RAYCA

Bioinformatics Assessment Task Instructions:

For this task, you are required to analyze single-cell RNA sequencing (scRNA-seq) data to evaluate your ability to process and analyze complex datasets, generate visualizations, and draw meaningful biological conclusions. The dataset you will use is publicly available at the following GEO repository link. Download the following files provided in the repository: matrix.mtx.gz (raw gene expression counts), barcodes.tsv.gz (cell barcodes), and features.tsv.gz (gene names/features).

Data Preprocessing and Seurat Object Creation

Utilize the R package Seurat, designed for quality control, analysis, and exploration of single-cell RNA-seq data https://satijalab.org/seurat/. Load the downloaded data into R and create a Seurat object. Perform initial quality control measures, including filtering out low-quality cells and genes, normalizing the data, and identifying highly variable features.

Dimensionality Reduction and Clustering:

With the Seurat object prepared, proceed to perform dimensionality reduction using Principal Component Analysis (PCA) to capture the primary sources of variation in the dataset. Following PCA, apply Uniform Manifold Approximation and Projection (UMAP) to visualize the data in a two-dimensional space, facilitating the identification of distinct cell populations. Subsequently, conduct clustering to group cells with similar expression profiles, which will aid in the annotation of cell types.

Cell Type Annotation

Annotate the identified clusters by comparing their gene expression profiles to known marker genes. This process involves assigning biological identities to each cluster based on the expression of canonical markers, thereby elucidating the cellular composition of the dataset. Seurat provides functions to assist in this annotation process.

Trajectory Inference and Pseudotime Analysis

To explore the dynamic processes within the data, employ **Monocle 3**, a tool for trajectory analysis in single-cell RNA-seq data https://cole-trapnell-lab.github.io/monocle3/. Construct a trajectory to model the progression of cells through different states and calculate pseudotime values, which represent the inferred temporal order of cells along the trajectory. This analysis will provide insights into the developmental pathways present in the dataset.

Kinetics Plot of Marker Genes:

Identify marker genes associated with the inferred trajectory, focusing on those relevant to the developmental processes under study. Generate a kinetics plot to display the relative expression levels of these marker genes across pseudotime, illustrating how their expression changes as cells transition through different states. This visualization is essential for understanding gene regulation dynamics during development.

RAYCA

Documentation and Reporting:

Throughout the analysis, maintain comprehensive documentation of your methodology, including code, parameters used, and rationale for analytical choices. Summarize your findings in a report that includes visualizations such as UMAP plots, trajectory plots, and kinetics plots, accompanied by interpretations of the results. This report should demonstrate your ability to derive and communicate the results in a scientific tune.

Submission Instructions

Please submit your results via email at internship@raycabio.com

You have 5 days to complete and submit your report. Ensure that your submission is professional, comprehensive, and well-organized.

Evaluation Criteria

Your submission will be evaluated based on:

Completeness: Whether the report includes all required elements.

Accuracy: The quality and precision of your analysis.

Clarity: How well the report is written and visualized.
