Theory and Practice in gene expression analysis

Teacher: Santiago Carmona
Assistant: Christian Halter
Dates:

- 14.11.2023
- 15.11.2023
- 17.11.2023

NOTE: the exercises in this part will use R. Please install R and RStudio, if you have not installed it already. If you need help with the installation, we can support you.

NOTE: For experienced R users, as well as R newcomers, there are some heavily recommended tips & tricks at the end of this word file. Giving a whole tutorial on R and RStudio is out-of-scope for this course. However, for R newcomers we added some links to beginner guides.

Goals

- **Using existing packages** which make your life easier and speed up analyses
- Creating a **reproducible** bioinformatics experiment
- **Cell type classification and annotation**
- **Combining datasets and correcting for batch effects**

Table of Contents

Introduction

Figure 1 shows an overview of the general single-cell RNA-seq (scRNA-seq) data analysis workflow. In this exercise, you will learn about **Seurat**, the most popular R package for scRNA-seq, and how you can use it for:

- Quality control (QC) & pre-processing
- Data exploration & visualization
- Downstream analysis (clustering, differential gene expression, etc.)

If you want to know more about Seurat, then there are great tutorials available on their website for a basic introduction as well as for more advanced topics. Seurat can also be used for spatial transcriptomics and integrative multimodal analysis, e.g. combining scRNA-seq and scATAC-seq. If you prefer working in Python, there is the *Scanpy* package which reproduces many but not all functions of Seurat.

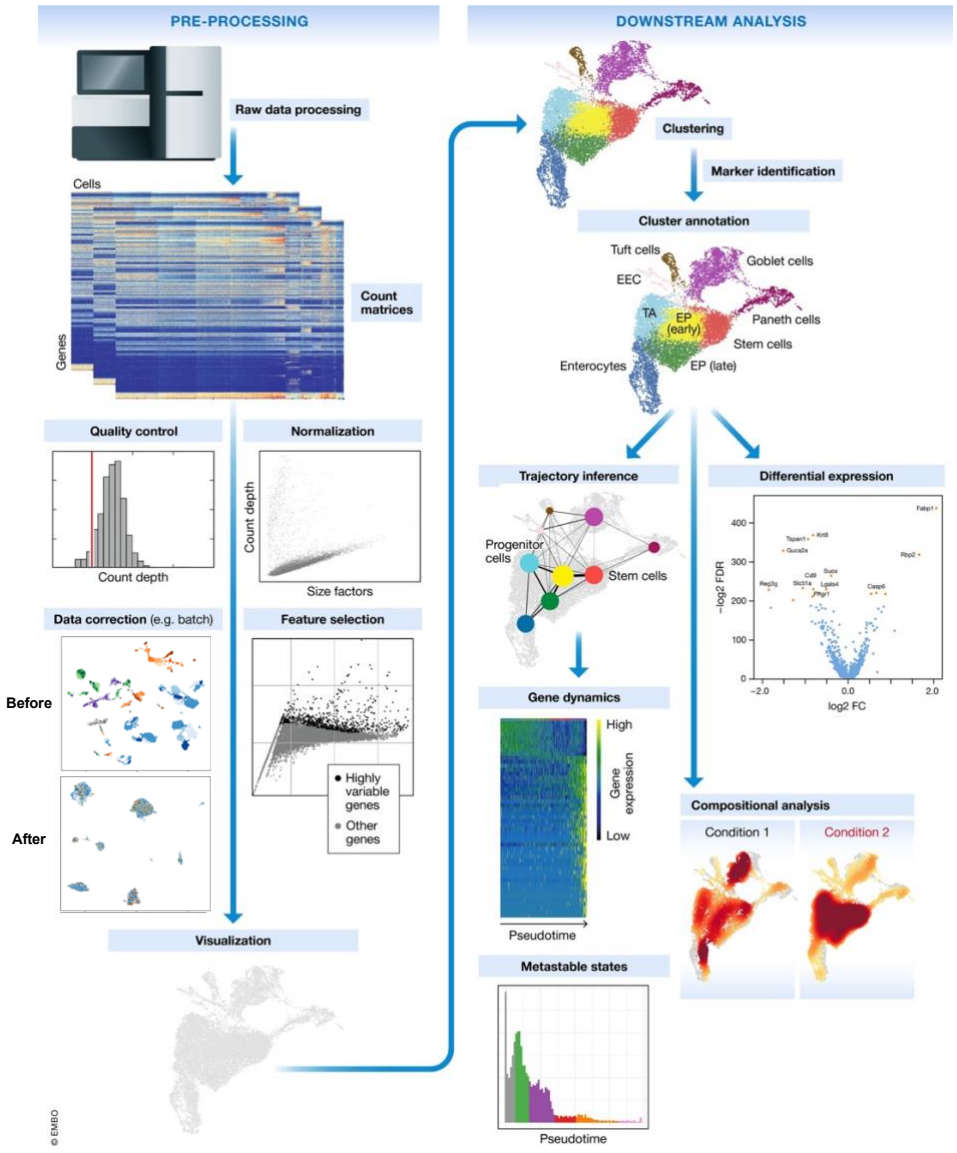


Figure 1: Overview of a standard scRNA-seq processing workflow. Modified from: Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019 Jun 19;15(6):e8746.

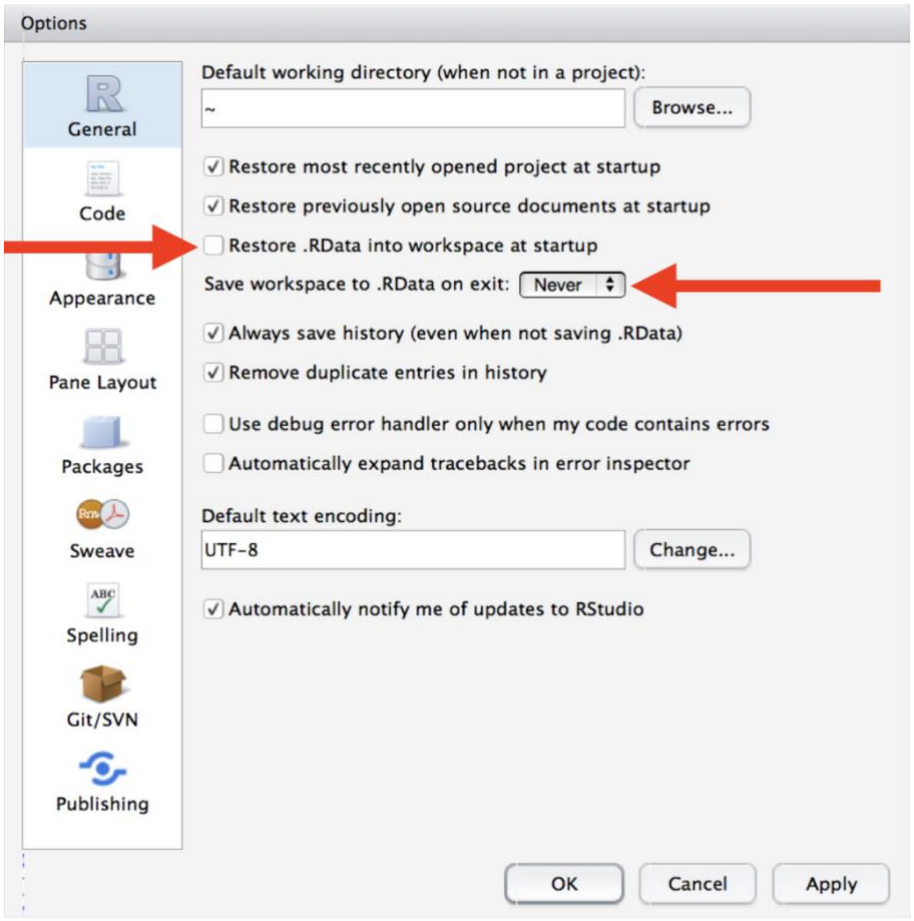
Links to R + RStudio beginner guides

https://www.datacamp.com/tutorial/r-studio-tutorial
https://moderndive.netlify.app/1-getting-started.html
https://education.rstudio.com/learn/beginner/
https://r4ds.had.co.nz/index.html

RStudio basic tips & tricks

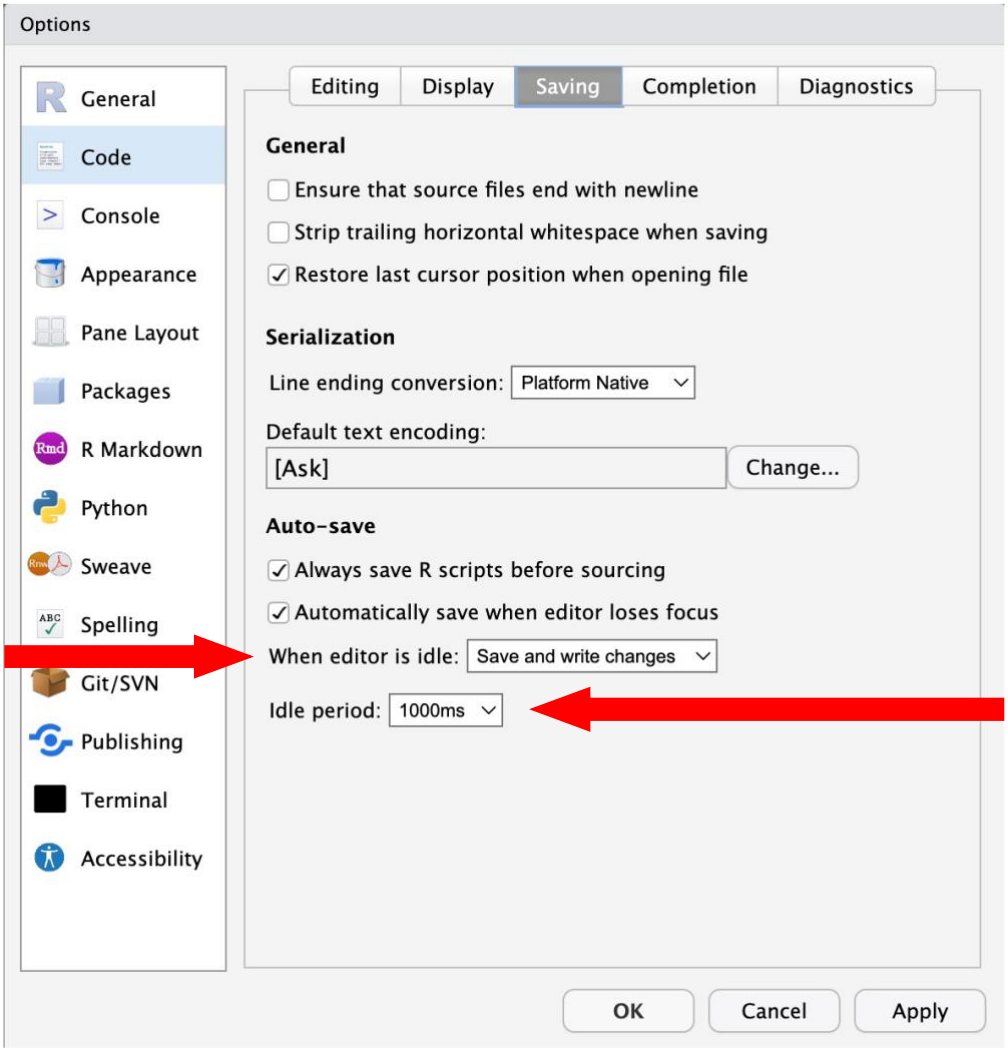
Turn of .RData

By default, RStudio has an undiserable functionality: it creates an .RData file, where it stores the workspace, such as variables you created during your work on an R script or project. When you close and re-open RStudio it will load this workspace from the .RData. However, **it can be very misleading to re-load some variable, rather than re-creating it from your script**, for example if you manually assigned a variable or it was created before you made some change to the source code and did not re-run the code to create the newly updated variable. To turn this behaviour off, go to:
Tools -> Global options ... -> and set the settings like in the figure below:



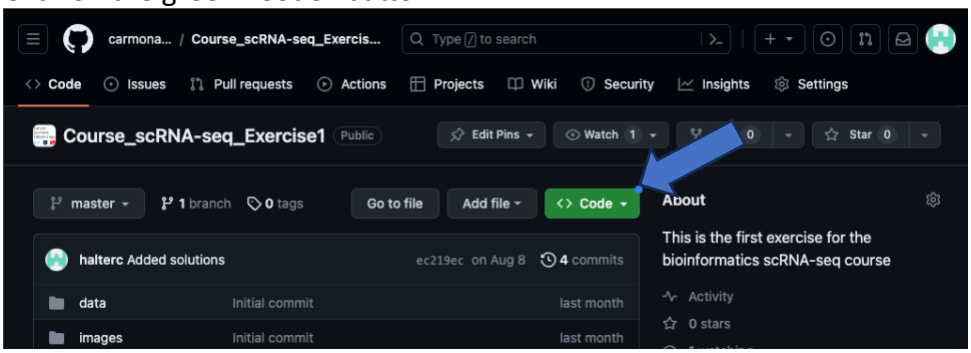
Auto-save

Another very useful function to prevent data loss in case R hangs itself up (e.g. due to out-of-memory or some other freezing error) is located here:



Exercise 1

- Clone the GitHub repository for exercise 1 to your computer
  - Open this link: https://github.com/carmonalab/Course_scRNA-seq_Exercise1
  - Click on the green “Code” button:



  - Select *Download ZIP*
  - Open the downloaded ZIP file containing the exercise repo and extract it anywhere you like
  - Open the folder and open the file “1_Course_scRNA-seq_Exercise1.Rproj”
  - If there is no script file open in RStudio yet, navigate to “Files” and open the R markdown file “1.1_Rscript.Rmd” containing the script for this exercise: