

Starting the project

UE Reprohackthon

Frédéric Lemoine / Thomas Cokelaer
Institut Pasteur

2025/10/10

Part 1

Introduction

1.1 Goals

Use case

- We propose to work on an RNA-Seq data analysis use case
- Goals:
 - Reproduce parts of the analysis
 - Using
 - A workflow management system (Nextflow or Snakemake)
 - Containers (Docker or Apptainer)
 - Git

1.2 RNA-Seq

Definition

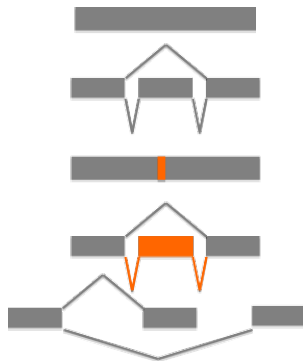


It is qualitative + quantitative

1.2 RNA-Seq

Applications

- Measuring gene expression
- Measuring alternative splicing
- Detecting expressed mutations
- Annotating genes (new exons)
- Detecting fusion transcripts



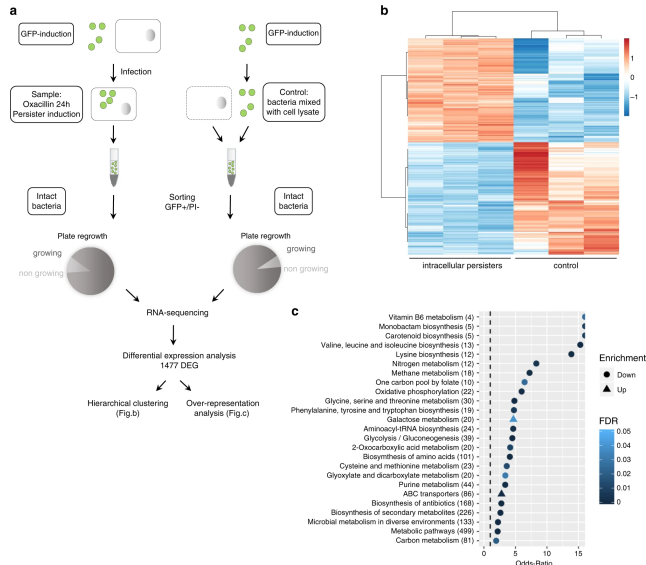
1.3 RNA-Seq

Use case

- Peyrusson et. al, Nature Communications, 2020 : Study of *Staphylococcus aureus* persisters (\neq resistance) transcriptomic profile (stress response to antibiotics). Intracellular *S. aureus* persisters may constitute a reservoir for relapsing infections and so therapeutic failures.
- The paper provides evidence of the presence of *S. aureus* persisters after antibiotic exposure, and monitor bacterial division at the single cell level.
- This tolerance may be a result of growth arrest, or an active process. What are the factors leading to antibiotic persistence and tolerance?
 - RNA-Seq : study changes in transcriptomic profile

1.3 RNA-Seq

Use case: experimental design



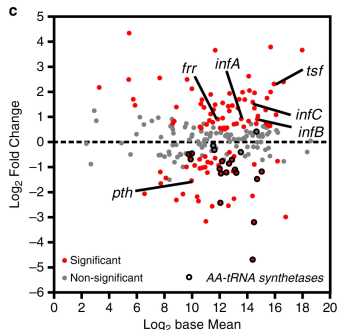
1.3 RNA-Seq

Use case: Results to reproduce

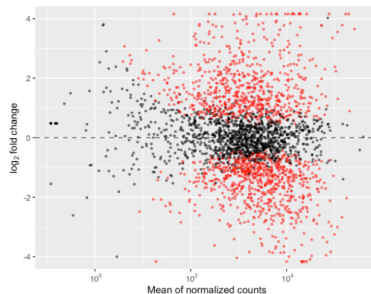
Goals:

- Reanalyze same dataset (available publicly): Find DEG
- Generate the figures

Fig. 3.c.: MA-plot of genes related to translation

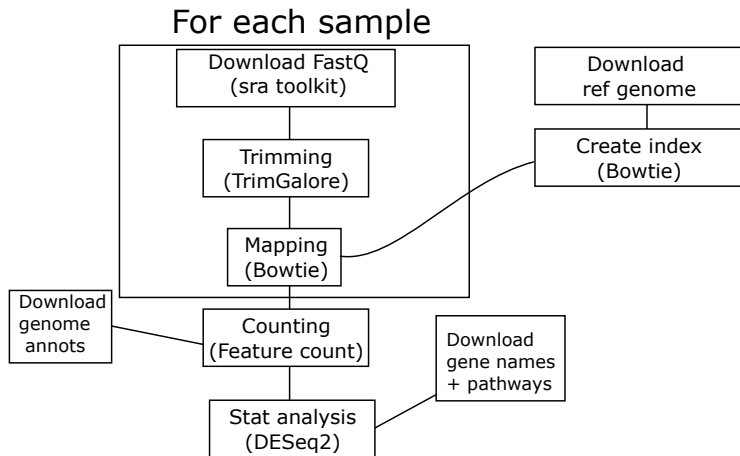


Supp. Fig. 3.: MA-plot of complete RNA-seq dataset



1.3 RNA-Seq

How? Workflow



1.3 RNA-Seq

How? Containers

Apptainer/Docker images to create (write recipes):

- Download FastQ: SRA Toolkit (Fastq-dump, Fasterq-dump)
- Mapping: Bowtie
- Reformating alignments: Samtools
- Counting reads: FeatureCount
- Stat analysis: R/DESeq2

You will have to write your own Docker/Apptainer image recipes!

1.3 RNA-Seq

How?

- Developed by groups of 4 students: Git + Vscode locally
- Executed on the Cloud
- Make the groups
- Choosing Nextflow/Snakemake

1.3 RNA-Seq

Deliverables

- Workflow code (on a git repository):
 - Readable
 - Tested
 - Commented
 - Documented
- Container recipes (on the same git repository)
- README.md + run.sh (on the same git repository) with all instructions for us to easily reproduce your analysis
- Report with:
 - Introduction (Reproducibility, biological topic of the papers to reproduce, etc.)
 - Material and Methods (Tools used, and setup)
 - Results (Workflow developed, results obtained after execution)
 - Conclusion / perspectives (Interpretation of the results, and conclusion about reproducibility)
- Dec. 12th: Final oral presentation
- Nov. 7th: First evaluation / progress report (see next slide)

1.3 RNA-Seq

Mid-Project deliverables

November 7th:

- 10 minutes oral presentation
- First version of the code (in the git repository) with:
 - All the image recipes (working) and
 - First steps of the workflow (data download, genome download and indexing) (working)

