
Bulk RNA-seq analysis

Bulk RNA-seq analysis

Aim of the analysis:

Identify DE genes across replicates of different conditions, and functionally characterize them

Dataset

Source: Recount3 repository

GTEX project data

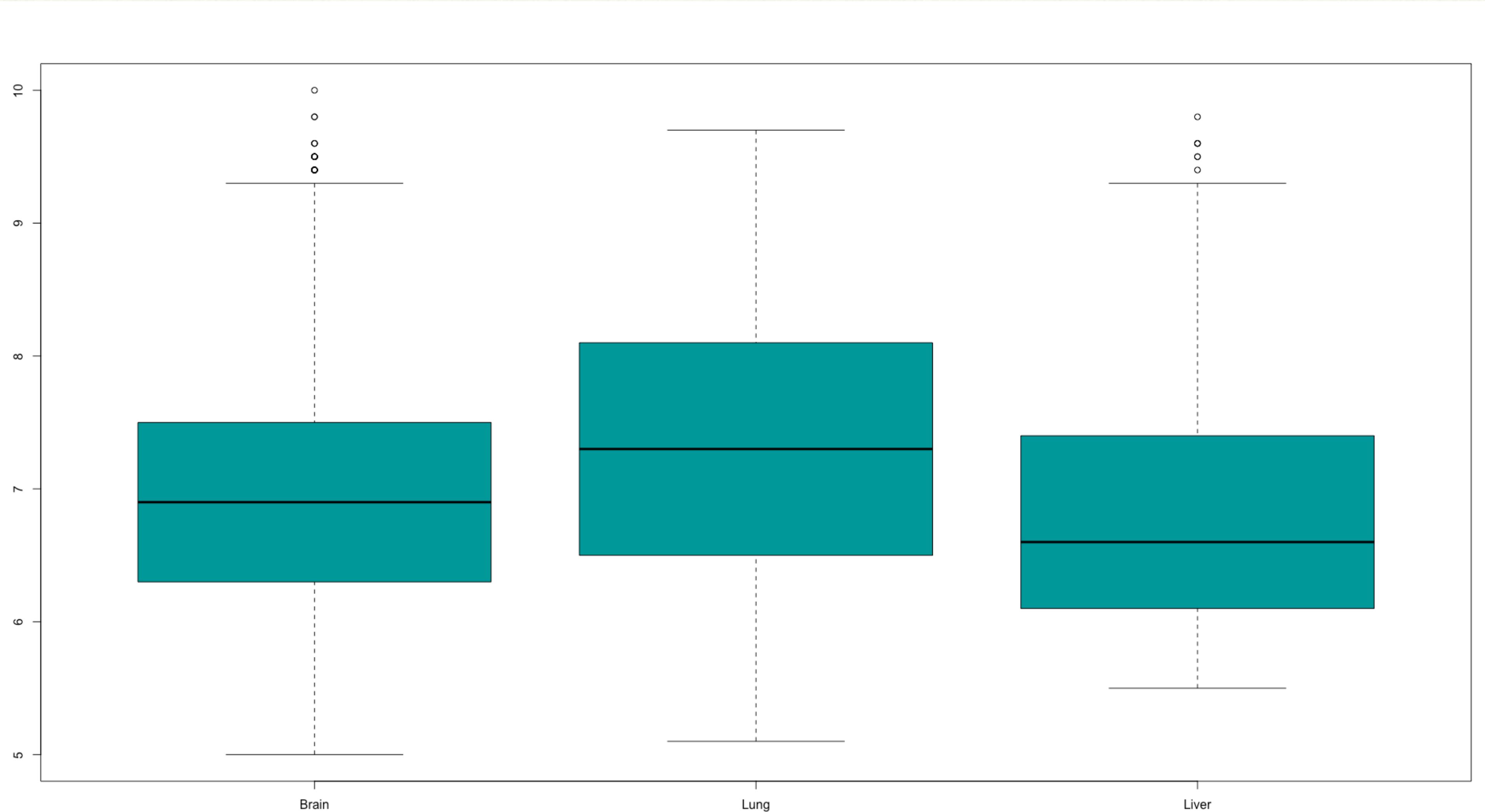
Conditions: Human Tissues

Brain, Lungs, Liver

Three replicates for each tissue

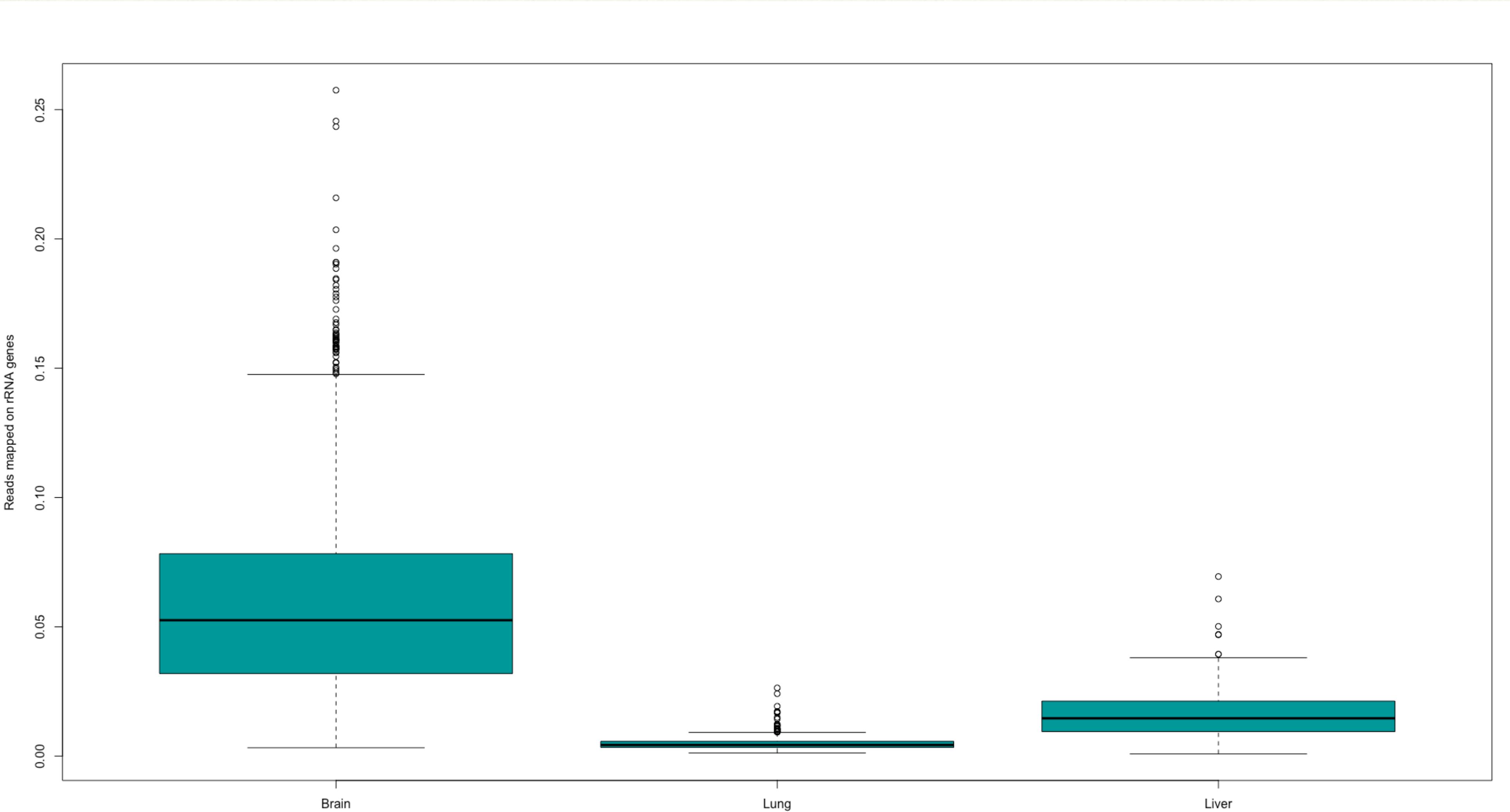
Rin Score

Value used to select the replicates: **7 or higher**



Reads mapped on rRNA genes

Value used to select the replicates: **0.1 or lower**



Uniquely mapped reads

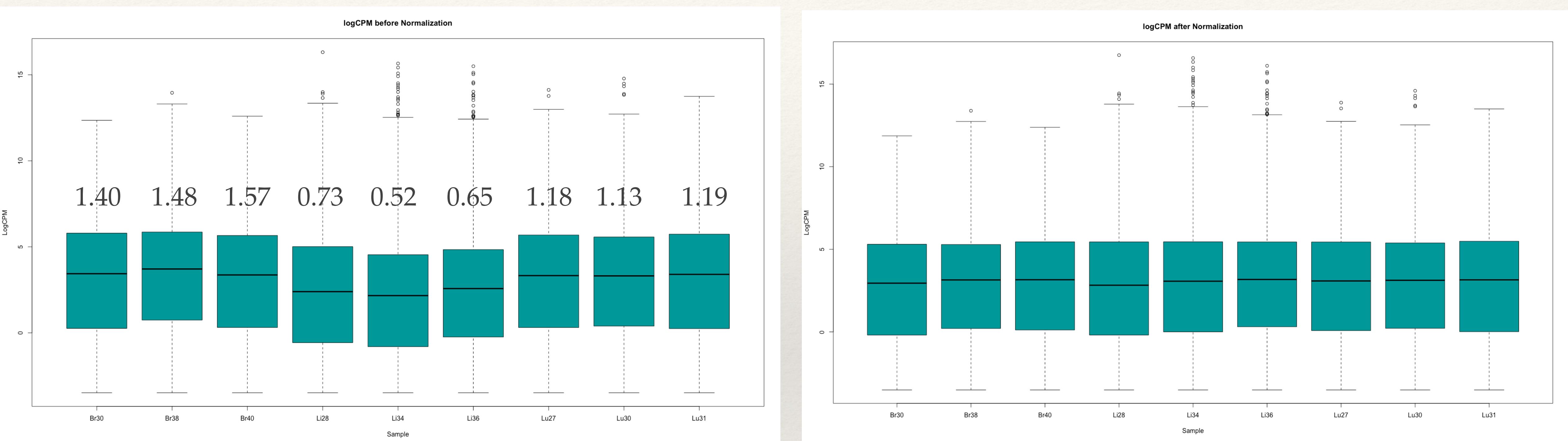
Value used to select the replicates: **85 or higher**



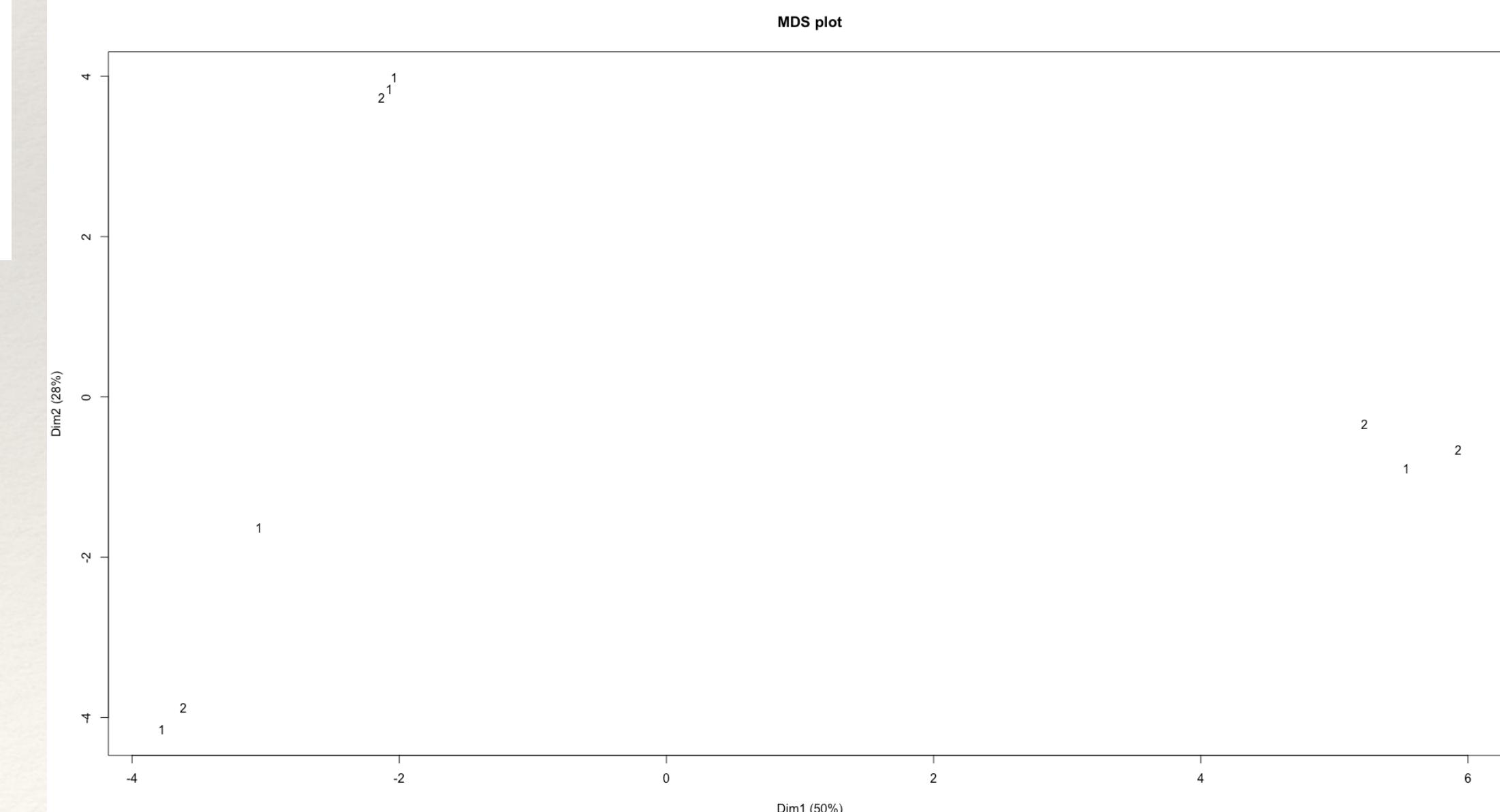
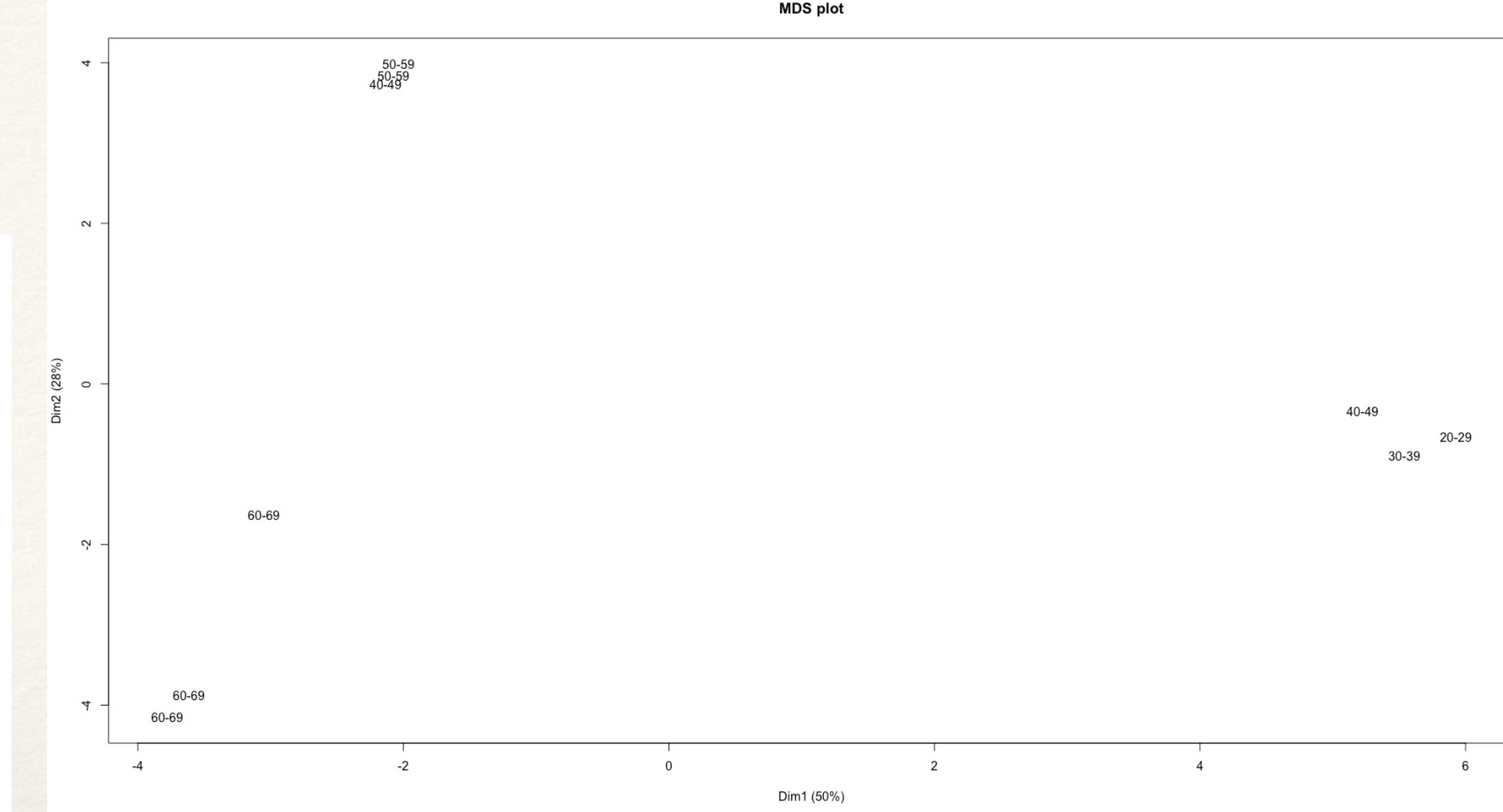
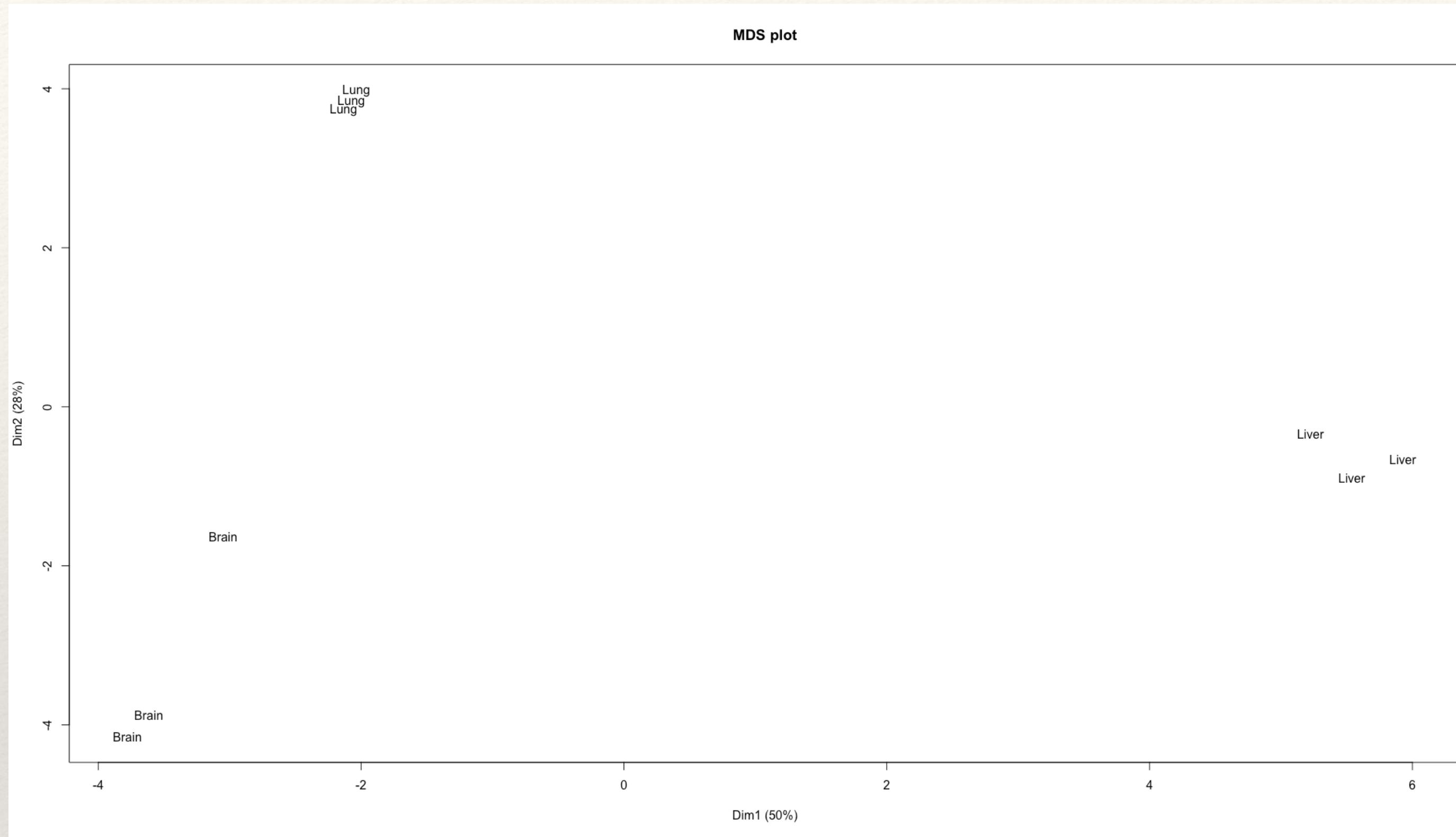
Selected replicates

	RIN	rRNA reads	% mapped reads
Brain			
30	7.4	0.02	87.9
38	7.2	0.08	91.3
40	7	0.03	91.6
Liver			
28	7.4	0.02	90.8
34	7.5	0.01	90.8
36	8.7	0.01	91.6
Lungs			
27	8.8	0.01	91
30	8.4	0.01	91.2
31	8.4	0.01	91.8

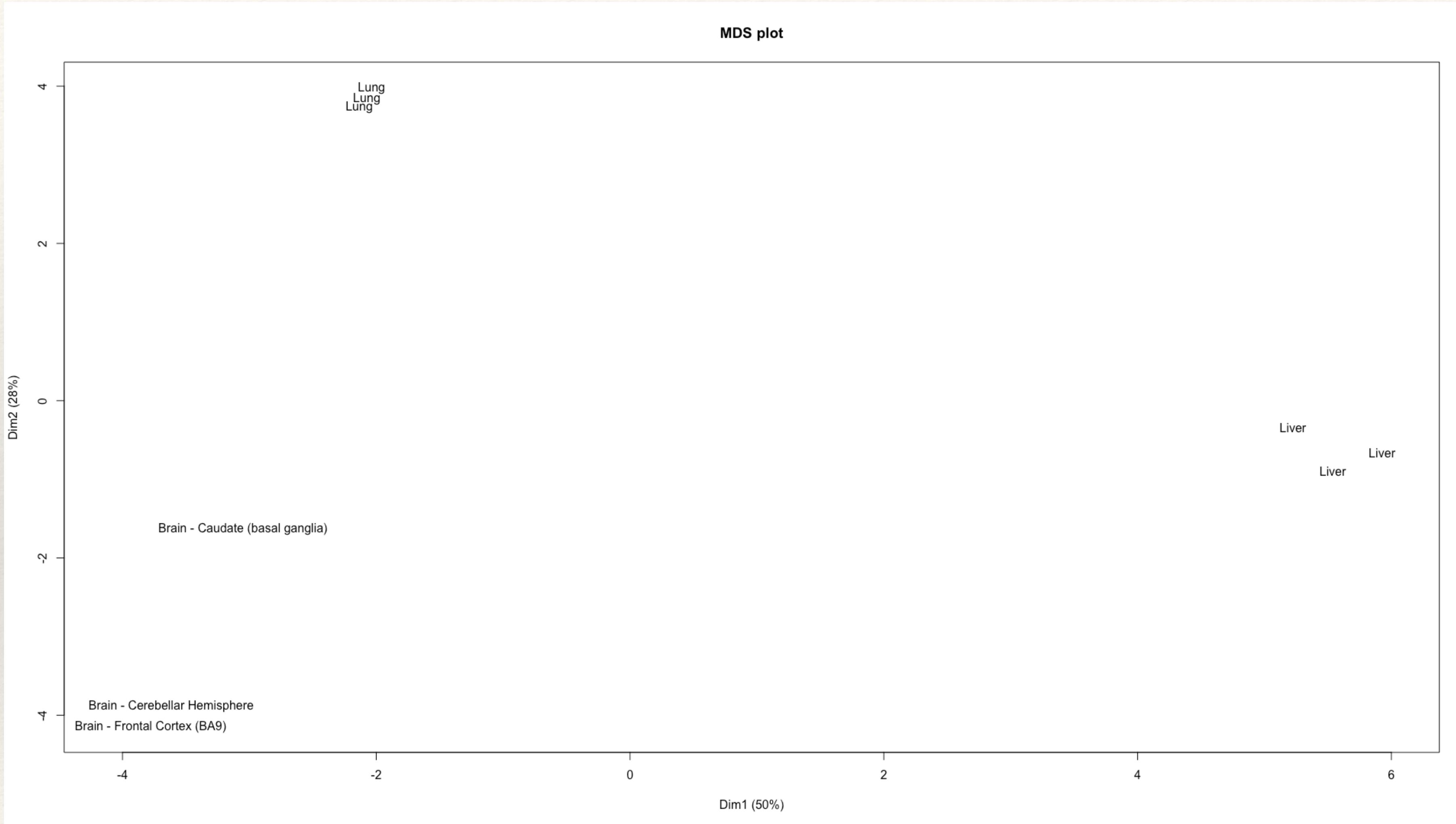
Normalization



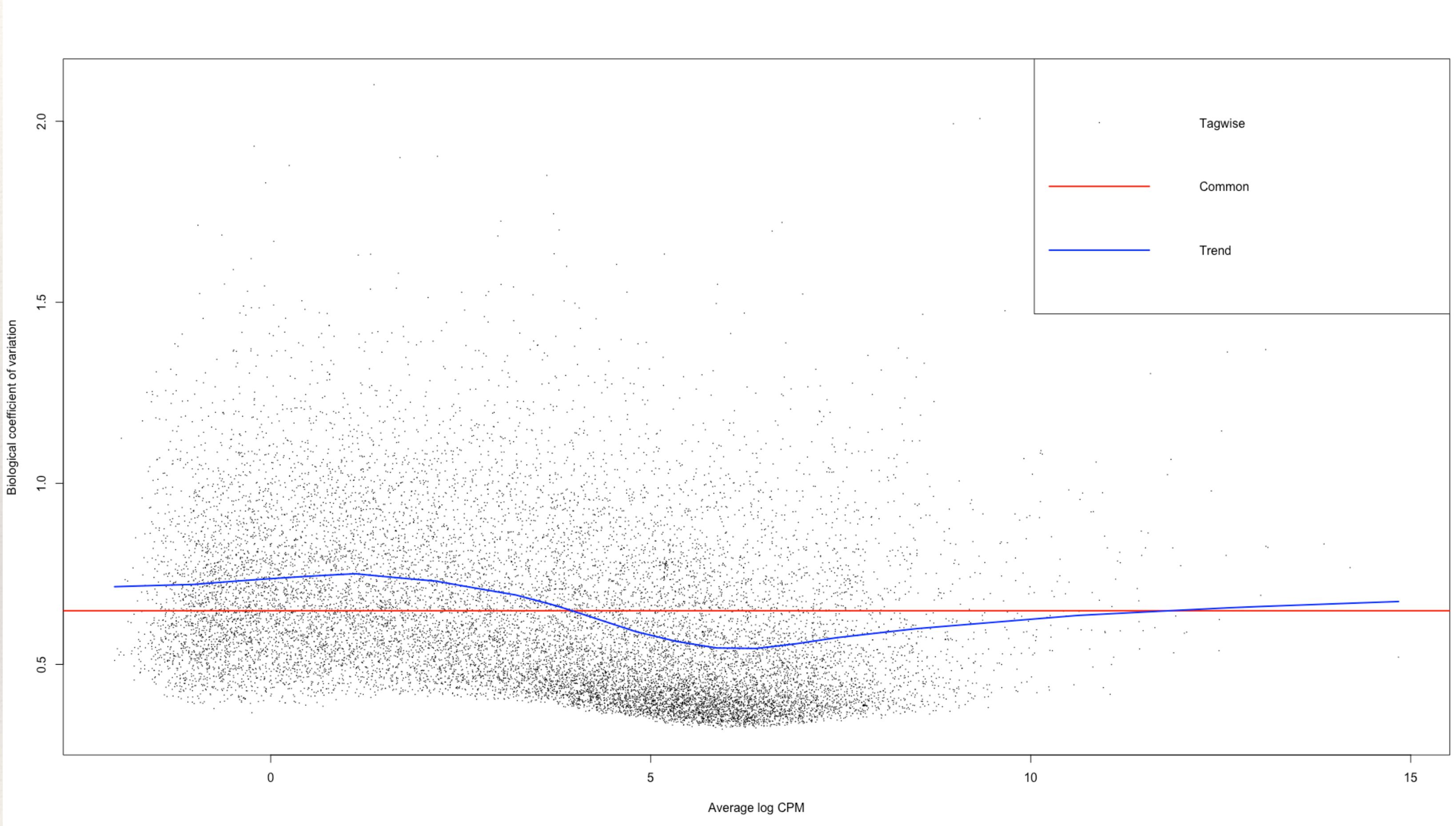
MDS plot



MDS plot - other sources of variability



Dispersion plot



DE genes

- Must be over-expressed in one tissue against the other two
- Must be filtered by statistical significance to reduce the possibility of errors
- Must check for the presence of ribosomal protein genes

Threshold used

- FDR < 0.01
- logCPM > 0
- Exclude genes that have no annotation, for example known short RNA

Data taken from Enrichr, considering:

- ❖ Pathways: Bioplanet 19, KEGG 2021 Human, WikiPathway 2021 Human
- ❖ Ontologies: GO Biological Process 2021, GO Molecular Function 2021, GO Cellular Component 2021

Brain



Pathways

- Oligodendrocyte Specification and Differentiation
- Glial Cell Differentiation
- Synaptic Vesicle Pathway
- Morphine Addiction

Ontologies

- Nervous system development
- Synaptic Transmission
- Calcium Ion Binding

Diseases

- Epilepsy

Enrichr

Liver



Pathways

- Steroid hormone biosynthesis
- Metabolism
- Ascorbate and aldarate metabolism
- Drug metabolism

Ontologies

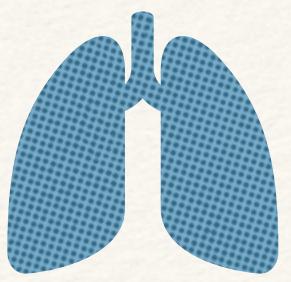
- Steroid metabolic process
- Glucoronosyltransferase activity
- Glucoronate metabolic process

Diseases

- Disorders of bilirubin excretion

Enrichr

Lungs



Pathways

- HNF3A pathway
- Lung fibrosis
- Cell Adhesion molecules

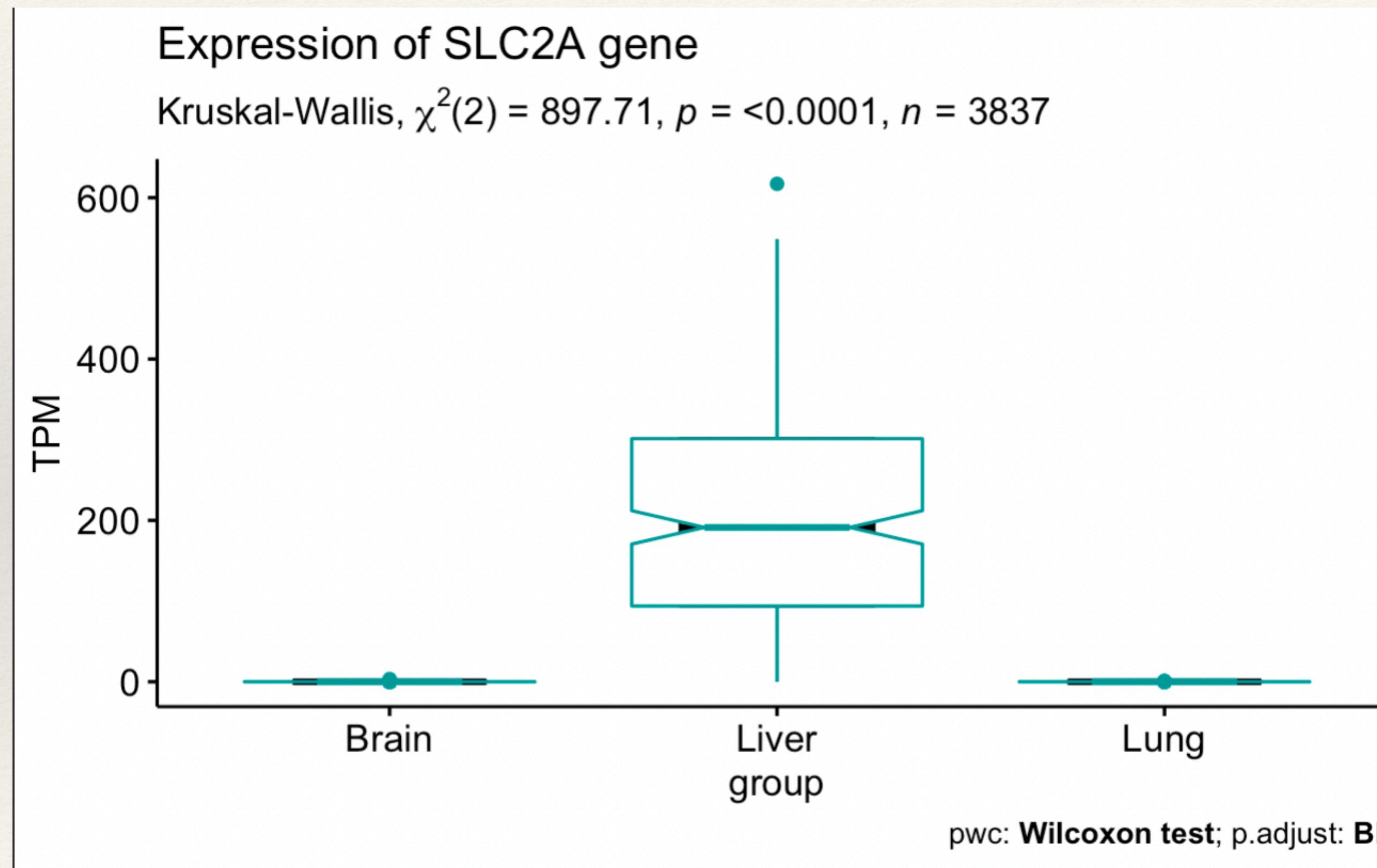
Ontologies

- Lung morphogenesis
- Alveolar lamellar body
- Surfactant homeostasis

Diseases

- Solitary pulmonary nodule

Example gene - Liver - SLC2A



Number of exons spanning from 10 to 11, while the introns are from 9 to 10.

SLC2A encodes for a protein that allows glucose transport

Conclusions

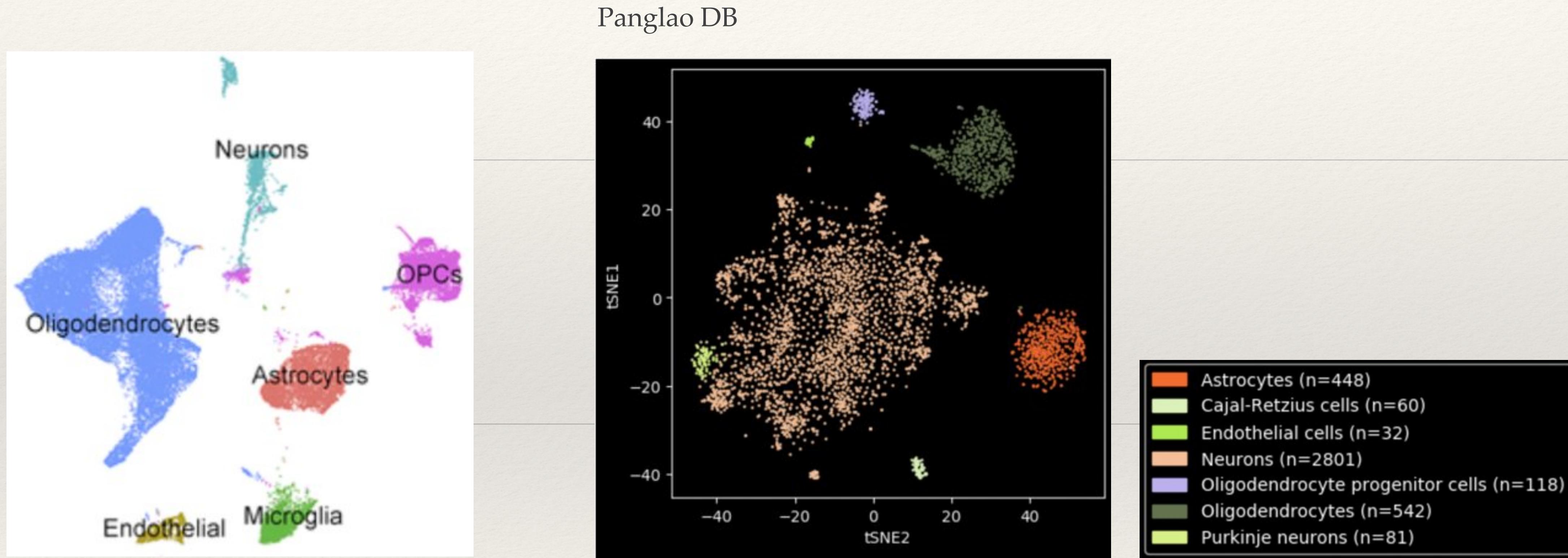
- ❖ Correctly identified significant genes for each tissue that were over expressed in comparison with the other samples

In particular the functional characterization confirmed the DE genes found, therefore the analysis results are reliable

Single cell analysis

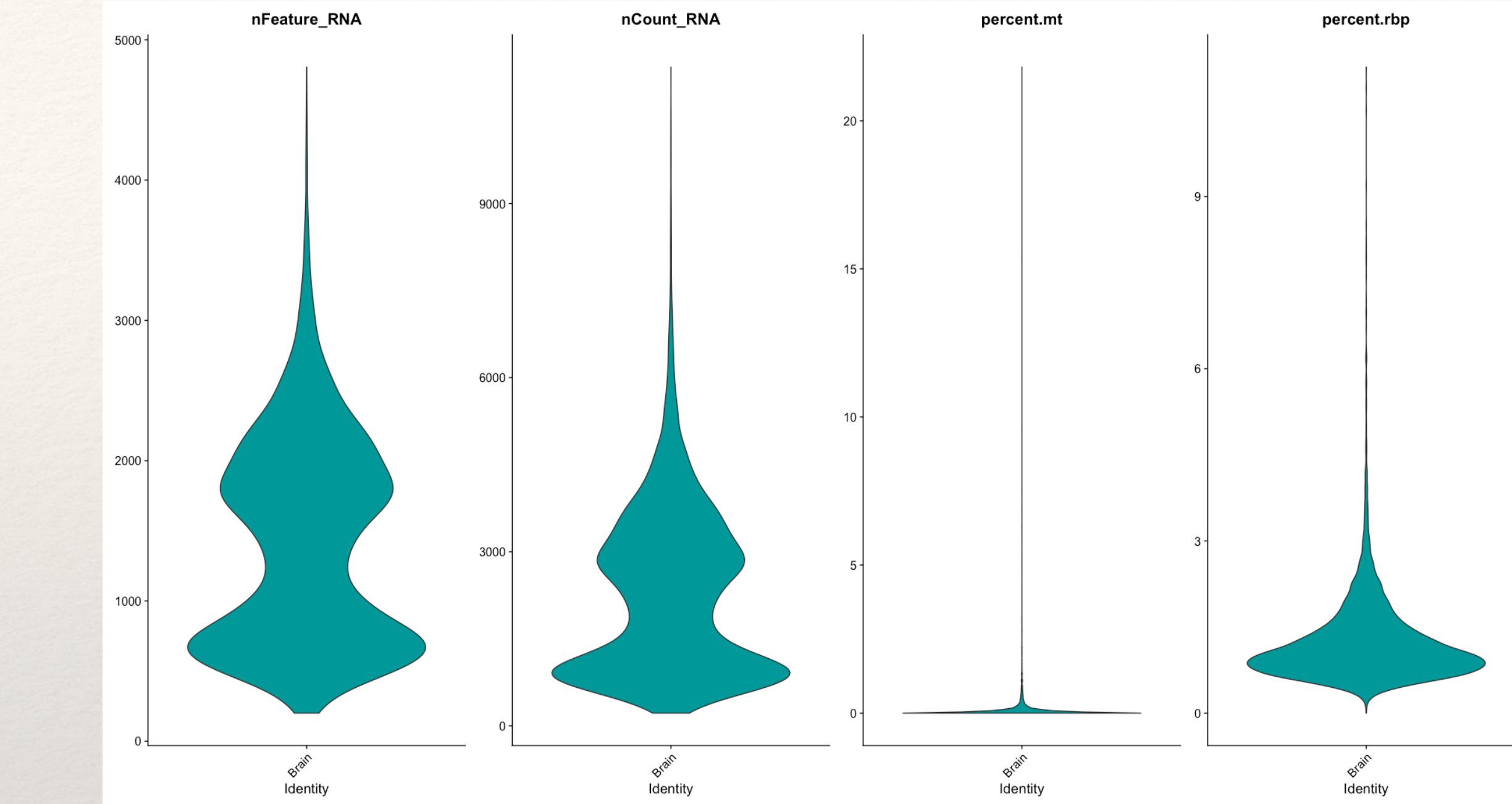
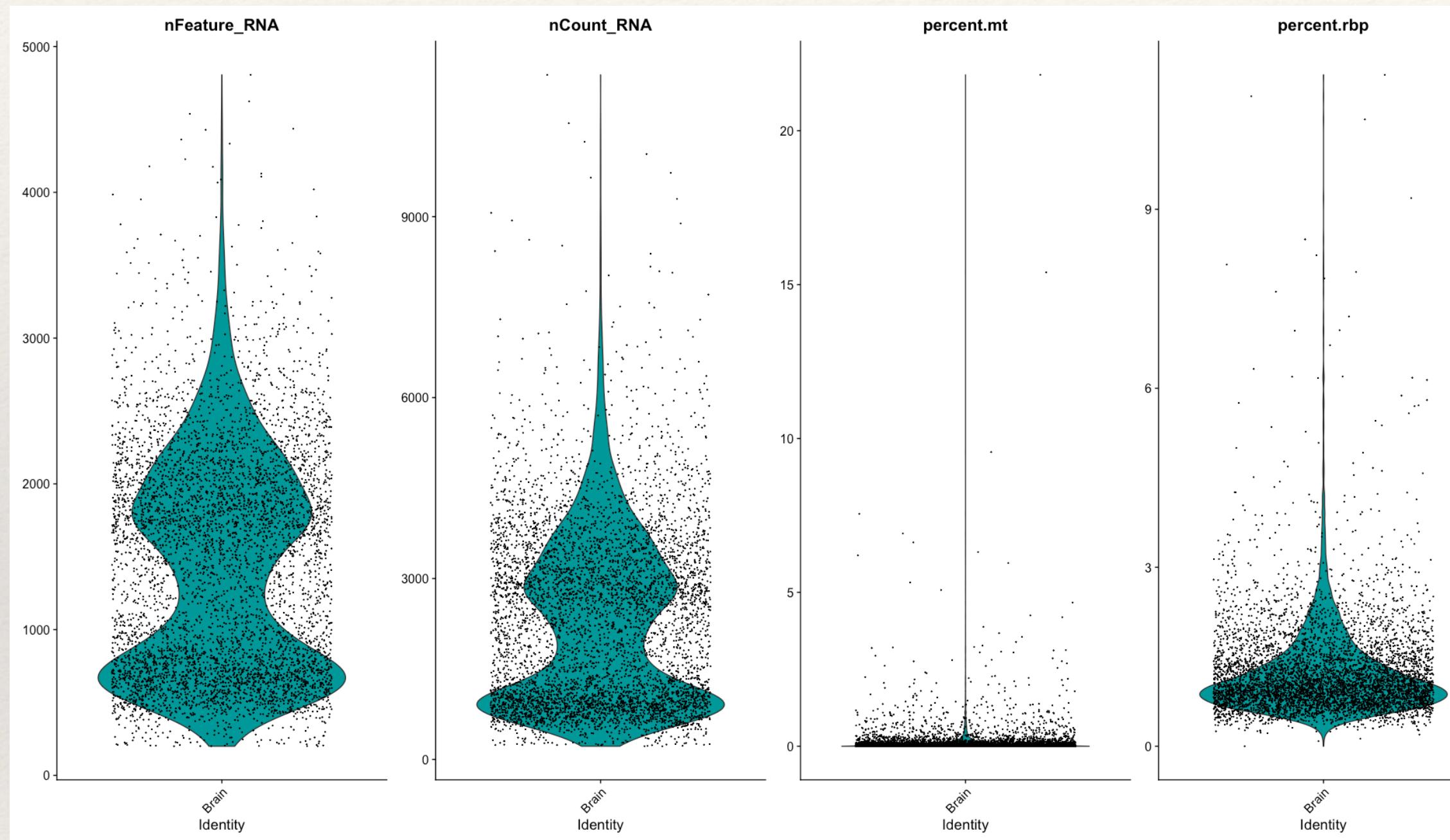
Single cell Analysis

THE OG: Welch *et al.* (2019) 'Single-cell multi-omic integration compares and contrasts features of brain cell identity'



CELL QUALITY CONTROL

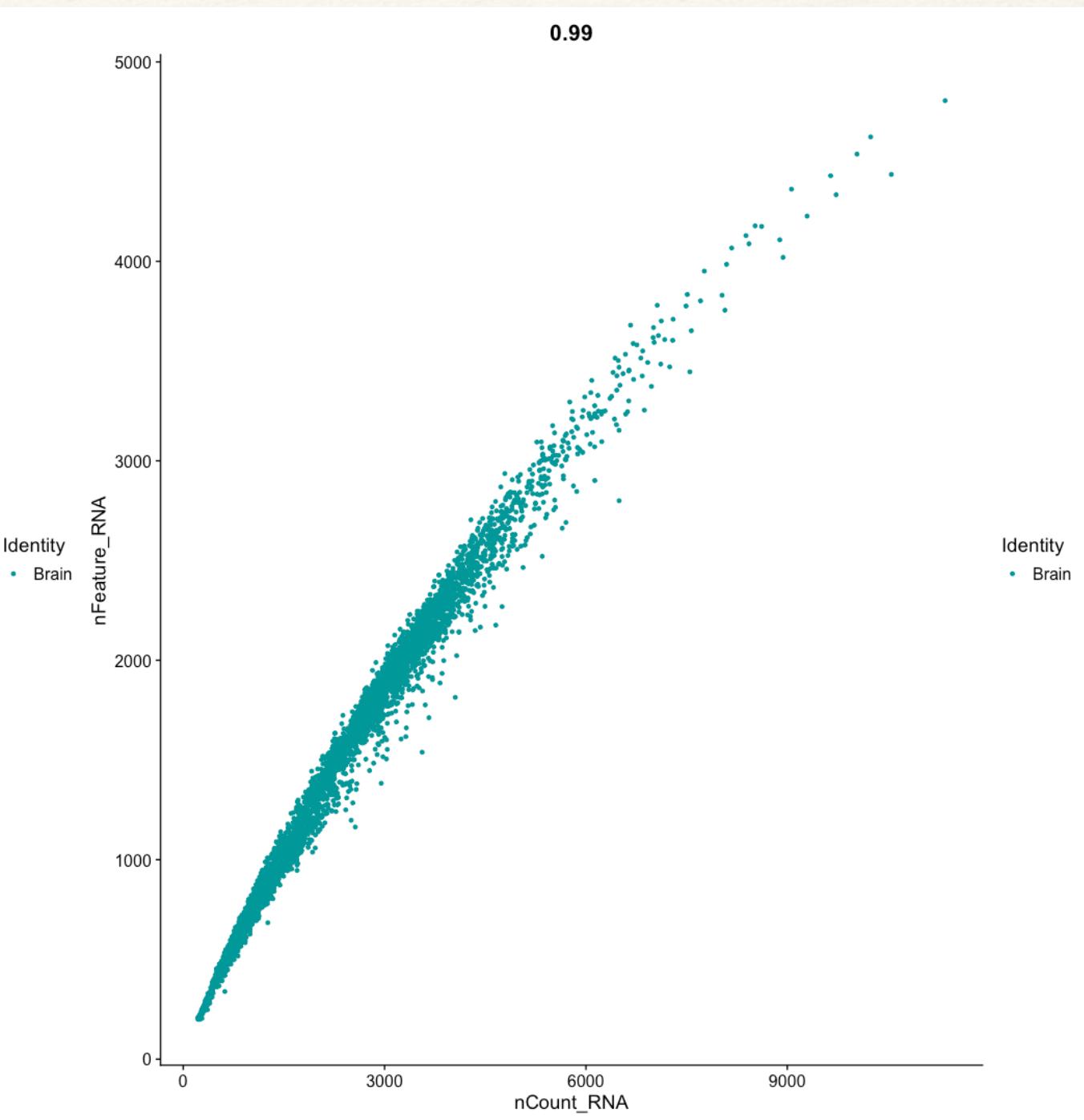
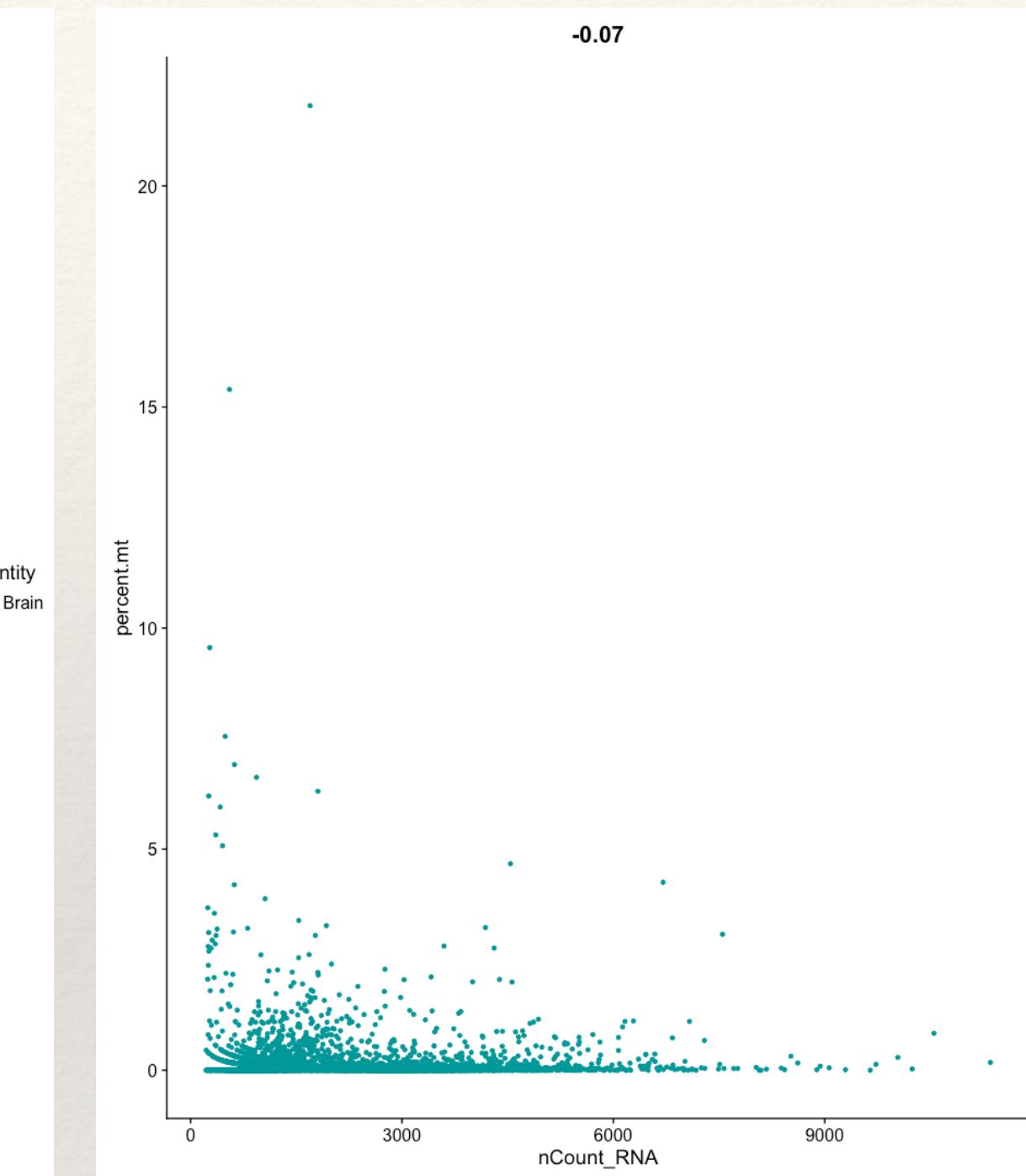
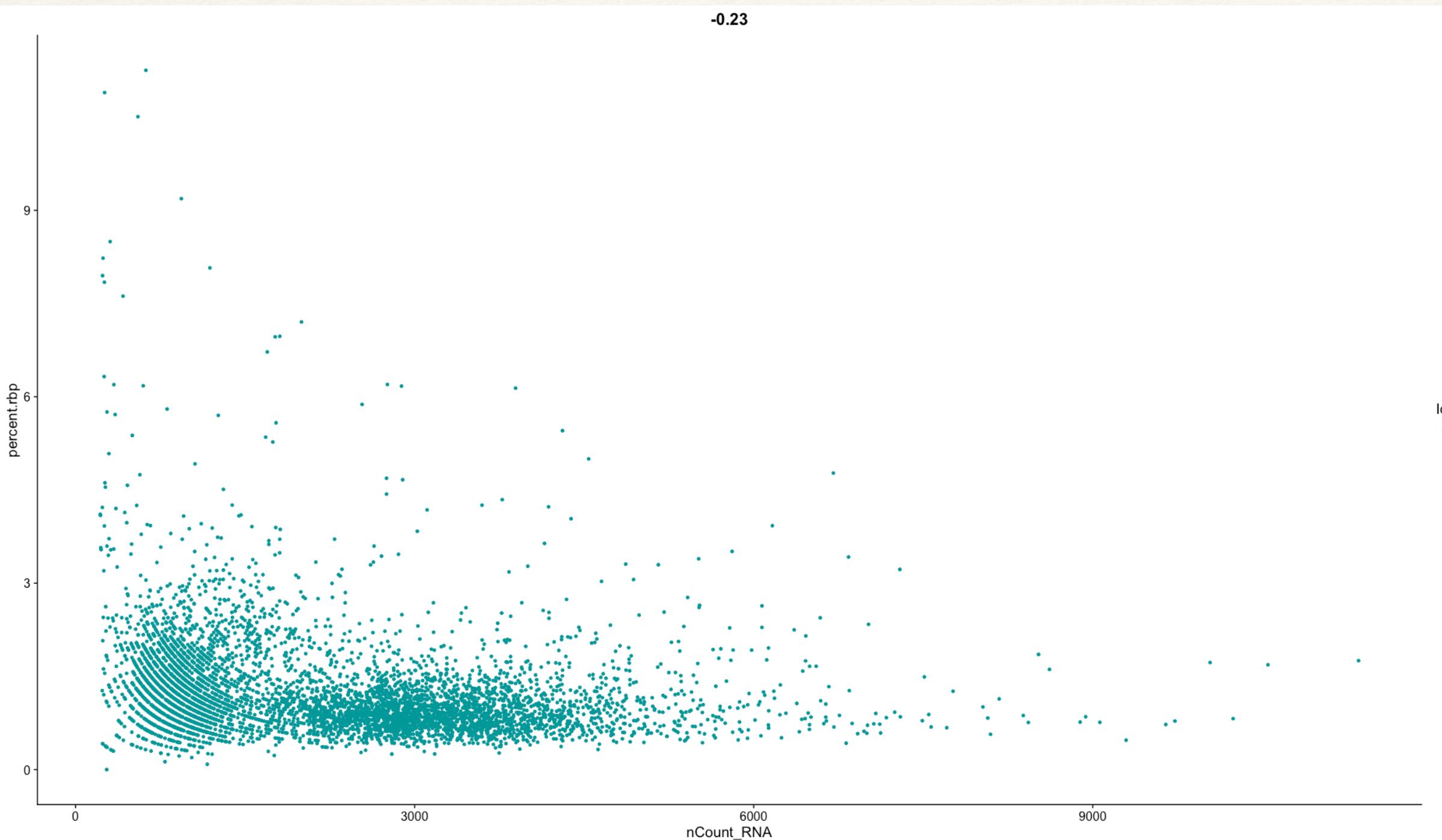
nFeature_RNA > 200
nFeature_RNA < 3800



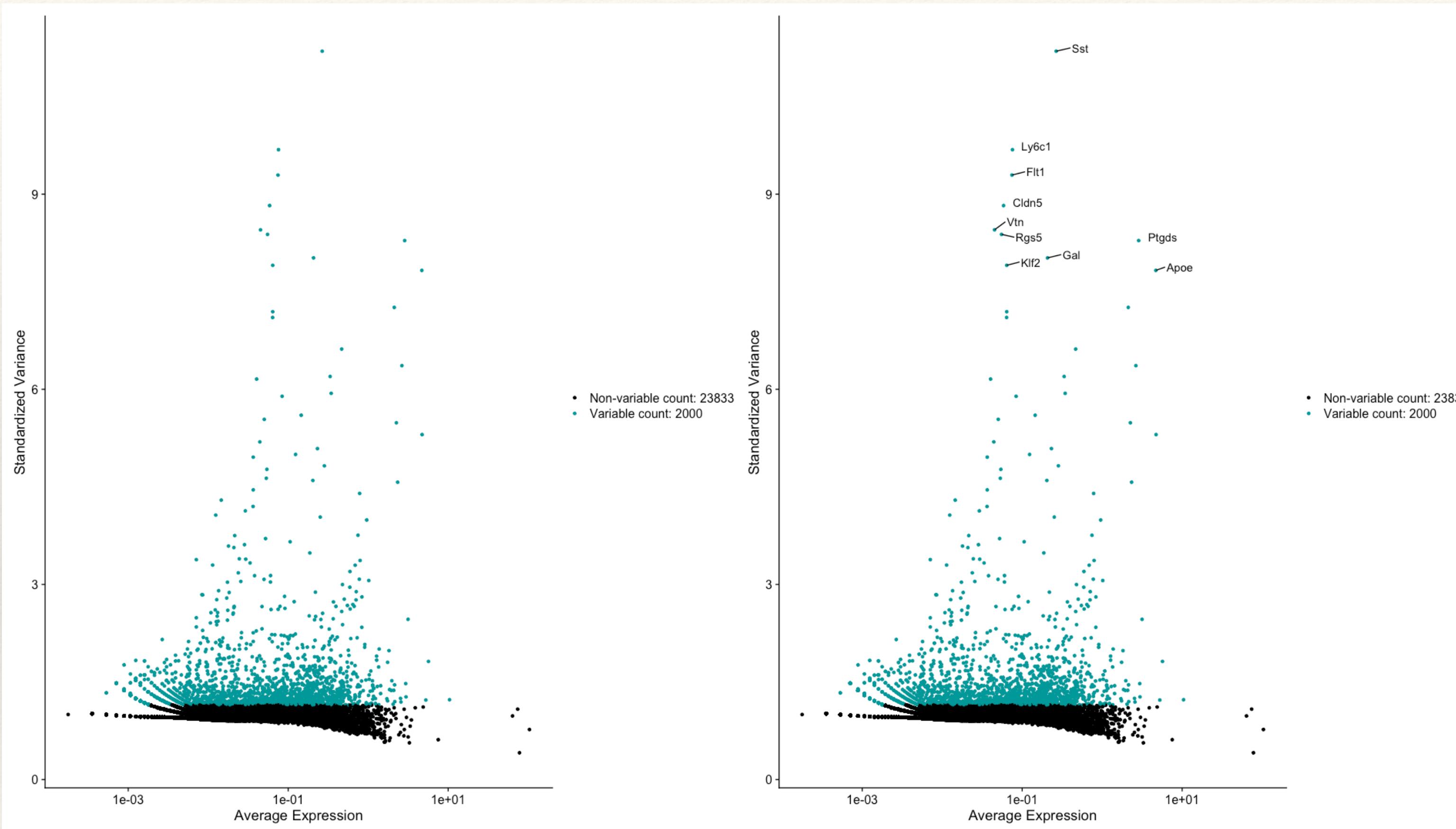
percent.mt < 2
From 5710 samples to 5635

CELL QUALITY CONTROL

The only visible correlation is between the number of reads and the number of genes detected



RESULTS: MOST VARIABLE GENES

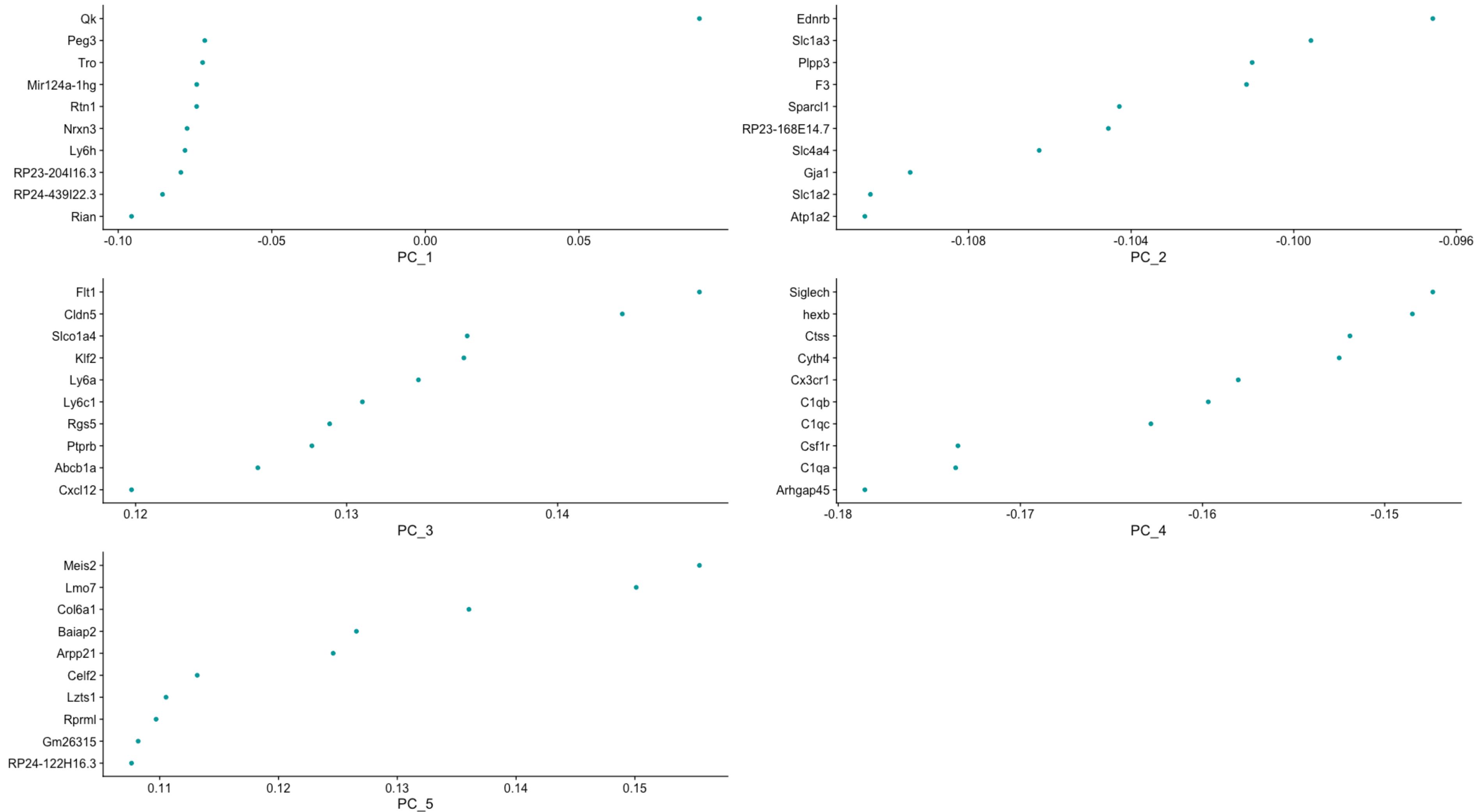


SST: Involved in hypothalamus developing

Flt1: FMS-like tyrosine kinase 1, the lack of causes neuron degeneration

Vtn: Involved in oligodendrocyte differentiation

RESULTS: Principal Component Analysis

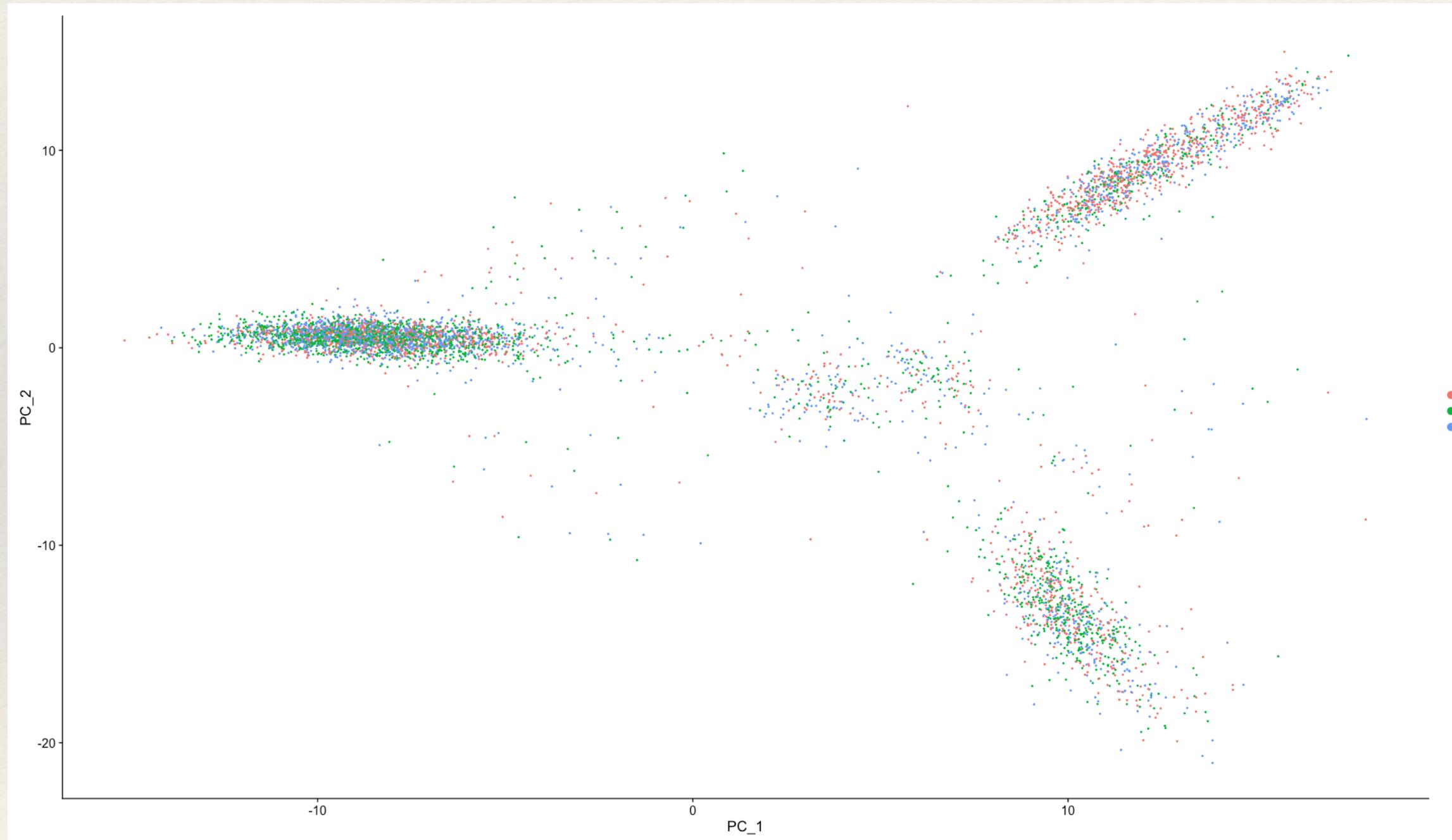


The first three genes in the first component encode for proteins necessary for the process of myelination, negative regulation of transcription (brain specific) and cell signaling

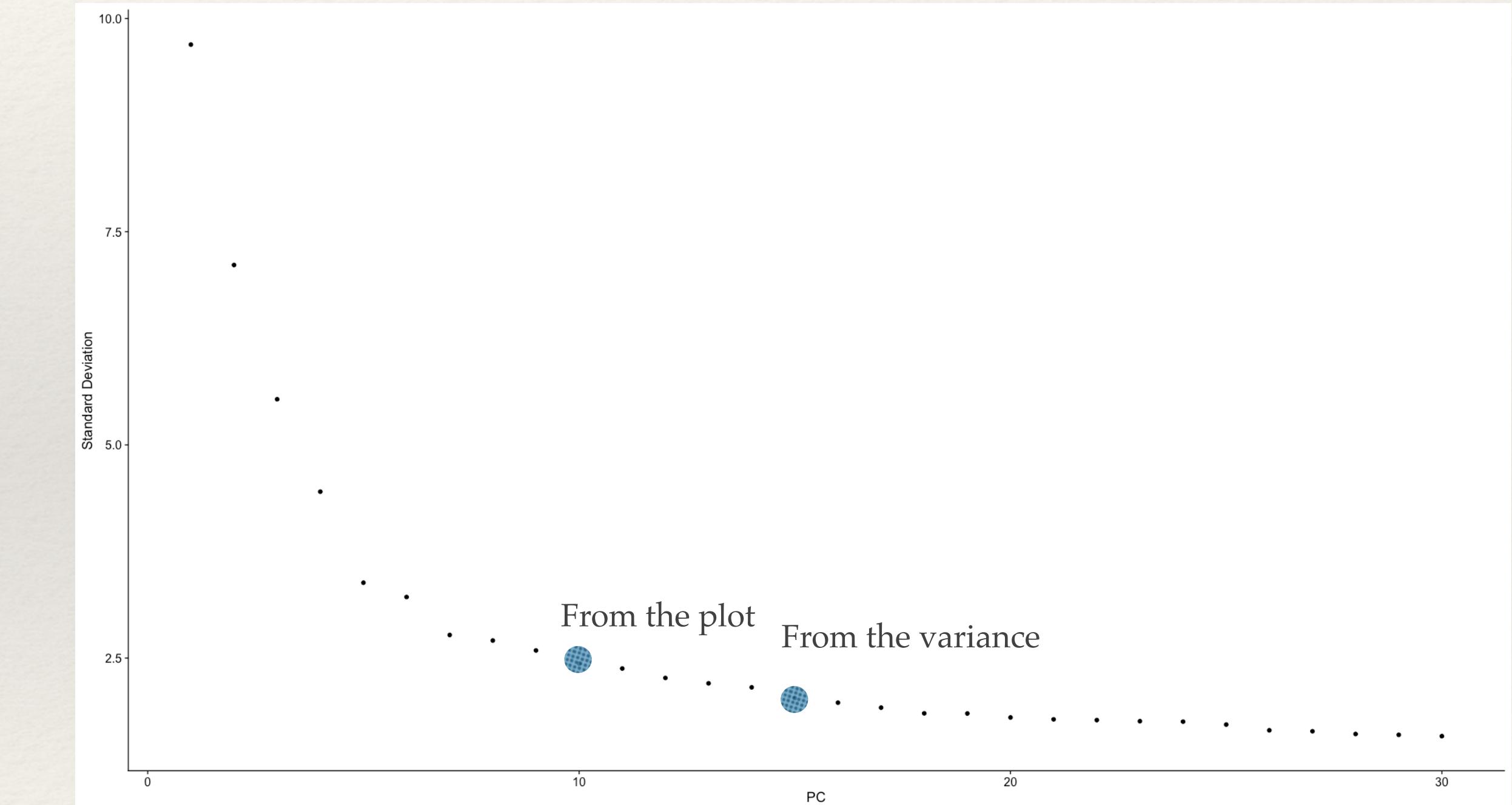
In the second component we can find genes associated with neuron development, neuron junction, glutamate binding activity, glial cell differentiation

RESULTS: Principal Component Analysis

Projection of the cells along the first two PCs according to Cell Cycle Phase

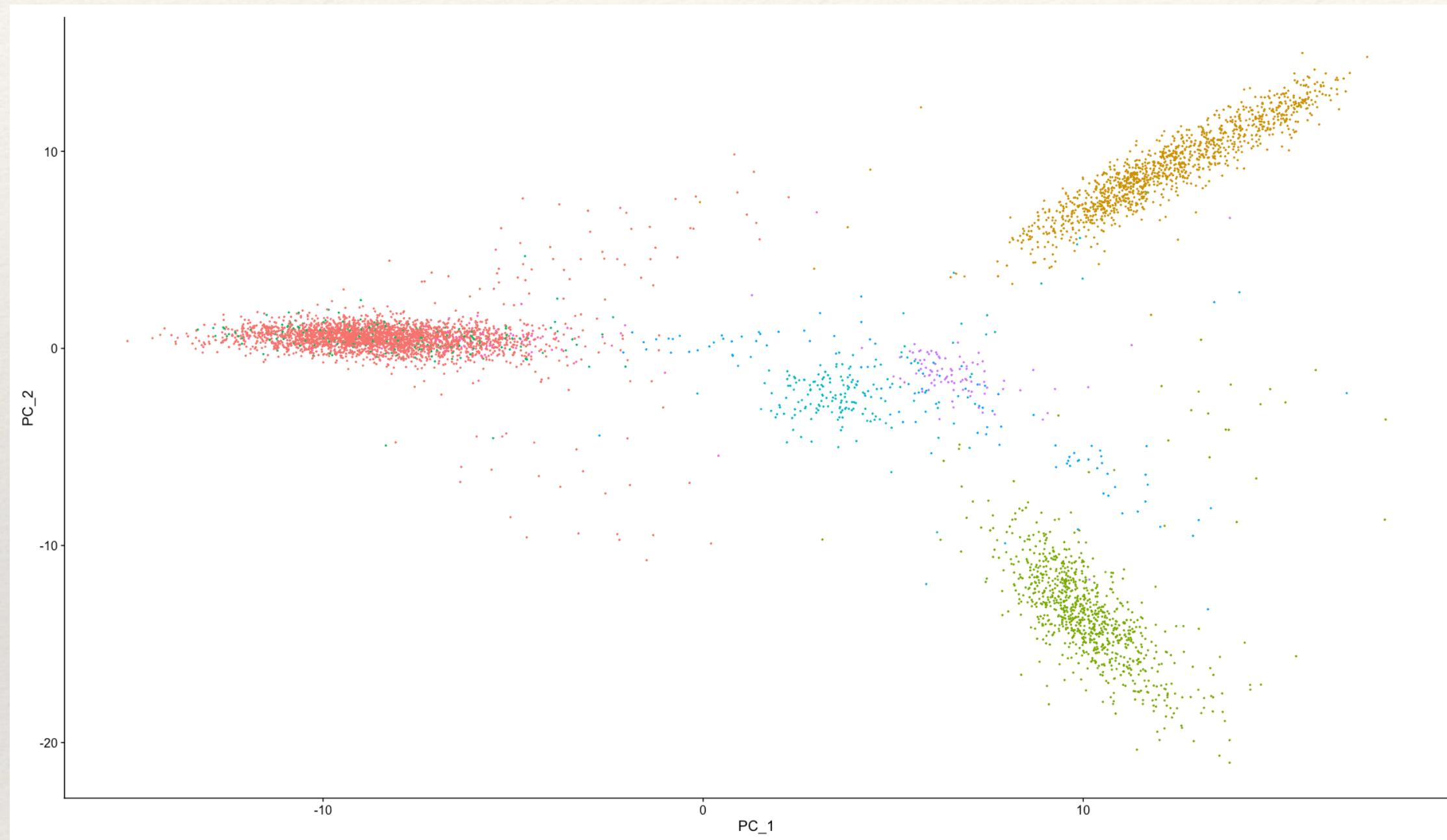


Elbow Plot

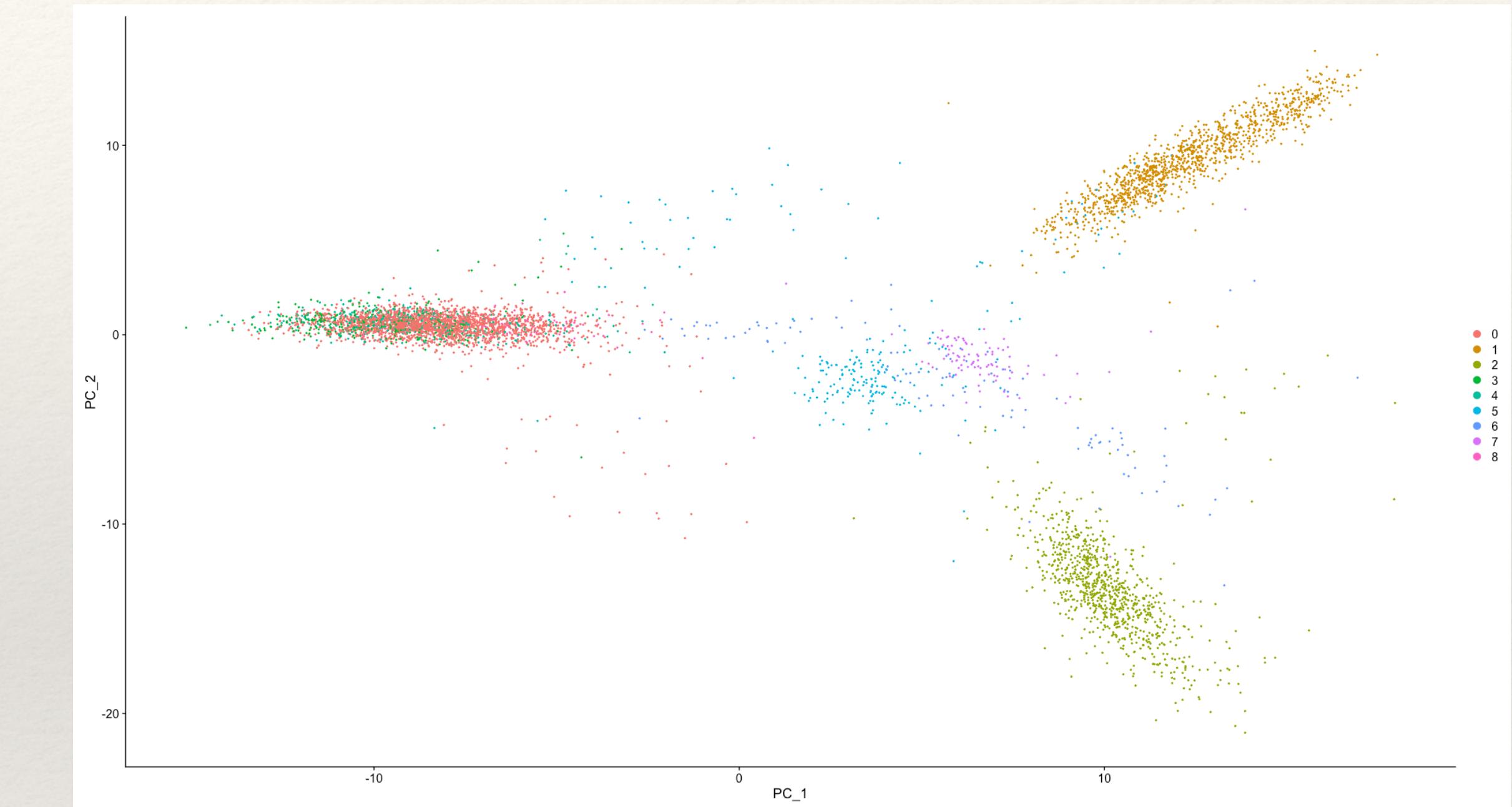


RESULTS: Principal Component Analysis

10 PCs



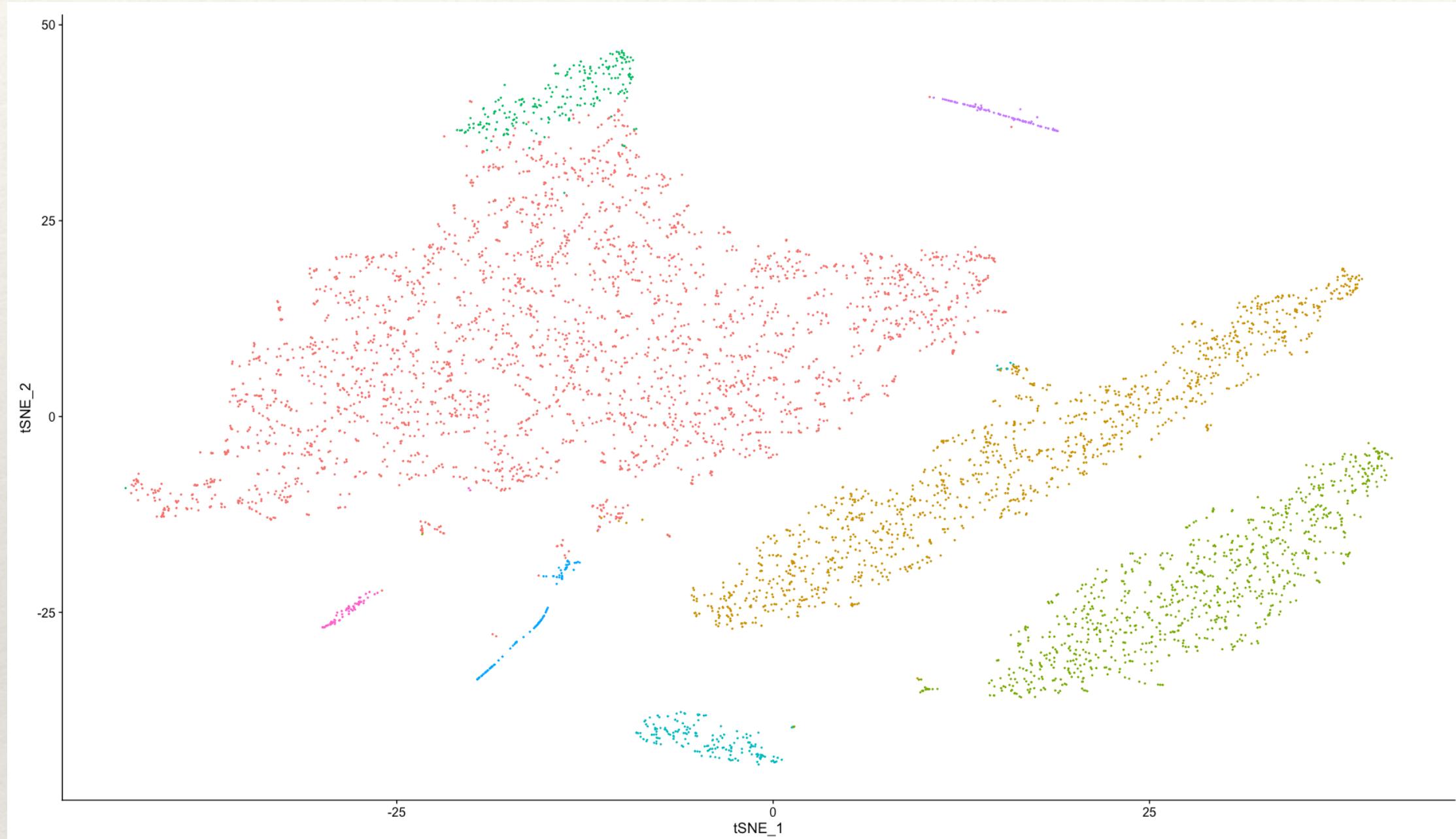
15 PCs



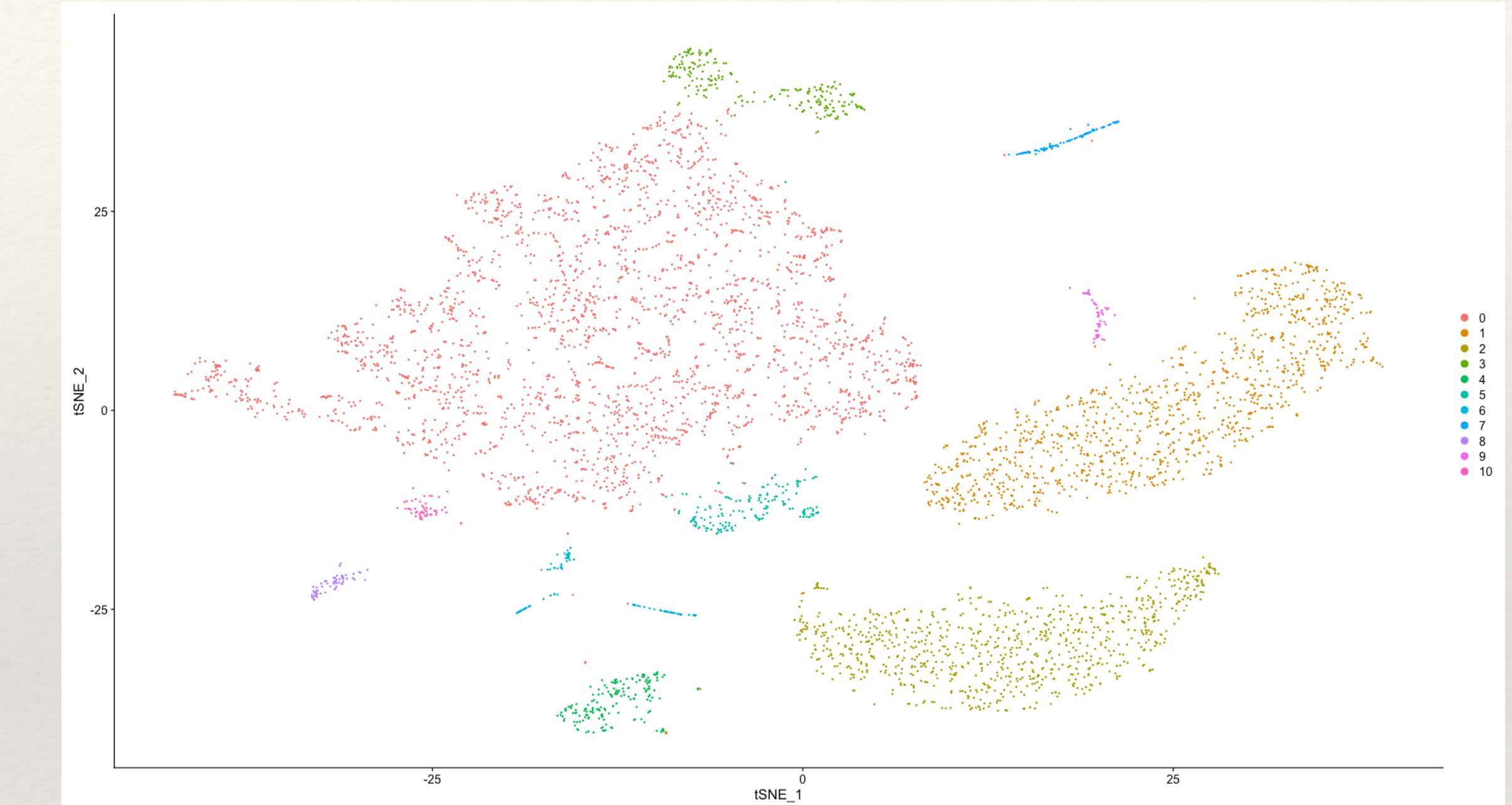
After obtaining the different clusters and plotting them with a t-SNE, I checked the presence of marker genes in the different clusters found

RESULTS: Alternative clustering comparison

10 PCs



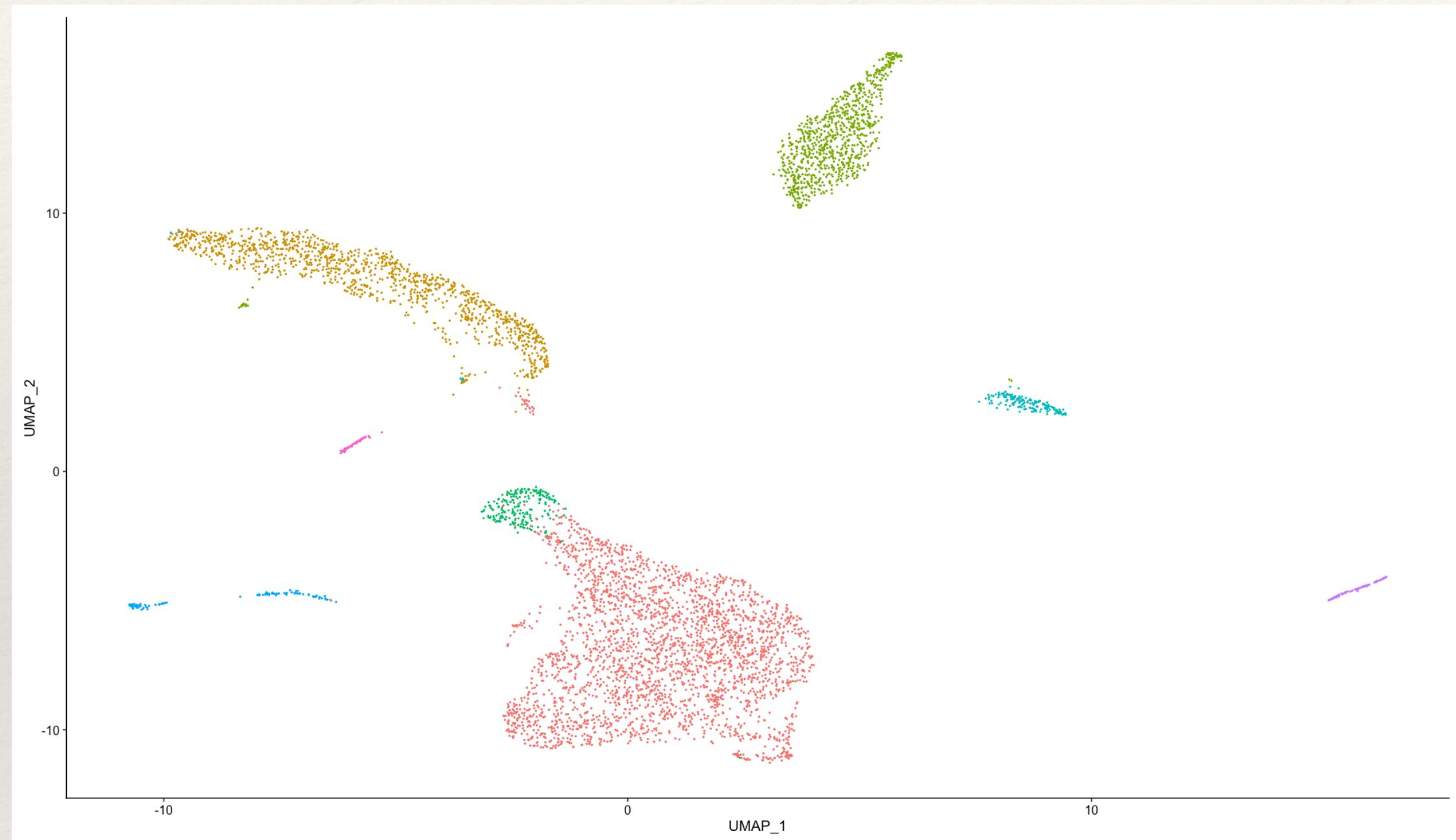
15 PCs 0.2 resolution



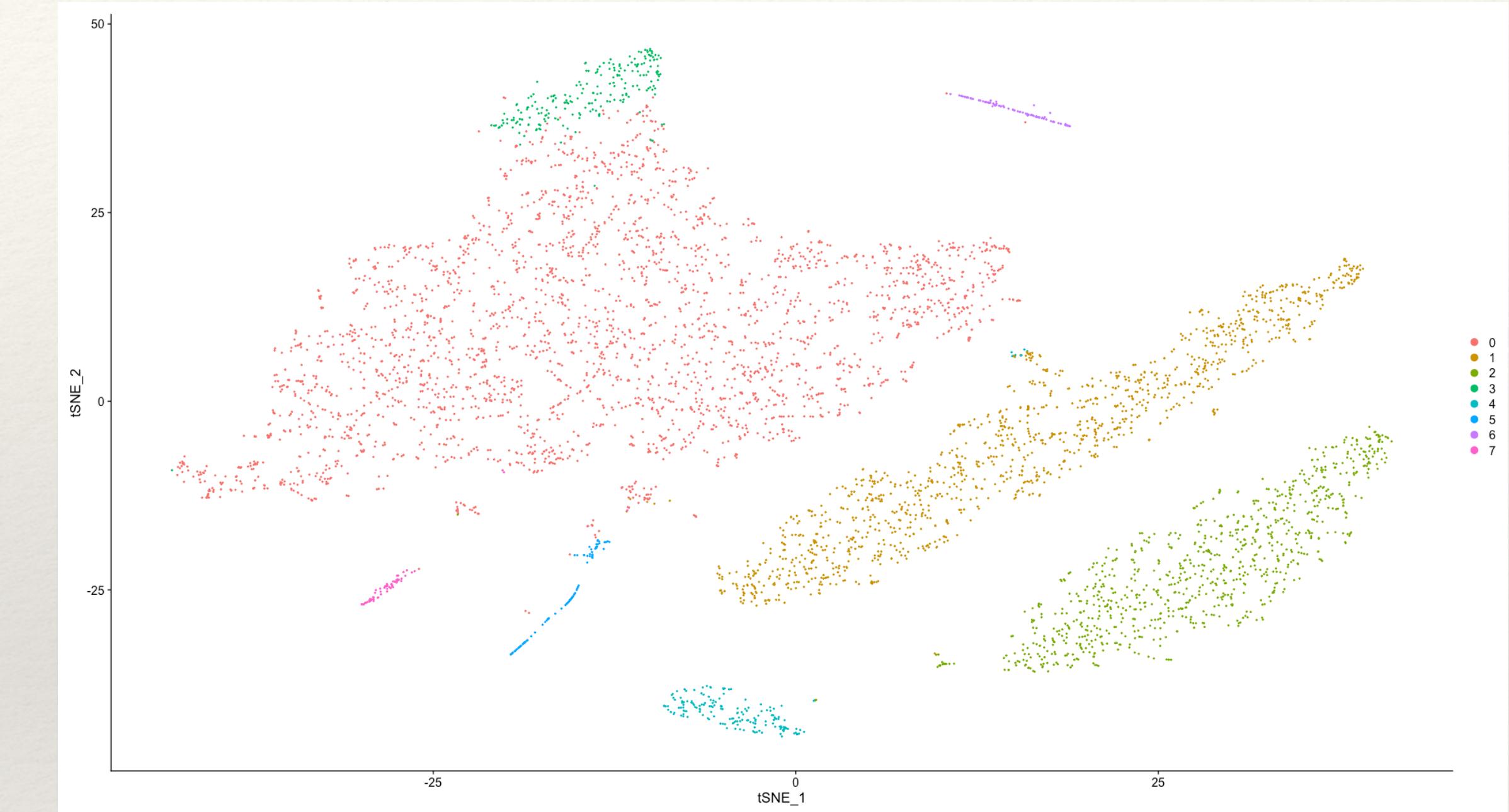
Considering the gene expression and the differentiation between the clusters, I decided to use 10 PC

RESULTS: 2D projection

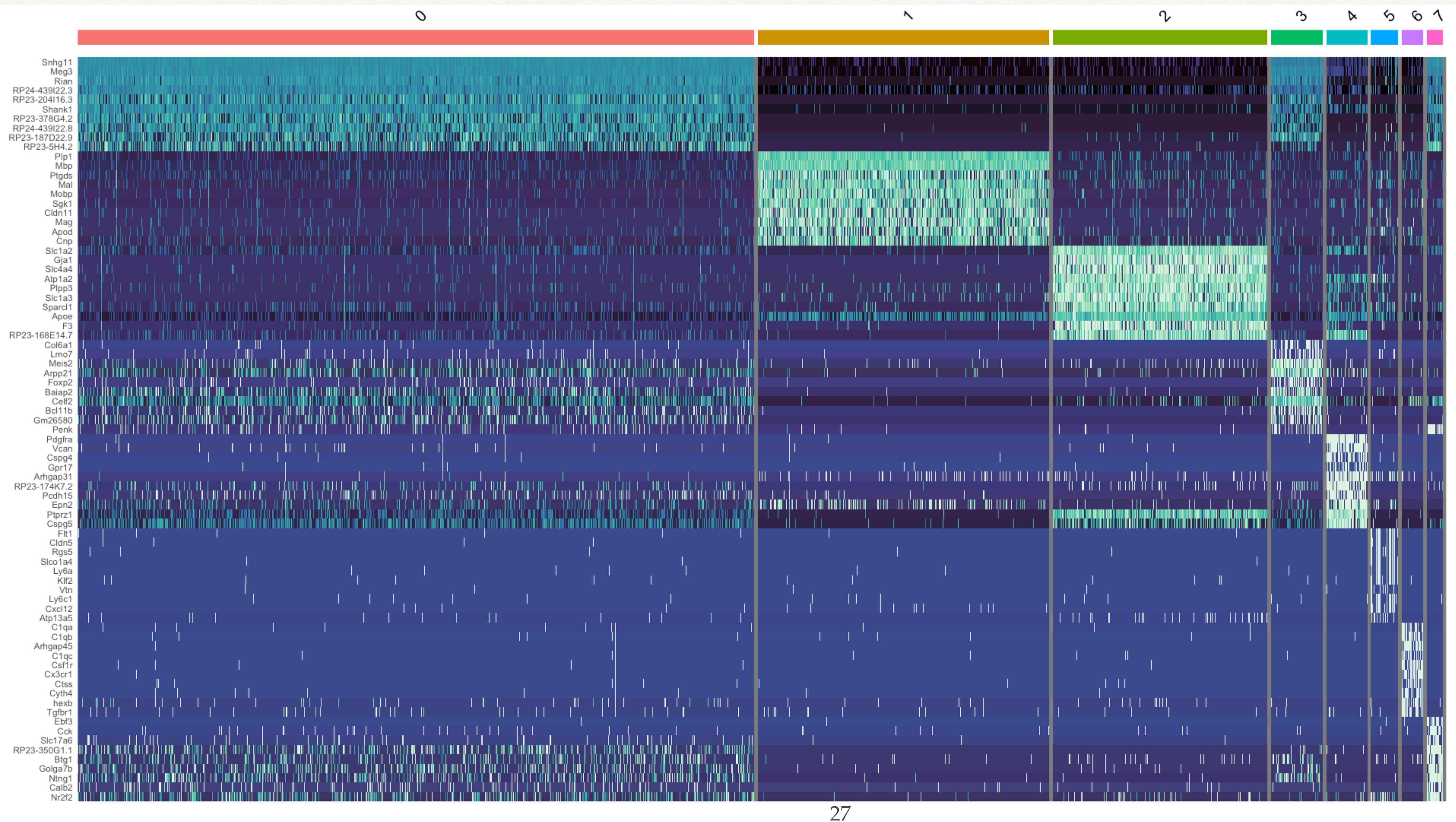
UMAP



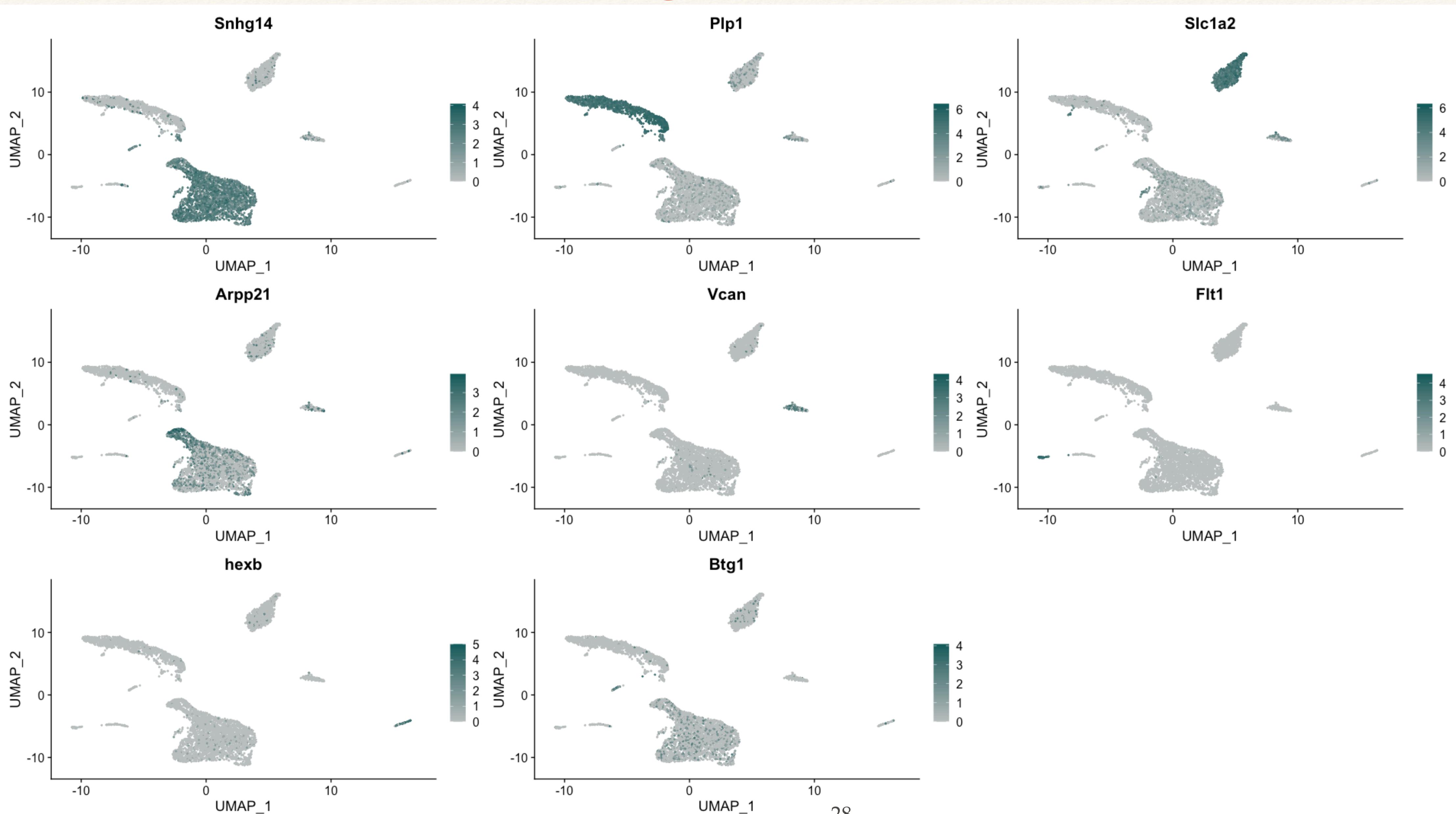
t - SNE



RESULTS: Heatmap



RESULTS: Marker genes



RESULTS: Final result

