Downstream Analysis Tutorial

Samuele Soraggi

Table of contents

This tutorial will give you the extensive basic commands and explanations for the single cell analysis of your own dataset.

- The **first part of the tutorial** (Section **??**) is focused on **preprocessing** the data, which means primarily filtering and normalizing it.
- The **second part of the tutorial** (Section **??**) is focused on **integrating** all sixteen datasets produced from the lab sessions (you will perform this integration analysis in groups), identifying cell types and find a population of cells expressing the HAR1 gene to analyze different conditions of mutant VS wild type Lotus japonicus.
- The **third part of the tutorial** (Section **??**) applies tipycal gene analysis to detect genes conserved and differentially expressed between conditions
- The **fourth part of the tutorial** (Section **??**) pivots around the study of groups of genes co-expressed in the data and in specific clusters and conditions

The first two parts follow the phylosophy of the best practices explained in Luecken and Theis (2019) and Heumos et al. (2023). The third part applies standard statistical tests on the average gene expressions in subsets of the data. The last part is based pulling cells transcripts together with different granularities to improve the statistical power of calculations based on their gene expression (as in Morabito et al. (2023)).

The tutorial is based on four samples of Lotus Japonicus (two rhizobia-infected and two wild types) from Frank et al. (2023). The last section follows some of the tutorials from hdWGCNA.

> Learning outcomes
>
> At the end of this tutorial **you will be able to use `R` to**
>
> - **Filter** your data selecting specific criteria