# singlecell_Seurat_pipeline

2023-02-12

## Single-cell RNA-Seq analysis pipeline (GEO dataset)

This is a pipeline to analyze single-cell RNA Seq data from GEO. In this script the single cell RNA-Seq results from this paper is regenerated: https://pubmed.ncbi.nlm.nih.gov/34956864/ Title: Bulk and Single-Cell Profiling of Breast Tumors Identifies TREM-1 as a Dominant Immune Suppressive Marker Associated With Poor Outcomes

## Initial library loading

## 0. Fetch data from GEO

**Note: No need to fetch the data if it is already created**

## 1. Create the counts matrix

**Find the files in data folder (i.e., barcodes, features, matrix)**

## 2. Create a seurat object

## 3. QC and filtering

**View Seurat object meta data**

## 3.1 Calculate mitochondrial percentage

**High percentage shows bad quality**

## 3.2 Explore QC

**Plot feature and RNA counts and mitochondial percentage**

## 3.3 Filter cells

**more than 800 RNA count and more than 500 genes and less than 10 mitochondrial%**

## 4. Finding Variable genes

**note: This data set contains one sample, if more than 1 sample was used correction for potential batch effects should be considered**

**4.1 Normalize data**

**4.2 Find variable genes**

**with visualization of top 10 variable genes**

## 5. Clustering the cells

**5.1 Scale the data**

**5.2 Perform linear dimensionality reduction**

**5.3 Select the PCA plots with elbow plot**

**5.4 Find Neighbors**

**5.5 Understand the resolution**

**5.6 Optimize the resolution**

**This resolution should be changed until the correct number of clusters are achieved**

**5.7 Set identity of clusters**

**5.8 umap observation**

**5.9 tsne observation**

## 6. Annotate the clusters

**6.1 Find the assay type**

**6.2 Find the differentially expressed markers**

**Find markers for every cluster compared to all remaining cells, report only the positive ones**

**6.2 Select top (4 -> can be changed) upregulated genes in each cluster**

**6.3 Visualize top upregulated genes in each cluster**

**6.4 Ridge plots - from ggridges.**

**Visualize single cell expression distributions in each cluster**

**6.5 Violin plot**

**Visualize single cell expression distributions in each cluster**

**6.6 Feature plot**

**visualize feature expression in low-dimensional space**

### 6.7 Check the individual feature

**Use for optimization of annotating**

### 6.8 Select the top features

**Feature identification based on up-reg genes and pangloadb**

### 6.9 Assign the new clusters to data

## 7. Visulization

### 7.1 View annotated clusters

### 7.2 Single cell heatmap of features expression

### 7.3 Feature plot

**Visualize feature expression in low-dimensional space**

### 7.4 Heatmap plot

**Select top (6 -> can be changed) upregulated genes in each cluster**

### 7.5 Dot plots

**the size of the dot corresponds to the percentage of cells expressing the feature in each cluster. The color represents the average expression level**