Computational Exploration of Single-Cell RNA-seq Data Using Seurat on Human Immune Cells

Objective: Students will perform an end-to-end single-cell RNA-seq analysis using Seurat in R, starting from an .h5ad dataset. The workflow includes quality control, clustering, marker discovery, and biological interpretation.

Dataset:

Input dataset: https://datasets.cellxgene.cziscience.com/36a22a49-8532-43d4-afa2-d90c894a6c7a.h5ad

Source portal: https://cellxgene.cziscience.com/datasets

Reference study: https://www.nature.com/articles/s41586-025-09686-5

Tutorial reference: https://bioinformatics.ccr.cancer.gov/docs/getting-started-with-scrna-seq/Seurat_QC_to_Clustering/

Research Questions (Students must answer all)

- Q1: What QC thresholds did you choose and why? Justify using plots and reasoning.
- Q2: How stable are your clusters when varying PCs or resolution?
- Q3: Assign cell type identities to clusters using marker genes and evidence.
- Q4: Perform differential expression between two biologically meaningful groups and interpret results.
- Q5: Provide one study-linked insight (rare cluster, doublet detection, gradient pattern, or gene program).

Required Figures

- Figure 1: QC violin plots before and after filtering
- Figure 2: QC scatter plots
- Figure 3: PCA diagnostics
- Figure 4: UMAP colored by clusters and cell types
- Figure 5: Marker heatmap/dotplot and feature plots
- Figure 6: Differential expression result

Submission Structure

- script/: analysis.R or analysis.Rmd

- figures/: all required figures
- tables/: marker and DE CSV files
- report/: 2–4 page PDF report
- session_info/: sessionInfo.txt

Final Instructions: Zip the project folder, upload to Google Drive, ensure viewer access, and submit the link via LMS.