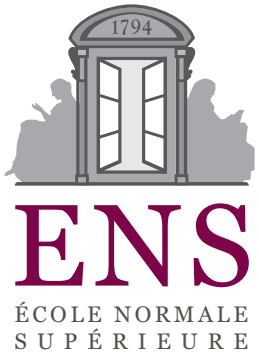


Single-cell RNA-seq analysis pipelines using the Eoulsan Workflow engine

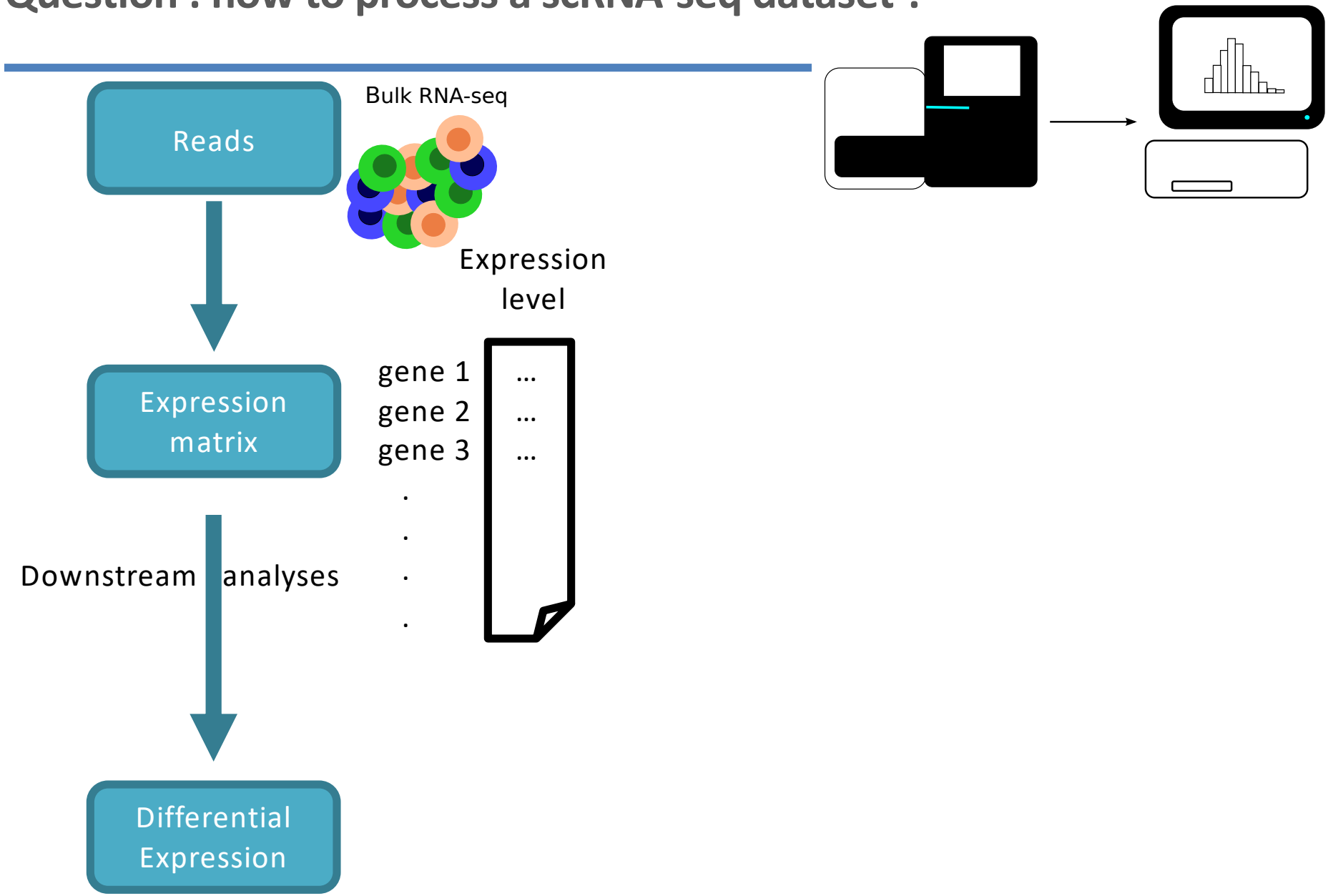
Laurent Jourdren, Nathalie Lehmann, Morgane Thomas-Chollier

Computational systems biology - IBENS

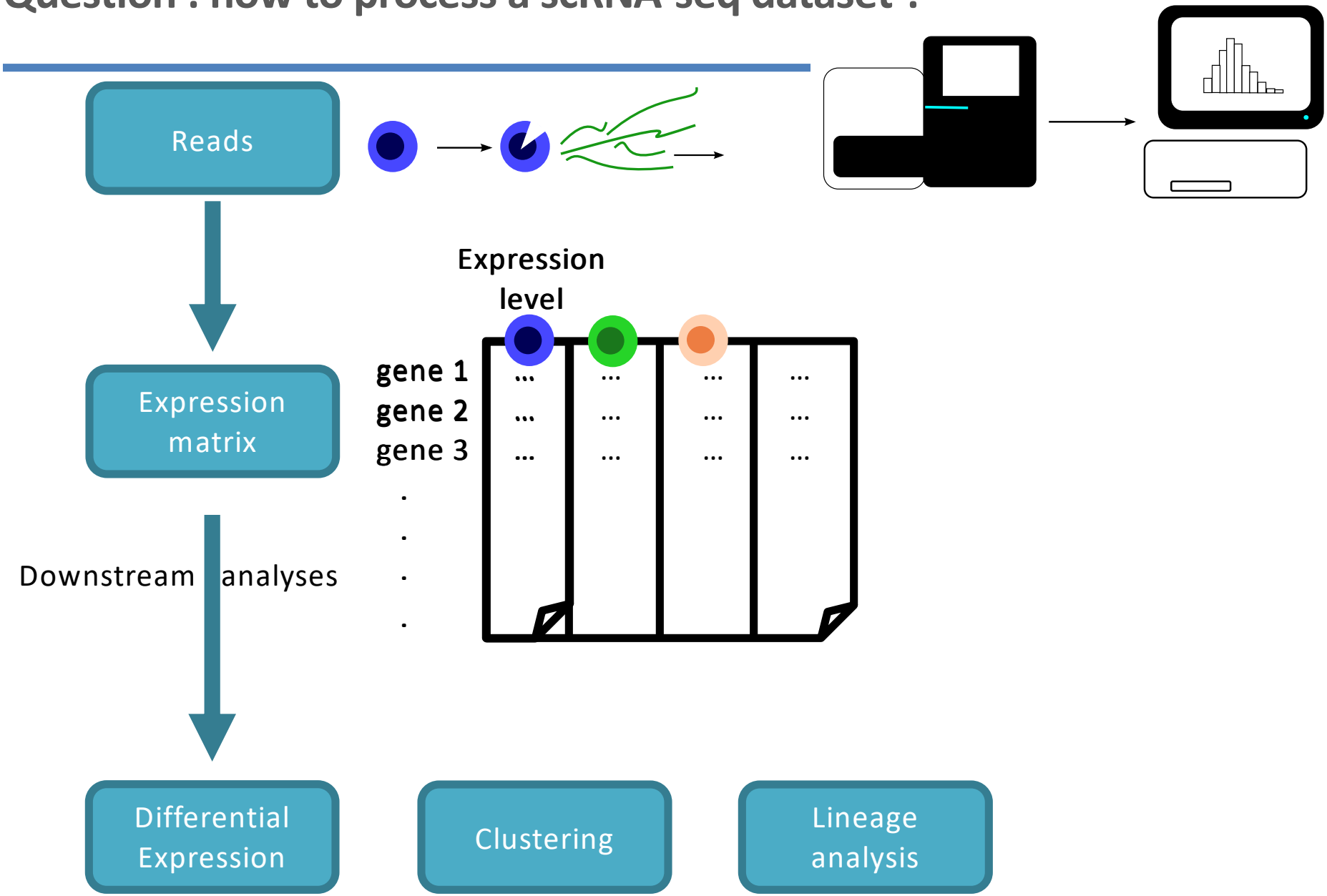
& Genomic Platform Paris Centre

(Stéphane Le Crom)

Question : how to process a scRNA-seq dataset ?



Question : how to process a scRNA-seq dataset ?



Eoulsan : a workflow engine designed for bulk RNA-seq

- Build on existing software

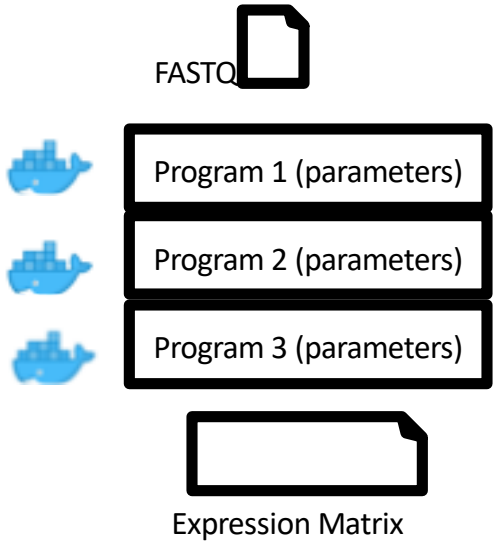
GENOMIC PARIS CENTRE

IBENS

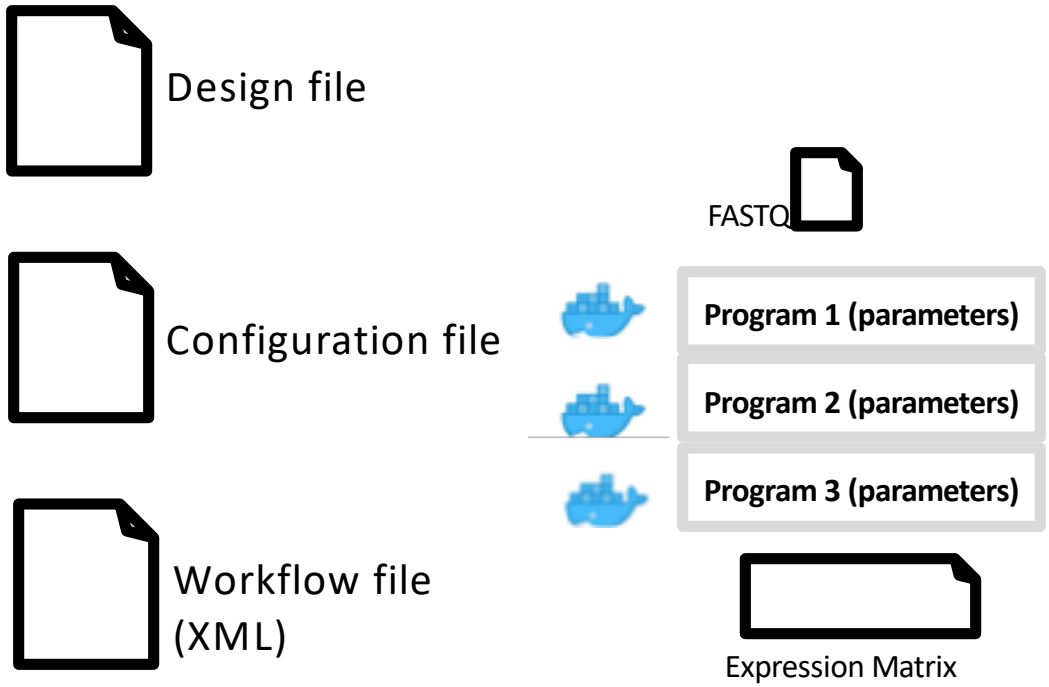
Eoulsan

Jourdren L, Bernard M, Dillies MA, **Le Crom S**. Bioinformatics. 2012

- Modularity and reproducibility (docker containers)



Users (bioinformatician)



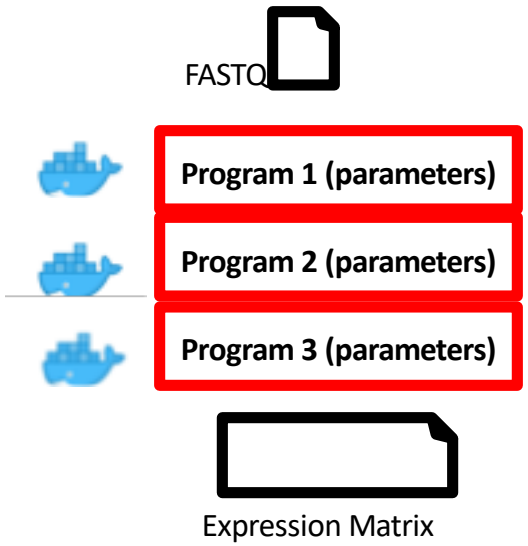
- Input/output complexity is hidden
- User just need to focus on the workflow design

Users (bioinformatician) | **Developper**

Design file

Configuration file

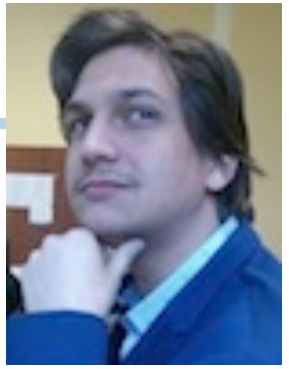
Workflow file (XML)



- Input/output complexity is hidden
- User just need to focus on the workflow design

- Input/output compatibility
- Ensures each program is encapsulated

Result : adaptation of Eoulsan for scRNA-seq



Geoffray Brelurut
M2 apprentice INSERM

- Identification and evaluation of state-of-the-art tools to integrate as modules to Eoulsan

Reads

FASTQ

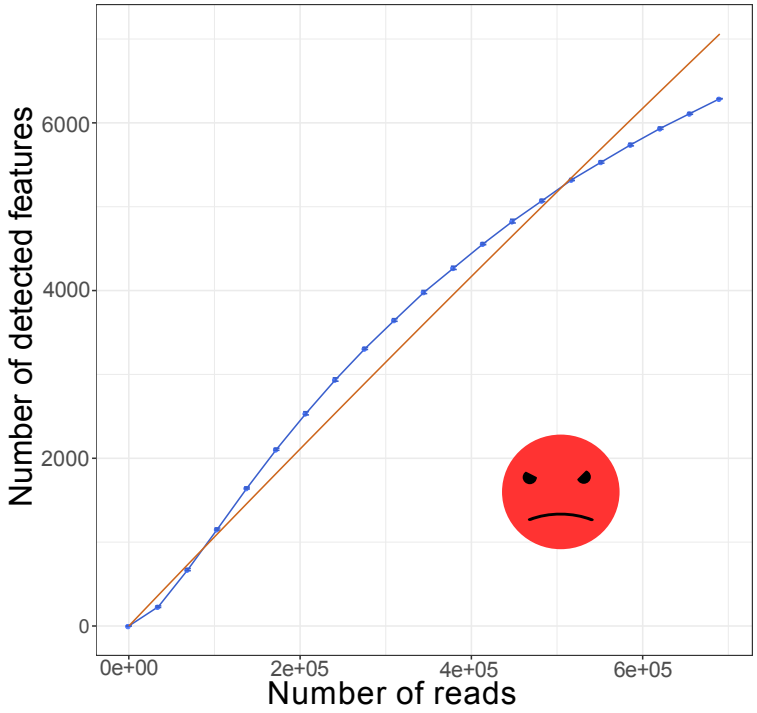
Read quality filtering	FASTQC
Alignment (mapping)	STAR
Alignment filtering	Picard, Samtools
Expression estimation	HT-Seq count
Expression merger	Original program
Cell quality filtering	Original program
Normalization across cells	Scran

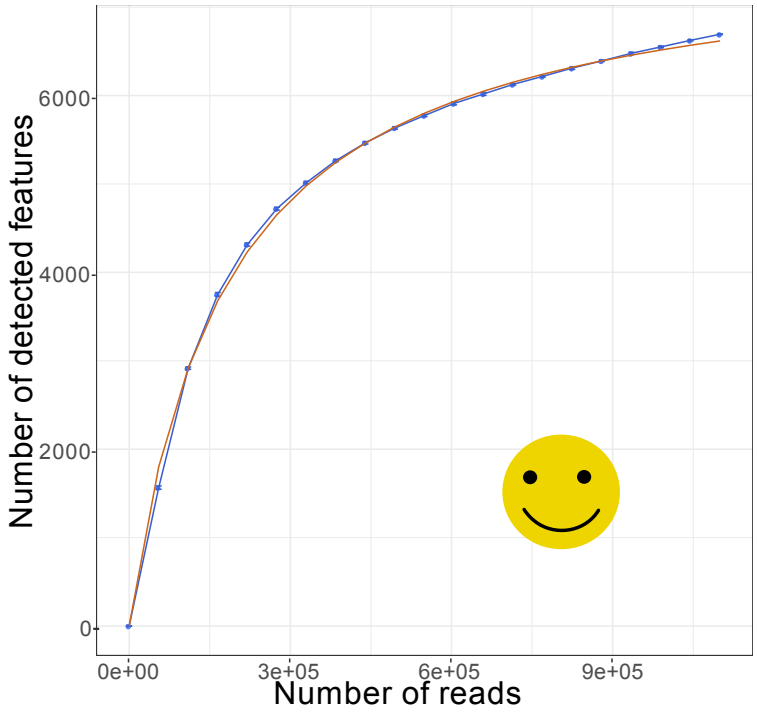
Expression matrix

TSV / SingleCellExperiment (R object) formats

Result: original development of a filter on cell quality

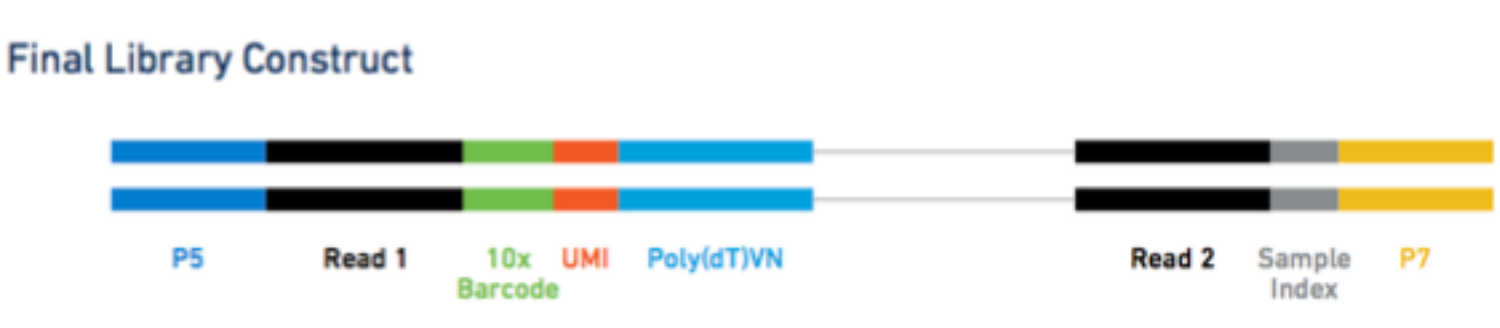
- Problem : detect and remove cells of low quality (=noise in data, artefactual cell variability)
  - Evaluate proportion of reads mapping to mitochondrial genes (<10%)
  - Threshold on mapped reads (>500,000/cell) and detected genes (>1500)
  - Filter on sequencing **saturation plots**





10x Genomics uses Unique Molecular Identifiers for 3' end transcriptomes

- 1 barcode per cell (16bp)
- 1 UMI per transcript (10bp random barcodes added to transcripts)

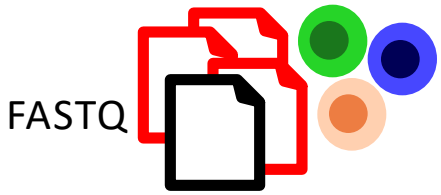


10x Genomics

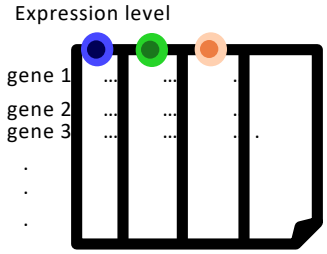
- In practice : don't get the full transcript (3' end only)

Aim : adapt Eoulsan scRNA-seq to support 10x GENOMICS

smart-seq2



Read quality filtering	FASTQC
Alignment (mapping)	STAR
Alignment filtering	Picard, Samtools
Expression estimation	HT-Seq count
Expression merger	Original program
Cell quality filtering	Original program
Normalization across cells	Scran



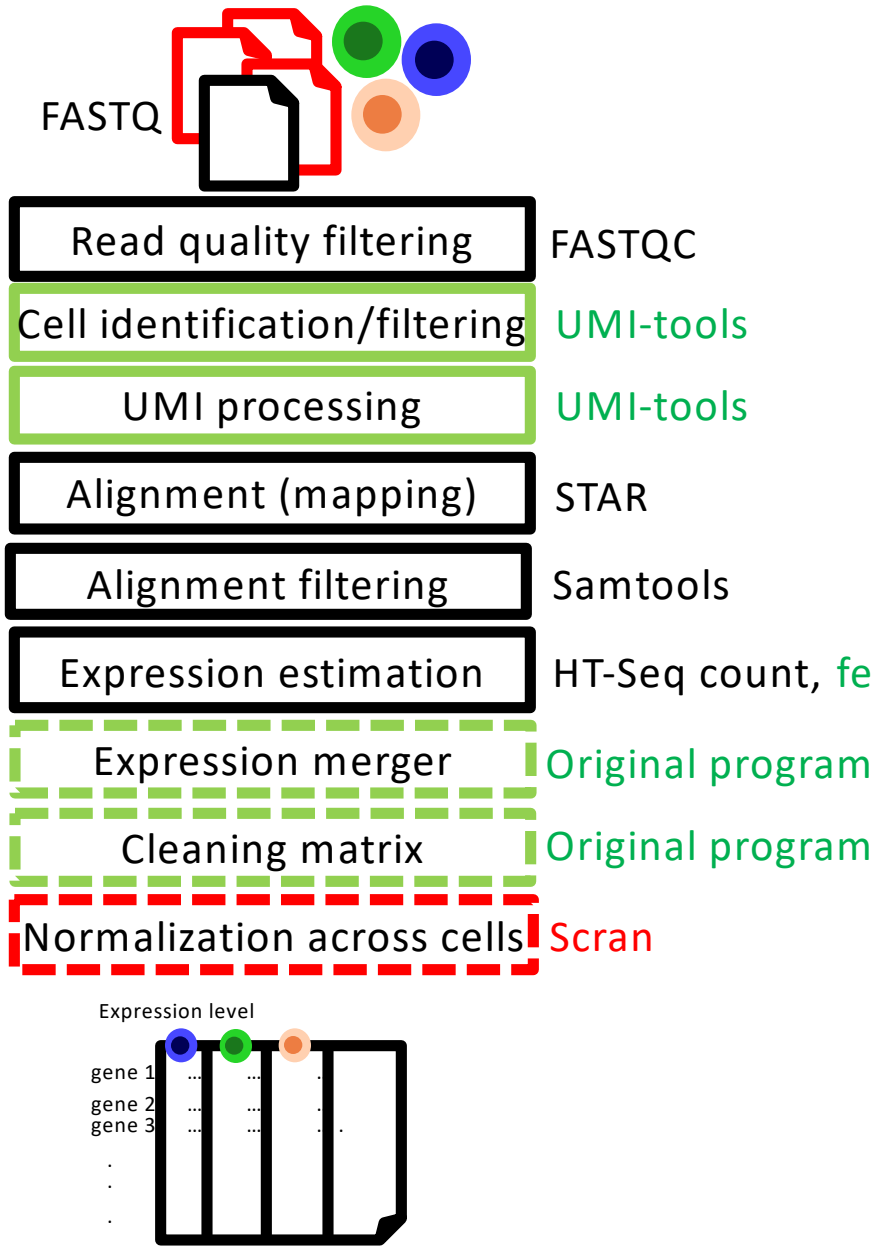
TSV / SingleCellExperiment (R object) formats

What should be changed ?

Barcode + UMI

Result : adapt Eoulsan scRNA-seq to support 10x GENOMICS

Nathalie Lehmann
Engineer CNRS

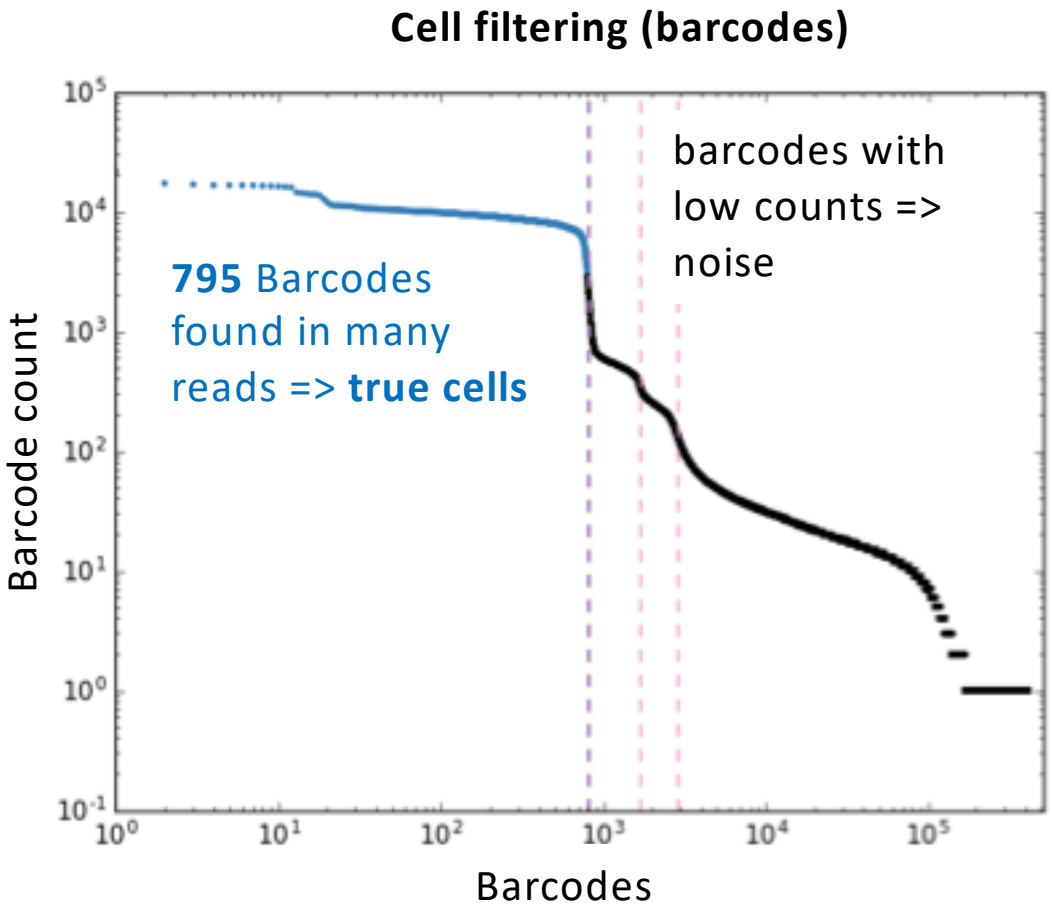


Application: Eoulsan scRNA-seq on 10x GENOMICS dataset

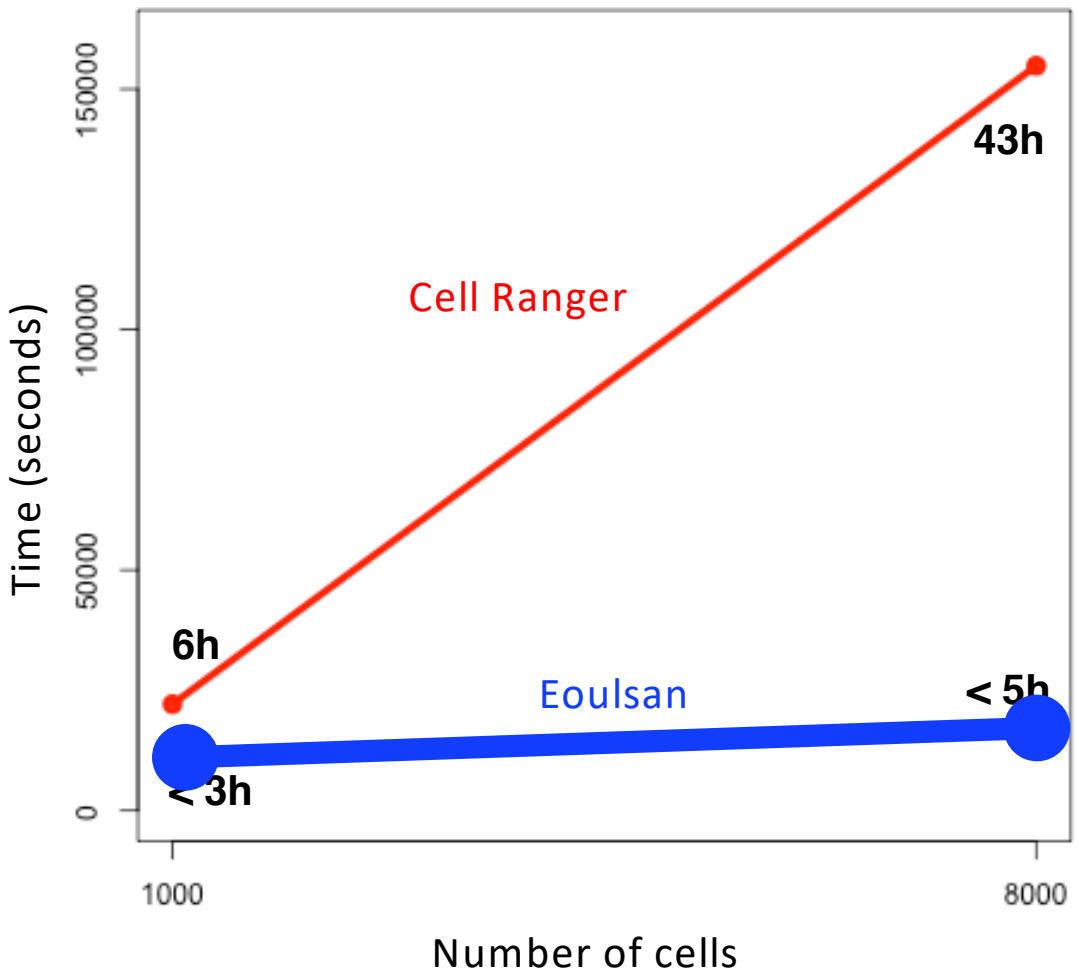
- Dataset from Vassili Soumelis (Institut Curie, Paris)
- Human immune cell

(plasmacytoid dendritic cell)

- 7 samples
- ~1.000 cells/sample

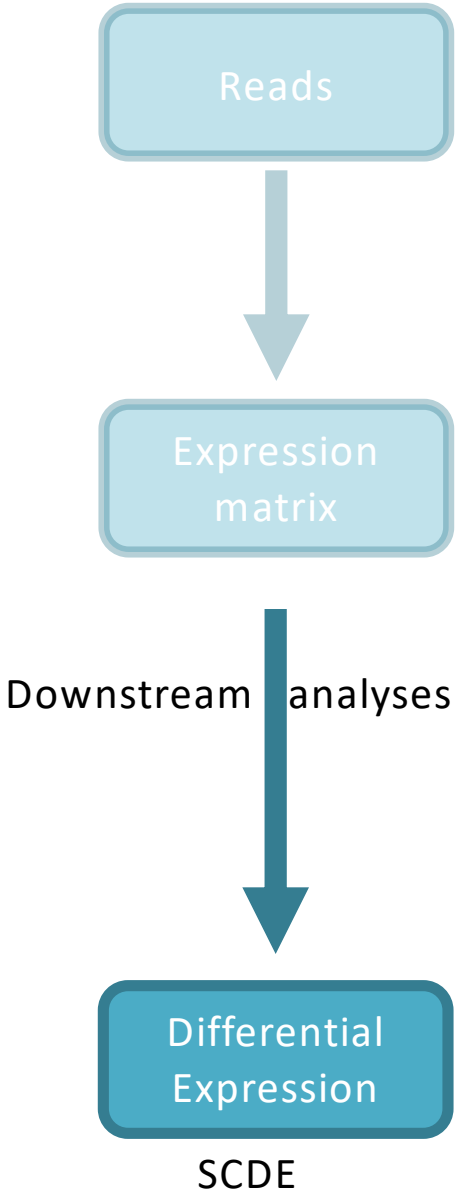


Eoulsan scRNA-seq is faster than Cell ranger



- Lower memory usage with Eoulsan

Eoulsan scRNA-seq also supports some downstream analyses



- Biological-driven analyses
- Still very active methodological developments
- Alternative options

Availability

- Eoulsan

https://github.com/GenomicParisCentre/eoulsan

- Install
- New user's tutorial (bulk, smart-seq, 10x)
- Complete documentation