Single-cell RNA-seq data analysis in Chipster 19.9.2018

Maria Lehtivaara, Eija Korpelainen
chipster@csc.fi

PART 2: Seurat with 10X Genomics data

In this hands-on session you will get familiar with Chipster's Seurat based tools, which filter the DGE table and cluster cells in order to find subpopulations.

The 10X Genomics data used in this tutorial is originally the example data for Seurat tools:
http://satijalab.org/seurat/get_started.html
It's a dataset of 2700 peripheral blood mononuclear cells (PBMCs).
We have imported in Chipster a tar package that contains the three 10X Genomics output files.

1. Open example session

Click **Open example session** and select the session **course_single_cell_RNAseq_Seurat**.

2. Setup Seurat object & quality control

Select the **files.tar.gz.** Select tool **Single cell RNA-seq / Seurat -Setup and QC.** Check the parameters, and **name your project** (for example as "PBMC"). **Run** the tool.
Open the **QCplots.pdf** in **external browser**. Look at both pages.

> Based on the plots, what would be the optimal upper limit for the number of genes and mitochondrial transcript percentage? Hint: check the default parameters used in the next tool.

3. Filtering, regression and detection of variable genes.

Select **seurat_obj.Robj** (this is an R-object, which can be exported and opened in R, or just passed to the next tool in Chipster, like we do now). Select the tool **Single cell RNA-seq / Seurat - Filtering, regression and detection of variable genes.** Check if the default parameters are good for this dataset, based on the QCplots? While the tool is running, click the **More help** button, and learn about the three steps this tool performs.
Once the tool is done, open the **Dispersion.pdf** and check also the second page.

> How are the cutoffs working for this data? How many variable genes are there?

4. Principal component analysis

Select **seurat_obj.Robj** from the previous step and run the tool **Single cell RNA-seq / Seurat -PCA.**
Open **PCAplots.pdf** in **external browser**. Look at the heatmaps and the standard deviation of PCs in the last two pages.

> How many principal components should we continue the analysis with (check the elbow in the standard deviation plot, inspect the heatmaps)? Would 10 be ok?

5. Clustering

Select **seurat_obj.Robj** from the previous step. Select tool **Single cell RNA-seq / Seurat - Clustering**. In the parameters, set **Number of principal components to use** = **10**.
While waiting for the tool to run, you can study the manual (click the **More help** button).

> What are the three main steps of this tool?

When the results are ready, study the **tSNEplot.pdf** (open in external browser).

> How many clusters are there in this data?

6. Markers for a specific cluster

Open **markers.tsv** as a **spreadsheet**. Notice that we have all the markers for different clusters in one table. Now, let's choose only markers for cluster 2. Choose **markers.tsv** and the tool **Utilities / Filter table by column value.** Fill in the parameters accordingly:

> **Column to filter by = cluster**
> **Does the first column have a title = no**
> **Cutoff = 2**
> **Filtering criteria = equal-to.**
> How many biomarkers were recognized for cluster 2?

7. Visualize markers

Choose **seurat_obj.Robj** generated in step 5. Select tool **Single cell RNA-seq / Seurat -Visualize markers**. Type a marker **gene name** to the parameter field (choose one of your own, or try for example with MS4A1, LYZ and PF4). You can run the tool several times for different genes.
Open a **biomarker_plot.pdf** in external browser.

Is your gene a good marker for that cluster?

BONUS: Try the Seurat tools for the DropSeq data used in Part 1. You can also analyze your own data.