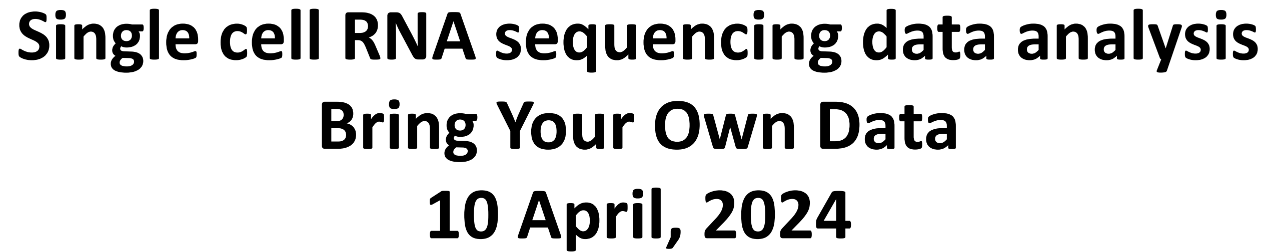


Åsa Björklund, Jennifer Fransson & Susanne Reinsbach

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Today

9:00 - Short introduction

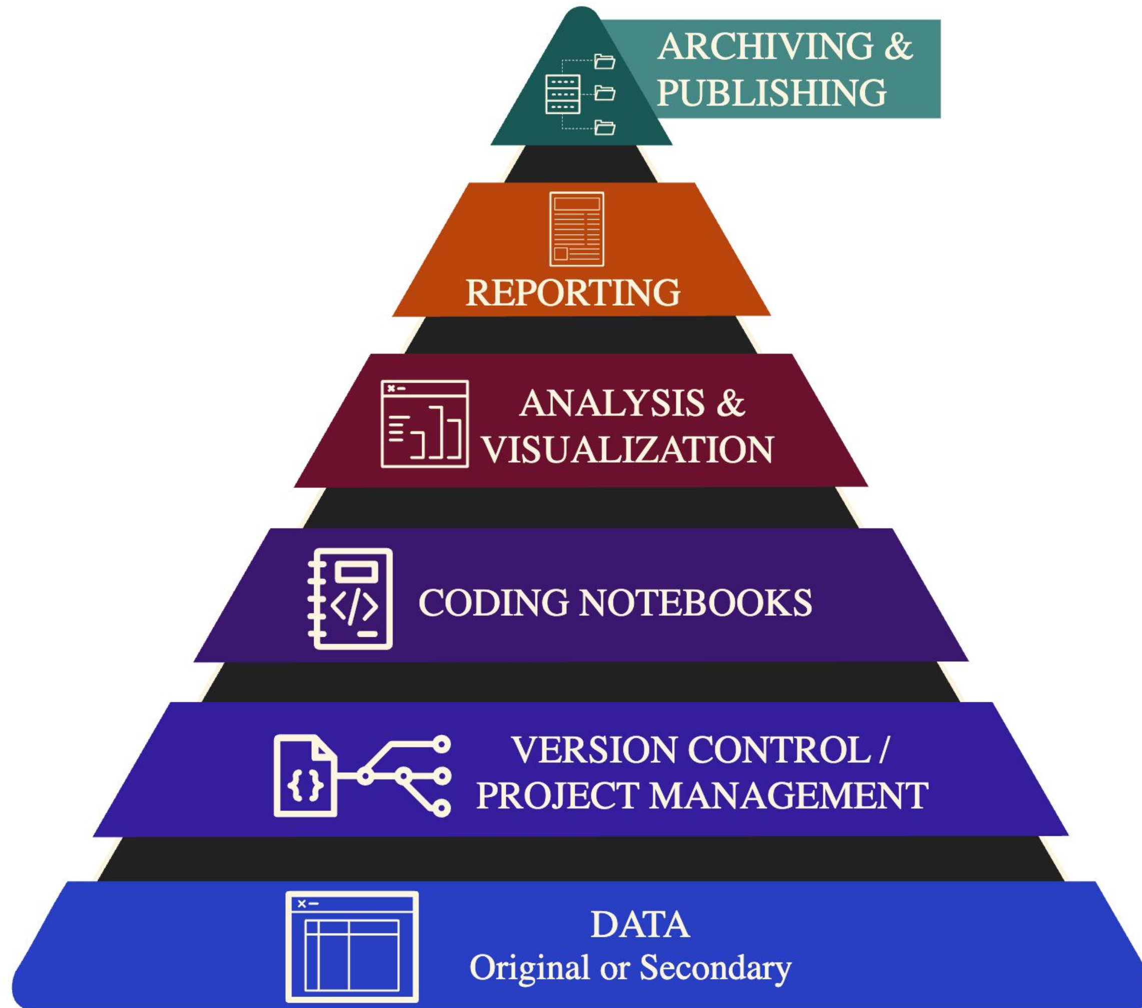
9:15 - Data managers present

You work on your own and we will be around to help.

14:30 - Presentations of all your projects.

16:00 - Summary of the day.

Reproducible research



(Little and Lafferty-Hess, 2023)

NBIS
NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

SciLifeLab

Reproducible code

- You should always be able to find and recreate the results.
  - Scripts should be able to run from input files to create the output.
  - Never work with saved R sessions!
  - Divide tasks in reasonable chunks and use separate scripts.
- Name your scripts with relevant names so you can find them 2 years later 😊
- Always backup code – good idea to use **github** that also gives you version control.

Reproducible software installations

Example:

1. I analyse dataset A with Seurat v4, have nice results, submit them to a journal.
2. I start analysing dataset B, I want to use Seurat v5 and update to a new version of R and all other packages.
3. Reviews in project A comes back. I need to replot some of the figures. Now I cannot reproduce the same figures since I have another version of R/Seurat.

- conda environments / renv is often good enough
- Containers is the optimal way, but also more work.

https://nbisweden.github.io/workshop-reproducible-research/

Memory issues

- scRNAseq datasets are often large, think about how you code. Avoid duplicating objects!
- Remove unused matrices and clear memory with gc() (in R) or gc.collect() (in python).
- Try to keep your matrices **sparse**!

- In Seurat – can use DietSeurat() function to remove assays, data slots etc.

Duplicating objects

```
counts = read10x…
sobj =
CreateSeuratObject(counts)
..
sobj_filt =
sobj[,keep.cells]
sobj_raw = <after analysis
pipeline>
sobj_int = <after
integration>
sobj_int2 = ..
```

6 copies of the count matrix. Many copies of data/scale data.

```
counts = read10x…
sobj =
CreateSeuratObject(counts)
rm(counts)
gc()
..
sobj = sobj[,keep.cells]
sobj= <after analysis
pipeline>
sobj@assay$INT = <after
integration>
sobj_int2@assay$INT2 = ..
```

Often better to save intermediate files than to have all in memory.

Memory issues

- If you still have issues with memory in R, test setting e.g. R_MAX_VSIZE=70Gb in the .Renviron file. Default is 16Gb. (check FAQ section)
- With docker - allocate enough resources to the container (https://nbisweden.github.io/workshop-scRNAseq/other/docker.html)

Breakout rooms

- We will assign you to breakout rooms
- Start with an introduction of yourself and your project.

- If you have a question or something you want to discuss with TAs, please write in the **#exercises** channel on slack.
- We will take questions as they come in order on slack.

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Presentation

We want a **very** short - max 5 minutes presentation of what you have done.

- 1 slide
- What dataset
- Which analysis steps did you perform?
- Which were your main results?
- What were the biggest problems you encountered?

We will send out a link for the presentation on slack.

QUESTIONS?