INSTITUTE OF BIOORGANIC CHEMISTRY
Polish Academy of Sciences

Interview exercise

Requirements:

1. Single cell data: https://github.com/jkubis96/Exercise_interview.git

2. R softwere (recommended IDE: RStudio)

2. Seurat library for analysis: https://satijalab.org/seurat/articles/install.html

3. Other useful libraries: tidyverse, readxl

Exercise I:

1. Download necessary data

2. Load data to R and run the analysis using the Seurat package

3. Conduct basic visualization: count ~ genes (scatter plot), number of genes (violin plot), percent [%] of mitochondrial genes, etc

4. Conduct data normalization [explain this type of normalization]

5. Find the most variable marker genes in the data set and visualize them by plot

6. Conduct scaling of the data and determine principal components (PC) [It will be necessary for the clustering]

7. Choose the right principle components for the next analysis (use Elbow plot, JackStraw plot, or both for visualization)

8. Conduct clustering using SNN/KNN algorithms based on previously chosen PCs

9. Run tSNE or UMAP and visualize obtained clusters

10. Find marker gene specific to obtained clusters based on average FC and p-val < 0.05

11. Select top 10 markers for each cluster based on avg_log2FC and save into *.xlsx format