Introduction to single-cell transcriptomics Mandatory TD

Pick one project and perform the requested analyses (and more if you can). A report (one PDF file) containing all your code as well as results, and a discussion of the latter must be returned to me via Moodle by December 22, 2024 the latest. You can work in groups of two.

Project 1

Download the human breast cancer spatial data from https://www.10xgenomics.com/resources/datasets/human-breast-cancer-whole-transcriptome-analysis-1-standard-1-2-0. Use filtered data.

Then, download 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: `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. Assign the cell types to all the cells. This data set is large; you need a powerful PC or to do the project on a compute server.

Then, annotate the regions of the spatial transcriptomic data using the single-cell atlas (check Seurat documentation) using the functions `FindTransferAnchors()` and `TransferData()` using the single-cell atlas cell type annotations. Discuss the result.

Next, select variable genes with the spatial-specific function `FindSpatiallyVariableFeatures()` and generate a figure showing their expression over the tissue. Discuss the correspondence with cell type annotations of the tissue regions and submit the variable genes to a gene set enrichment analysis tool to discover if specific pathways were spatially variable.

Project 2

Load the serialized single-cell Seurat object `myeloid-cell.RDS` with `readRDS()` function. The data were already normalized and must be downloaded from the following link: https://filesender.renater.fr/?s=download&token=9a5ca826-643a-4606-8af1-5d71c2c9b42c

Perform PCA and UMAP dimension reductions and identify the present immune cell populations. To help, start by doing the clustering by adjusting its parameters to find 4 clusters. The search for marker genes and/or the use of known immune population gene markers will give you the populations. Hint: one population is macrophages.

Once this is done, install the `SingleCellSignalR` Bioconductor package and determine the ligand-receptor interactions these cell populations have between them. There will be many.

Lastly, limit your analysis to macrophages and check the expression of the genes *CXCL9* and *SPP1*. What do you observe? What happens in the other cell populations?