Integrative Cellular Biology & Bioinformatics

CHAIR FOR CLINICAL BIOINFORMATICS
SAARLAND UNIVERSITY

Project 3

The dataset that will be analyzed in this project is a 10X Visum dataset that includes the spatial gene expression data of a mouse brain. For the analysis, the Seurat package will be used.

Deadline: 28.01.2022 23:59

The same rules apply as for the previous projects.

Tutorials:

- If you have any Problems, please contact Omar (omla00001@stud.uni-saarland.de) or visit the tutorials.
- During the project, there will be tutorials on some of the tasks.

Publishing of the project	03.01.2022	Introduction
	10.01.2022	Spatial Transcriptomics in Seurat
	17.01.2022	Differential Expression Analysis
	24.01.2022	Data integration and Cell type identification
Deadline Project	28.01.2022	

Submission:

- You will have to submit **one** tar.gz file that includes
  - The code
  - PDF file including all images and responses to the questions
- The code must be well commented and must run without an error to obtain any points
- If you use any other sources for your answers, don't forget to give the reference

Introduction in Seurat

The tool that you will mainly use in this project is Seurat. It is a tool that combines many functionalities for the analysis of single cell data. For the start, there is a good documentation here for the first steps of the analysis.

Seurat includes the functionality to explicitly analyse spatial transcriptomics data. A tutorial and good introduction can be found here.

Download the data

You can download the dataset for this project here.

For each sample you will get 2 files: A '.tar.gz' file and a '.h5' file. You will need both files for the analysis.

Task 0: Spatial Transcriptomics (Bonus: 5P)

10X genomics Visum technology. The given dataset is a spatial transcriptomics dataset of a Mouse Brain. It has been aligned and prepared using the Space Ranger software.

Task 0.1: Properties of the Slides

Give the spot size, the distance between spots and the number of spots for this technology.

Task 0.2: Resolution of the spatial transcriptomics technology

Compare the resolution of the technology with the size of an average eucaryotic cell and explain what this means when dealing with the data.

Task 0.3: Output of Space Ranger

Have a first look at the given data. Give the image taken of the sample, the coordinates of the spots and one gene-expression matrix.

Task 1: Spatial transcriptomics data in Seurat (5P)

Have a look at this tutorial: https://satijalab.org/seurat/articles/spatial_vignette.html

Task 1.1: Loading the data

You can load the data using the Load10X_Spatial function that will require both files (the H5 file and a folder with the image data).

Task 1.2: Inspecting the Seurat-object

Inspect the Seurat object and which data is stored in it. Where are the gene-expression data and the tissue image stored? How can you access the data?

Task 1.3: Visualization of a feature

You can visualize features in the data as you already did in the scRNA-seq data or visualize it directly in the tissue slide. Pick two random genes and visualize the gene expression in the tissue.

Task 2: Data preprocessing (7P)

Now you have learned how to handle spatial transcriptomics data in Seurat, we can move on to the data analysis. The preprocessing is very similar to the first project.

Task 2.1: Filtering

Filter the data using the number of features per spot and the number of mitochondrial features.

Choose the cut-off values by plotting the distribution of both variables in the sample and selecting a cut-off value that removes the outliers. Be careful to not set these values too low.

Compare the found thresholds to the thresholds used for the scRNA-seq data in the first project. Explain the differences also with regard to how spatial transcriptomics data are produced.

Task 2.2: SCTransform

Use SCTransform for the further preprocessing of the data.

Which steps of the preprocessing from project 1 are replaced by this function?

Task 3: Dimensionality reduction, clustering and visualization (5P)

Task 3.1: Dimensionality reduction

Use PCA to do a dimensionality reduction of the data. Use a plot (e.g. elbow plot) to determine the number of dimensions to use in the next steps.

Task 3.2: Clustering

Perform a clustering of the data based on the result of the PCA.

Task 3.3: UMAP

For the visualization of the clustering use UMAP to project the data into the 2-dimensionaly space.

Show the clustering in the 2-dim UMAP space and in the tissue slide (1 plot each).

Task 4: Differential Expression Analysis (10 P)

For this task pick one of the samples to do the analysis on.

Task 4.1: DEG (Differentially Expressed Genes) analysis based on the clustering

Compare the gene-expression between the different clusters. Therefore, perform a DEG analysis only based on the gene expression data. (see the “Find differentially expressed features” section in this tutorial).

Save the differentially expressed genes as you will need it in the next task.

Task 4.2: DEG analysis based on the spatial patterning

Another method to find differentially expressed genes in the spatial transcriptomics data also includes spatial information. Use Seurat to find the top 3 spatially variable features.

Visualize the gene expression in the tissue slide.

Compare the result to the result from Task 4.1. Are those genes also differentially expressed between clusters?

Task 5: Merging the data (8P)

Task 5.1: Merging without Batch-correction

Merge the two datasets without Batch-correction.

Task 5.3: Dimensionality-reduction and clustering

Repeat the dimensionality reduction and clustering steps from Task 3 for the merged dataset and show the visualization in the UMAP and the Tissues.

Compare the clustering within the two tissues. Which clusters can you observe in both samples? And which clusters can you find only in one sample?

Task 5.2: Merging with Batch-correction

Use Data Integration to combine the two datasets. Again, repeat the dimensionality-reduction and clustering steps.

Task 5.3: Detection of Batch-effects

Show the results in the 2-dimensional UMAP space. Compare the results of both methods. Are there Batch-effects that have been resolved?

Decide and argue if Batch-correction is needed or not and use the according data for the further processing.

Task 6: Cell-type identification (8P)

Download the reference dataset here. Prepare the data using SCTransform, Dimensionality-reduction and clustering. Tansfer the labels of the scRNA-seq dataset to the spatial transcriptomics dataset. Plot the Annotation in the UMAP space.

Hint: You can also have a look at this tutorial: https://learn.gencore.bio.nyu.edu/seurat-integration-and-label-transfer/

Task 7: Manual Annotation (7P)

Perform a DEG analysis as in task 4.1 on the whole dataset. Pick three celltypes from the annotation from task 6 and use the CellMarker database to find one marker-gene for each of the cell types.

Visualize the gene expression of the marker-genes in a UMAP plot and in each of the tissue slides.