

# Exercise 10

Dena Leshkowitz & Gil Stelzer

## Analyzing transcript assembly of RNA-Seq data

### Introduction

In this assignment, you will analyze outputs from scRNA-Seq analysis.

The data that you will analyze was obtained using 10X chromium single cell RNA sequencing a research on the neurohypophysis (NH). This research will be released soon, this is the work of: Qiyu Chen, Dena Leshkowitz, Janna Blechman and Gil Levkowitz, title: Single-cell molecular and cellular architecture of the mouse neurohypophysis, Accepted to eNEURO.

The NH is located at the posterior lobe of the pituitary, is a major neuroendocrine tissue, which mediates osmotic balance, blood pressure, reproduction, and lactation by means of releasing the neurohormones oxytocin and arginine-vasopressin from the brain into the peripheral blood circulation. However, despite the physiological importance of the NH, the exact molecular signature defining neurohypophyseal cell types and in particular the pituicytes, remains unclear.

### Part I - Analysis of the CellRanger report

Open the html report (web\_summary.html) located at -

~/course\_2019/scRNA\_ex10/cellranger with Chrome or any other web browser

This is the report for the first sample (batch1) out of two samples.

#### Question 1

From the SUMMARY sheet - How many barcodes were assigned to cells?

#### Question 2

From the SUMMARY sheet - What is the range of UMI expression per cell?

#### Question 3

From the SUMMARY sheet - What could be the reason for cells with low and high extreme levels of expression?

#### Question 4

From the ANALYSIS sheet - Observe the clustering which is Graph-based.

Using the table that shows the Log<sub>2</sub> fold change and p-values, what is the problem with the genes that are enriched in cluster 2 (the table may be sorted by clicking the desired table column header)?

### Question 5

From the ANALYSIS sheet sequencing saturation plot - Do you think that sequencing more reads will result in more transcripts recovered?

## Part 2: Analysis of Seurat outputs

The Seurat analysis is located in `~/course_2019/scRNA_ex10/Seurat`

### Question 6

Open the tSNE - How many clusters were created?

### Question 7

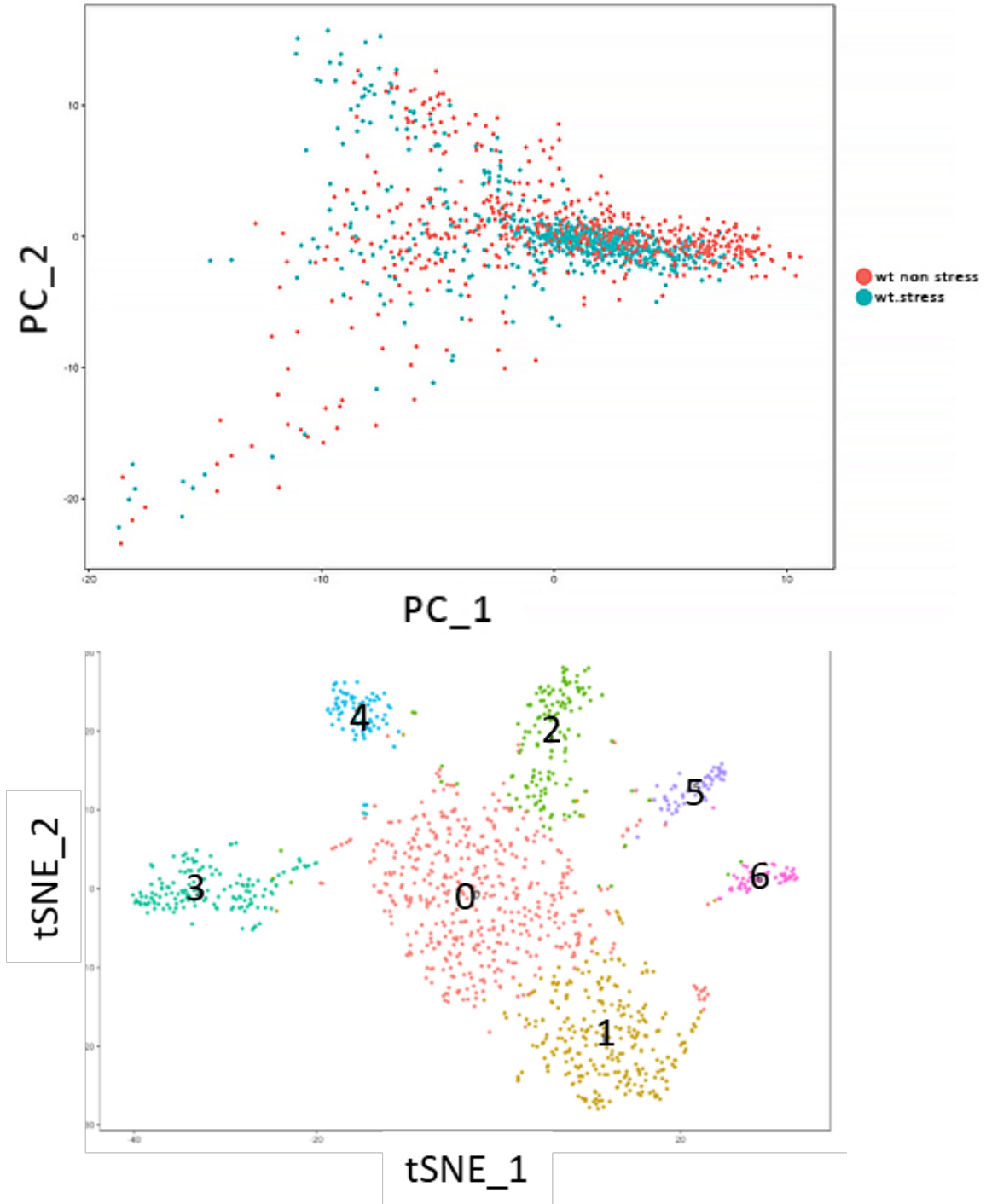
Open the clusterHeatmap.png file. The plot shows the top 20 differentially expressed genes per cluster that are up-regulated. Do you think there are clusters that are redundant and can be merged?

### Question 8

Open with excel the file differential\_genes\_report\_per\_cluster.xlsx file. Select the top 12 genes of cluster1 and search Panglaodb (<https://panglaodb.se/>) to find which cells they most resemble.

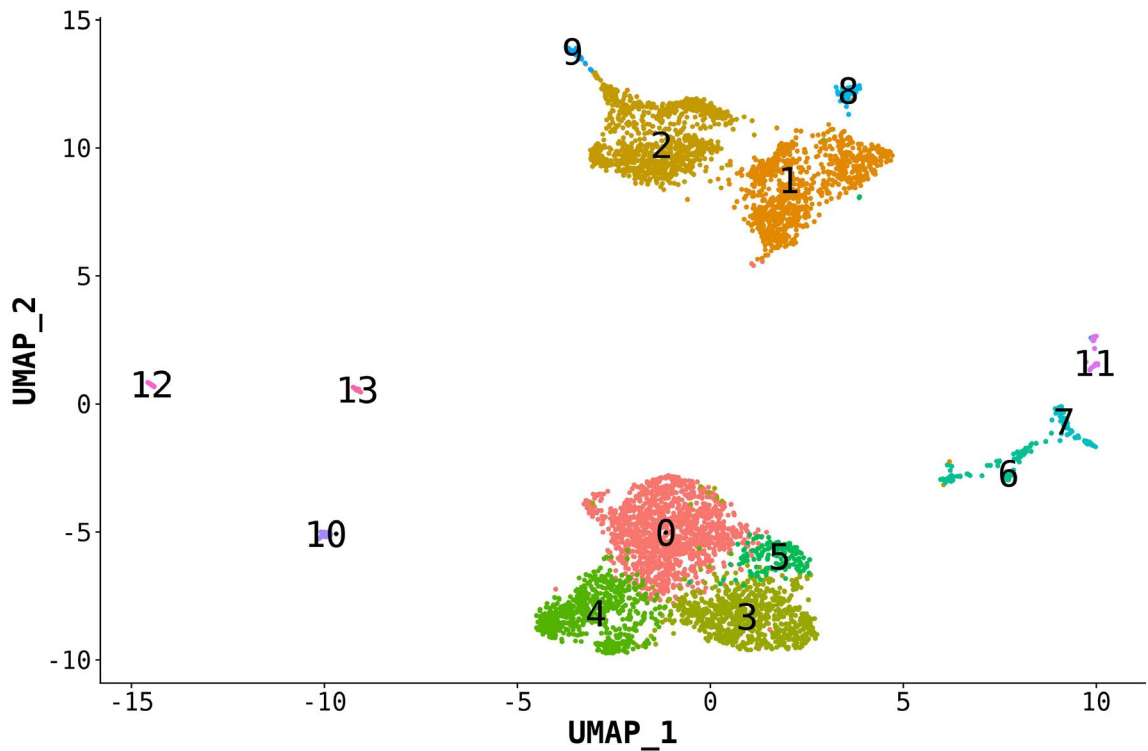
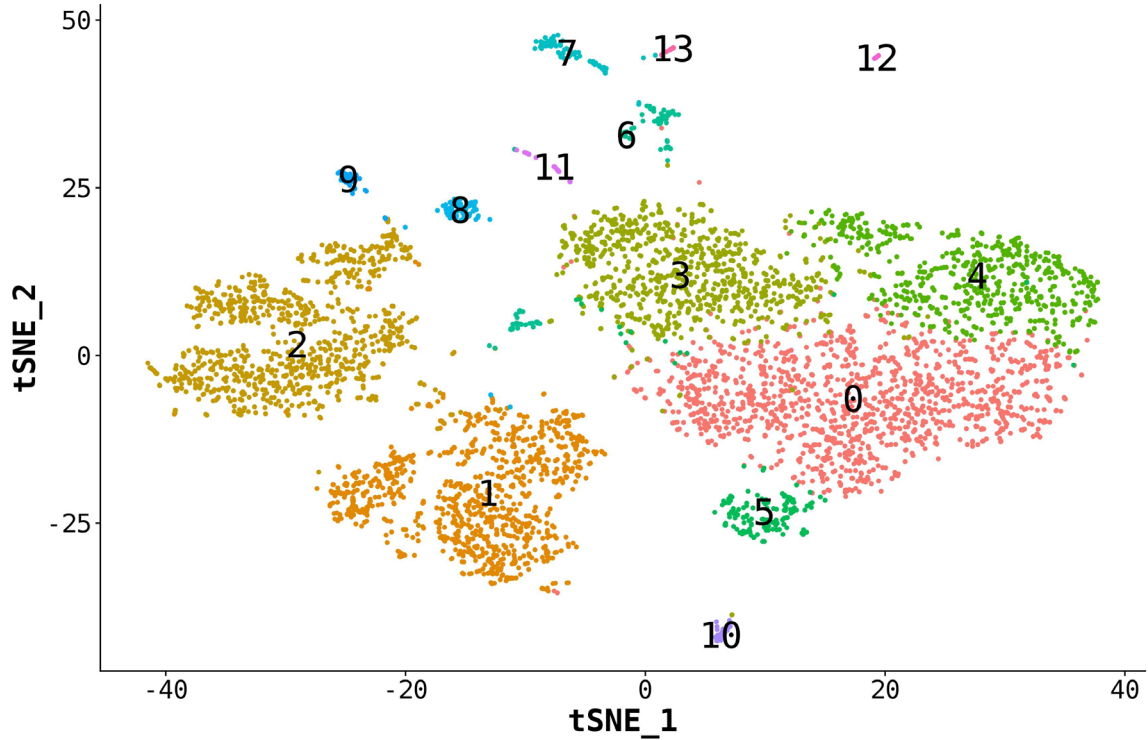
### Question 9

Look at the PCA and tSNE plots that appear below. What is the main reason that the cells appear as one cluster in the PCA plot whereas they appear as 7 clusters in the tSNE plot?



Question 10

Which cluster has the most similar expression pattern compared to cluster 8, based on the plots below?



### Question 11

Two genes were differentially expressed when comparing the clusters in the green circle compared to the one in the blue circle. We are satisfied with the parameters used for determining differential expression. Since the expression of both genes is similar in both

clusters as shown in the plots below, what could be the reason they were separated into two clusters and not included into one cluster?

