Introduction to single-cell transcriptomics Mandatory TD

Pick one project and perform the requested analyses (and more if you can). A report containing all your code as well as results (plots and/or tables) and a discussion of those must be returned to me via Moodle **by December 12, 2026** the latest. You can work in groups of 2 students. R Markdown is a convenient format to both analyze data and generate your report. Alternatively, you can use a Jupyter notebook. In both cases, you can generate an HTML or PDF output with code, plots and your comment & analysis.

Both projects rely on the breast cancer cell atlas (Wu, Nat Genetics, 2021) from https://singlecell.broadinstitute.org/single_cell/study/SCP1039/a-single-cell-and-spatially-resolved-atlas-of-human-breast-cancers (you need to log in to download). You will need the three usual files for Chromium (we are only interested in the single-cell data): `barcodes.tsv.gz`, `features.tsv.gz`, and `matrix.mtx.gz`. In addition, the table `Whole_miniatlas_meta.csv` gives you the cell type annotations assigned by the authors. All these files are available from the download page.

This **data set is rather large**; you need a good PC or you need to **sub-sample** it (*e.g.*, keep 25% of the cells through random selection).

Project 1: Bioconductor SCT tools

Bioconductor includes a comprehensive set of packages meant to analyze scRNA-seq data. A selection of these tools to perform all the common analyses we introduced is presented in a tutorial at

https://training.bioconductor.org/

under the module “bioc-scrnaseq”.

Apply the Bioconductor training workflow up to step 3, cell type annotation, and compare with the annotations by the authors.

Project 2: Cellular networks

Load data using Seurat and export the count matrix. Use a tool to infer the ligand-receptor interactions (LRIs) between all pairs of cell populations as defined by the authors. Manage to produce one or several graphic representations of these LRIs (total numbers and eventually some more limited selections of cell populations or LRIs).

Restrict the LRIs to immune checkpoints and report how many immune checkpoints are potentially active between each pair of cell population in a second plot.

To perform LRI inferences, you can either use BulkSignalR available from Bioconductor (https://www.bioconductor.org/packages/3.23/bioc/html/BulkSignalR.html) or CellChat from GitHub (https://github.com/jinworks/CellChat). In case you go for BulkSignalR, use the old LR-score through a SCSRNoNet class.

A reference listing immune checkpoints, you have to find by yourself.