SinCellTE 2022

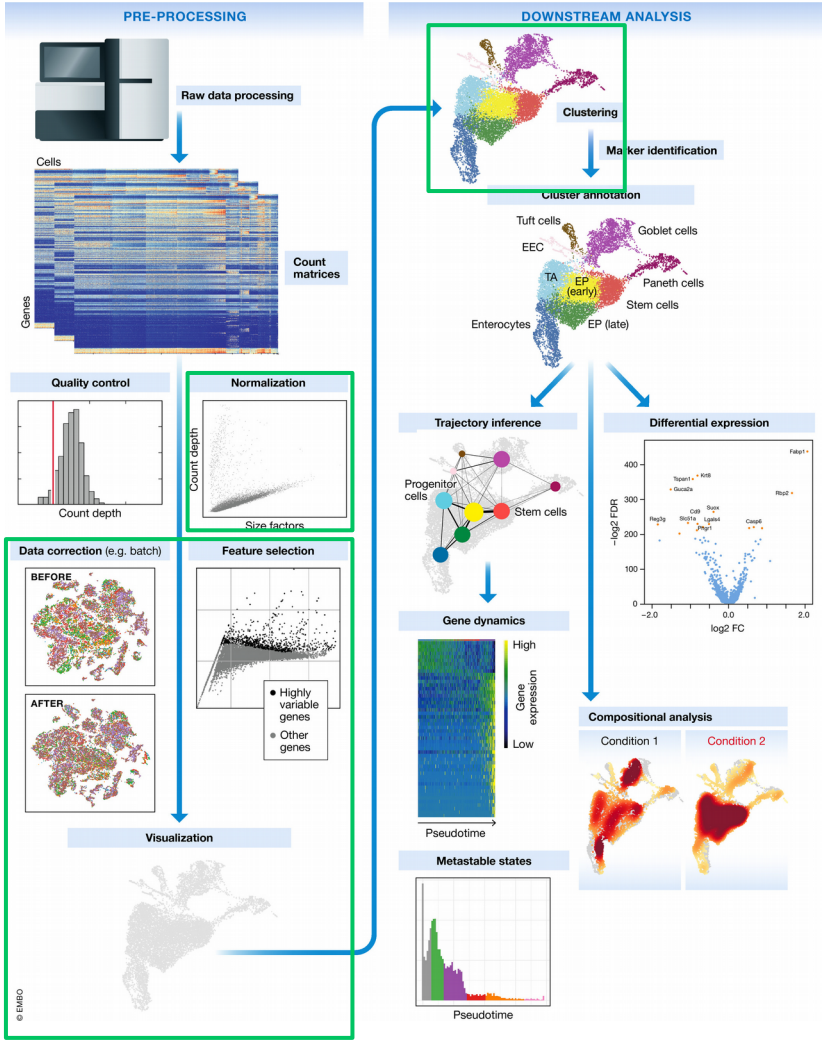
Practice : Statistical models and analyses

2022-01-11

Marine AGLAVE, Bioinformatics Core Facility, Gustave Roussy
Bastien JOB, Bioinformatics Core Facility, Gustave Roussy

Main steps of single cell data processing

- *this Practical Work*



Malte D Luecken & Fabian J Theis
Molecular Systems Biology (2019)

Prepare your work environment

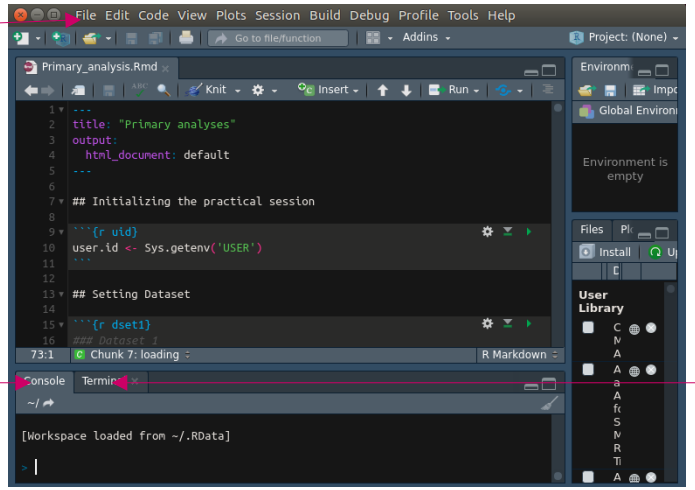
In a bash terminal :
1) Create your working folder
> mkdir -p /shared/projects/sincellte_2022/${USER}/Statistical_models_and_analyses/

2) Copy scripts
> cp -r /shared/projects/sincellte_2022/Courses/Statistical_models_and_analyses/scripts /shared/projects/sincellte_2022/${USER}/Statistical_models_and_analyses/scripts

3) Link datasets
> ln -s /shared/projects/sincellte_2022/Courses/Statistical_models_and_analyses/input /shared/projects/sincellte_2022/${USER}/Statistical_models_and_analyses/input

In the RstudioServer console :
1) Go to your working directory
> user.id <- Sys.getenv('USER')
> setwd(paste0("/shared/projects/sincellte_2022/", user.id, "/Statistical_models_and_analyses"))

2) Open script



Console

Terminal

Script

1) Setting Dataset, Parameters, Seed and Loading R packages
2) Loading data
3) Find variable genes + Normalization
    - a) SCTransform
    - b) LogNormalize
4) Check variable genes
5) Dimension reduction
    - a) pca
    - b) scbfa

yes — 6) Need bias correction ?

7) How to choose the right amount of reduction dimensions to keep ?
    - a) Strategy 1 : the elbow plot
    - b) Strategy 2 : algorithm method (Jackstraw)
    - c) Strategy 3 : Genes contribution
    - d) Strategy 4 : Evaluate and compare effect on clustering (using Louvain)

8) Final clustering at selected parameters

yes — 9) Checking bias on final clustering : need of a bias correction ?

10) Differential Gene Expression Analyses
11) Save

Dataset : description

Goal:
Identify the different cell types.

Data information :
Organism: Human
Type of tissue collected: Peripheral blood mononuclear cells.
Origin: patient.
Cells treatment: no treatment, healthy cells.
Technology : 10X Genomics Chromium

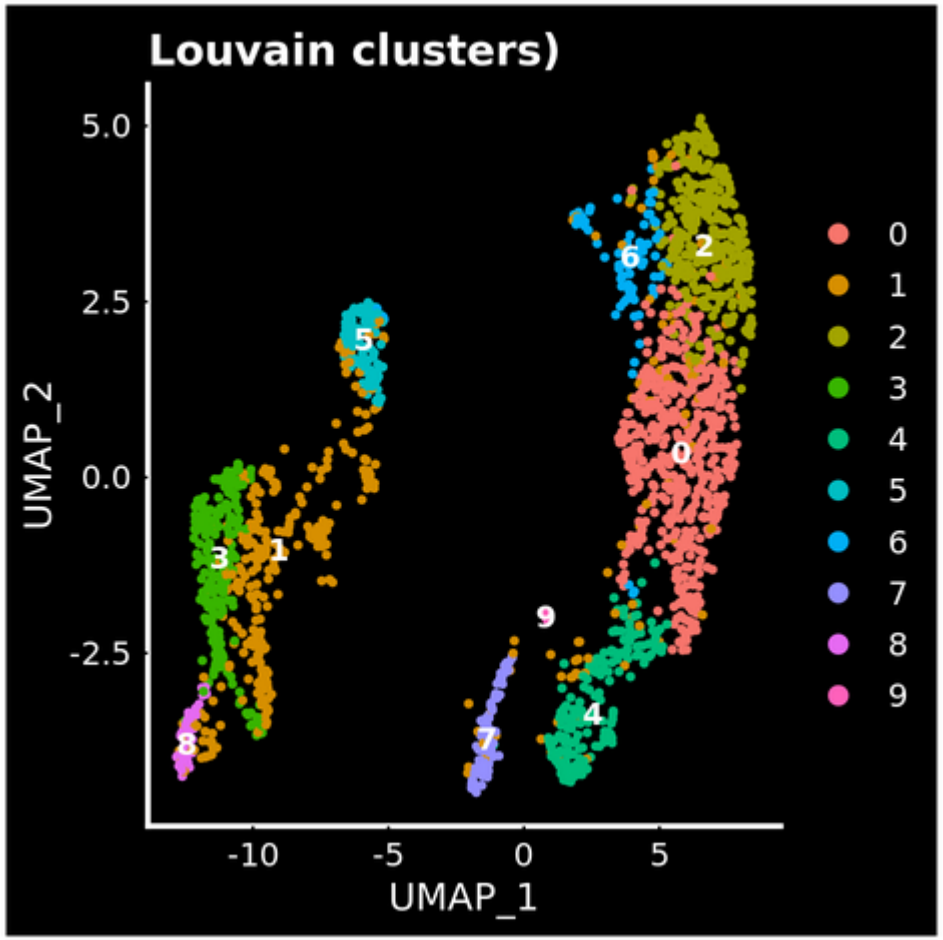
Expected Result:
We expect about ten cell types.

Input type:
Raw counts table from CellRanger.

Same as yesterday !

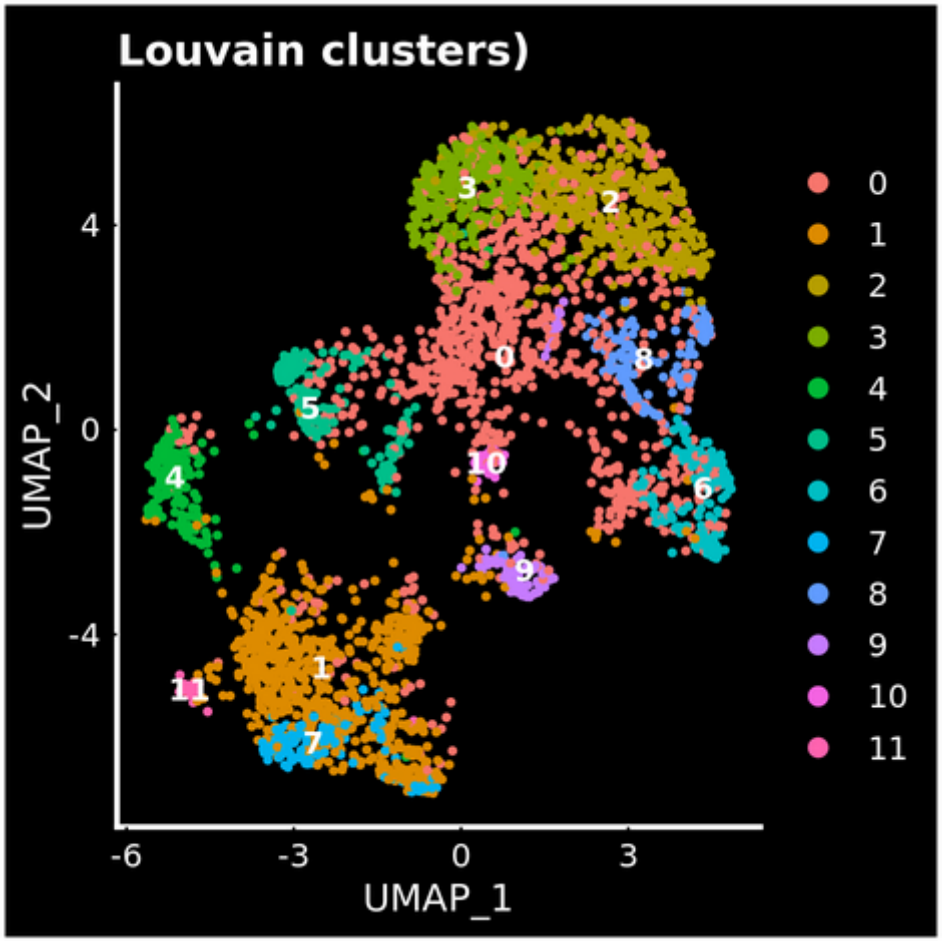
WITH or **WITHOUT**
the small population

Dataset : results (without small cells)



Cluster 0 : T cells
Cluster 1 : ?
Cluster 2 : T cells
Cluster 3 : Monocytes
Cluster 4 : NK
Cluster 5 : B cells
Cluster 6 : T cells ?
Cluster 7 : NK
Cluster 8 : Monocytes
Cluster 9 : Platelet

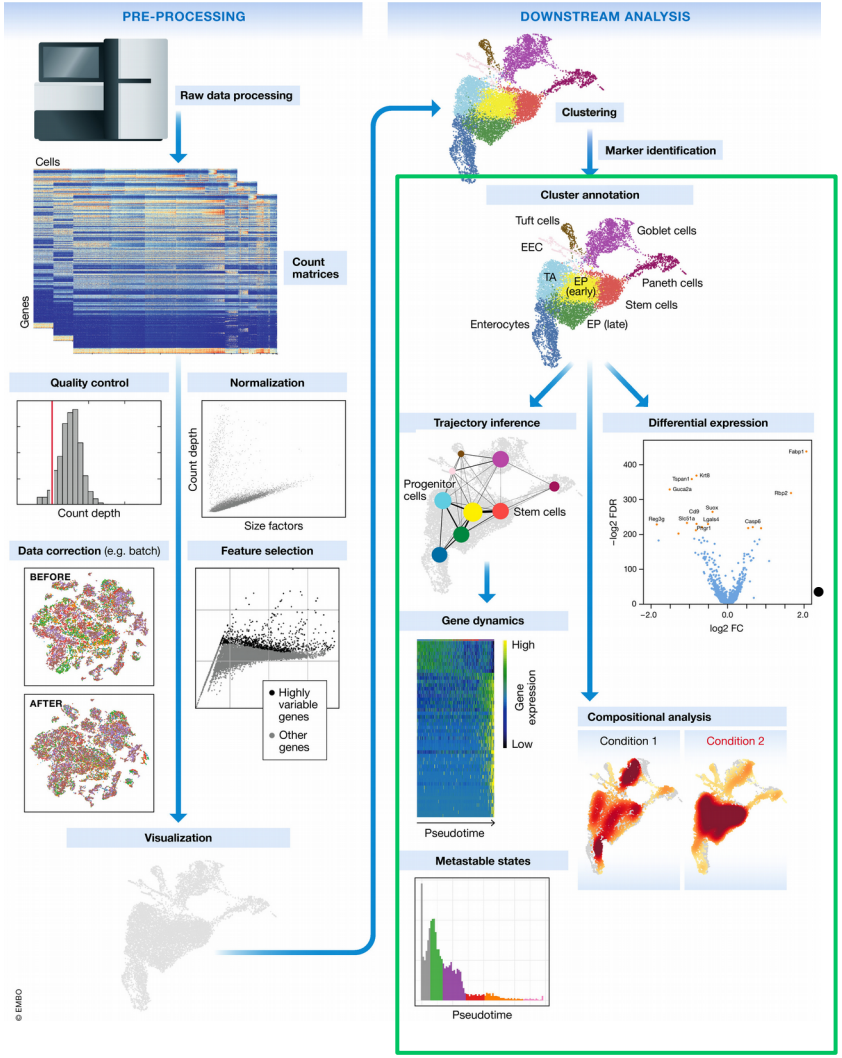
Dataset : results (with small cells)



Cluster 0 : Our *small cells*
Cluster 1 : Dendritic cells
Cluster 2 : T cells
Cluster 3 : T cells
Cluster 4 : ?
Cluster 5 : B cells
Cluster 6 : NK
Cluster 7 : Monocytes
Cluster 8 : T cells
Cluster 9 : Platelet
Cluster 10 : *Our small cells*?
Cluster 11 : Dendritic cells ?

Our population of **small cells is between T cells, B cells and NK cells.**
So, it's probably **Stem Cells : Common Lymphoid Progenitor**

Main steps of single cell data processing



this Afternoon

Malte D Luecken & Fabian J Theis
Molecular Systems Biology (2019)