STARTING THE PROJECT

UE REPROHACKTHON

Frédéric Lemoine / Thomas Cokelaer
Institut Pasteur

2021/10/01

Institut Pasteur

PART 1

Introduction

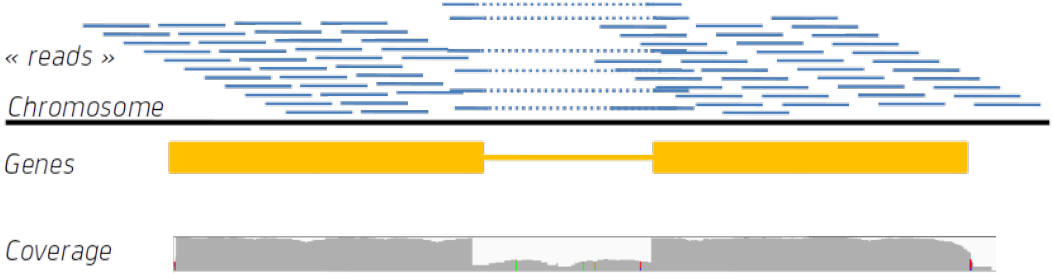
1.1 Goals

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 Singularity)
    - Git

1.2 RNA-Seq

Definition

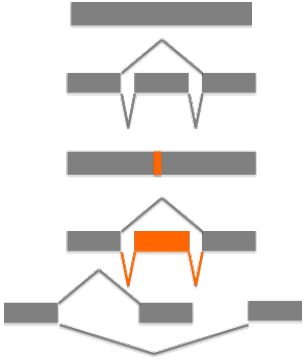


It is qualitative + quantitative

1.2 RNA-Seq

Applications

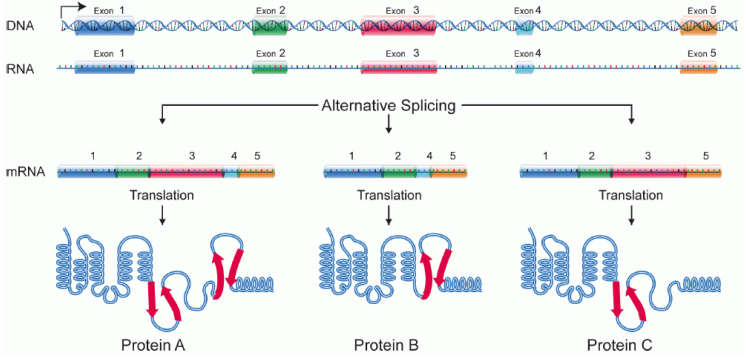
- Measuring gene expression
- Measuring alternative splicing
- Detecting expressed mutations
- Annotating genes (new exons)
- Detecing fusion transcripts



1.2 RNA-Seq

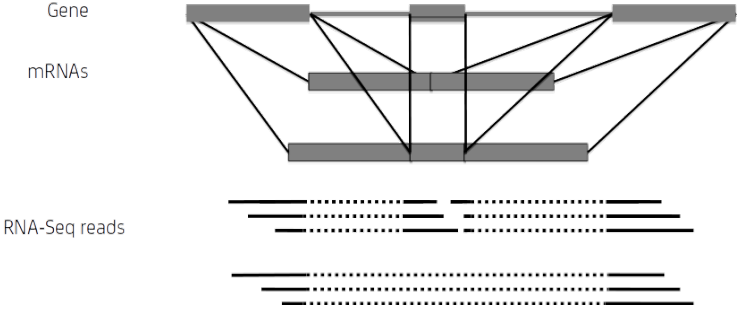
Alternative splicing

Wikipedia: ”Alternative splicing, or differential splicing, is a regulated process during gene expression that results in a single gene coding for multiple proteins. In this process, particular exons of a gene may be included within or excluded from the final, processed messenger RNA (mRNA) produced from that gene.”



Alternative splicing and RNA-Seq

Different mRNAs from the same gene



1.3 RNA-Seq

Use case

- Harbour et al. (Nat. Genet. 2013) sequenced RNAs from uveal melanoma patients with mutated SF3B1 gene or not. Though SF3B1 is a splicing factor, they did not find any splicing difference between patients;
- Furney et al. (Cancer Discov. 2013) reanalyzed the same dataset and found differential splicing between the two groups

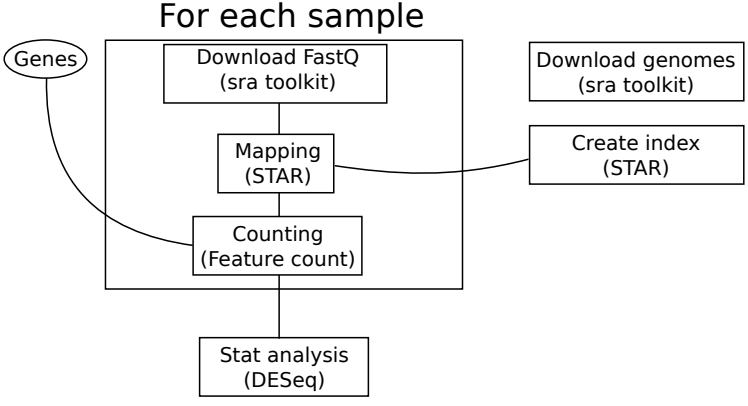
1.3 RNA-Seq

Goals of the project

- Reanalyzing the same dataset and try to find differentially **expressed genes**
- Datasets are publically available on SRA

1.3 RNA-Seq

How? Workflow



1.3 RNA-Seq

How? Containers

- Download FastQ: SRA Toolkit: `evolbioinfo/sratoolkit:v2.5.7`
- Mapping: STAR: `evolbioinfo/star:v2.7.6a`
- Reformating alignments: Samtools: `evolbioinfo/samtools:v1.11`
- FeatureCount: `evolbioinfo/subread:v2.0.1`
- Stat anlaysis: R/DESeq: `evolbioinfo/deseq2:v1.28.1`

If tools are missing: Ask us or create your own containers!

1.3 RNA-Seq

How?

- Developed by groups of 3/4 students: Git + Vscode locally
- Executed on the Cloud
- Make the groups
- Choosing Nextflow/Snakemake

1.3 RNA-Seq

Evaluation

- Code:
  - Readbility
  - Tested
  - Commented
  - Documented
- Execution: Can be executed/reproduced by us
- Report with:
  - Introduction
  - Methods
  - Findings
  - Conclusion
- Oral presentation

Institut Pasteur
25-28, rue du Docteur Roux
75724 Paris Cedex 15 - France
www.pasteur.fr