

Attached is a RDS object (PhD_test_obj.rds), a list containing the following

- Count
A count matrix of single cells against genes
- Metadata
The metadata of the corresponding count matrix containing cell type, label and replicate

This dataset has only 2 conditions (Disease vs Control), 15 cell types, and 5 replicates (or samples).

Task 1

The first task is to evaluate if you are able to extract relevant biological insights from a single-cell data with perturbations.

You are to

1. Identify the genes that are differentially expressed between the 2 conditions
For this task, please provide a final table with at least the following columns
 - a. Cell type
 - b. Gene
 - c. Effect size (log-2 fold change)
 - d. P-value
 - e. Adjusted p-value
2. Rank each cell type based on how much they are perturbed by the condition (from most perturbed to least perturbed)

You can decide on any available methods or provide your method to perform these analyses. You are also welcome to include additional analyses that you think is important as a biological insight.

Task 2

The second task is from a more translational standpoint – if we can use single-cell sequencing data for phenotypic classification.

You are to

1. Build a classifier that can predict disease versus control from the single-cell data
Below are some hints
 - a. Decide if the classifier needs to be cell-type specific (each cell type respond differently to the perturbation)
 - b. Decide on the train-test split; is it more useful to split train-test based on replicate or cells?
2. Please provide the following of your best performing classifier:
 - a. Accuracy
 - b. Area under the ROC
 - c. Confusion matrix

For **Task 1** and **Task 2**, please provide all the scripts and any data visualization figures you deem necessary.