Single-cell RNA-seq data analysis in Chipster 19.9.2018

Maria Lehtivaara, Eija Korpelainen
chipster@csc.fi

PART 3: Two sample analysis with Seurat

With these exercises, we demonstrate the use of Seurat tools for joint analysis of two samples. The session uses the same data and follows the steps in Seurat tutorial for integrated analysis: https://satijalab.org/seurat/immune_alignment.html

We begin with two expression matrixes: one for control PBMC cells, and another for PBMC cells stimulated with interferon beta. So now the cells will cluster based on cell type, but also based on the treatment, which makes the analysis a bit more complex.

1. Open example session

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

2. Setup Seurat object & quality control

Select the **immune_control_expression_matrix.txt.gz.** Select tool **Single cell RNA-seq / Seurat -Setup and QC.** Check the parameters, and **name your sample** as **CTRL**. You can name the project and give a bit stricter parameters for filtering (5 and 500 for example). Make sure that you have **assigned the file correctly**: this is an expression matrix (a **DGE table**). **Run** the tool.
Repeat this step for the **immune_stimulated_expression_matrix.txt.gz**, except now **name the sample** as **STIM**.

3. Filtering, regression and detection of variable gene

Select **seurat_obj.Robj**. Select the tool **Single cell RNA-seq / Seurat - Filtering, regression and detection of variable genes.** Adjust the parameters so that you are doing as loose filtering as possible –the QC plots from the previous tool can help you with this. (So, for example, "filter out cells that have at least 1 gene" and so on). Repeat for the other sample as well.
Once the tool is done, open the **Dispersion.pdf** and check also the second page.
How many variable genes are there?

4. Combine two samples and perform CCA

Select **both seurat_obj.Robj**ects from the previous step and run the tool **Single cell RNA-seq / Seurat – Combine two samples and perform CCA.**
Open **CCAplot.pdf** in **external browser**. Look at the plots.
Is there a correlation between the samples? Is there a shift between the samples? How many CCs should we continue the analysis with (check the elbow in the correlation strength plot)?

5. Integrated analysis of two samples

Select the combined **seurat_obj.Robj** from the previous step. Select tool **Single cell RNA-seq / Seurat – Integrated analysis of two samples**. In the parameters, set **number of CCs to use** to **20**. Run the tool.
While waiting, you can study the manual (click the **More help** button).
What are the main steps of this tool?
When the results are ready, study the **integrated_plot.pdf** (open in external browser).
Do the samples seem aligned now? How many clusters are there in this data?

6. Find conserved cluster markers and DE genes in two samples

Select the **seurat_obj.Robj** from the previous step. Run **Find conserved cluster markers and DE genes in two samples** for a cluster of your interest. Inspect the tables generated by the tool.
Select **de-list.tsv** and tool **Utilities / Filter table by column value**, and filter the table by column **p_val_adj**, so that you get only genes whose adjusted p-value is **smaller-than 0.05**.
How many differentially expressed genes were there between the two samples in this cluster? Write down few interesting genes from the list for the visualization exercise 7.
Select **conserved_markers.tsv**, and run the tool **Utilities / Filter table by column value twice** to get the list of **p_val_adj <0.05** for columns **CTRL_p_val_adj** and **STIM_p_val_adj.** Select the resulting tables, and draw a **Venn diagram**. Select the intersect area. On the right, go to **Selected tab** and click **Create dataset from selected.**

How many conserved biomarkers were recognized for the cluster? Write down few interesting genes from the list for the next tool.

7. Visualize markers and differentially expressed genes

Choose **seurat_obj.Robj** generated in step 5. Select tool **Single cell RNA-seq / Seurat - Visualize genes with cell type specific responses in two samples**. Type the gene names to the parameter field (the ones you listed in previous step), use comma (,) as a separator. You can run the tool several times for different gene lists.

Open **split_dot_plot.pdf** in external browser.

Are the differentially expressed genes expressed differently also in other clusters? Are the conserved markers expressed in other clusters?