

Reproducibility

Learning outcomes

After having completed this chapter you will be able to:

- Understand the importance of reproducibility
- Apply some basic rules to support reproducibility in computational research

Material



Download the presentation

Some good practices for reproducibility

During today and tomorrow we will work with a small *E. coli* dataset to practice quality control, alignment and alignment filtering. You can consider this as a small project. During the exercise you will be guided to adhere to the following basic principles for reproducibility:

1. **Execute the commands from a script** in order to be able to trace back your steps
2. **Number scripts** based on their order of execution (e.g. `01_download_reads.sh`)
3. Give your scripts a **descriptive and active name**, e.g. `06_build_bowtie_index.sh`
4. Make your scripts **specific**, i.e. do not combine many different commands in the same script
5. Refer to **directories and variables** at the beginning of the script

By adhering to these simple principles it will be relatively straightforward to re-do your analysis steps only based on the scripts, and will get you started to adhere to the **Ten Simple Rules for Reproducible Computational Research**.

By the end of day 2 `~/project` should look (something) like this:

```
.
├── alignment_output
├── reads
├── ref_genome
├── scripts
│   ├── 01_download_reads.sh
│   └── 02_run_fastqc.sh
```

```
| |— 03_trim_reads.sh
| |— 04_run_fastqc_trimmed.sh
| |— 05_download_ecoli_reference.sh
| |— 06_build_bowtie_index.sh
| |— 07_align_reads.sh
| |— 08_compress_sort.sh
| |— 09_extract_unmapped.sh
| |— 10_extract_region.sh
| |— 11_align_sort.sh
|— trimmed_data
```