



SINGLE-CELL MULTIOmICS TO UNCOVER FOUNDATIONAL REGULATORY MECHANISMS OF LENS DEVELOPMENT AND PATHOLOGY

Based on Tangeman et al., 2024

22102 Applied Single Cell Bioinformatics Jan 24

Rodrigue Lefèvre-Morand - s222659

Javier Lopez Bort - s223237

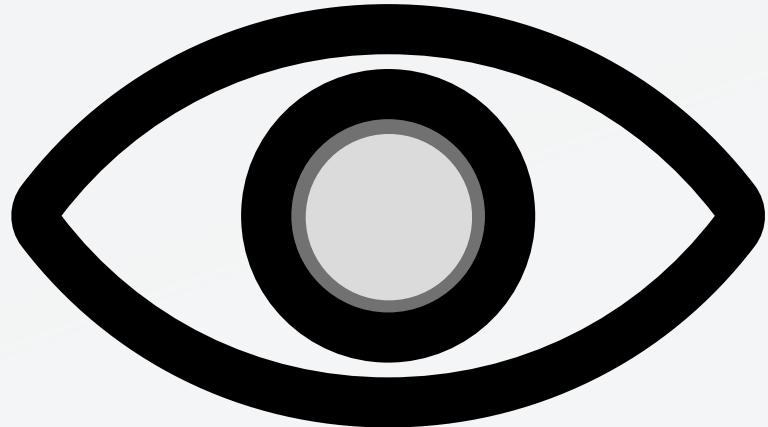
Martynas Baltušis - s222858

Alexander Zubov - s223162

CONTENT

- 
- 01** INTRODUCTION
 - 02** PREPROCESSING
 - 03** CELL TYPE ASSIGNMENT
 - 04** TRAJECTORY ANALYSIS
 - 05** DIFFERENTIAL ANALYSES
 - 06** MULTIMODAL DATA INTEGRATION
 - 07** DISCUSSION AND PERSPECTIVES

INTRODUCTION - TOPIC



CATARACT

- Severe chronic disease
- Not much understood about its causes
- **What is known:** Changes in cell types and cell-communication are involved

MULTIOMICS

Most studies focused on **bulk transcriptomics**

- Not able to capture details specific for singular cells differentiation patterns

*Integrated single-cell multiomics uncovers foundational regulatory mechanisms of lens development and pathology,
Tangeman et al., 2024*

- Multimodal single cell study
- Aimed to investigate lens development via snRNAseq, snATACseq and CUT&RUNseq sequencing



INTRODUCTION - DATASETS

Conditional snRNA-seq data

- Conditions:
 - Control/reference
 - Removal of retina
 - Removal of retina + treatment with FGF
- Embryos with different ages (day 4 or 5)
- 14,249 features & 58,747 cells

Multimodal snRNA/snATAC-seq data

- No conditions

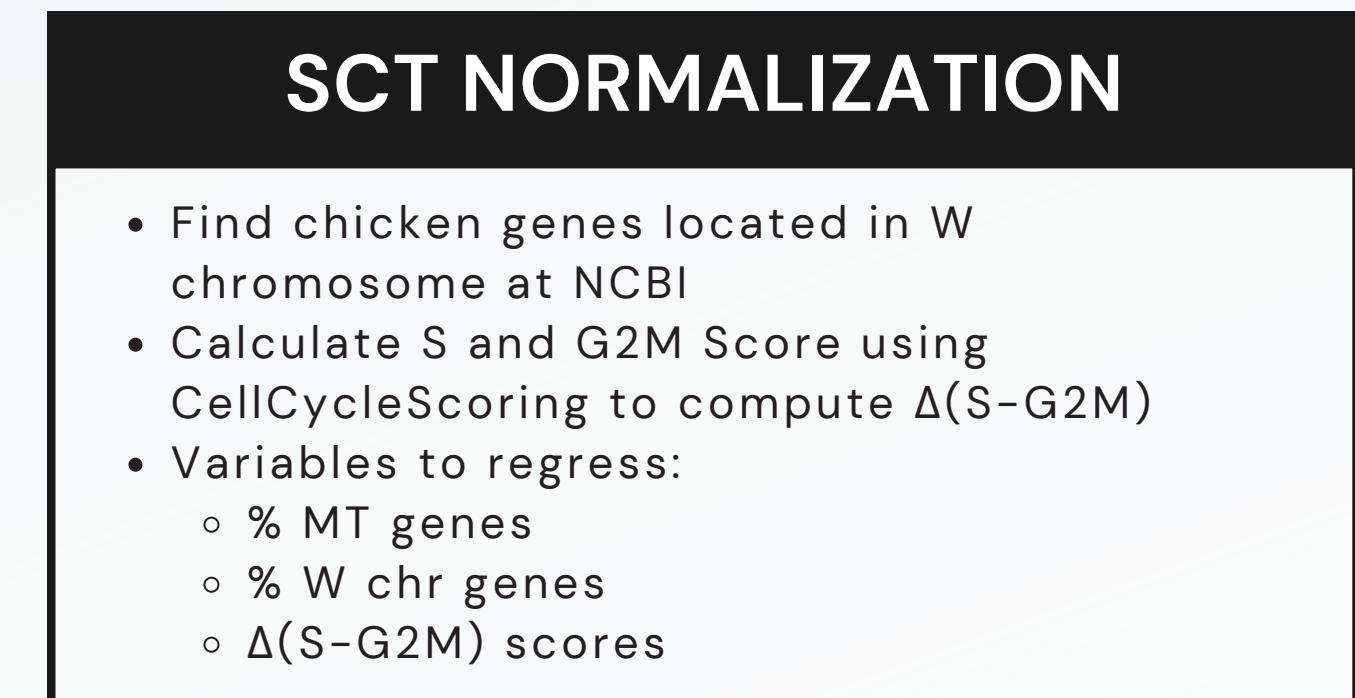
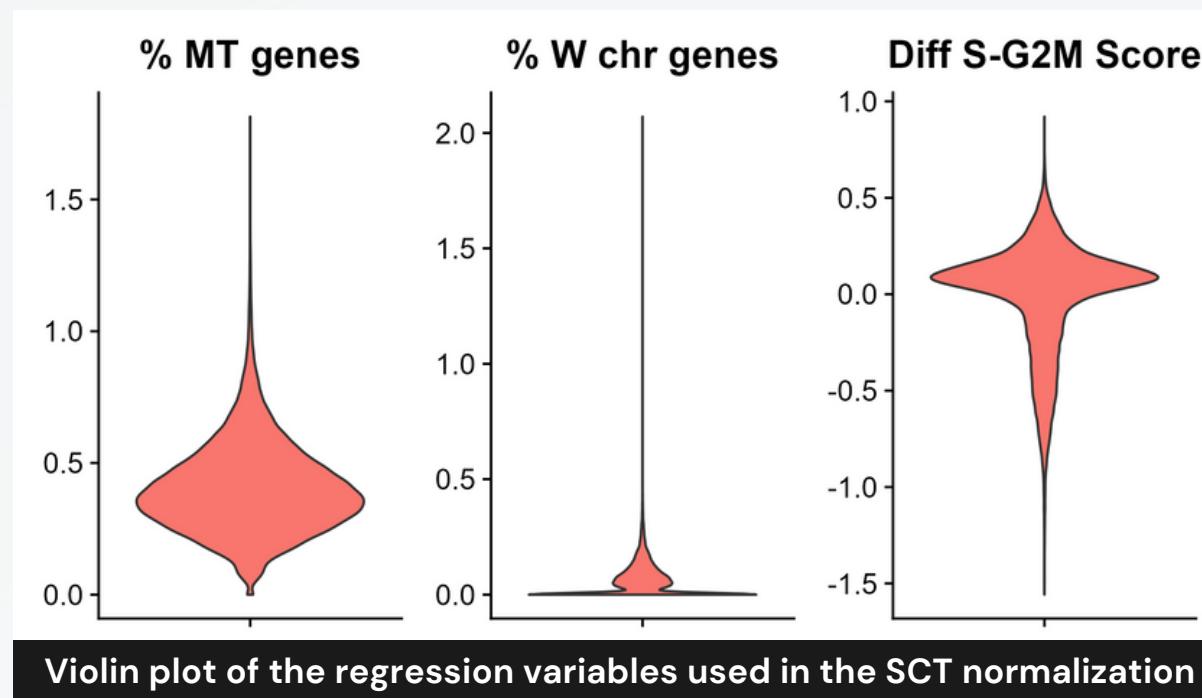
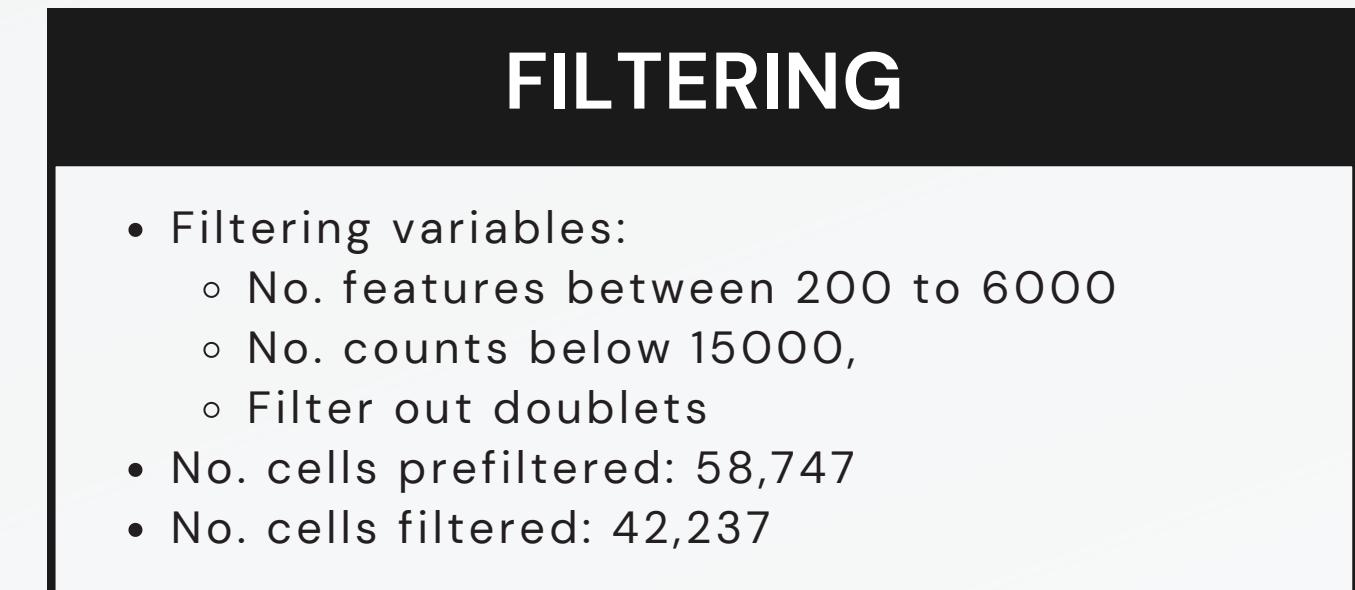
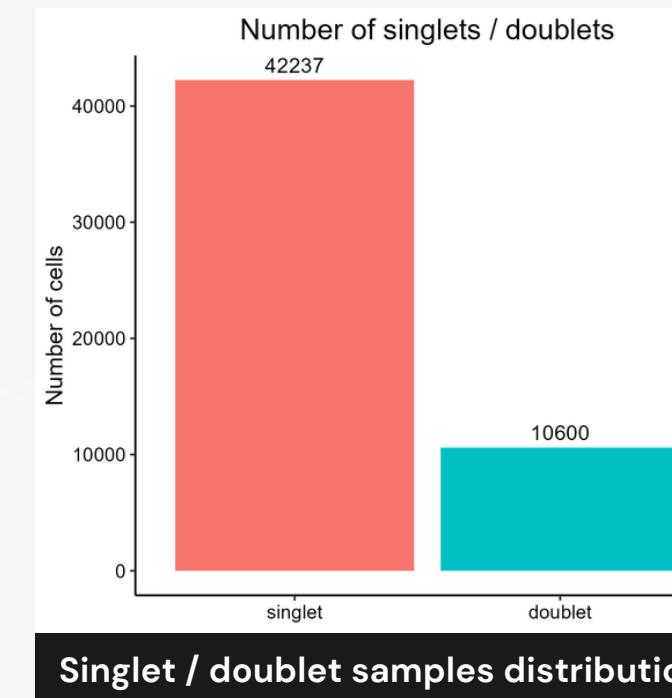
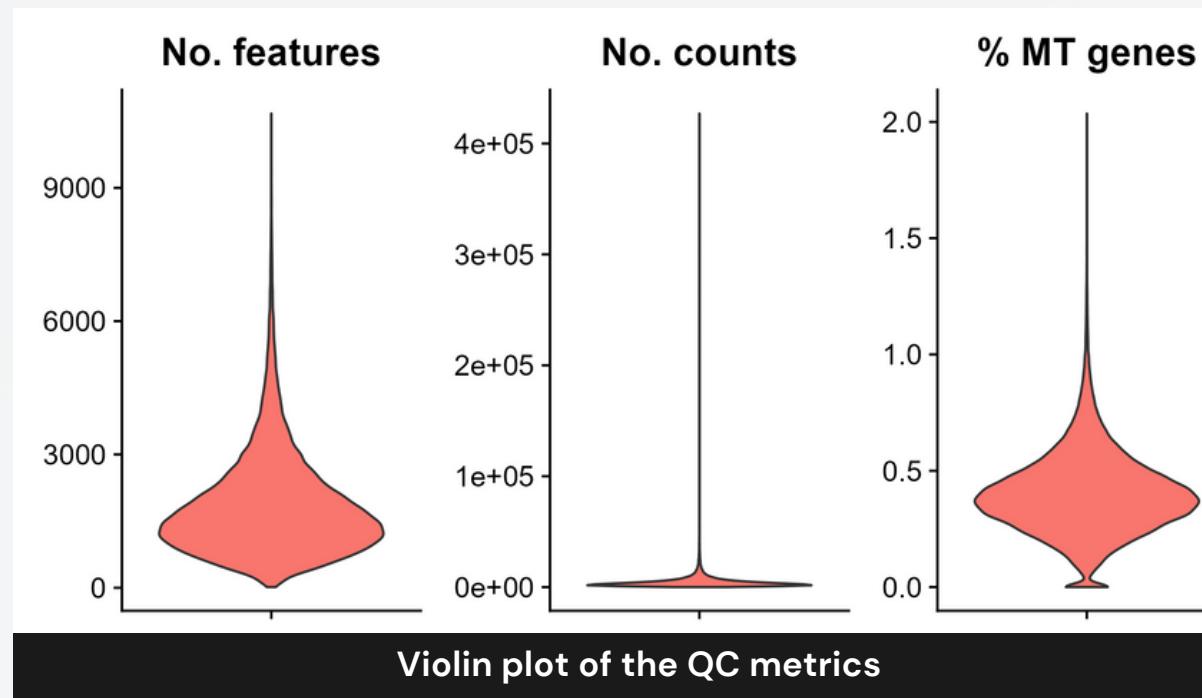
Why snRNA-seq instead of scRNA-seq?

- snRNA-seq allows to study sensitive cells
- Cells embedded in extracellular matrix difficult to dissociate

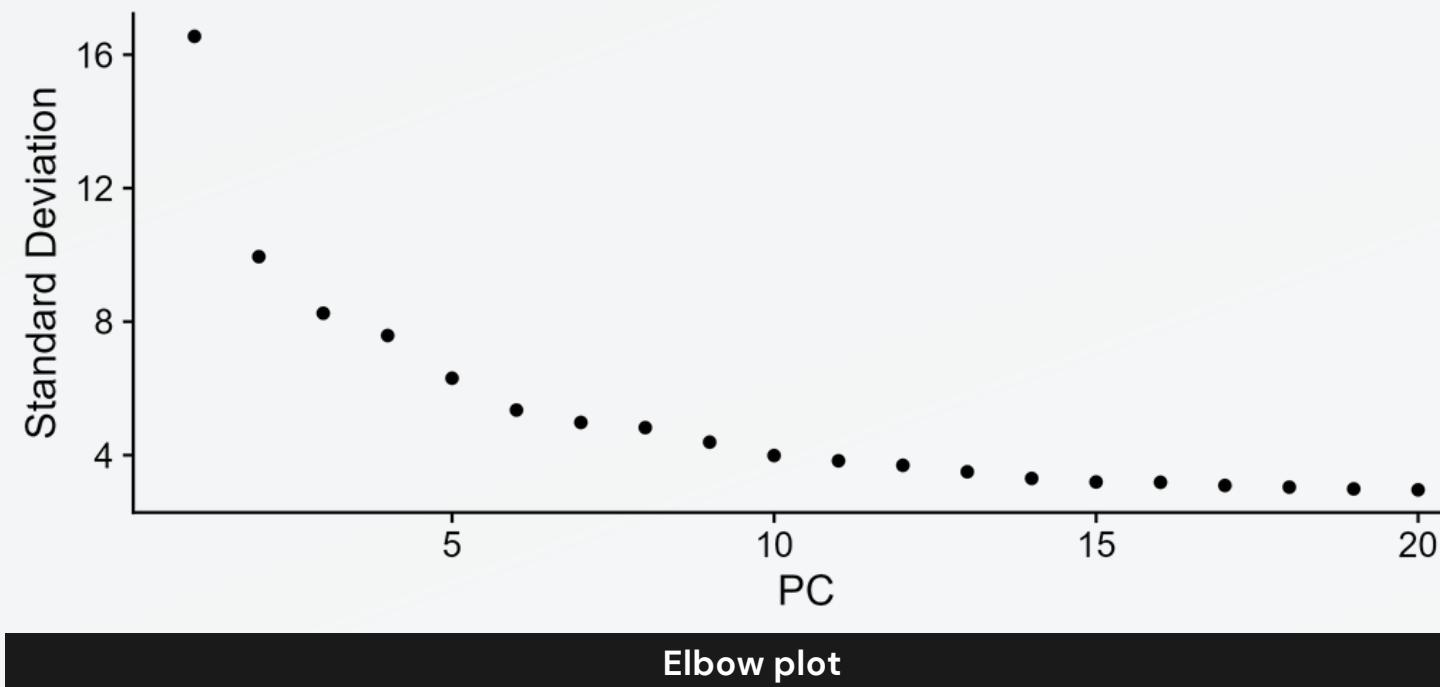
PREPROCESSING

Filtering, normalization, clustering, subset & integration

FILTERING & SCT NORMALIZATION

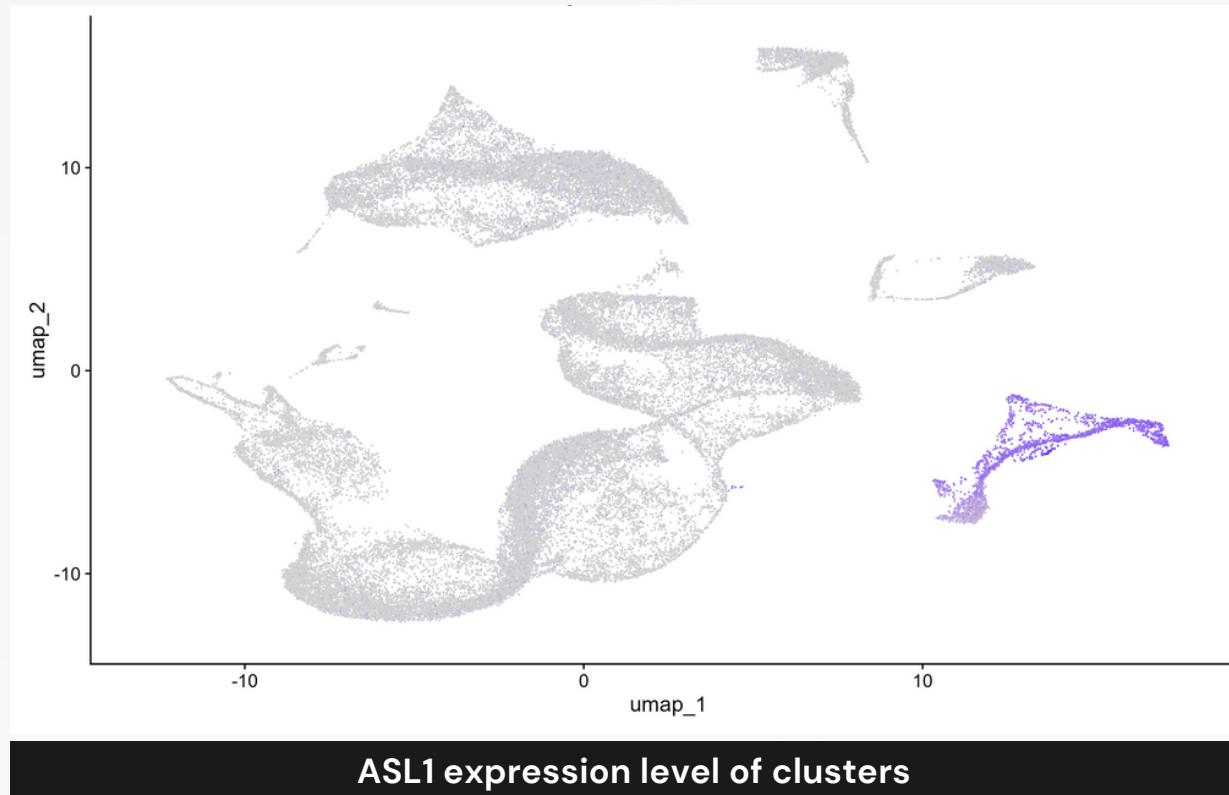
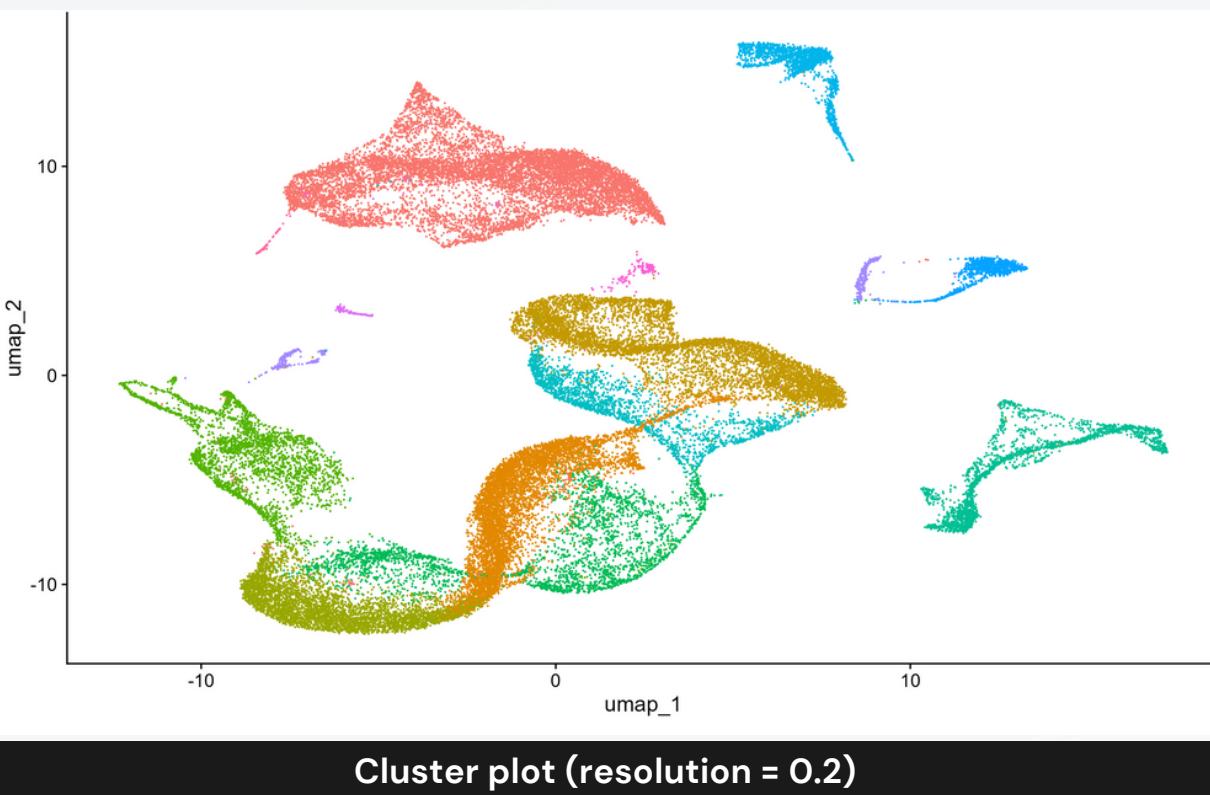


CLUSTER & SUBSET



CLUSTERING

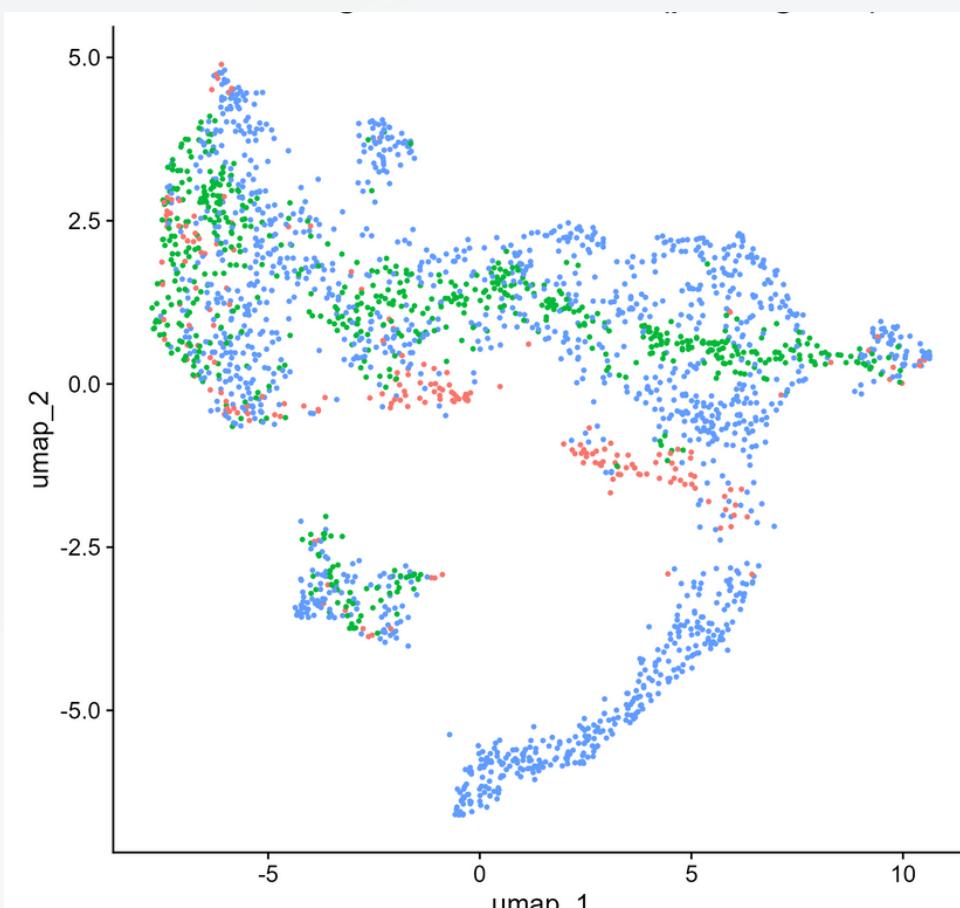
- PCA reduction
- UMAP reduction (15 PCA dimensions)
- Find neighbours (15 PCA dimensions)
- Clustering (resolution = 0.2)



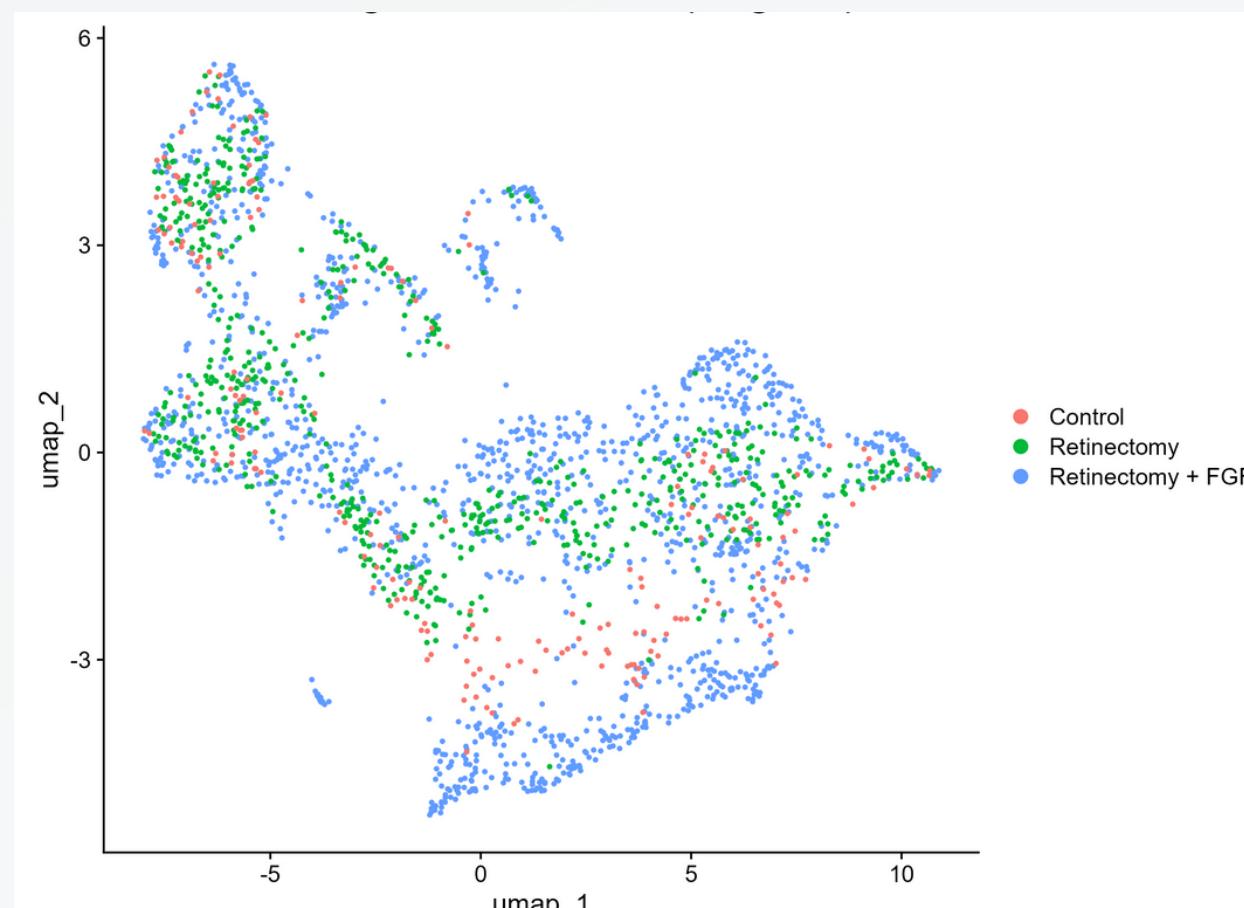
SUBSET

- ASL1 is a gene marker for lens cells
- Cluster 6 shows ASL1 high level expression
- Assign cluster 6 to lens cells cluster
- Subset lens cell cluster
- No. lens cells: 2,324

INTEGRATION ON SUBSET



Clustering based on condition (preintegrated)



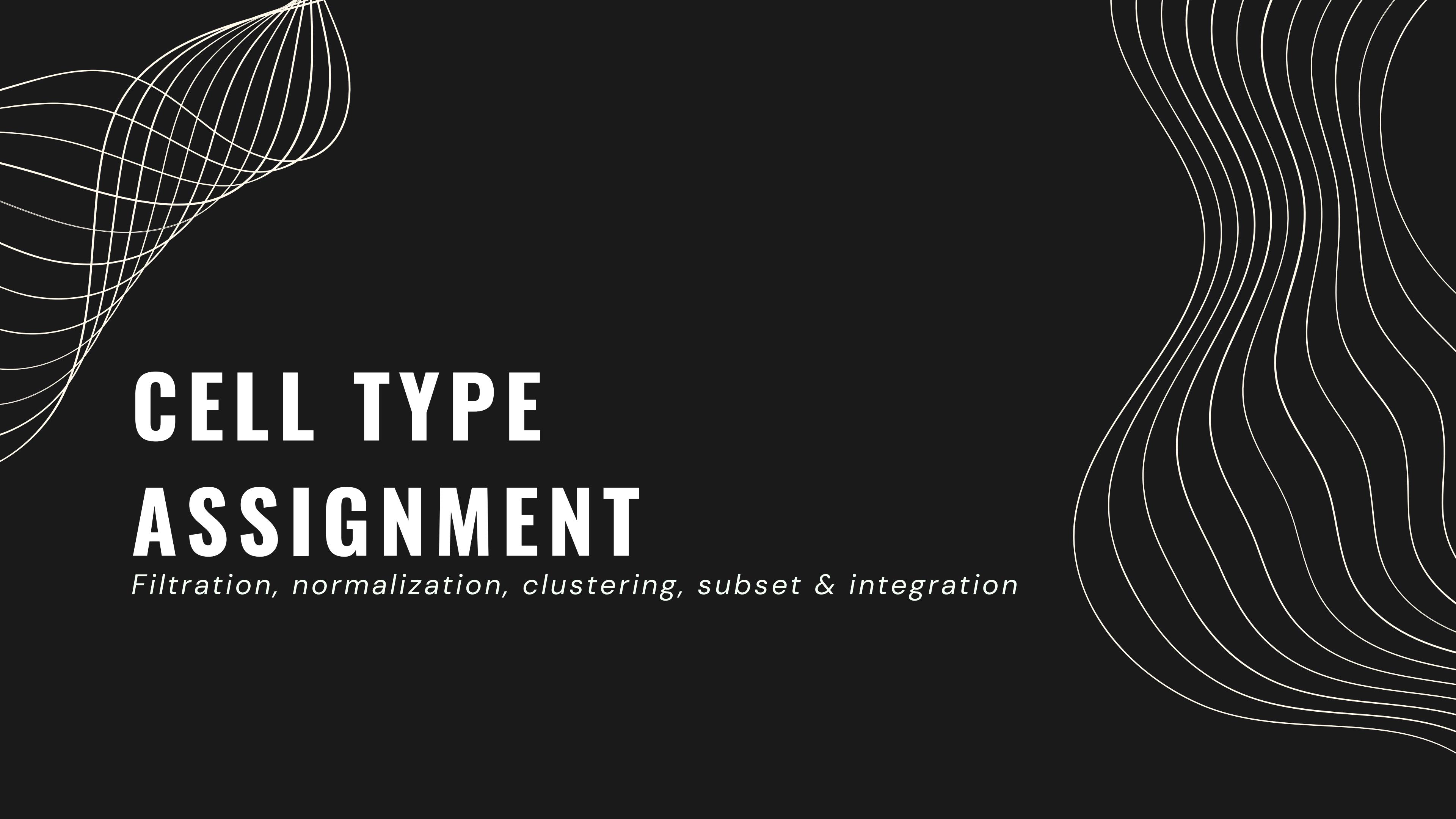
Clustering based on condition (integrated)

INTEGRATION

- Separate dataset by condition
- SCT normalization on each dataset
- Select integration features
- Find integration anchors
- Integrate data

UMAP REDUCTION (ORIGINAL & INTEGRATED)

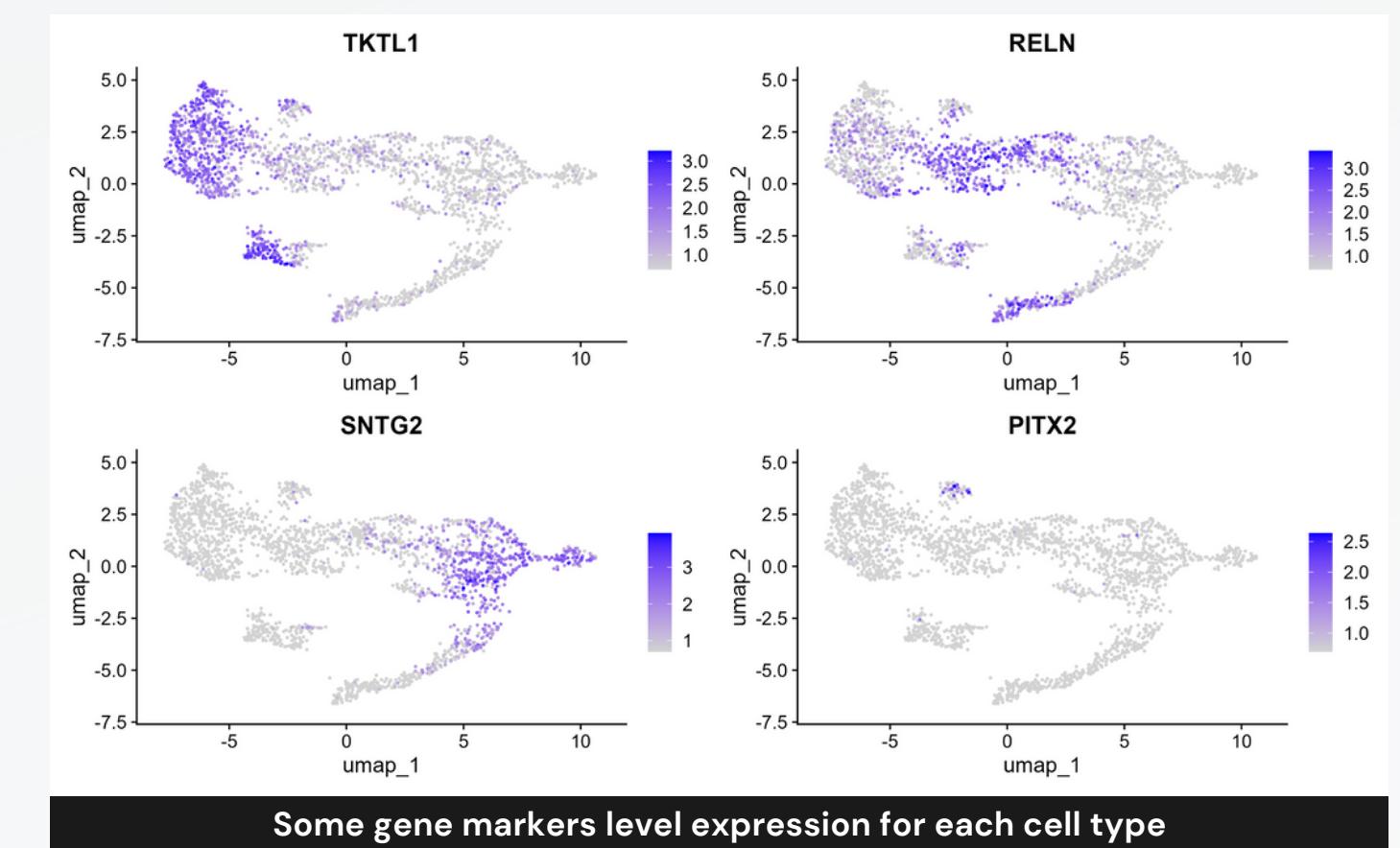
- SCT normalization (% MT genes, % W chr genes, Diff S-G2M Score)
- Scale data
- PCA reduction
- UMAP reduction (15 PCA dimensions)



CELL TYPE ASSIGNMENT

Filtration, normalization, clustering, subset & integration

CELL TYPE ASSIGNMENT



CLUSTERING (INTEGRATED)

- Find neighbours (15 PCA dimensions)
- Clustering

FIND CONSERVED MARKERS

- Grouping variable: condition
- Select top 4 most significant genes
- Find gene markers associated with each cell type

CELL TYPE ASSIGNMENT

- Assign each cluster to a cell type based on gene markers:
 - 0: epithelial
 - 1: intermediate
 - 2: fiber
 - 3: intermediate



TRAJECTORY ANALYSIS

Trajectory construction and DGE over pseudotime

TRAJECTORY ANALYSIS

INPUT

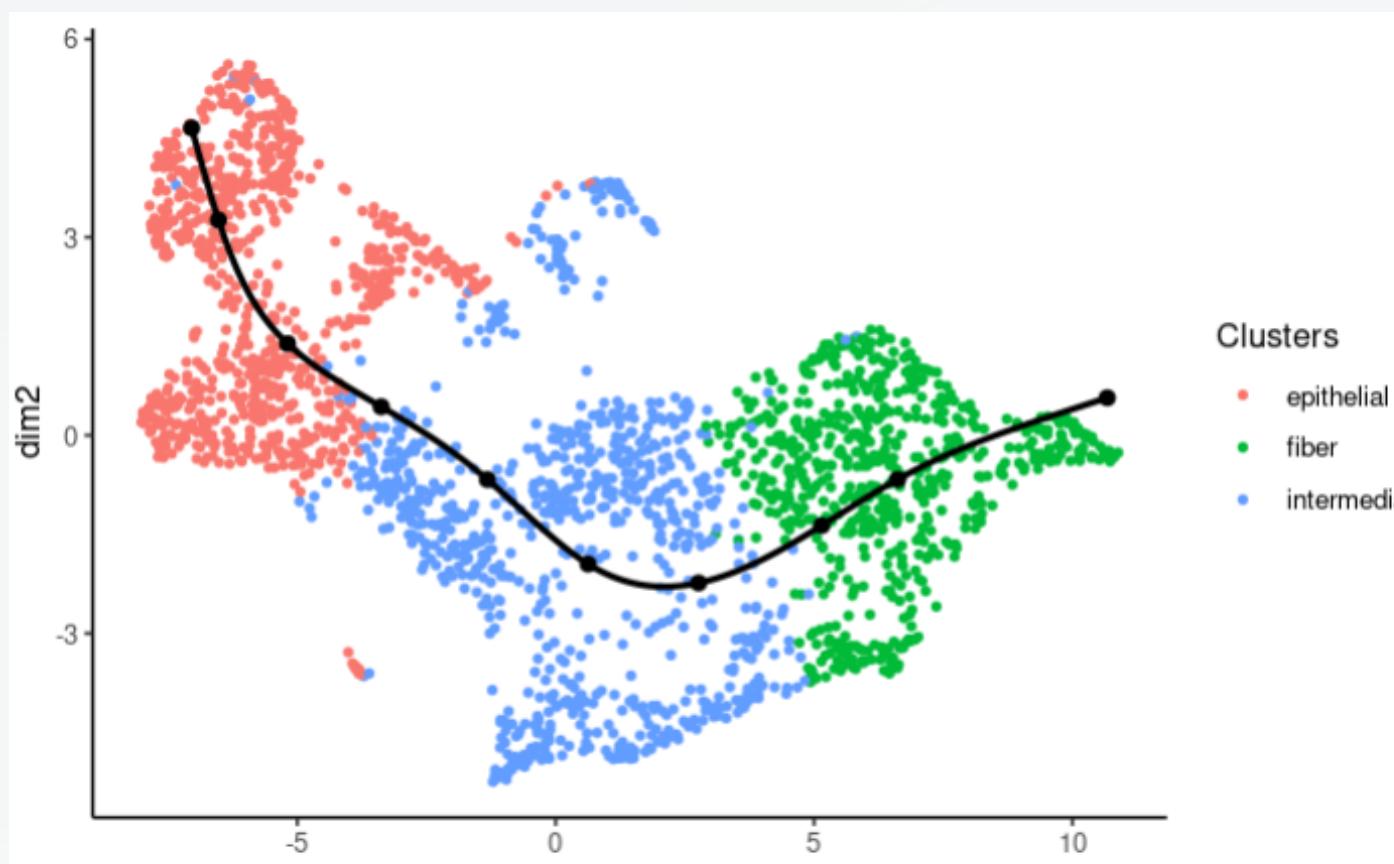
- Integrated, reclustered, and annotated Seurat object
- Dimension reductions, clusters and counts extracted

SLINGSHOT

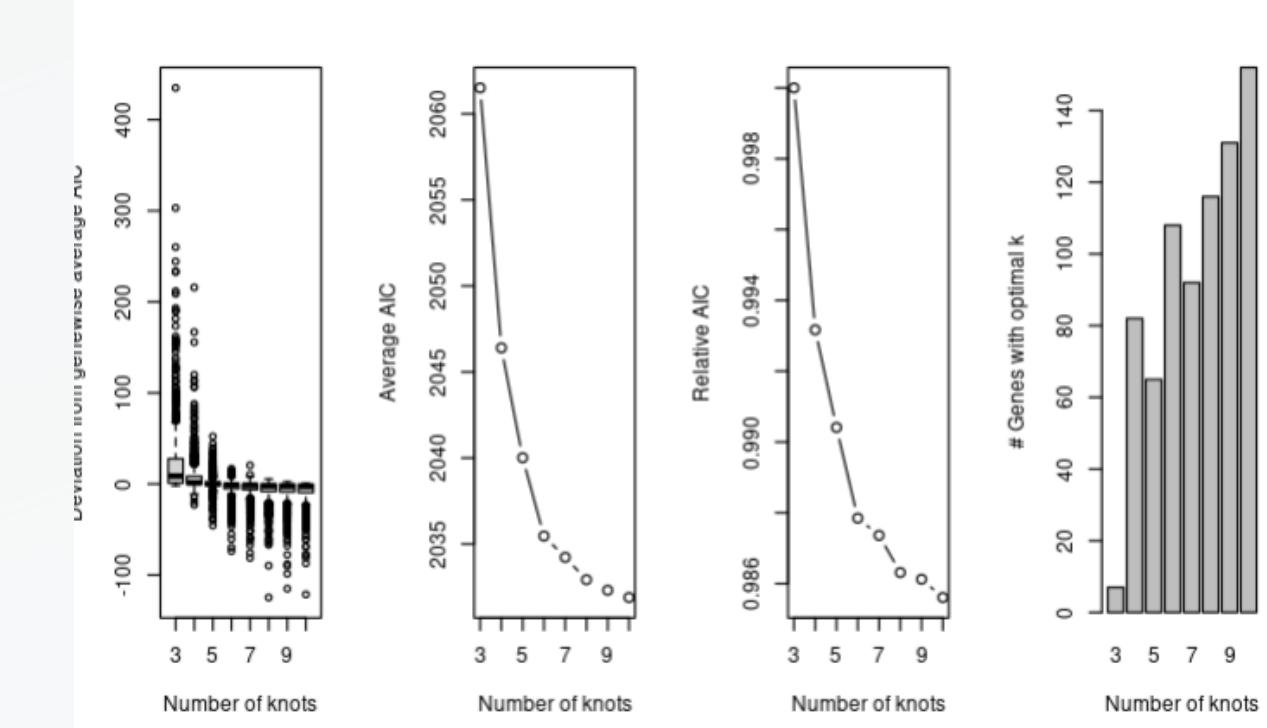
- Standard Slingshot run:
- `getLineages()`
 - Starting cluster: epithelial cells
 - Ending cluster: fiber cells
 - `getCurves()`

TRADESEQ

- `evaluateK()`:
 - 3 to 10 knots
 - 1000 gene subset
- `fitGAM()`
 - 10 knots
- `startVsEndTest()`
 - Single lineage
 - DE between progenitor and differentiated cell type



Plotted trajectory across the 3 cell type clusters

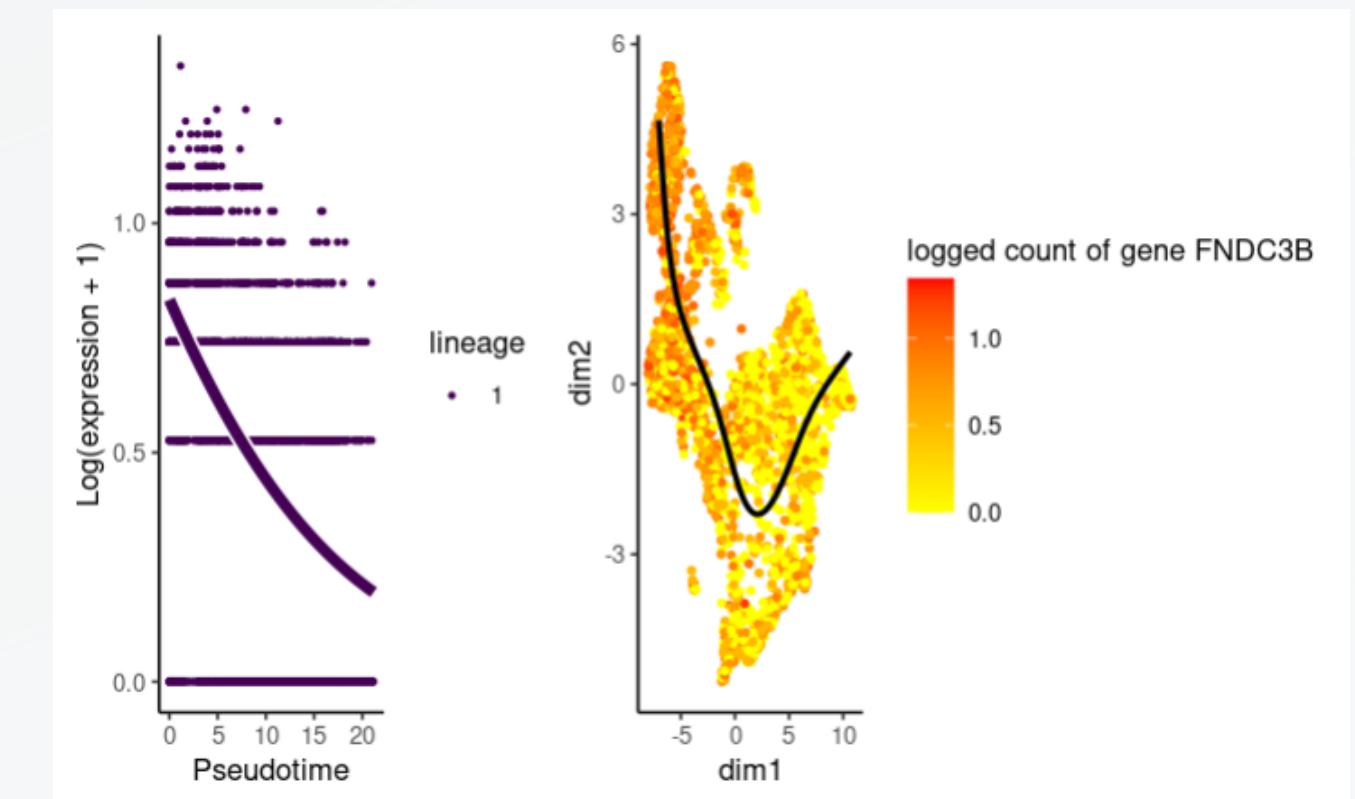
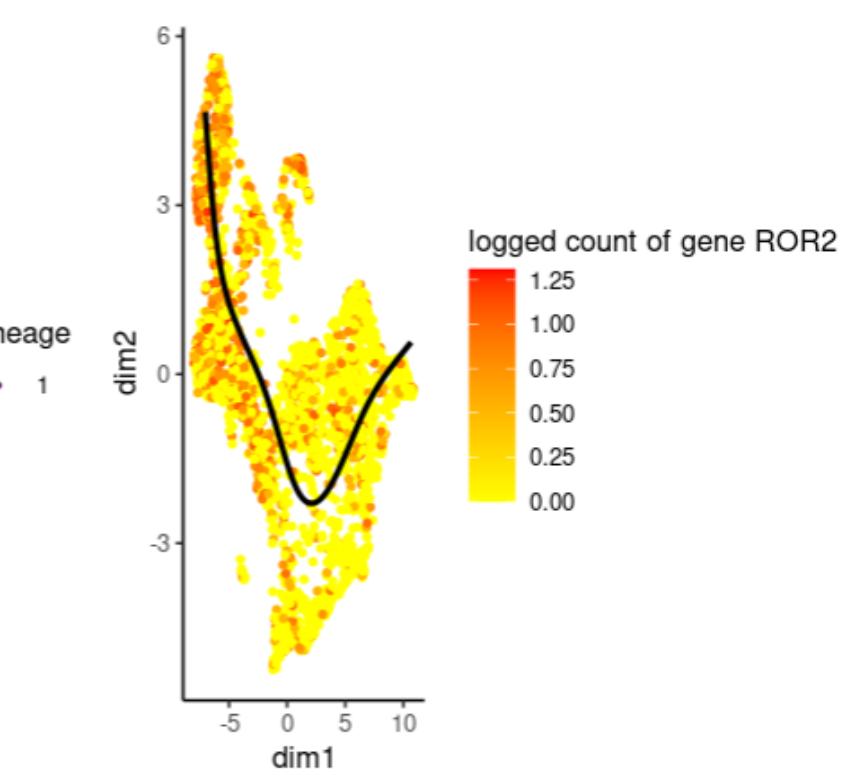
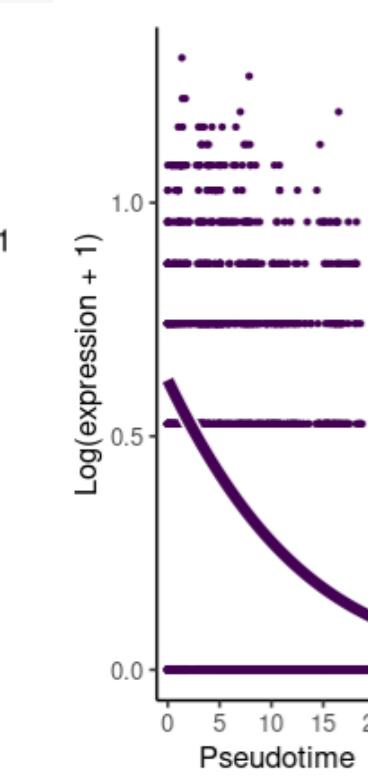
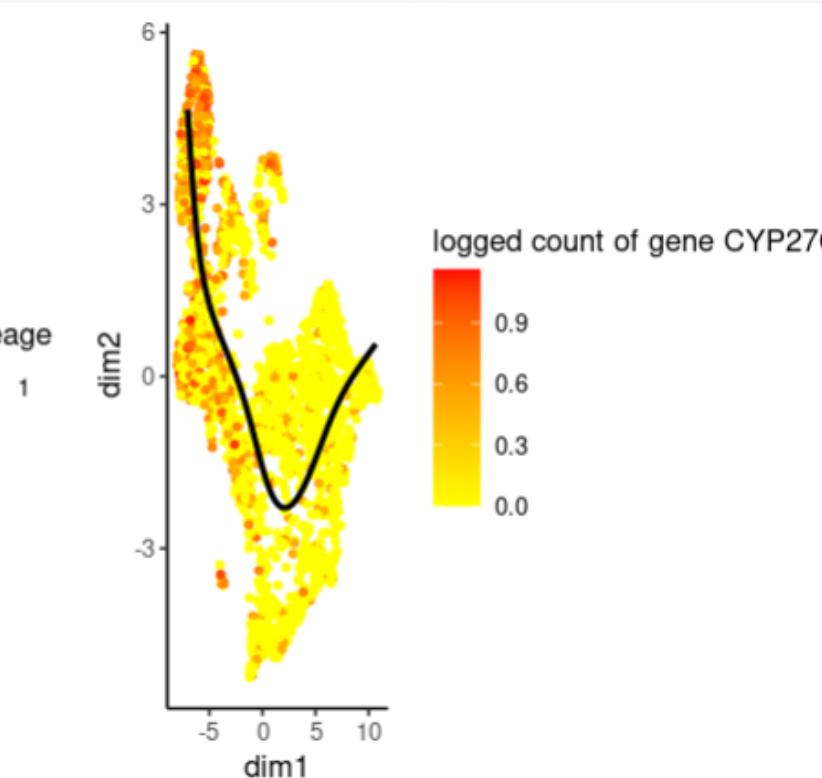
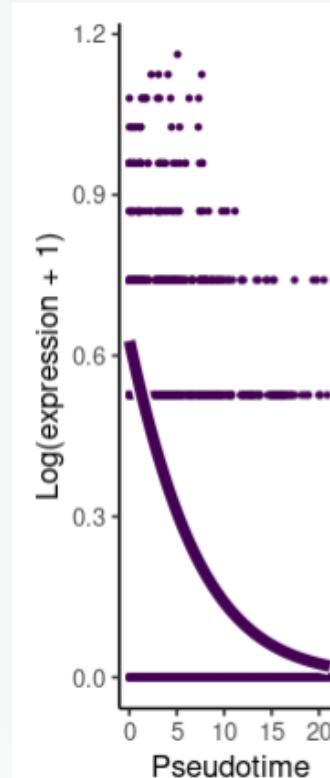


evaluateK output used to select knot value

TRAJECTORY ANALYSIS

Decreased expression over pseudotime

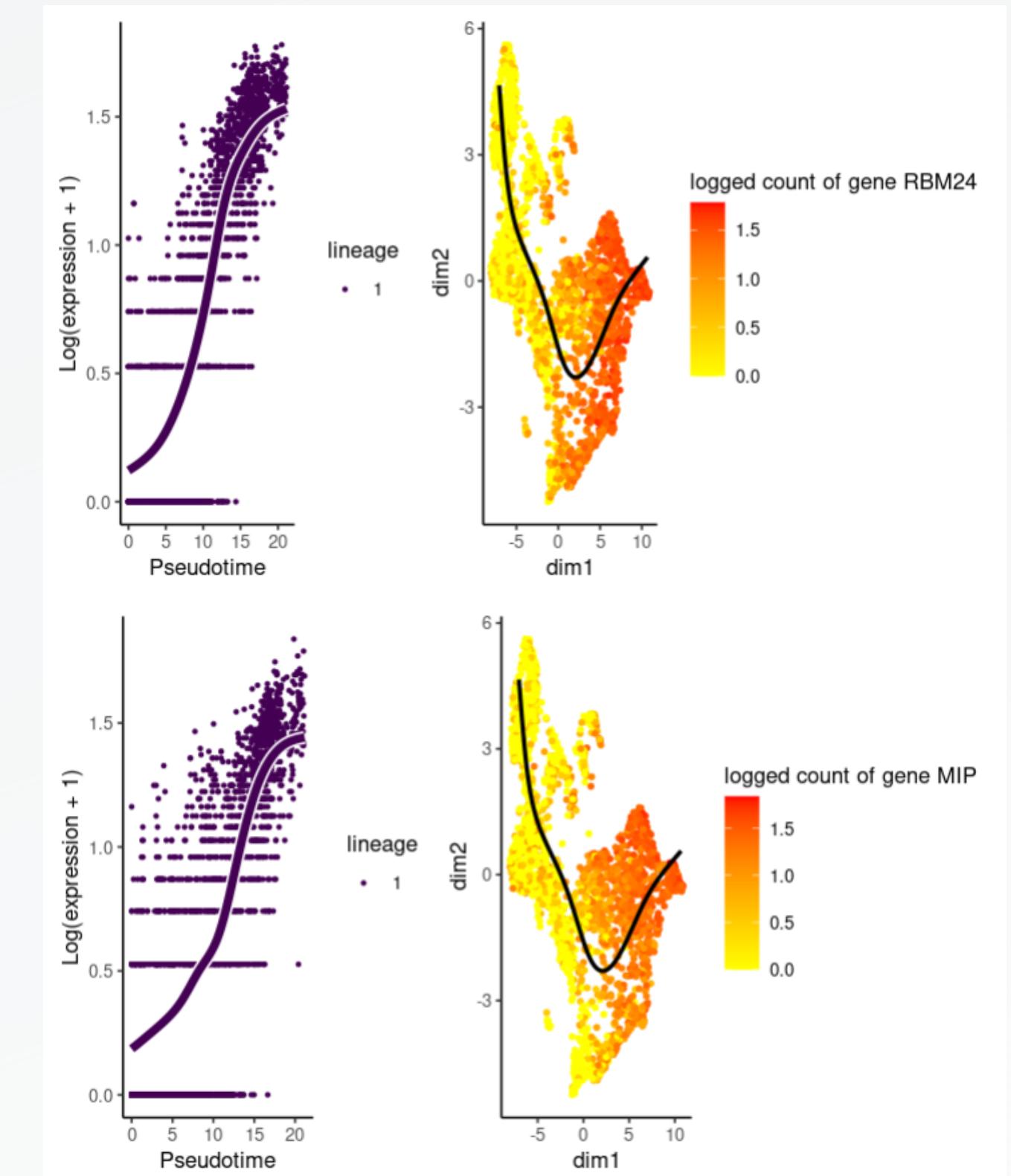
- Pseudotime represents cell differentiation from epithelial to fiber
- Among most significant genes are:
 - **FNDC3B** – Involved in protein binding
 - **CYP27C1** – Cytochrome P450
 - **ROR2** – receptor tyrosine kinase

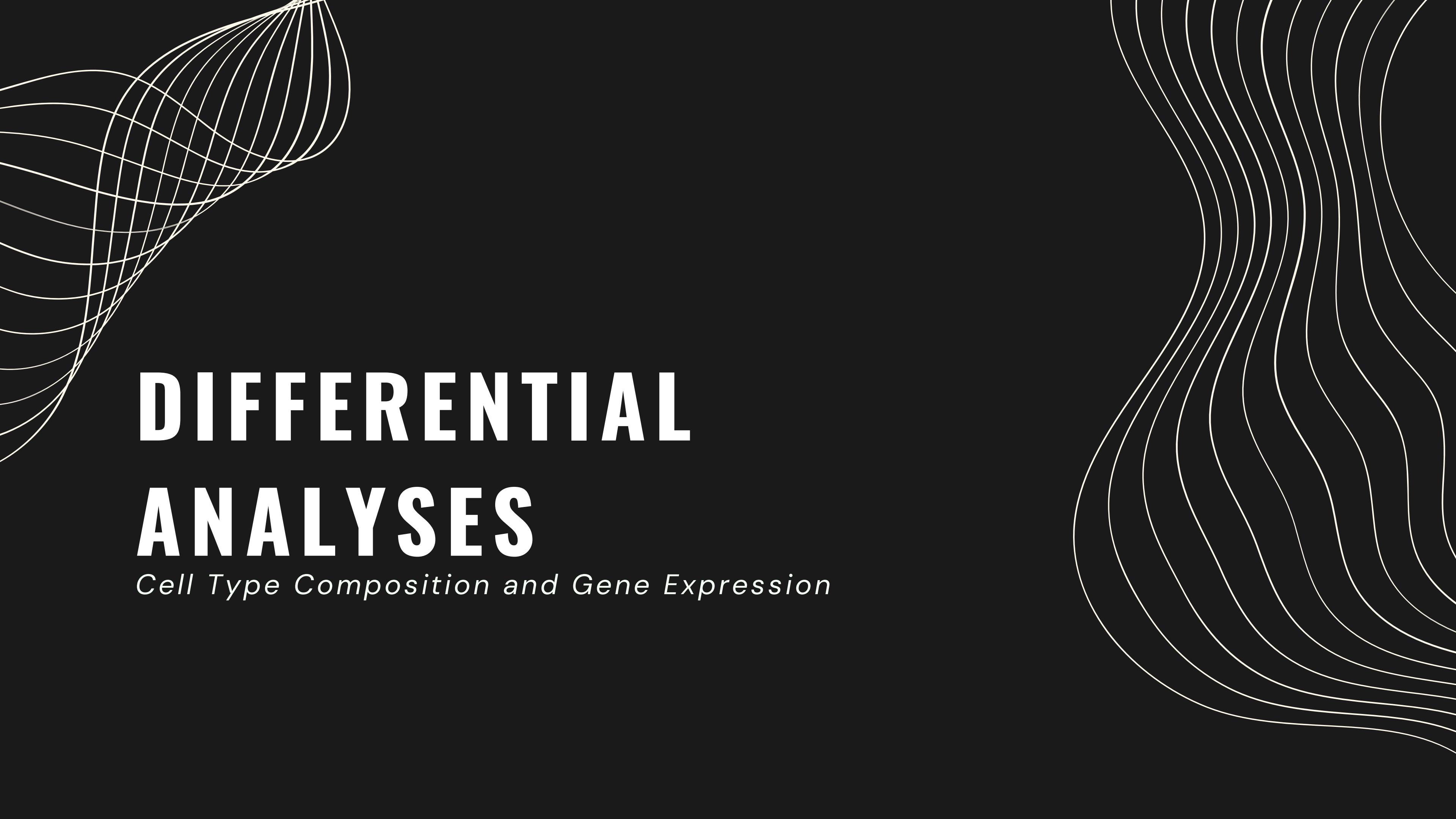


TRAJECTORY ANALYSIS

Increased expression over pseudotime

- Pseudotime represents cell differentiation from epithelial to fiber
- Among most significant genes are:
- **RBM24**
 - RNA binding motif
 - Negative regulation of cytoplasmic translation
 - Positive regulation of cell differentiation
 - Regulation of mRNA metabolic process
- **MIP**
 - Expressed in lens
 - Membrane channel protein





DIFFERENTIAL ANALYSES

Cell Type Composition and Gene Expression

BACKGROUND

- Based on differentiation results from *Tangeman et al* that aimed at analyzing the role of certain signaling factors in lens development
- Fibroblast growth factor (FGF) putatively associated with fiber cell development
 - Analyzing effect of FGF on embryos with removed retinas

Foundations

- Cell types annotated manually based on marker genes, as found in trajectory analysis (in correspondence with paper)
- Pooling across embryos of different ages
- Discarding samples from 24 h after retinectomy

DIFFERENTIAL CELL TYPE COMPOSITION

Foundations

- Using integrated Seurat object containing only ASL1-expressing cells
- Package: DCATS

Workflow

1

- Similarity matrix from SNN graph
- Count matrix for 3 cell types

2

- Design matrix with 2 columns
- Retinectomy or not
- FGF treatment or not

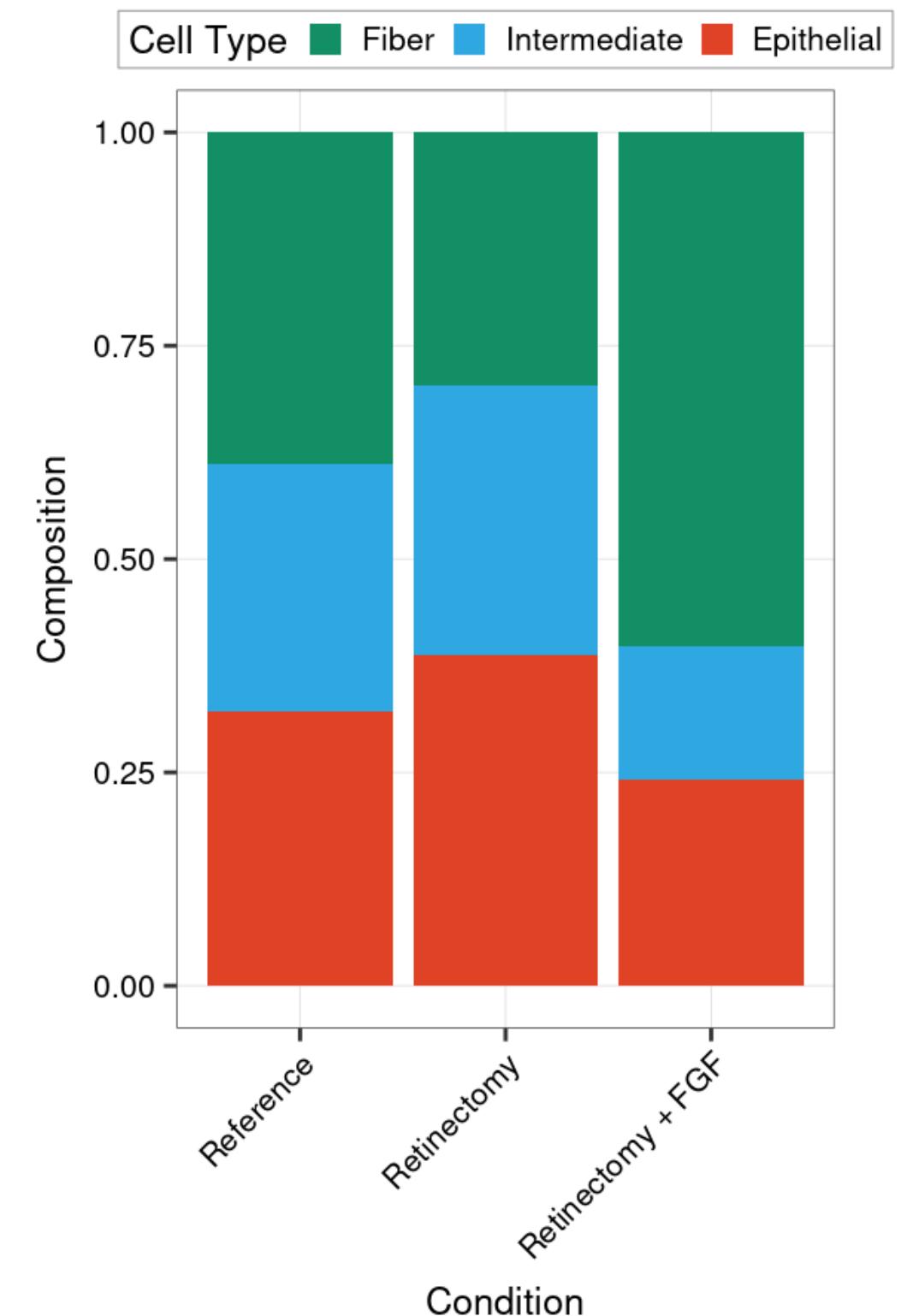
3

Fit of GLM

DIFFERENTIAL CELL TYPE COMPOSITION

Results

- Significant compositional change for epithelial & fiber cells
- Reduction of fiber cells via retinectomy
 - Recovery and even increase with FGF treatment



DIFFERENTIAL GENE EXPRESSION

Foundations

- Using integrated Seurat object with only ASL1-expressing cells
- Only 6h after retinectomy (no 24 h)
- Packages used: *Seurat*, *DESeq2*

Workflow

1

Pseudobulking per
cell type and sample
(embryo, condition)

2

Design matrix with 3 distinct
values:

- Reference
- Retinectomy
- Retinectomy + FGF

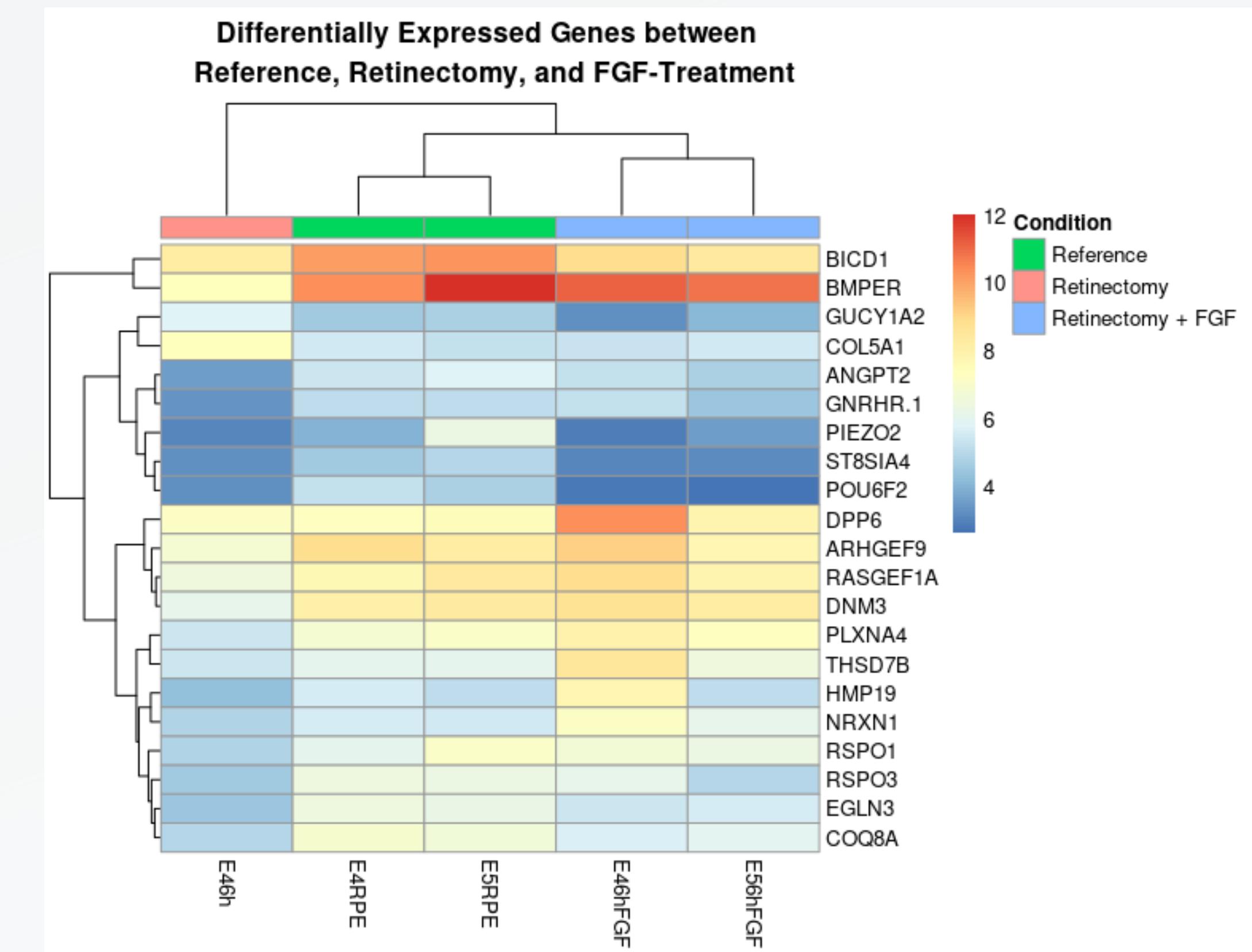
3

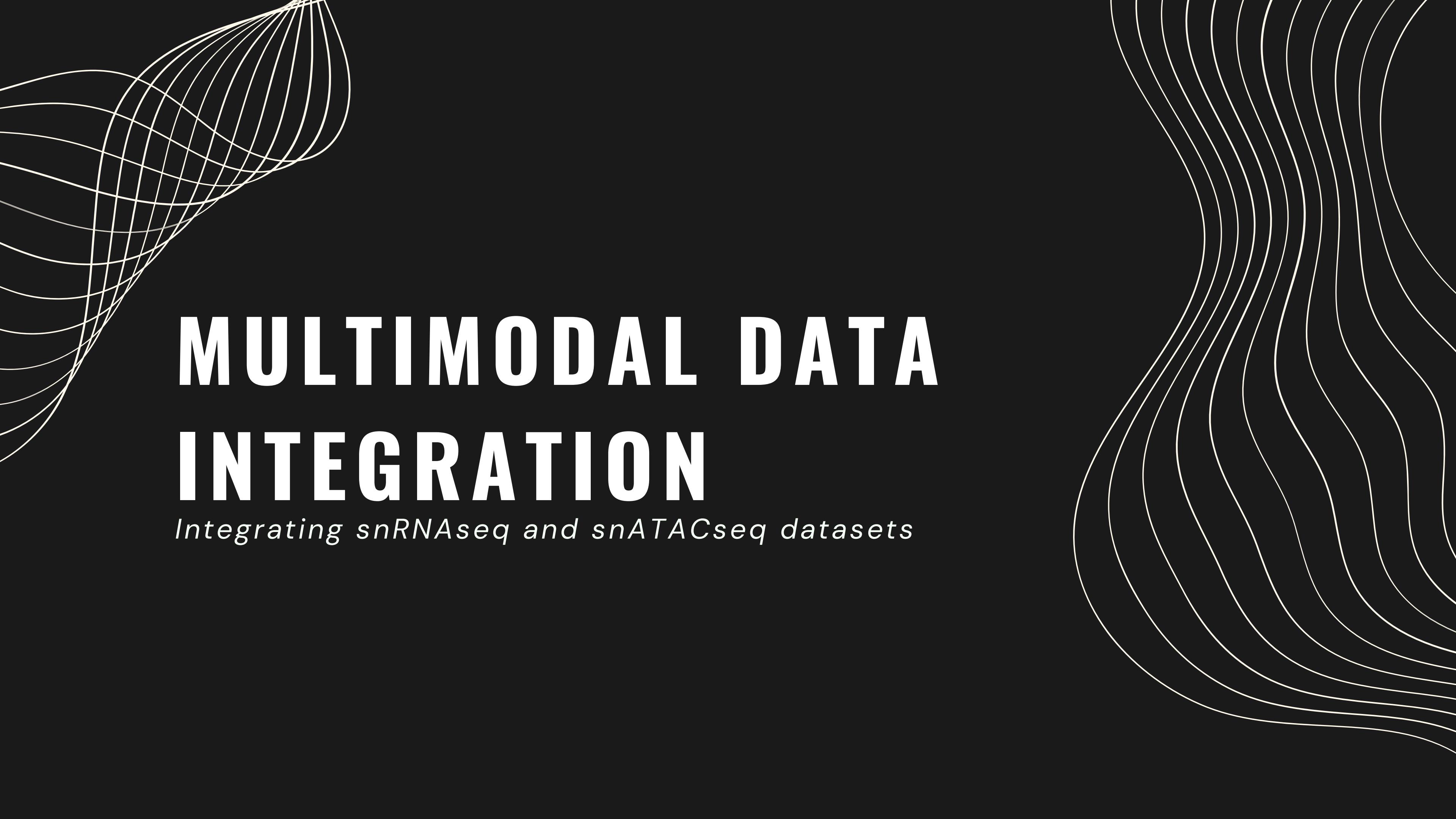
- Running *DESeq2*
- Good fit based on QC
criteria (PCA,
dispersion)

DIFFERENTIAL GENE EXPRESSION

Results

- Only few dozen significant genes found
 - Mostly fiber cells (focus)
- Good correspondence with known mechanisms/paper
- Recovery of genes lost to retinectomy by FGF induction





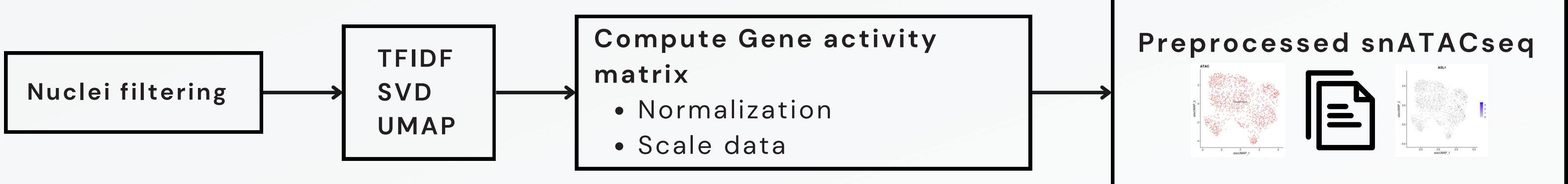
MULTIMODAL DATA INTEGRATION

Integrating snRNAseq and snATACseq datasets

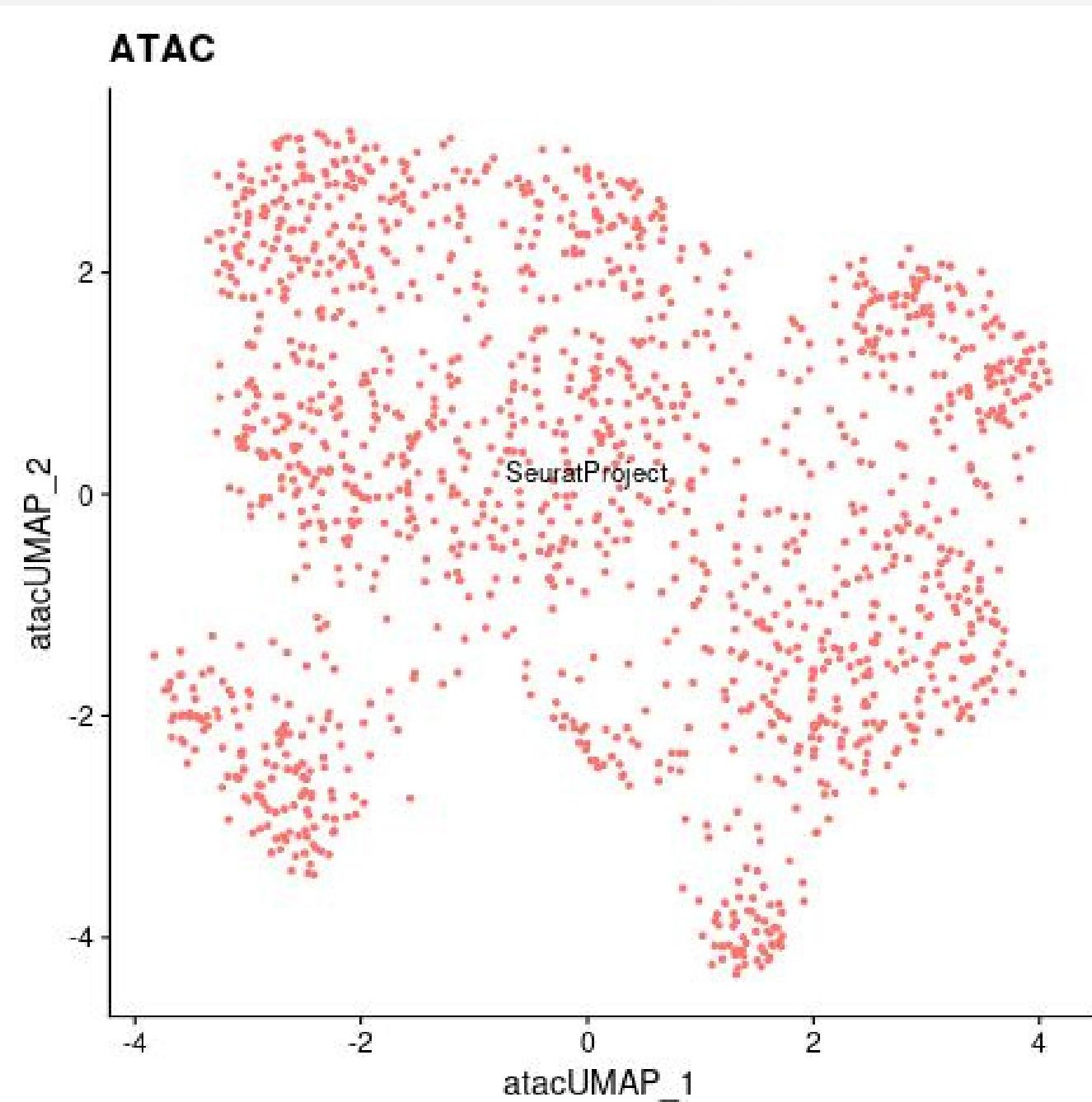
I. PREPROCESSING - WORKFLOW



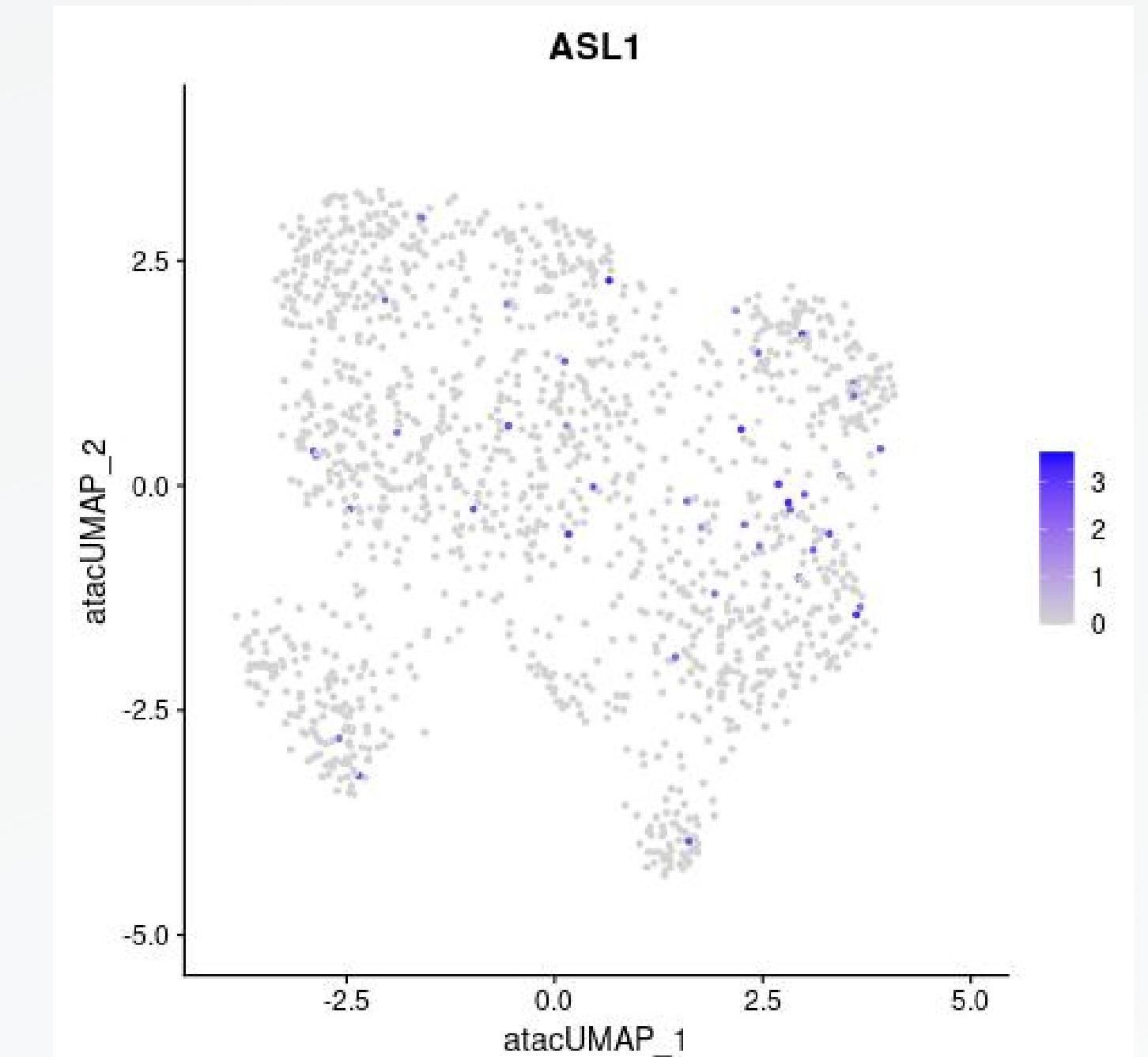
Workflow



I. PREPROCESSING - VISUALIZATION



Vizualitation of the UMAP results from the preproceesed
snATACseq data



Feature plot of cells expressing ASL1 as detected by
snATACseq according to the Gene activity matrix

II. MULTIMODAL DATA INTEGRATION

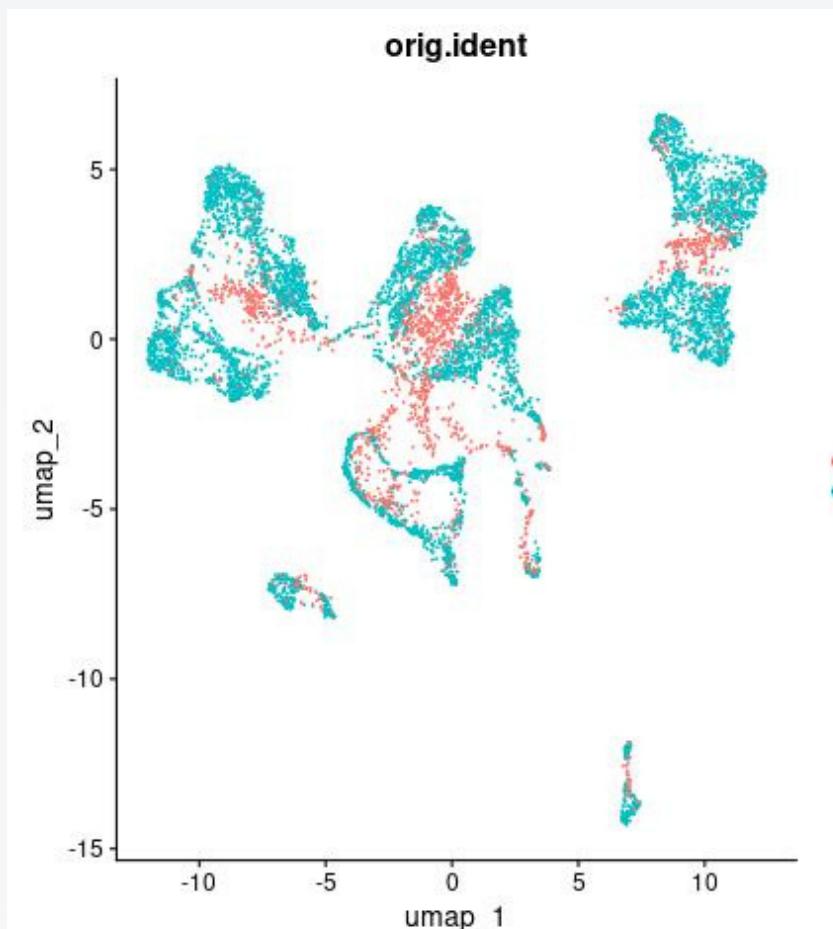
Workflow – part 1



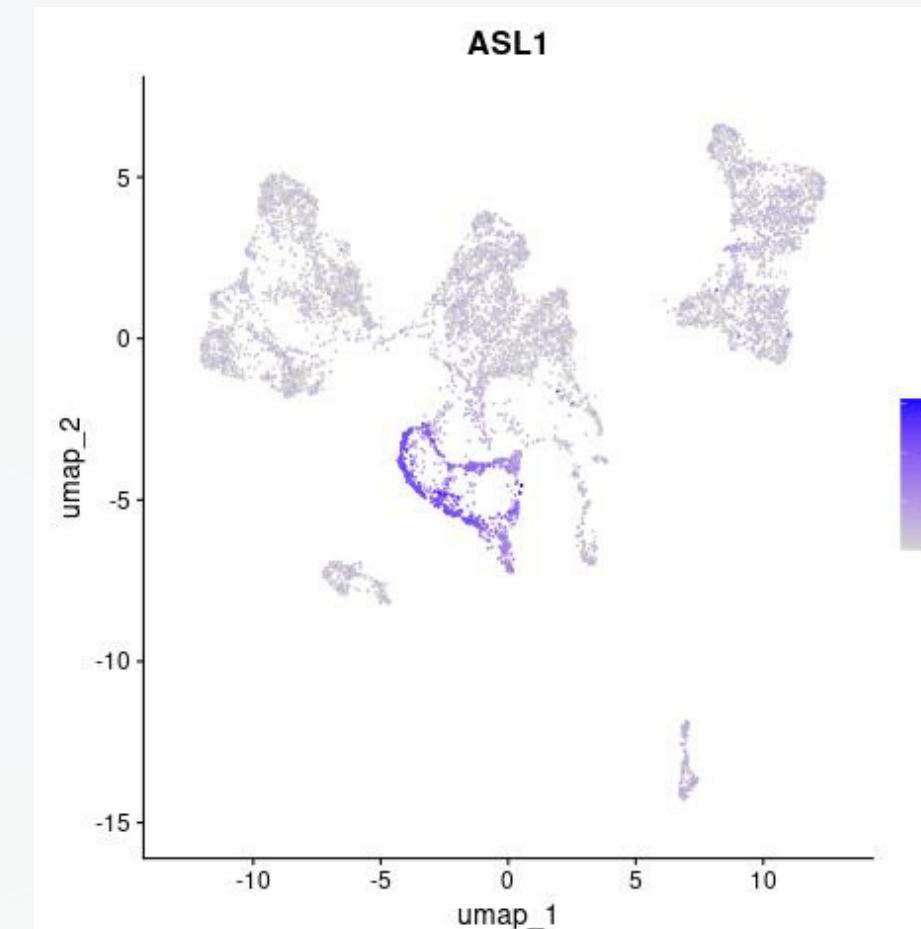
II. DATA INTEGRATION

Workflow - part 2

Visualize coembed data



Vizualitation of the UMAP results from the integrated seurat object (E4-6h)

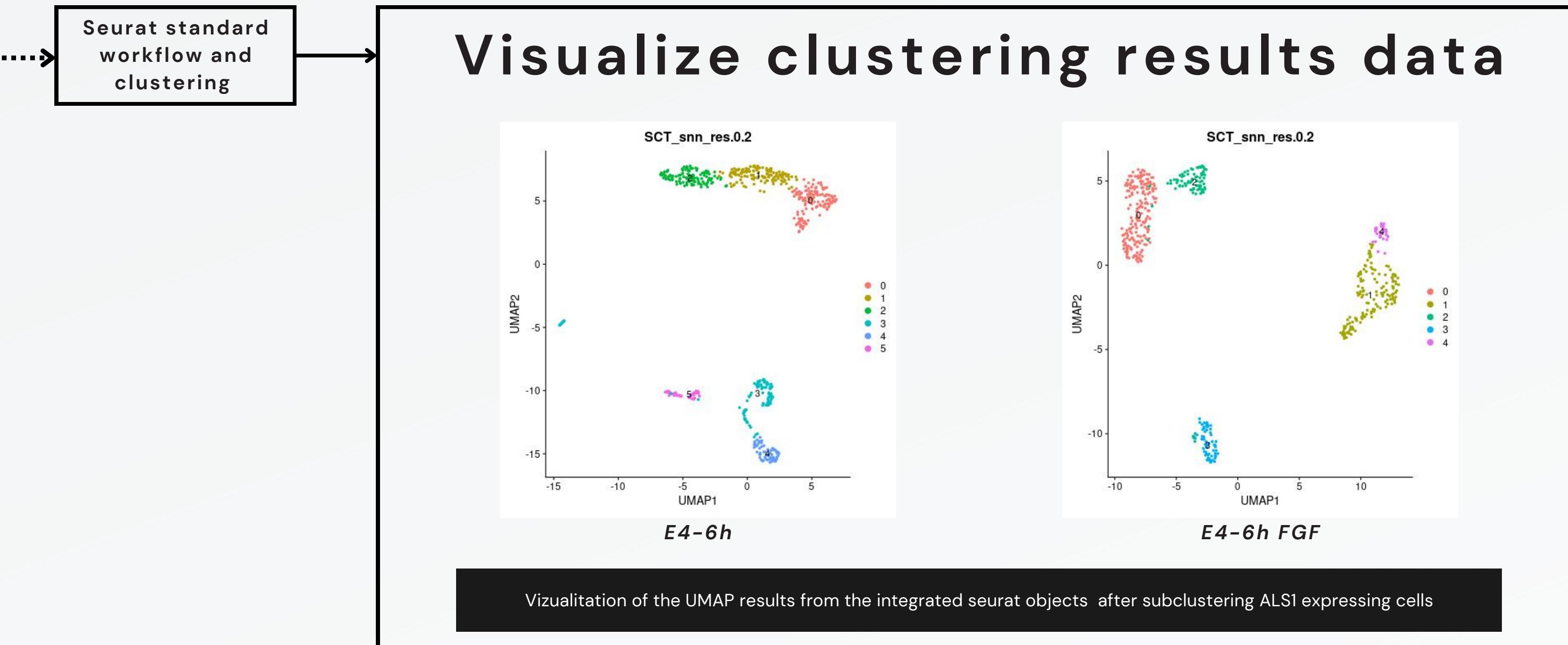


Feature plot of cells expressing ASL1 in the integrated seurat obejct (E4-6h)

Subset by clusters expressing ASL1

II. DATA INTEGRATION

Workflow - part 3





DISCUSSION & PERSPECTIVES

Review of what was done and what could be next

DISCUSSION

- Main findings from *Tangeman et al* replicated
 - Differentiating cell-type identified
 - Adding to the understanding of gene expression in lens tissue related to cataract
- Potential issues with filtering affecting downstream analyses
 - Not many genes with significant diff. expression
 - Solution: optimizing preprocessing, better cell type annotation
- snATAC data doesn't strong clustering patterns on its own
 - Solution: Fragment file could help refining the data

OUTLOOK

- Adding cell-cell communication analysis to further examine effect of retinectomy and FGF on signaling between cells
- From the integrated dataset, we would like to perform:
 - Cell label transfer
 - Trajectory analysis - QC for the trajectory analysis performed in this project
 - Gene regulatory network reconstruction to compare control and stimulated samples

BIBLIOGRAPHY

- Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & Satija, R. (2023). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*. <https://doi.org/10.1038/S41587-023-01767-Y>
- Lin, X., Chau, C., Ma, K., Huang, Y., & Ho, J. W. K. (2023). DCATS: differential composition analysis for flexible single-cell experimental designs. *Genome Biology*, 24(1), 1–21. <https://doi.org/10.1186/S13059-023-02980-3/FIGURES/4>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9>
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Iny Stein, T., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., & Lancet, D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*, 54(1), 1.30.1-1.30.33. <https://doi.org/10.1002/CPBI.5>
- Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1), 1–16. <https://doi.org/10.1186/S12864-018-4772-0/FIGURES/5>
- Tangeman, J. A., Rebull, S. M., Grajales-Esquivel, E., Weaver, J. M., Bendezu-Sayas, S., Robinson, M. L., Lachke, S. A., & Rio-Tsonis, K. Del. (2023). Integrated single-cell multiomics uncovers foundational regulatory mechanisms of lens development and pathology. *BioRxiv*. <https://doi.org/10.1101/2023.07.10.548451>
- Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., & Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications* 2020 11:1, 11(1), 1–13. <https://doi.org/10.1038/s41467-020-14766-3>