

# Transcriptomics Exam

# Bulk RNAseq

# Aim of the project

## FINDING DIFFERENTIALLY EXPRESSED (DE) GENES

We selected 3 samples for **Brain**, **Colon** and **Liver** from the **Gtex** repository.

After a Quality control check we used **edgeR** to perform the DE analysis.

## EVALUATE DIFFERENT WAYS TO PERFORM DIFFERENTIAL EXPRESSION ANALYSIS.

Furthermore we want to asses if the methods used for identifying DE genes are **robust** enough to be reliable even without cleaning **pseudogenes**, **non canonical chromosomes**, **mitochondrial genes**, **rRNA genes**.

# Brain Quality Control

	RIN	rRNA	%Mapping
Sample 90	<b>6.8</b>	<b>0.065</b>	<b>90</b>
Sample 91	<b>7.2</b>	<b>0.059</b>	<b>89</b>
Sample 92	<b>7.4</b>	<b>0.067</b>	<b>89</b>

# Colon Quality Control

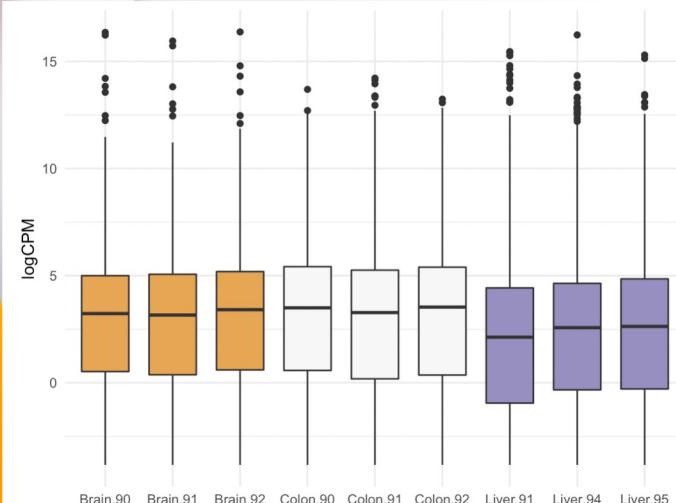
	RIN	rRNA	%Mapping
Sample 90	<b>7.8</b>	<b>0.01</b>	<b>89</b>
Sample 91	<b>6.5</b>	<b>0.01</b>	<b>90</b>
Sample 92	<b>7.8</b>	<b>0.006</b>	<b>92</b>

# Liver Quality Control

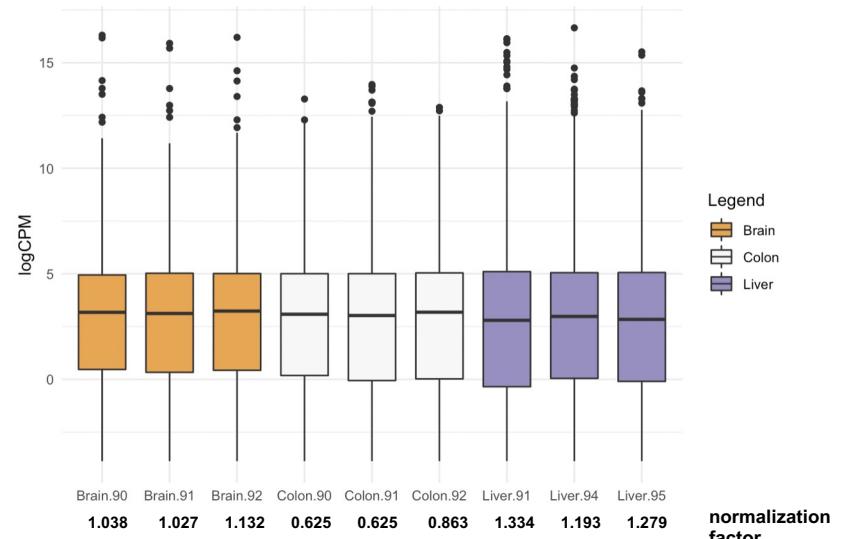
	RIN	rRNA	%Mapping
Sample 91	<b>9.1</b>	<b>0.007</b>	<b>90</b>
Sample 94	<b>6.9</b>	<b>0.012</b>	<b>89</b>
Sample 95	<b>7.4</b>	<b>0.03</b>	<b>91</b>

# Normalization of raw data

Before



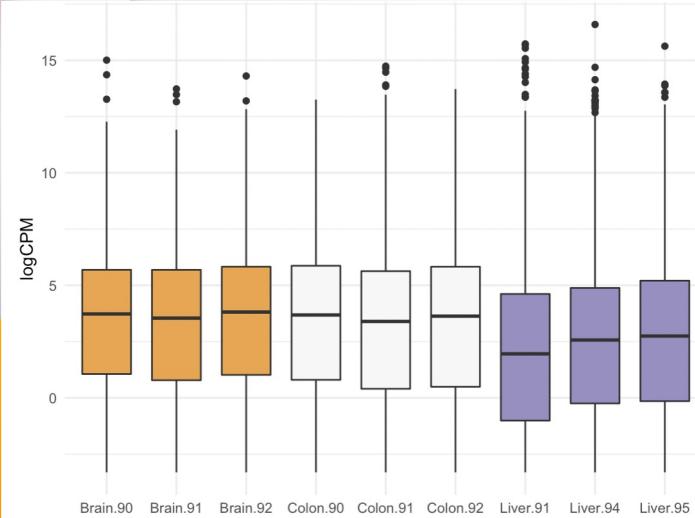
After



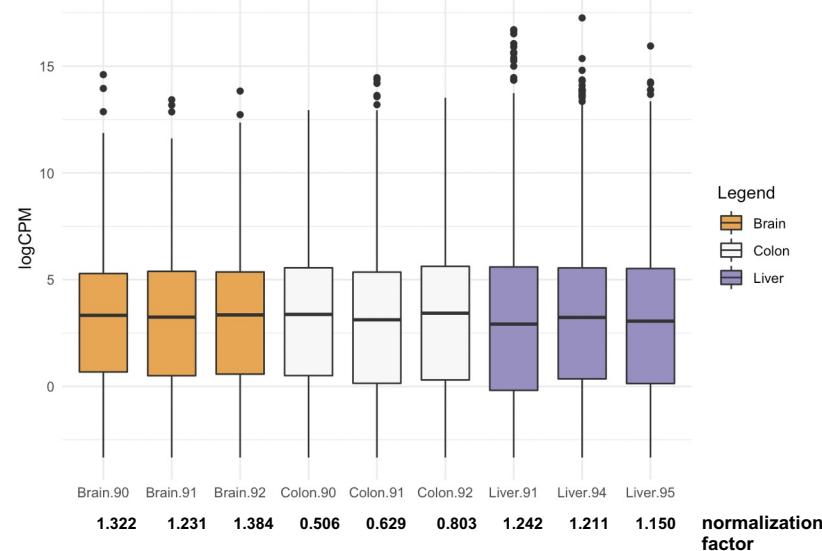
Normalization was performed using the **Trimmed Means of M values (TMM)** method with default parameters ( $M = 30\%$ ,  $A = 5\%$ ). We are assuming that **most of the genes** will have **equal or nearly equal counts** in the pairwise comparisons of the different samples.

# Normalization of clean data

Before



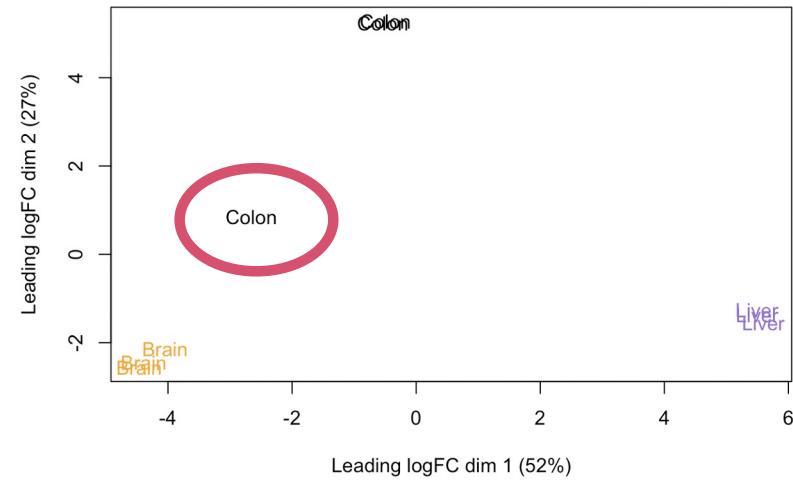
After



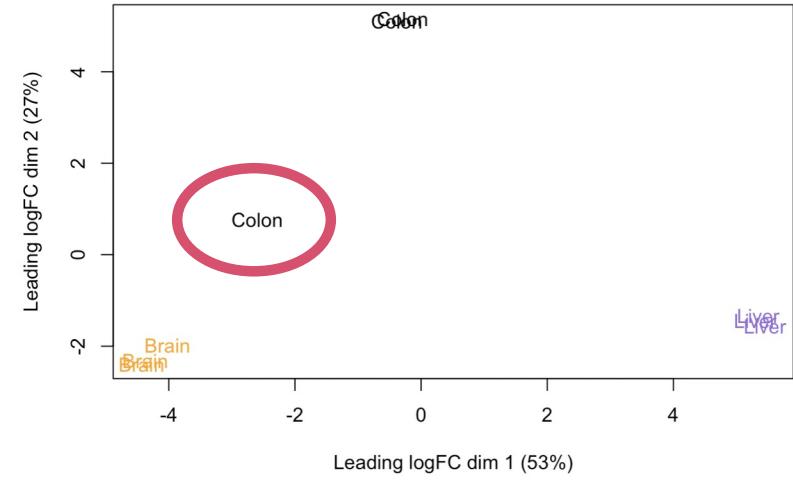
As before we used **TMM** method with default parameters ( $M = 30\%$ ,  $A = 5\%$ ).

# Multidimensional Scaling plot

Raw



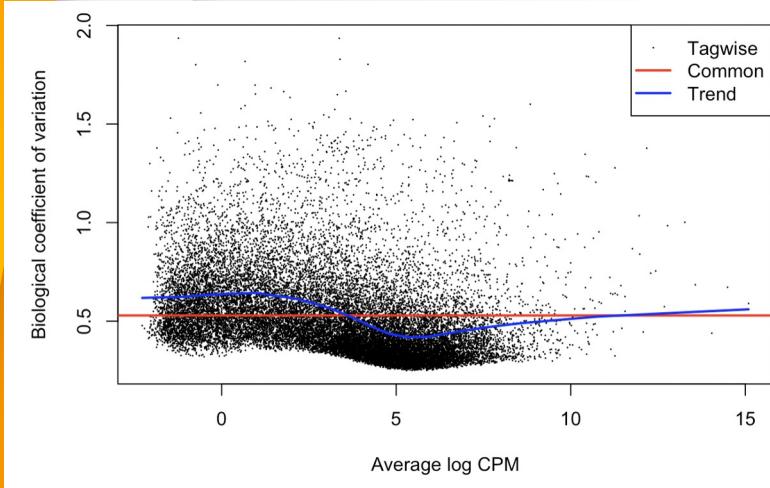
Clean



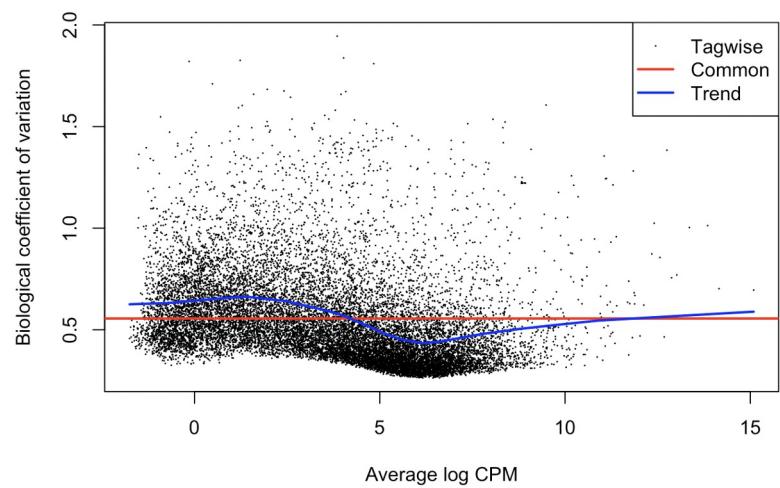
The **MultiDimensional Scaling (MDS) plot** enable us to visualize the degree of similarity between different samples . The distance used is **log-fold ratio** computed accounting for the **top 500 variable genes**.

# Biological Coefficient of Variation

Raw



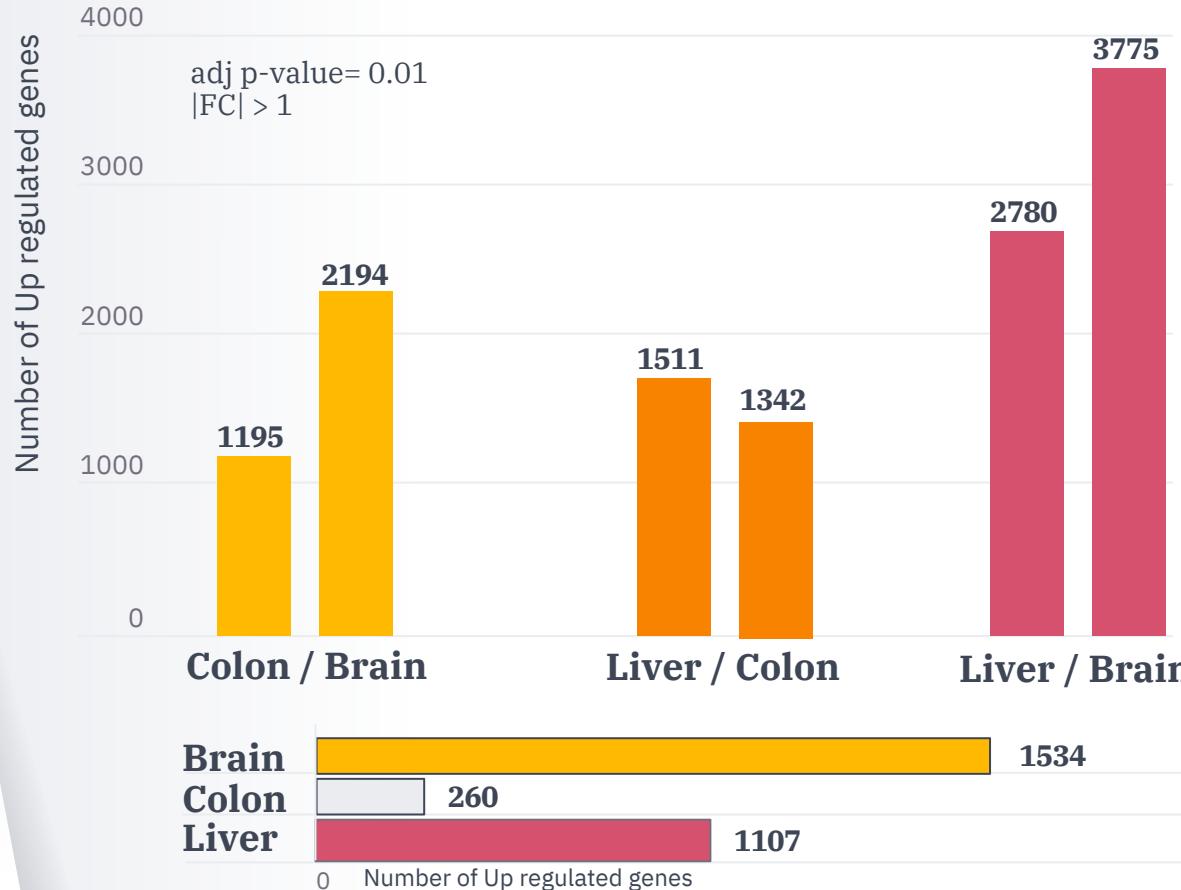
Clean



The **Biological Coefficient of Variation (BCV)** plot is a key step of DE analysis since the number of **counts** is the result of a **random process**. So it has to be described by a random variable with a Negative Binomial distribution to account for both **technical and biological variation**.

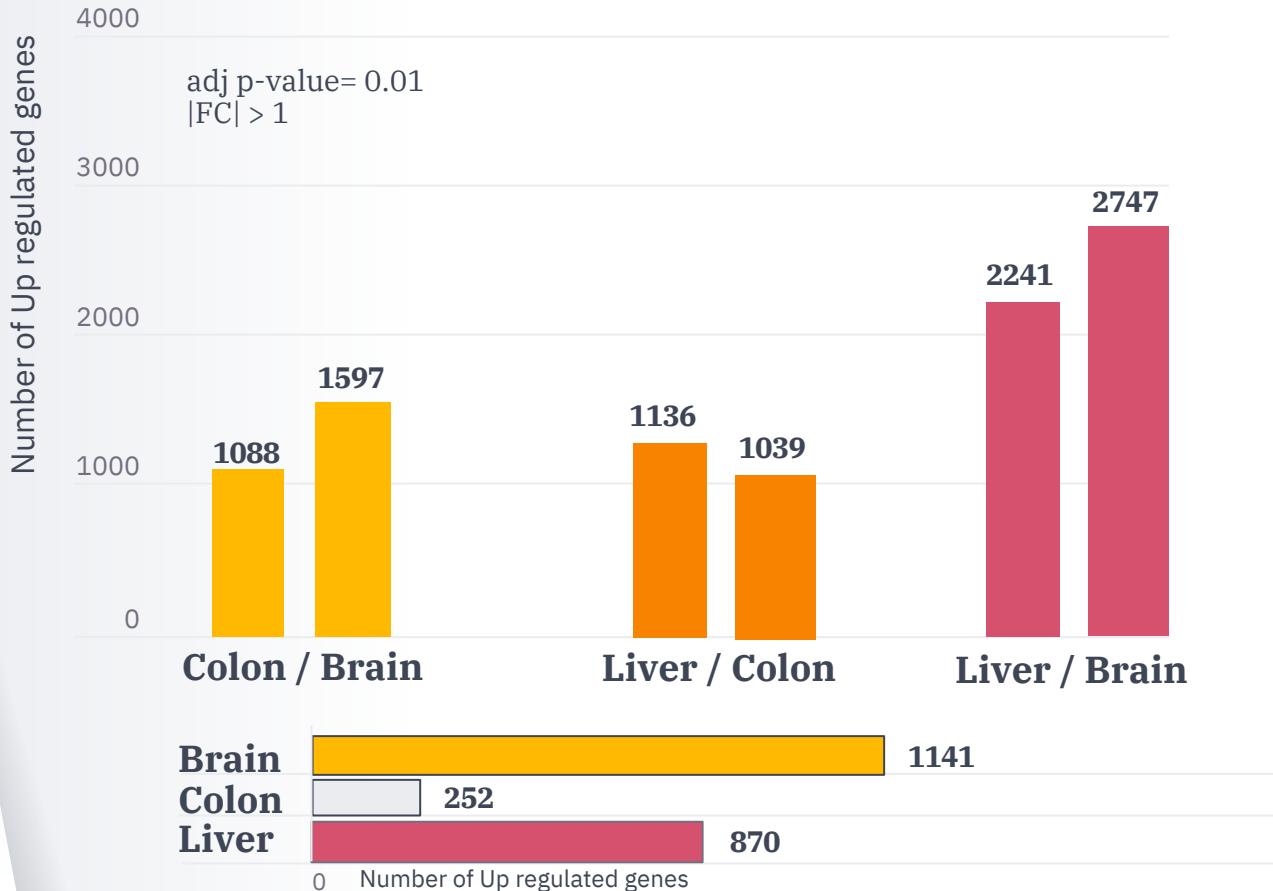
# Differentially Expressed Genes

## Raw data



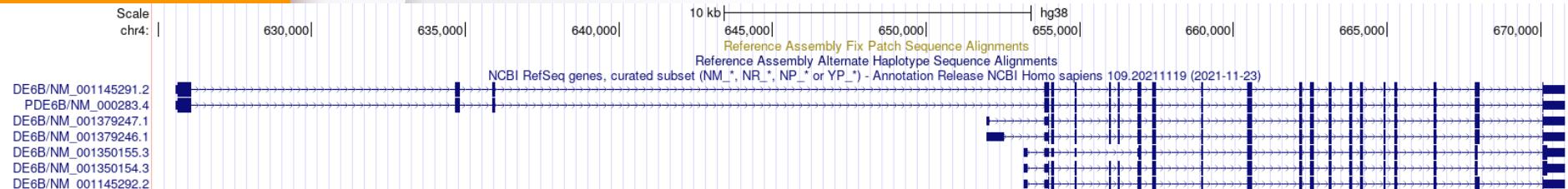
# Differentially Expressed Genes

## Clean data



# Example Gene: PDE6B

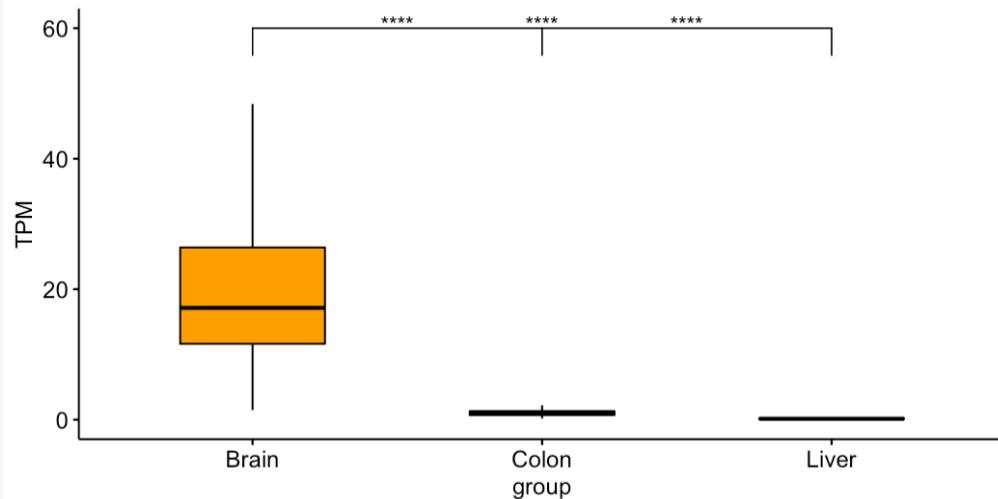
Transmission and amplification of visual signal



UCSC Genome Browser on Human (GRCh38/hg38)

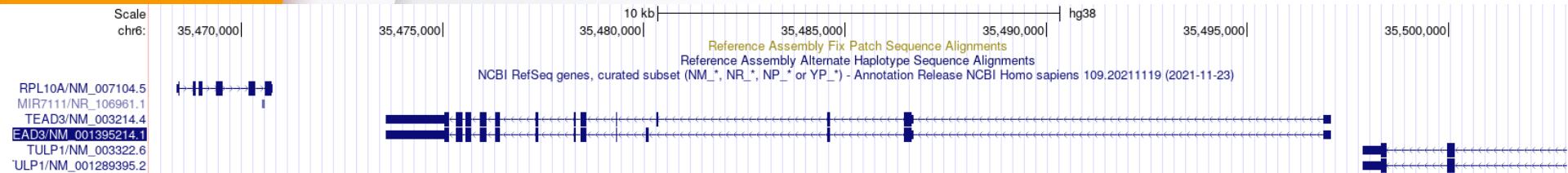
## PDE6B expression across organs

Kruskal-Wallis,  $\chi^2(2) = 2392.37$ ,  $p = <0.0001$ ,  $n = 4004$



# Example Gene: TEAD3

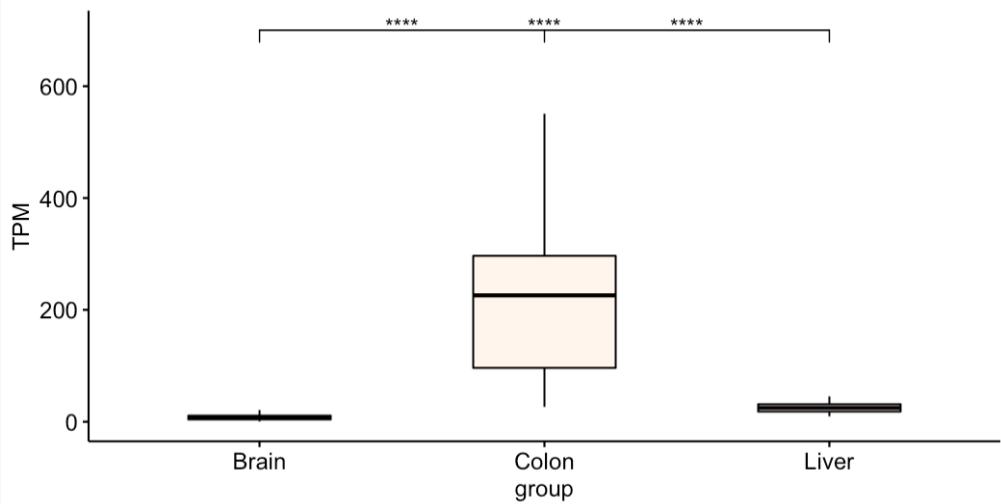
Transcriptional enhancer factor (TEF) family of transcription factors  
Translation of this protein is initiated at a **non-AUG (AUA)** start codon



UCSC Genome Browser on Human (GRCh38/hg38)

TEAD3 expression across organs

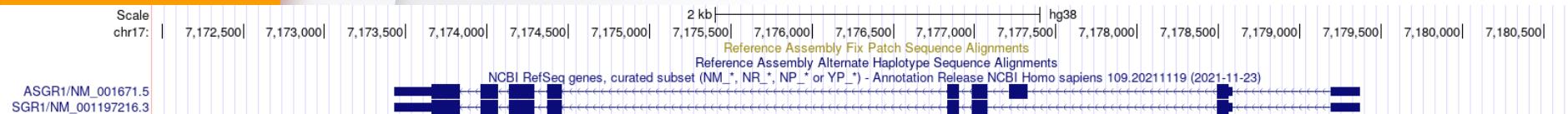
Kruskal-Wallis,  $\chi^2(2) = 2317.46, p < 0.0001, n = 4004$



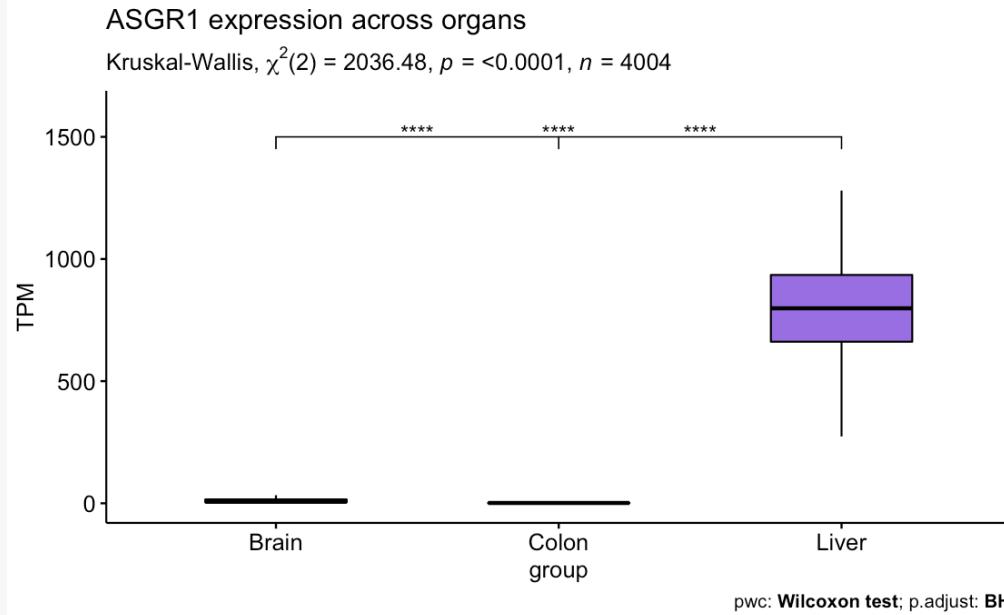
# Example Gene: ASGR1

Encodes a subunit of the **asialoglycoprotein receptor**.

The asialoglycoprotein receptor may facilitate **hepatic infection** by multiple viruses including hepatitis B, and is also a **target for liver-specific drug delivery**.



UCSC Genome Browser on Human (GRCh38/hg38)



# Functional enrichment analysis

## BRAIN: Raw data

### BioPlanet 2019

- NEURONAL SYSTEM
- TRANSMISSION ACROSS CHEMICAL RECEPTORS

### GO biological process

- CHEMICAL SYNAPTIC TRANSMISSION

### WikiPathway

#### 2021 Human

- SYNAPTIC VESICLE PATHWAY WP2

### GO Molecular Function 2021

- VOLTAGE-GATED CATION CHANNEL

### KEGG 2021

#### Human

- NICOTINE ADDICTION

### GO Cellular Component 2021

- NEURON PROJECTION

# Functional enrichment analysis

## BRAIN: Clean data

### BioPlanet 2019

- NEURONAL SYSTEM
- TRANSMISSION ACROSS CHEMICAL RECEPTORS

### GO biological process

- CHEMICAL SYNAPTIC TRANSMISSION

### WikiPathway

#### 2021 Human

- GABA RECEPTOR SIGNALING WP4
- SYNAPTIC VESICLE PATHWAY WP2

### GO Molecular Function 2021

- VOLTAGE-GATED CATION CHANNEL

### KEGG 2021

#### Human

- NICOTINE ADDICTION

### GO Cellular Component 2021

- NEURON PROJECTION

adj p-value < 1e-7

Note: We removed LOC..., LIN..., MIR..., SNORD..., RPL... gene symbol

# Functional enrichment analysis

## COLON: Raw data

### BioPlanet 2019

- ECM-RECEPTOR INTERACTION

### GO biological process

- EXTRACELLULAR STRUCTURE ORGANIZATION
- EXTERNAL ENCAPSULATING STRUCTURE

### WikiPathway

#### 2021 Human

- FOCAL ADHESION WP306

### GO Molecular Function 2021

- PLATELET-DERIVED GROWTH FACTOR

### KEGG 2021

#### Human

- FOCAL ADHESION
- ECM-RECEPTOR INTERACTION
- PROTEIN DIGESTION AND ABSORPTION

### GO Cellular Component 2021

- COLLAGEN-CONTAINING EXTRACELLULAR MATRIX

# Functional enrichment analysis

## COLON: Clean data

Pathways  
Ontologies

### BioPlanet 2019

- TGF-beta  
**REGULATION OF ECM**
- ECM-RECEPTOR  
INTERACTION

### GO biological process

- **SUPRAMOLECULAR FIBER ORGANIZATION**
- EXTRACELLULAR STRUCTURE ORGANIZATION
- EXTERNAL ENCAPSULATING STRUCTURE

### WikiPathway 2021 Human

- FOCAL ADHESION  
WP306

### GO Molecular Function 2021

- PLATELET-DERIVED GROWTH FACTOR

### KEGG 2021

#### Human

- PROTEOGLYCANS IN CANCER
- FOCAL ADHESION
- PROTEIN DIGESTION AND ABSORPTION
- ECM-RECEPTOR INTERACTION

### GO Cellular Component 2021

- COLLAGEN-CONTAINING EXTRACELLULAR MATRIX

adj p-value < 1e-7

Note: We removed LOC..., LIN..., MIR..., SNORD..., RPL... gene symbol

# Functional enrichment analysis

## LIVER: Raw data

### BioPlanet 2019

- METABOLISM

### GO biological process

- CELLULAR AMINO ACID CATABOLISM

### WikiPathway 2021 Human

- COMPLEMENT AND COAGULATION

### GO Molecular Function 2021

- STEROID HYDROXYLASE ACTIVITY

### KEGG 2021

#### Human

- COMPLEMENT AND COAGULATION
- DRUG METABOLISM
- METABOLISM OF XENOBIOTICS

### GO Cellular Component 2021

- INTRACELLULAR ORGANELLE LUMEN

# Functional enrichment analysis

## LIVER: Clean data

### BioPlanet 2019

- METABOLISM

### GO biological process

- STEROID METABOLIC PROCESS
- CELLULAR AMINO ACID CATABOLISM

adj p-value < 1e-7

Note: We removed LOC..., LIN..., MIR..., SNORD..., RPL... gene symbol

### WikiPathway 2021 Human

- COMPLEMENT AND COAGULATION
- METAPATHWAY BIOTRANSFORMATION

### GO Molecular Function 2021

- STEROID HYDROXYLASE ACTIVITY

### KEGG 2021 Human

- COMPLEMENT AND COAGULATION
- DRUG METABOLISM
- METABOLISM OF XENOBIOTICS

### GO Cellular Component 2021

- INTRACELLULAR ORGANELLE LUMEN

## Results

- ▶ In the end we were able to **identify** correctly all the **tissues** from the Up regulated genes in the respective tissues. And to demonstrate that the methods employed are **robust** enough to identify DE genes also in raw data.



# scRNAseq

# Retrieving the data:

PanglaoDB:

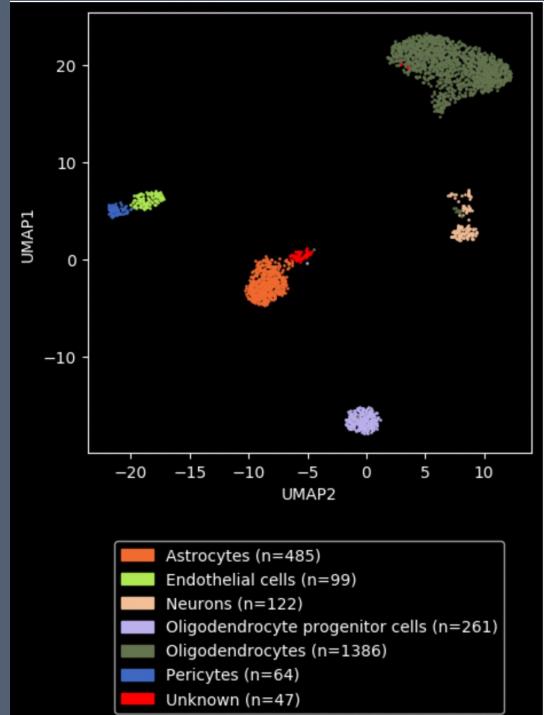
SRA: SRA850958

Species: **Homo Sapiens**

Tissue: **Substantia Nigra**

Number of cells: **2,432**

Protocol: **10x chromium**

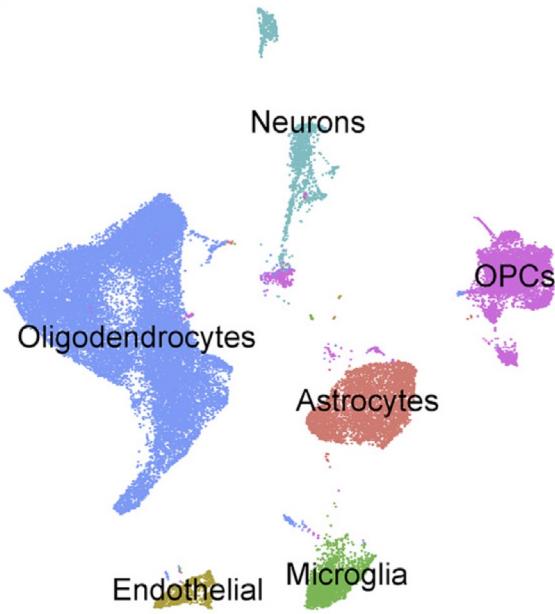


# Understanding the data

Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. Welch et al., 2019, Cell 177, 1873–1887 June 13, 2019 a 2019 Elsevier Inc.

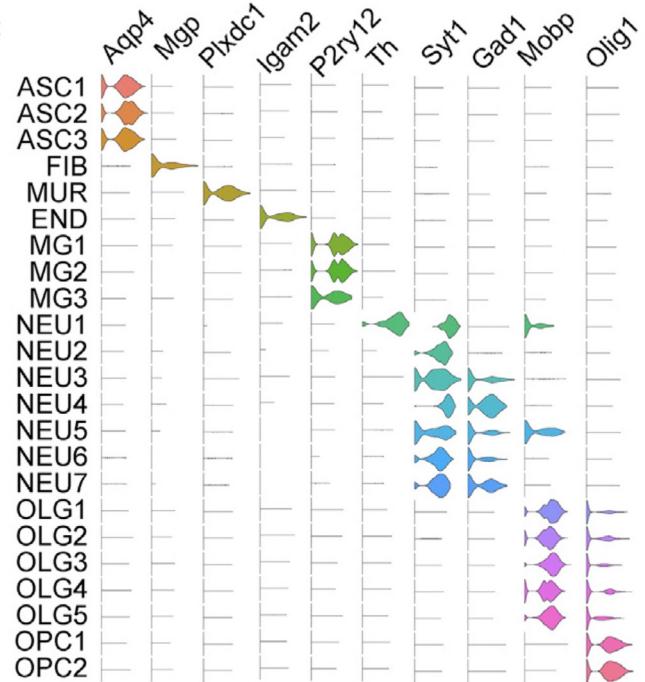
UMAP

B

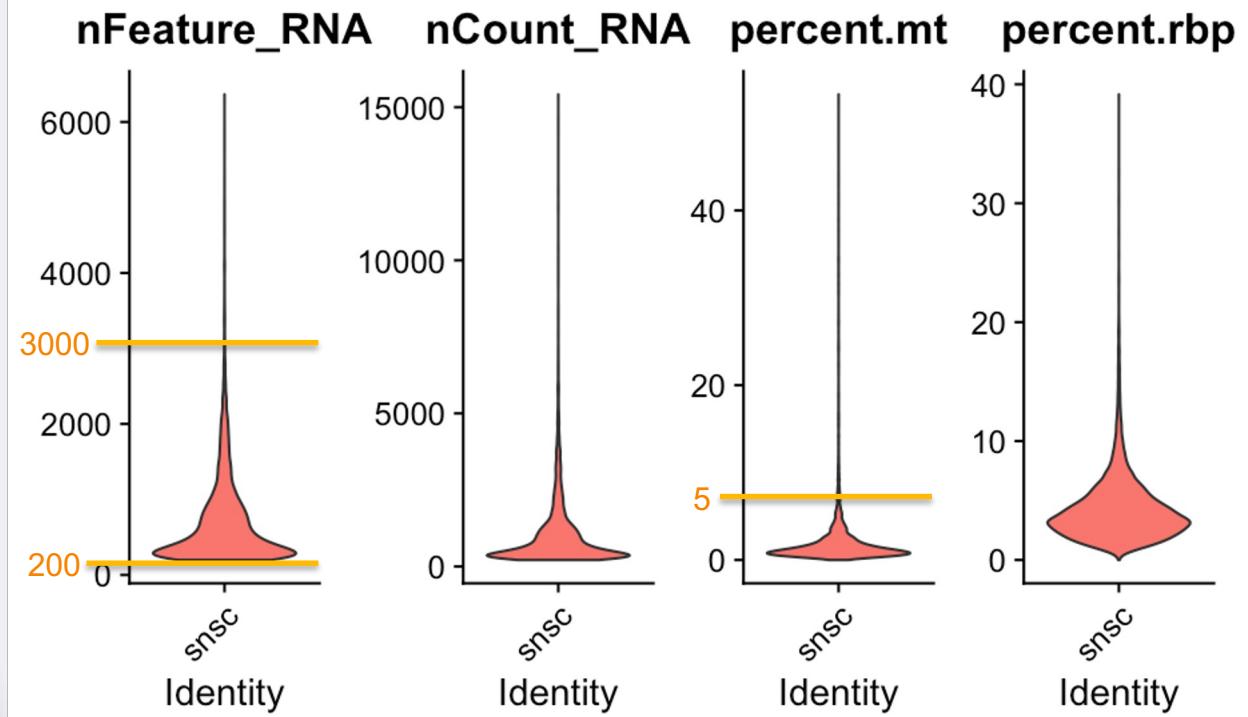


Marker genes

C

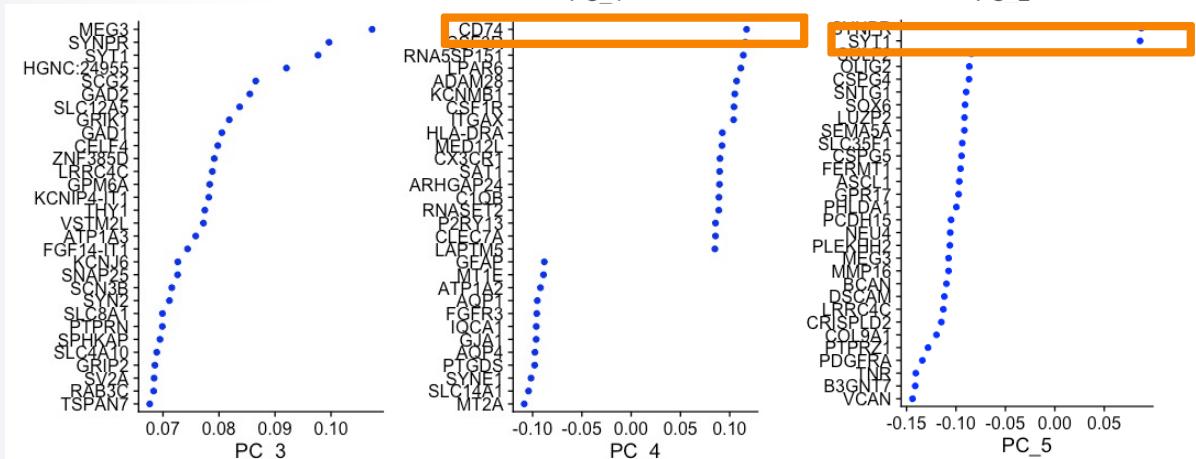
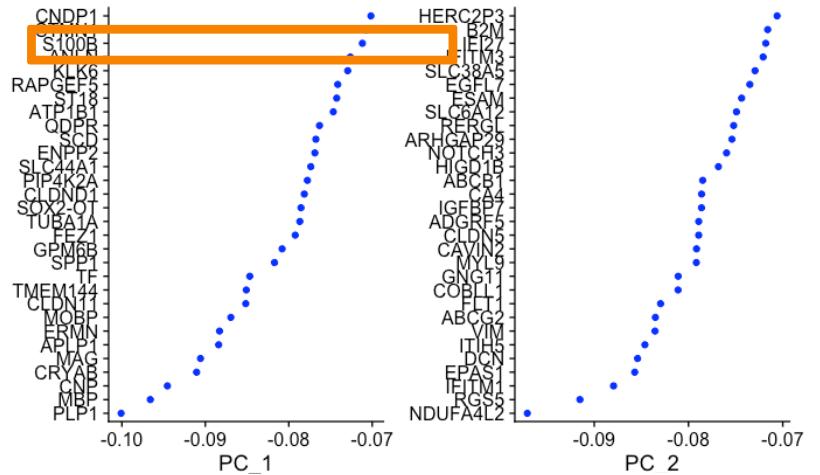


# Cell Quality Control

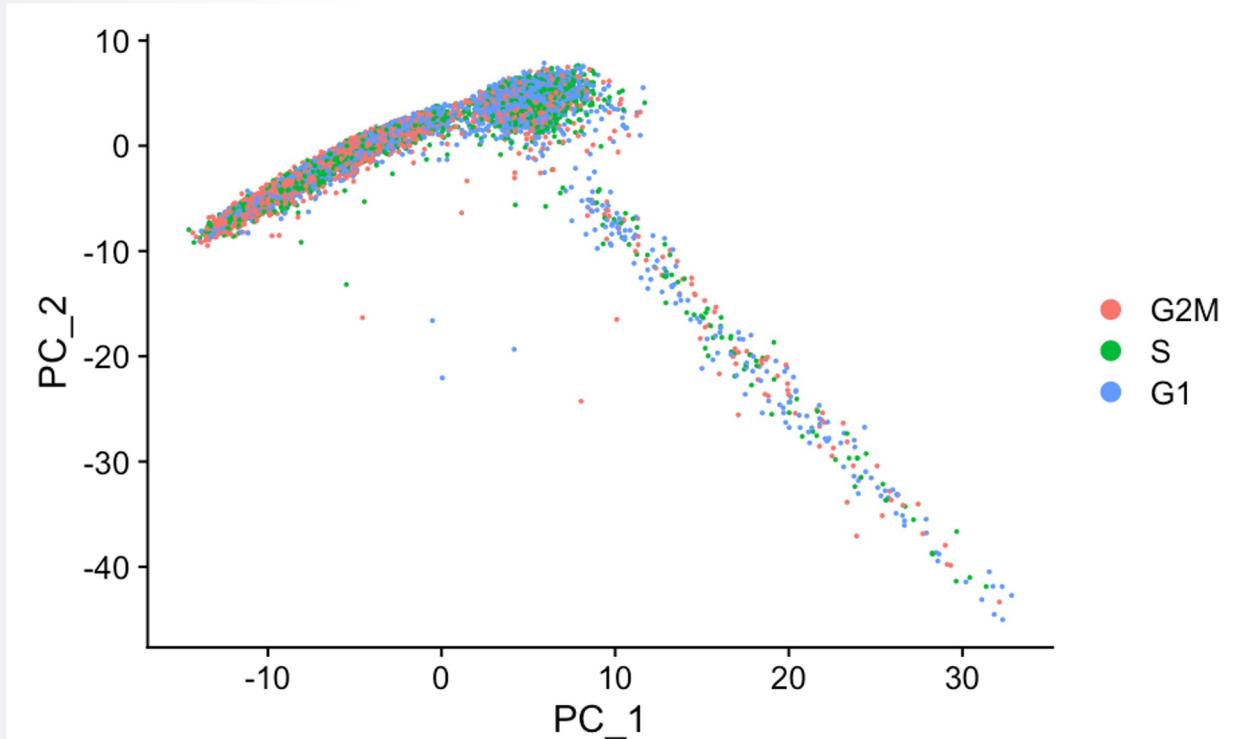


Discarded 621 cells (8,8%)

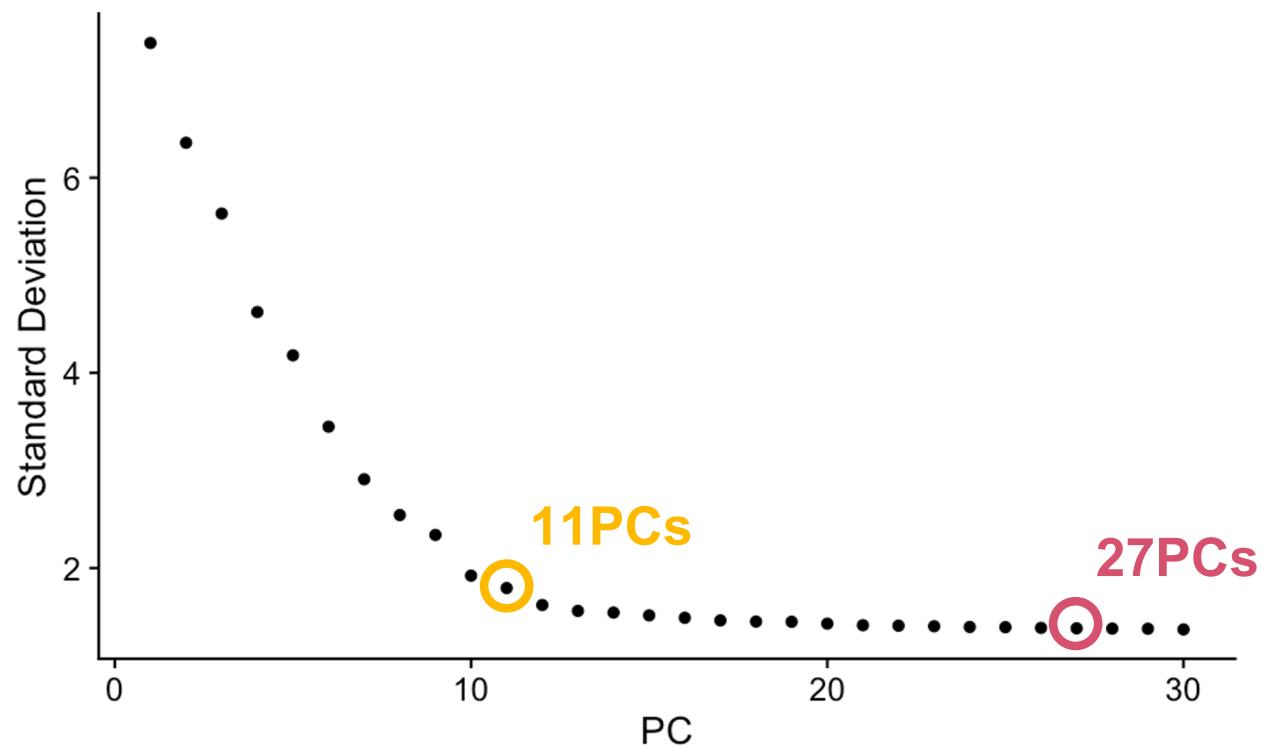
# The most variable genes



# Assessing the role of Cell Cycle

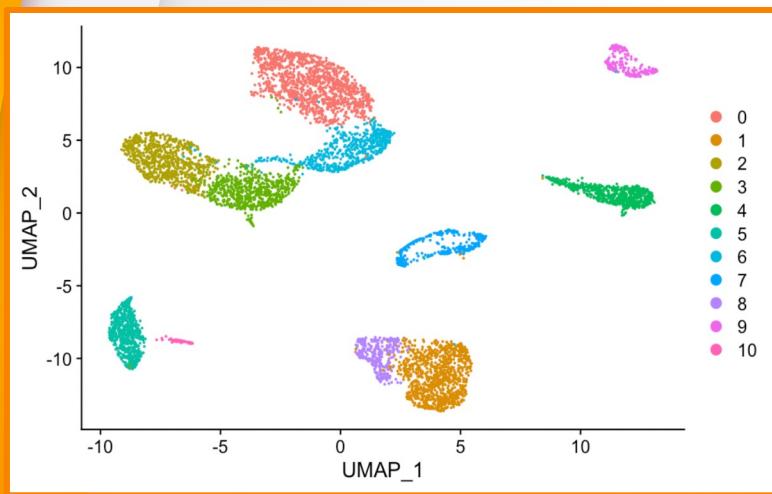


# The Elbow Plot

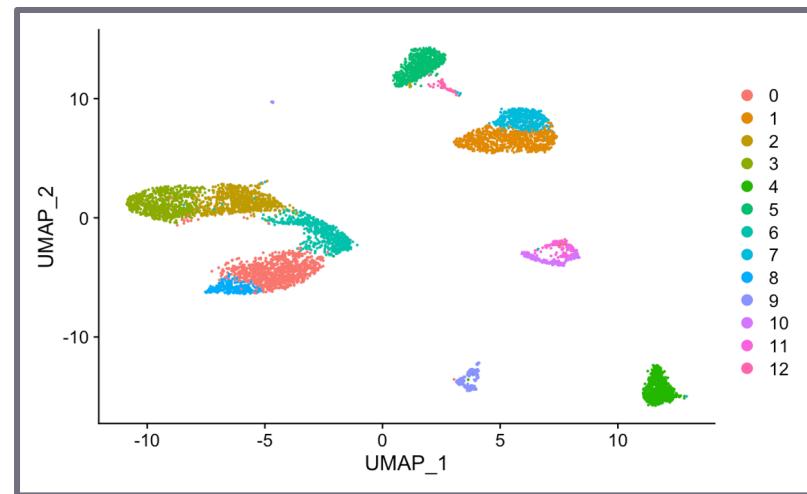


# Clustering results

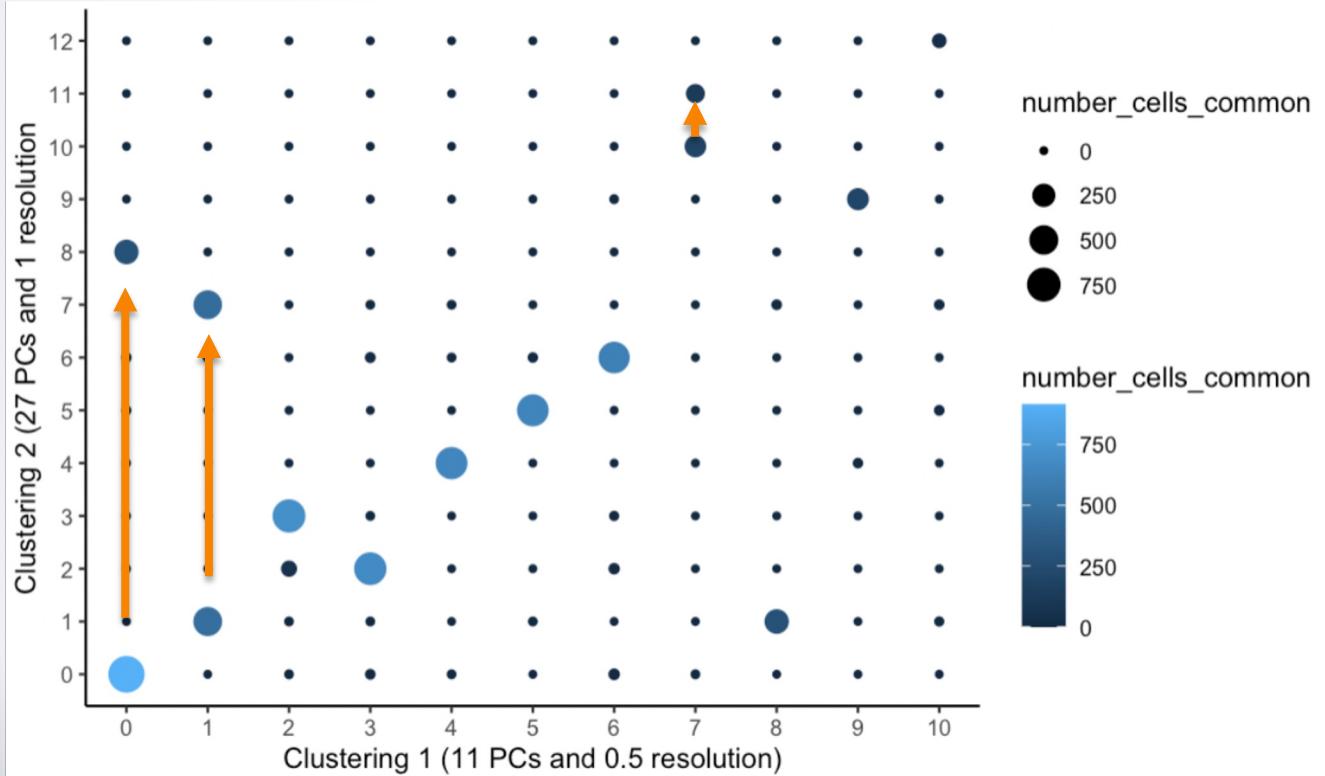
11 PCs and 0.5 resolution



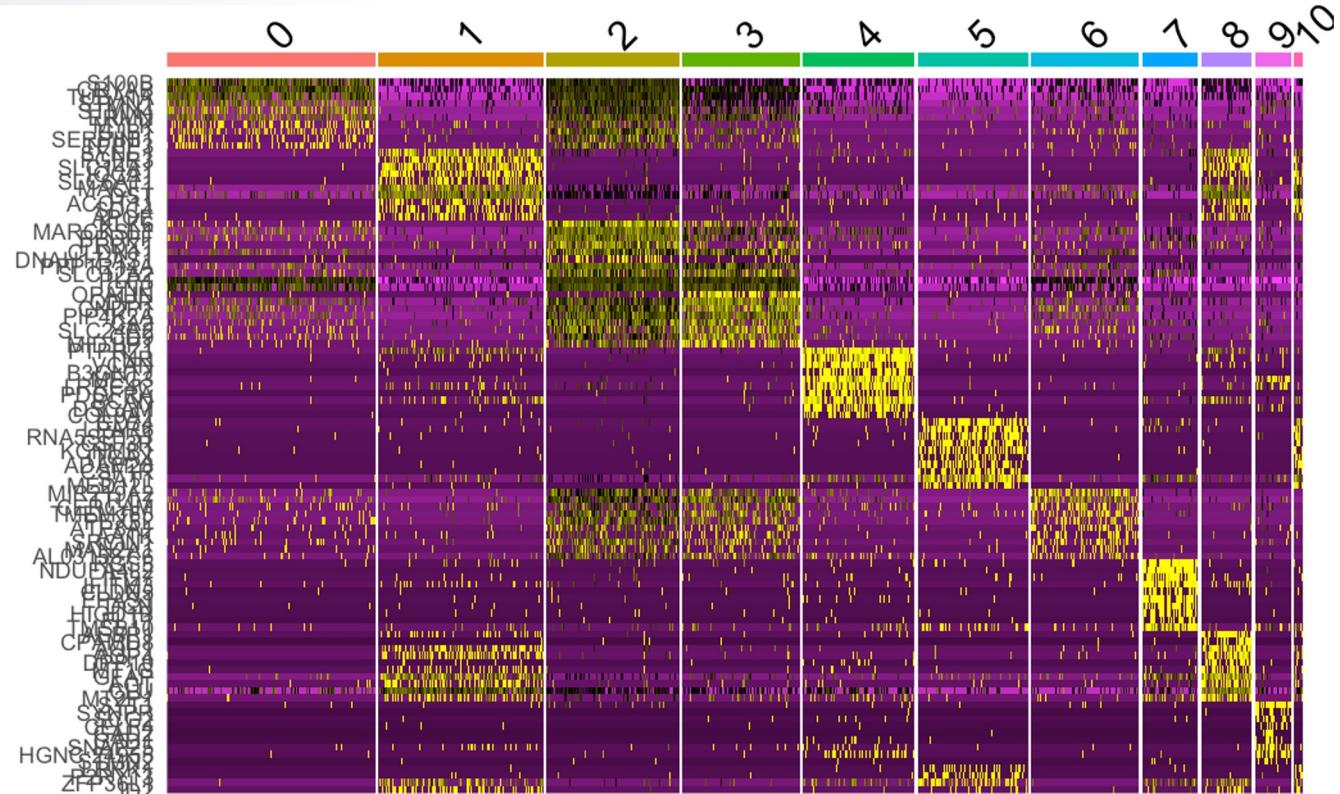
27 PCs and 1 resolution



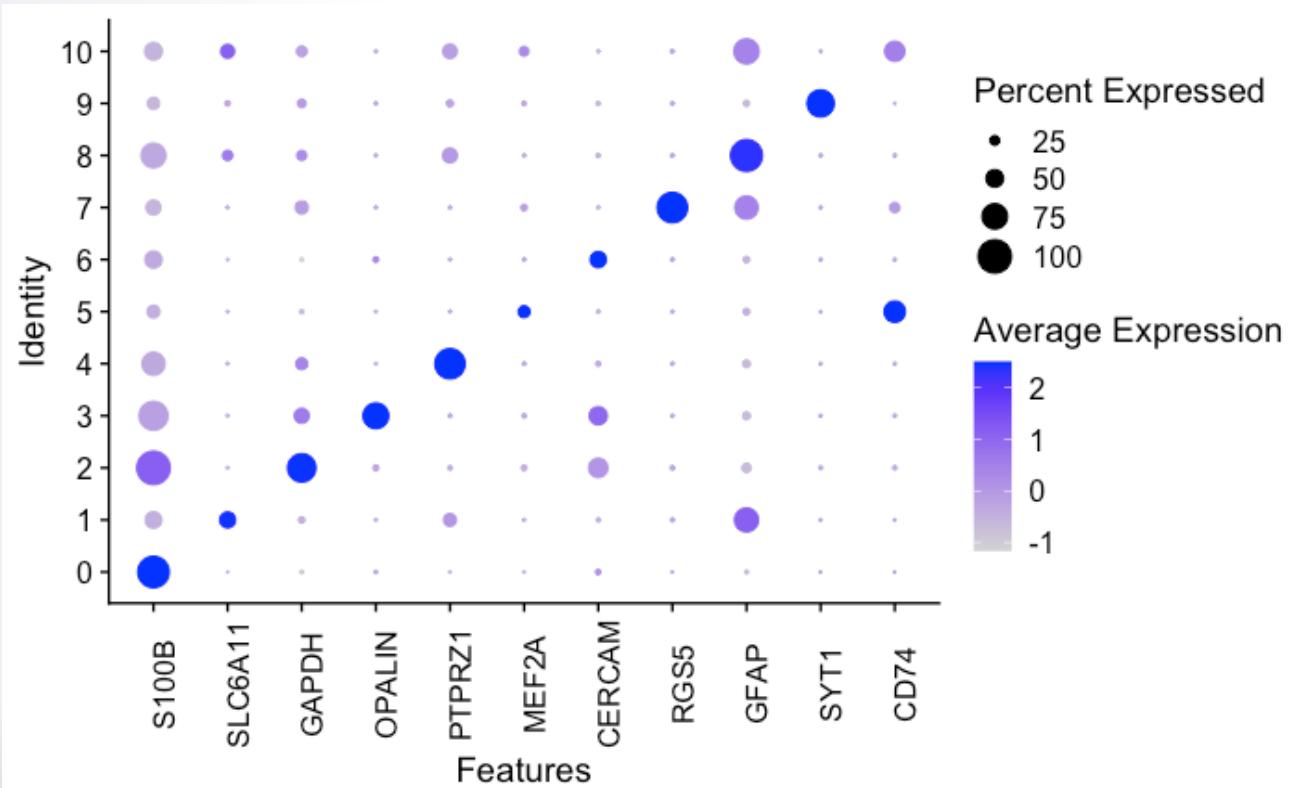
# Overlapping cells between clusters



# Possible subtypes

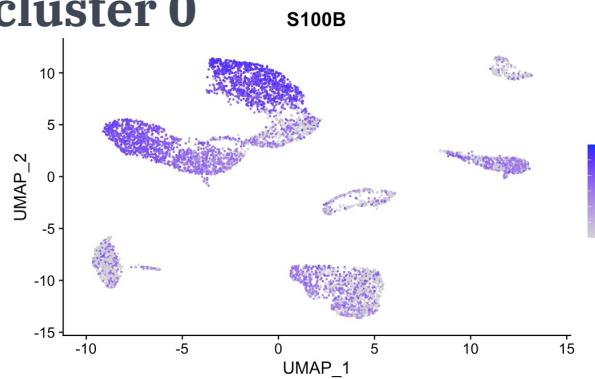


# Marker genes

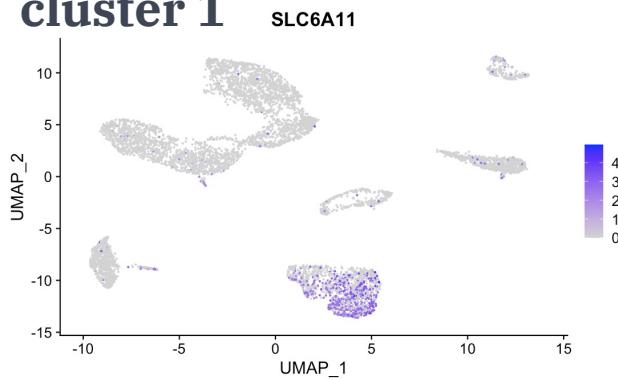


# Expression of marker genes

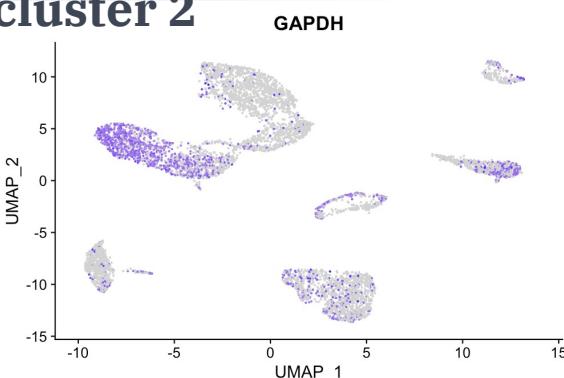
cluster 0



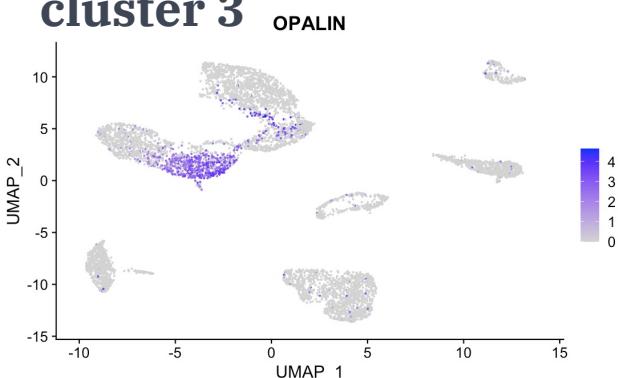
cluster 1



cluster 2

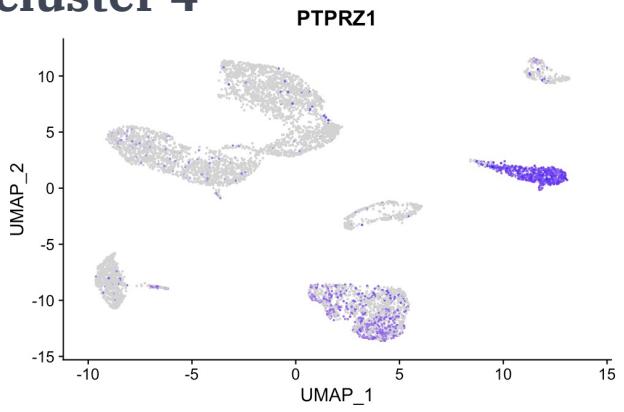


cluster 3

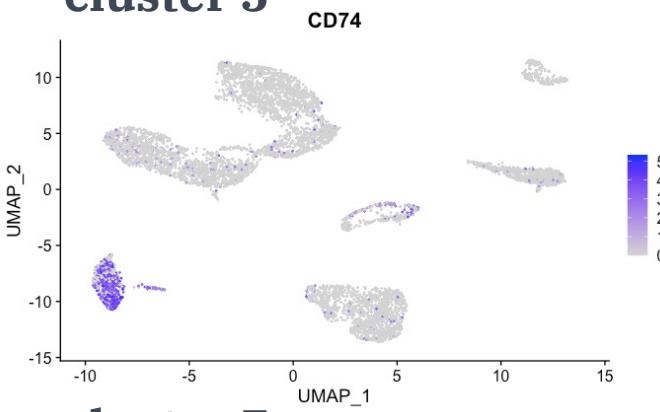


# Expression of marker genes

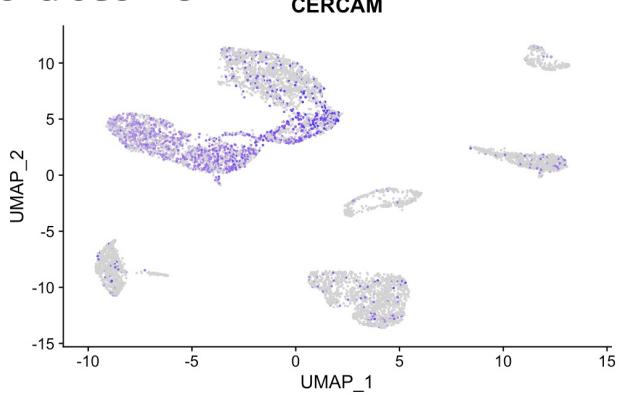
## cluster 4



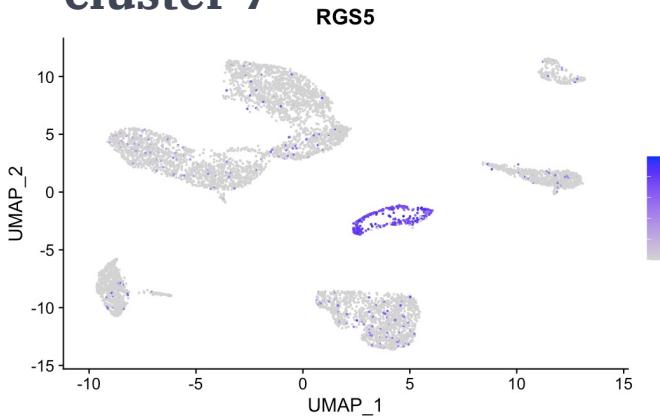
## cluster 5



## cluster 6

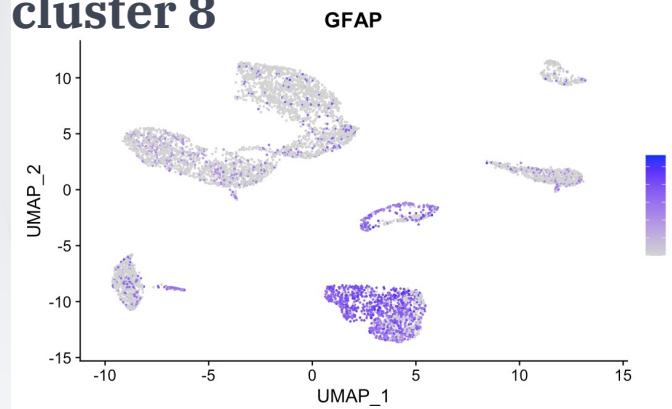


## cluster 7

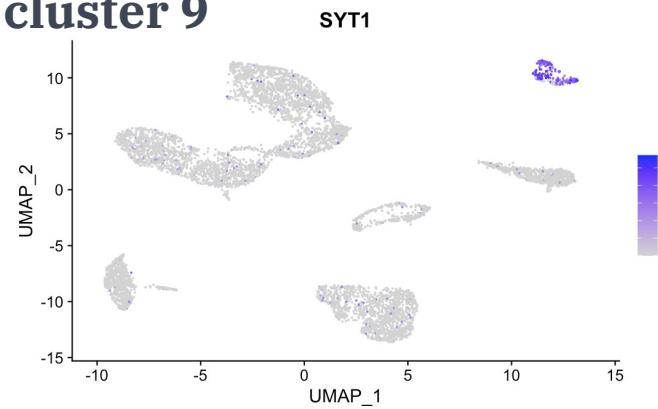


# Expression of marker genes

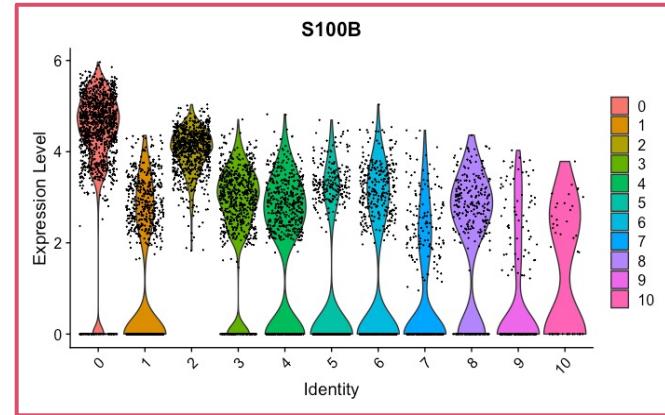
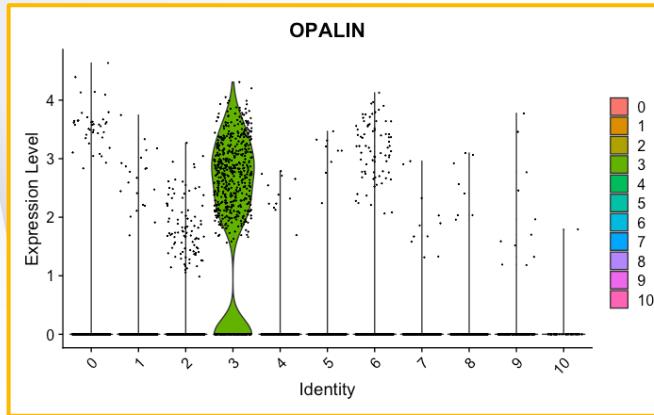
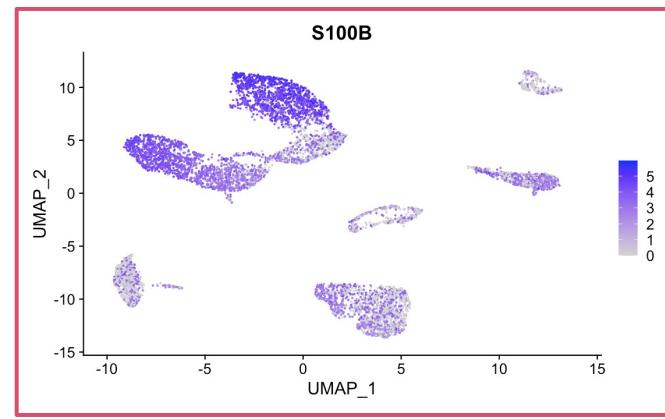
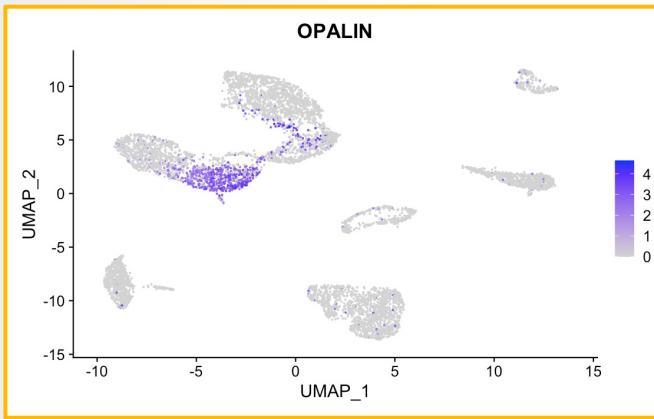
**cluster 8**



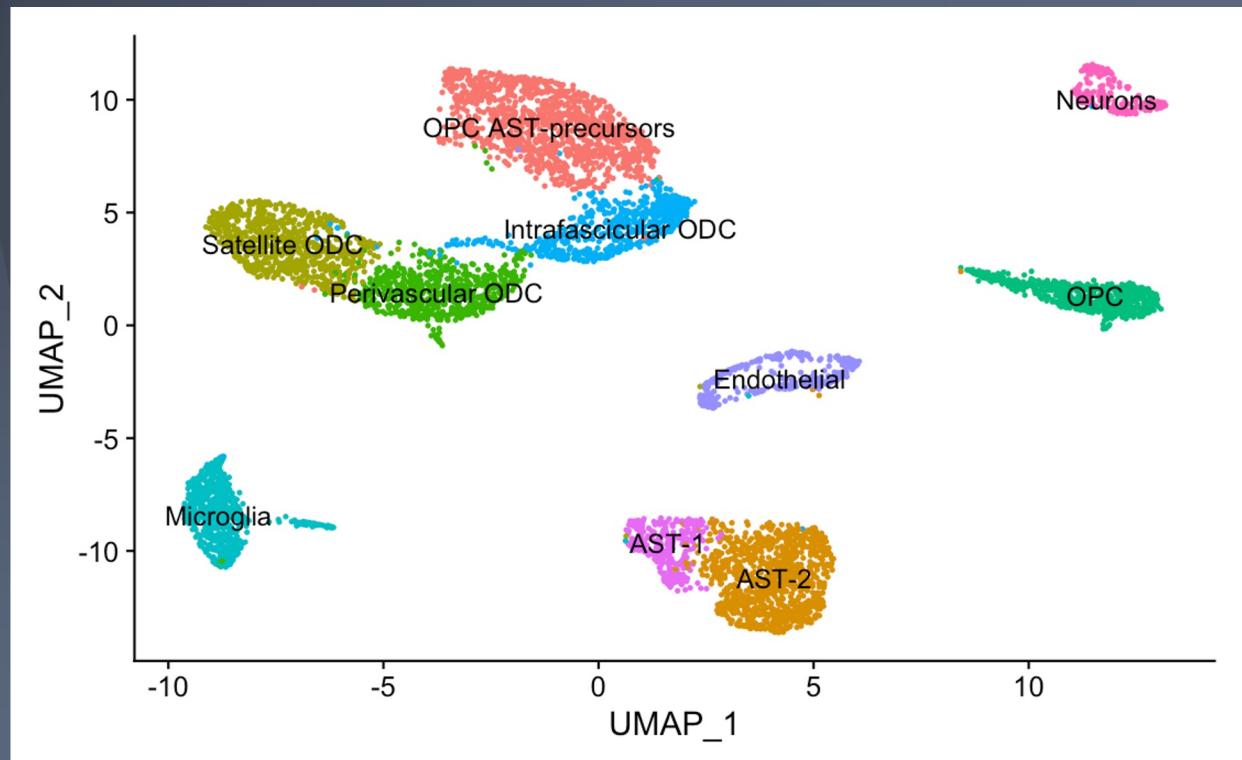
**cluster 9**



# Subtypes of Oligodendrocytes

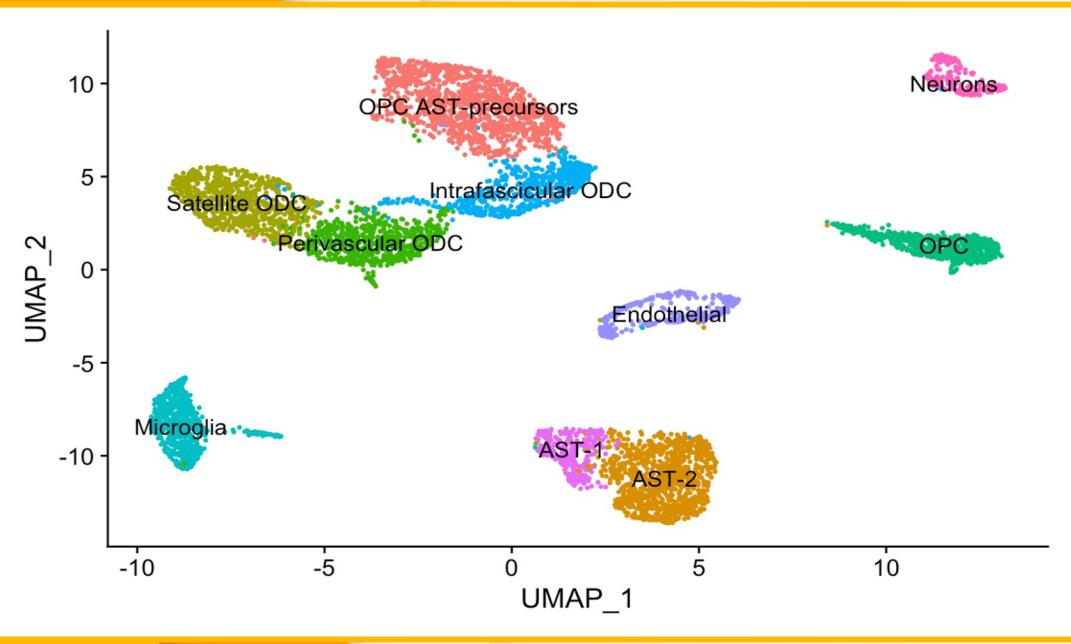


# Final result

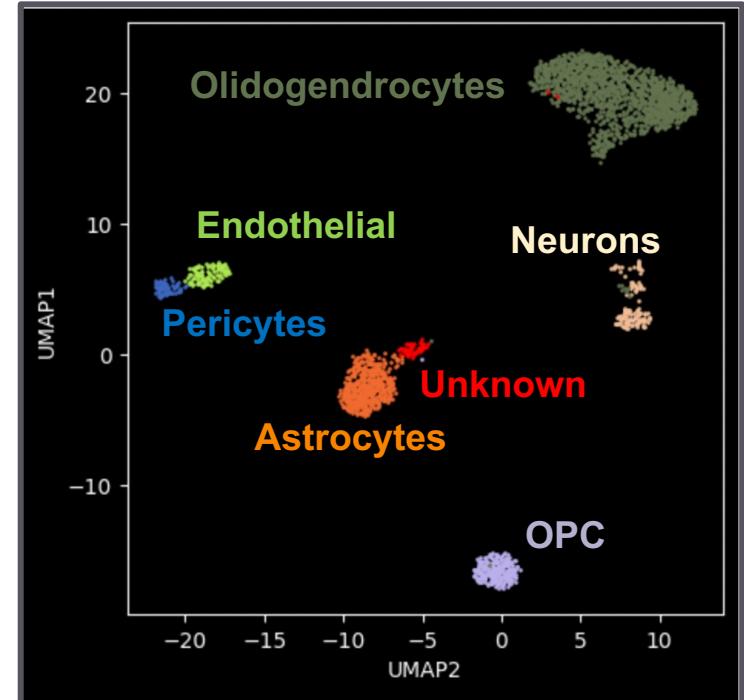


# Evaluating the Result

## Our analysis



## PanglaoDB



# Bibliography

- Agarwal, D., Sandor, C., Volpato, V., Caffrey, T. M., Monzón-Sandoval, J., Bowden, R., Alegre-Abarrategui, J., Wade-Martins, R., & Webber, C. (2020). **A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders.** *Nature Communications*, 11(1), 4183. <https://doi.org/10.1038/s41467-020-17876-0>
- Ludwin, S. K. (1979). **The perineuronal satellite oligodendrocyte: A role in remyelination.** *Acta Neuropathologica*, 47(1), 49–53. <https://doi.org/10.1007/BF00698272>
- Seeker, L. A., & Williams, A. (2022). **Oligodendroglia heterogeneity in the human central nervous system.** *Acta Neuropathologica*, 143(2), 143–157. <https://doi.org/10.1007/s00401-021-02390-4>
- Stratoulias, V., Venero, J. L., Tremblay, M., & Joseph, B. (2019). **Microglial subtypes: Diversity within the microglial community.** *The EMBO Journal*, 38(17). <https://doi.org/10.15252/embj.2019101997>
- Tabata, H. (2015). **Diverse subtypes of astrocytes and their development during corticogenesis.** *Frontiers in Neuroscience*, 9. <https://doi.org/10.3389/fnins.2015.00114>



Thank you for the attention.