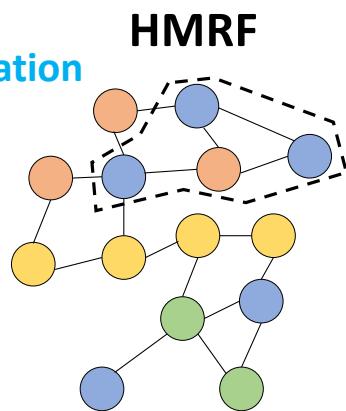
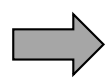
Training set: scRNA-seq

Test set: seqFish



Spatial representation

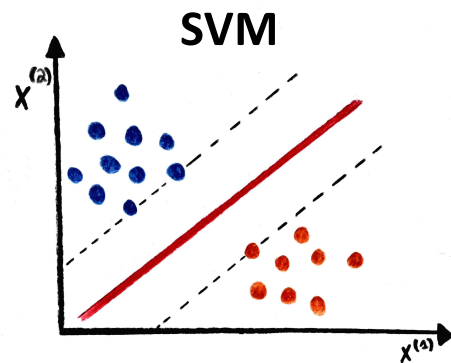
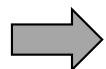
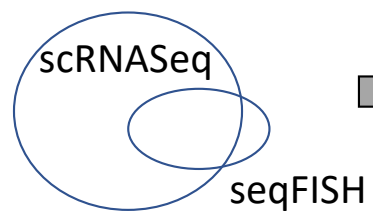
spatial coherent genes



Spatial representation

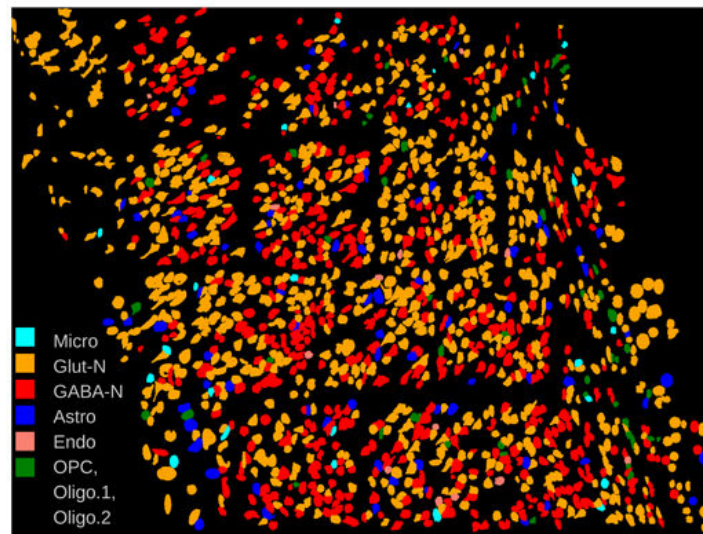
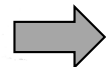
- How to best represent cells in a spatial context?
 - Include morphology and count for differences (e.g. size)?
 - Abstract:
 - cell centroids
 - polygonal coordinates
 - Delaunay network and voronoi tessellation

gene features



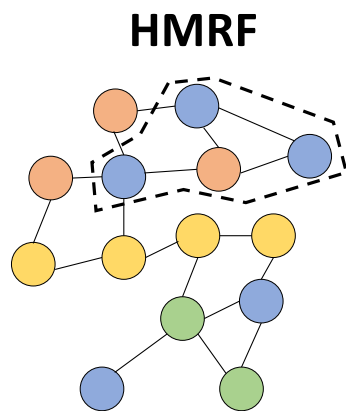
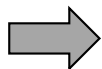
Training set: scRNA-seq

Test set: seqFISH

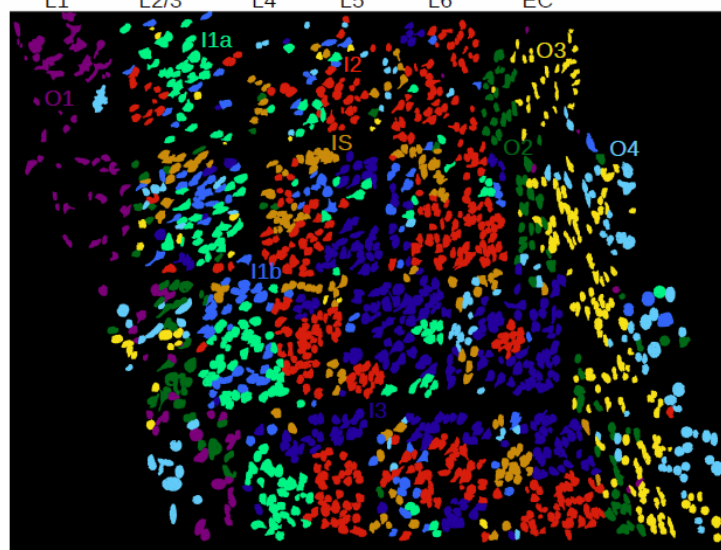
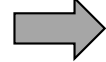


L1 L2/3 L4 L5 L6 EC

spatial coherent genes



Spatial information



Spatial information

- Cell types and spatial structure are different challenges
- Different information at different spatial resolution (zoom in / zoom out)
 - need for hierarchical models that are network based?
- Can Visium (large area / low resolution) be linked with seqFISH (small area / high resolution)?
 - deconvolution problem? [#deconvolution](#) (Luca Marconato)
- How do we test for statistical significance when the test statistic is a distance matrix? (Shila Ghazanfar)
 - How to incorporate biologically driven null hypotheses?
- Be aware of Gene Ontology enrichment
 - Introduction of biases through annotation (very old) and correlation structure due to integration
 - Give certain genes different weights (e.g. marker genes)

Action items:

See #seqfish_theme

Goals:

figures: 2 key figures per analysis

text: main findings and suggestions

