# Reproducibility in NGS data analysis

# Practical session : Literate programming

This session will be divided into three parts:

* General observations on Markdown
* RMarkdown / RStudio
* Jupyter

Table of Contents

Reproducibility in NGS data analysis 1

Practical session : Literate programming 1

A. General information 2

B. Introduction to Markdown 3

1. What is Markdown? 3

2. Where is Markdown ? 3

C. RMarkdown / RStudio 4

1. Quick introduction to RStudio 4

2. RMarkdown (DEseq2 analysis) 4

D. Jupyter 5

1. Quick introduction to Jupyter 5

2. Jupyter with Python 5

3. Jupyter with R engine 6

Version : 0.1 (18/08/2020)

Contacts :

* Thomas Denecker
* Celine Hernandez [celine.hernandez@i2bc.paris-saclay.fr](mailto:celine.hernandez@i2bc.paris-saclay.fr)
* Claire Toffano-Nioche

## General information

Websites

Resources

RMarkdown

Jupyter

Note that the commands you have to enter are inside a blue box. The first ‘$’ character should **not** be copied.

## Introduction to Markdown

### What is Markdown?

Markdown is a lightweight markup language widely used on the Internet to specify formatting in a user-readable manner.

Titles

Emphase

Lists

Image

Link

Citation

Tasks

Table

### Where is Markdown ?

Everywhere !

We will have a look at two renown websites : Wikipedia and Github

Wikipedia

Go to the Wikipedia documentation

GitHub

Go to the GitHub documentation

Q: How do you specify an italic formatting?

Q: How do you write a table?

The Markdown specification had ambiguities. By consequence, its implementation can differ from one tool to another.

## RMarkdown / RStudio

### Quick introduction to RStudio

* Local versus server
* And what about Docker?

**This is not an exercise and is provided for your information. You skip this part and come back to it later.**

Docker can be used to install and run RStudio on your computer. You can find a very nice tutorial here:

<https://ropenscilabs.github.io/r-docker-tutorial/02-Launching-Docker.html>

This tutorial describes how to:

Launch RStudio inside of a Docker container

Connect to the interface

Link a volume to a Docker container

Note that if you want to install new packages you might need to build a new images built on top of this one. To do so, you will need to write a Dockerfile.

<https://community.rstudio.com/t/rocker-rstudio-container-update/23106>

* General organization of the interface

RMarkdown and knitR libraries

You don’t need RStudio.

### RMarkdown (DEseq2 analysis)

**Goal: Transform an R script into an RMarkdown file.**

* Execute the provided R script
* Create a new Rmd notebook

RMarkdown files vs R Notebooks

(look at structure)

Header

Markdown

Chunk

* Knit the Rmd file
* Adapt the Rmd file

Replace code in chunk by content of the R script. Do not modify the header.

Congratulations!

You just successfully created your first reproducible R analysis !

* Improve the Rmd file

Create separate chunks. Name the chunks.

Chunk options

Move the documentation into the text part. Use different Title levels.

Insert a value inside the text.

Text can be generated on-the-fly using inline variables.

Change the size of the plot.

What is missing in this file for the analysis to be completely reproducible?

Do you know how to install a specific library version?

* Other chunk options

Cache (be careful!)

Shiny elements

* Knit from the command line.

Use different engines, for instance Bash. Can variables be accessed from one engine to another?

How would you do ?

## Jupyter

### Quick introduction to Jupyter

* Jupyter and JupyterLab
* And what about Docker?
* Jupyter engines

Engines cannot be mixed.

Specialized blocks (no code inside a Markdown block).

### Jupyter with Python

Open a provided file

Interactive elements

### Jupyter with R engine

Popularity of Jupyter. Integration with GitHub.

Momix repository

$ docker run genomicpariscentre/samtools samtools

Program: samtools (Tools for alignments in the SAM format)

Version: 0.1.19-96b5f2294a

This is the end of the practical session. We hope you enjoyed it.

Don’t hesitate to ask any questions and feel free to contact us any time after the session (email addresses on the front page).