A Shiny application To Explore Clusterings of Single-Cell RNA-seq data

ASTEC-sc

Parcours

```
Bac S  →  Mathématiques      →  Statistiques et       →  Doctorat ? ...
           appliqués et science    Informatique
           économique              Décisionnelle
```

Méthodes statistiques

Apprentissage automatisé

Récolte et gestion des données

Différents stages dont :

2020 : "Interface shiny pour l'étude de données scRNAseq"

- Sébastien Déjean
- Cathy Maugis Rabusseau

Apprentissage

Méthode Post Hoc
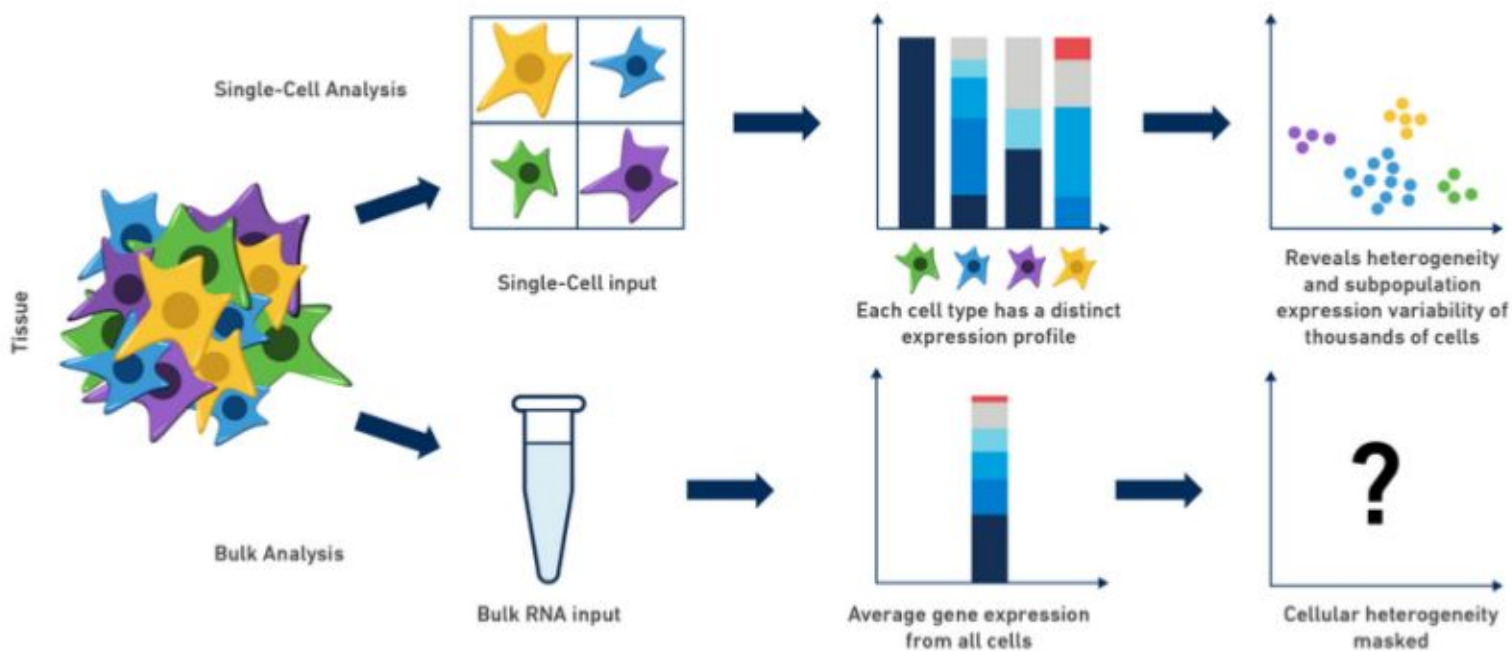
- Pierre Neuvial
- Laurent Risser
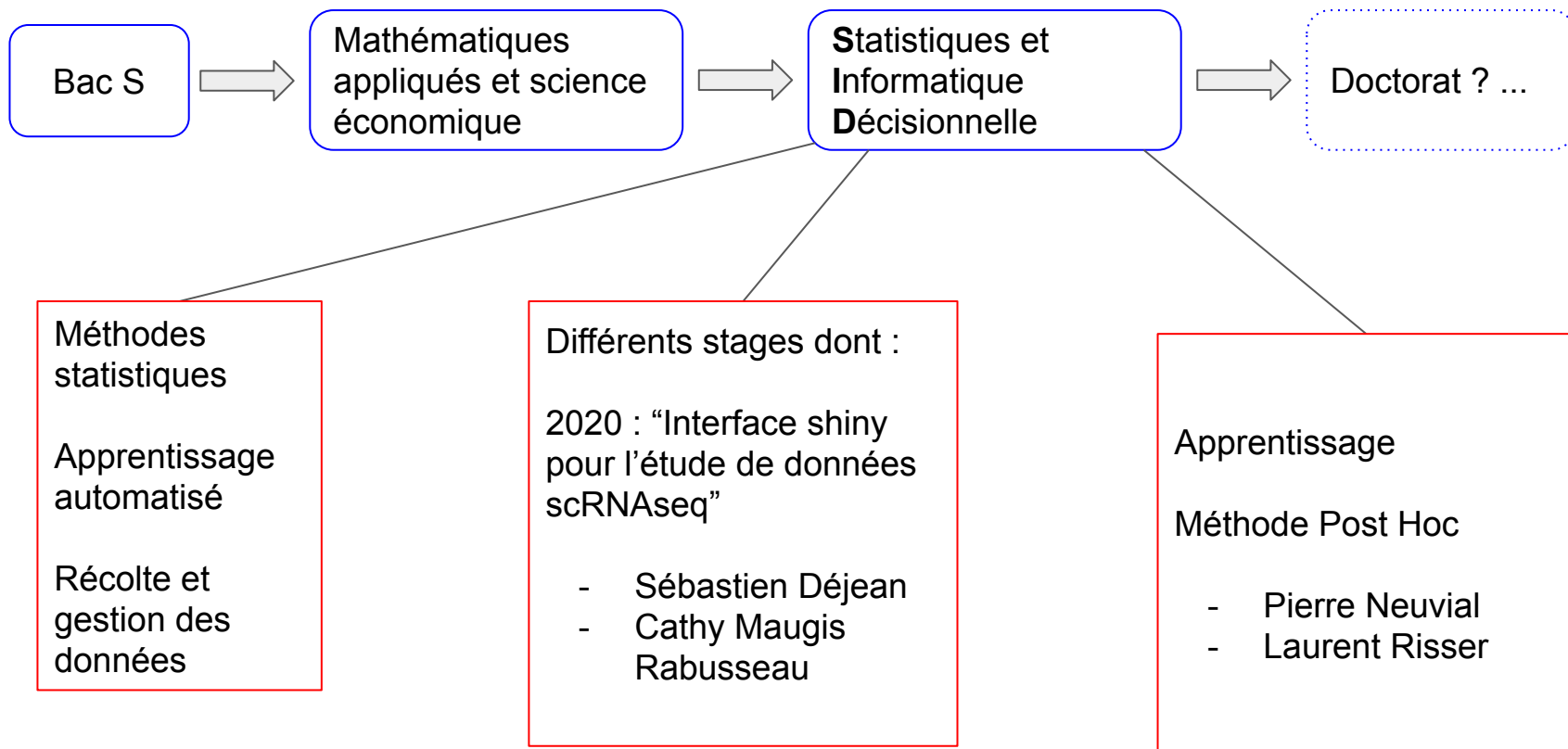
Contents

1. Context
2. Overview of the app ASTEC-sc
3. What's next ?

Contents

1. Context
2. Overview of the app ASTEC-sc
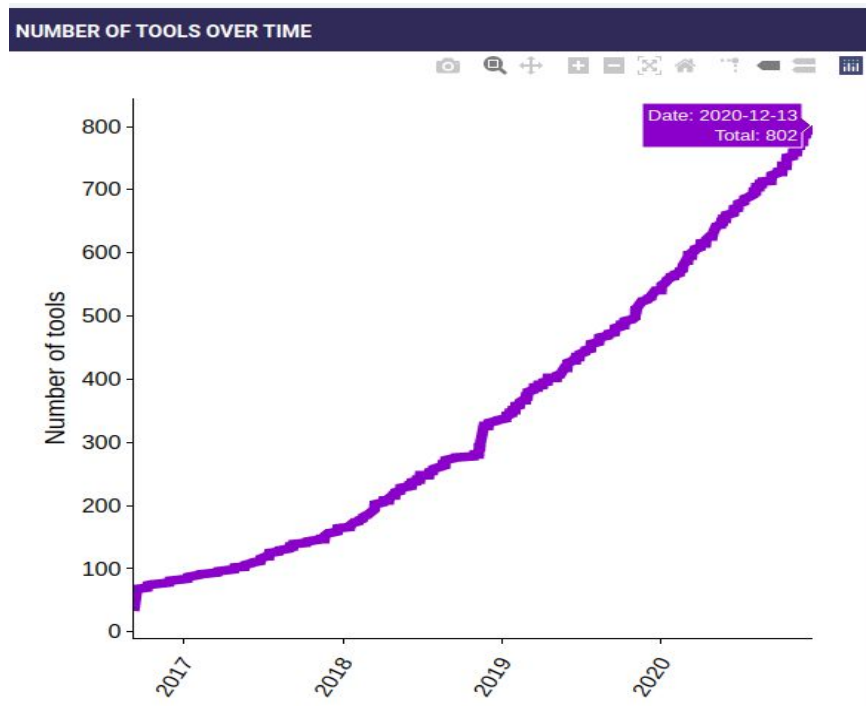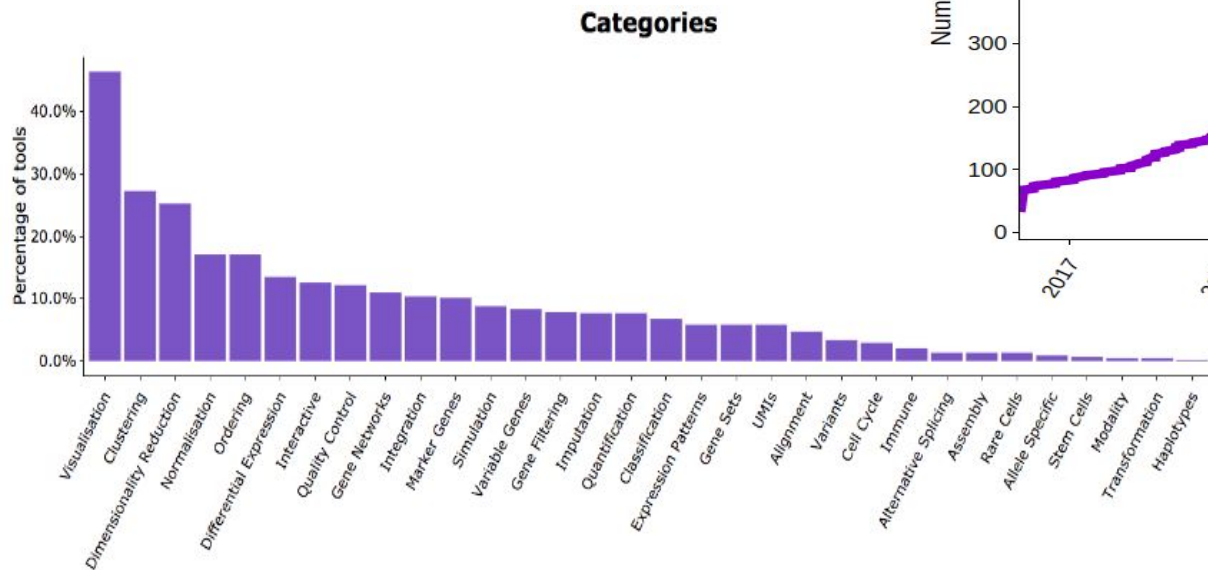3. What's next ?

Data : single cell RNA seq

(https://community.10xgenomics.com)

Main steps in statistical analysis
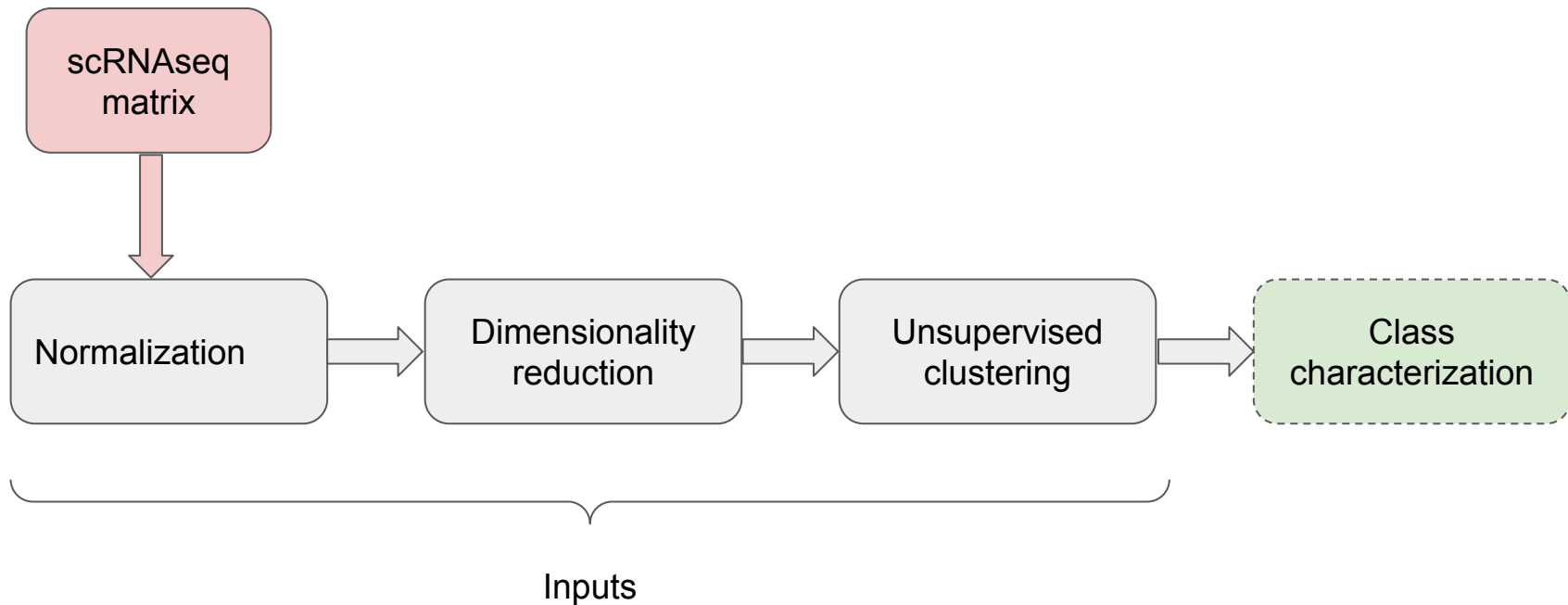
Single cell RNA seq in R

Packages : *SingleCellExperiment (SCE),*
Seurat, …

NUMBER OF TOOLS OVER TIME

Date: 2020-12-13
Total: 802

Categories

https://www.scrna-tools.org/

7

Structure of SCE object

list(clustering1, clustering2, …)

X
Xnorm
Xlog

count(sce)
logcount(sce)
normcount(sce)

reducedDims(sce)

sce@metadata[['clustering']]

sc@metadata[['a']]

sc@metadata[['b']]

...

SCE@int_element Metadata$a

SCE@int_element Metadata$b

...

SCE object

SimpleList(PCA, tsne, …)

Cell information

vectorCell1
vectorCell2
vectorCell3

Biological functions

vectorfeature1
vectorfeature2
vectorfeature3

8

Structure of SCE object

SCE = SingleCellExperiment(assays = SimpleList(counts=X, normcounts=Xnorm, logcounts=Xlog)) #assays

reducedDims(SCE) = SimpleList(...)

SCE@metadata[["clustering"]] = list(...)

SCE@metadata[["cellType1"]] = vector1

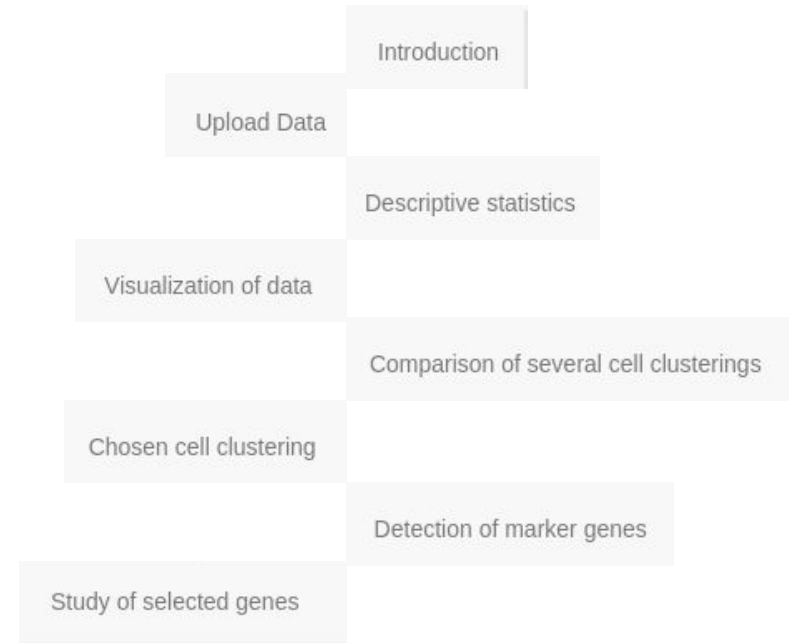SCE@int_elementMetadata$NameType1 <- vectorfeature1

Contents

Tabs of the app

Introduction
★ Upload data
Descriptive statistics
★ Visualization of data
★ Comparison of cell clusterings
★ Analysis of one cell clustering
★ Detection of marker genes
★ Analysis of genes

Introduction

Upload Data

Descriptive statistics

Visualization of data

Comparison of several cell clusterings

Chosen cell clustering

Detection of marker genes

Study of selected genes

sc-ASTEC

| Introduction | Upload Data | Descriptive statistics | Visualization of data | Comparison of several cell clusterings | Chosen cell clustering | Detection of marker genes |

Study of selected genes

Welcome to this Shiny App

The aim of this shiny application is to exploit clustering results on single-cell RNA-seq data.

This application only uses SingleCellExperiment (SCE) object. The structure of this SingleCellExperiment object, denoted by SCE, is described below.

Structure of SCE

Please follow these recommendations for building your SCE object :

1. **sce@assays** : contains the different types of data (row counts (counts), normalized counts (normcounts), lognormalized counts (logcounts), ...). By default: 'counts', 'logcounts' and 'normcounts' are required. The rows correspond to features (e.g. genes), the columns to cells.
2. **sce@metadata** contains supplementary information for cells (qualitative variables).
3. **sce@metadata$clustering** contains a list of the different cell clusterings.
4. **reducedDims(sce)** contains a list of coordinate matrices for each considered dimensionality reduction method.
5. **sce@int_elementMetadata** contains the supplementary information for features. In particular, sce@int_elementMetadata$KnownFunc may contain a binary dataframe where each column correspond to a biological function and some features in row.

This app is compatible with R version >= 3.6.3 (2020-02-29) and requires the following packages:
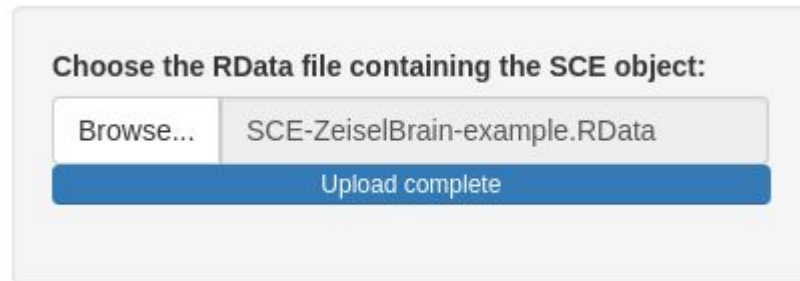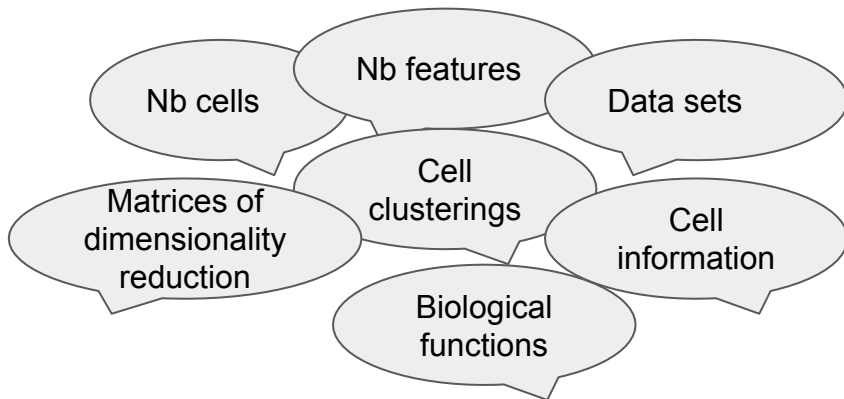
- scales 1.1.0
- Seurat 3.1.5
- dplyr 0.8.5
- scater 1.14.6
- SingleCellExperiment 1.8.0

- DT 0.13
- plotly 4.9.2.1
- ggplot2 3.3.0
- shiny 1.4.0.2

Upload data

Upload a SCE object in a Rdata file

Does the application detect the information?

User can control if his object is well constructed :

Choose the RData file containing the SCE object:

Browse...   SCE-ZeiselBrain-example.RData

Upload complete

Nb cells

Nb features

Data sets

Matrices of dimensionality reduction

Cell clusterings

Cell information

Biological functions

sc-ASTEC

Introduction   Upload Data   Descriptive statistics   Visualization of data   Comparison of several cell clusterings   Chosen cell clustering   Detection of marker genes

Study of selected genes

Choose the RData file containing the SCE object:

| Browse... | SCE-ZeiselBrain-example.RData |

Upload complete

The SCE object contains:

- the study of 3005 cells expressed in 2072 features.

- the following data sets:

  counts, normcounts, logcounts

- the following cell clusterings:

Name	Nb of clusters
ClustZeisel	9
ClustPcared	9
ClustSC3	9
ClustSeurat	15

- the following coordinate matrices of dimensionality reduction methods:

Name	Nb of dimensions
PCAlogcounts	10
tSNElogcounts	2
UMAPlogcounts	2
PCAnormcounts	10

Visualization in dimensionality reduction

User choices :
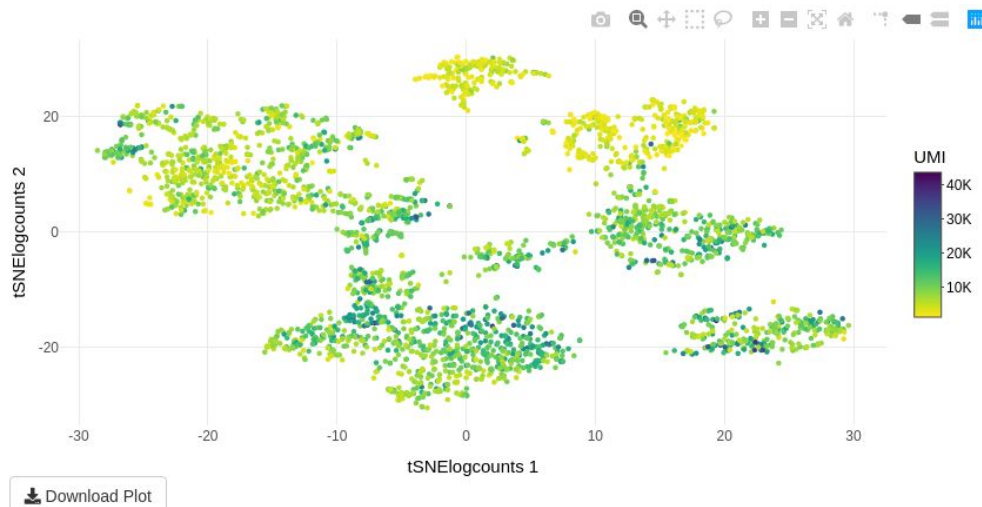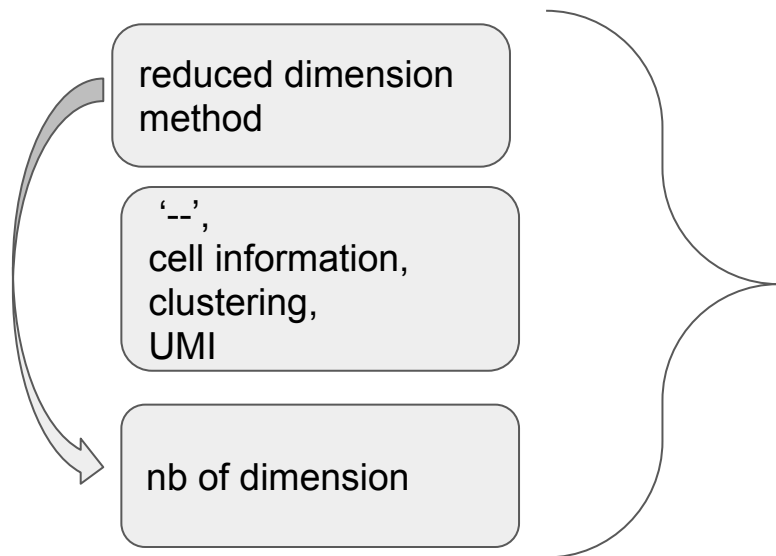
sc-ASTEC

Introduction   Upload Data   Descriptive statistics   **Visualization of data**   Comparison of several cell clusterings   Chosen cell clustering   Detection of marker genes
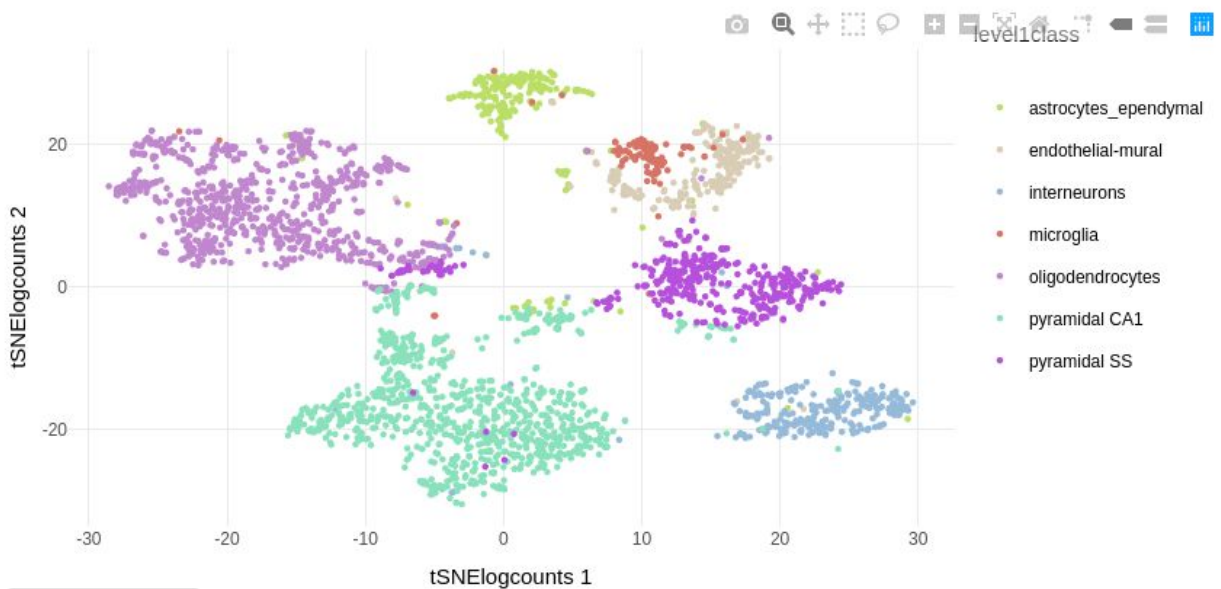
Study of selected genes

Choose the coordinates of dimension reduction:

tSNElogcounts   ▾

Cell information for colour:
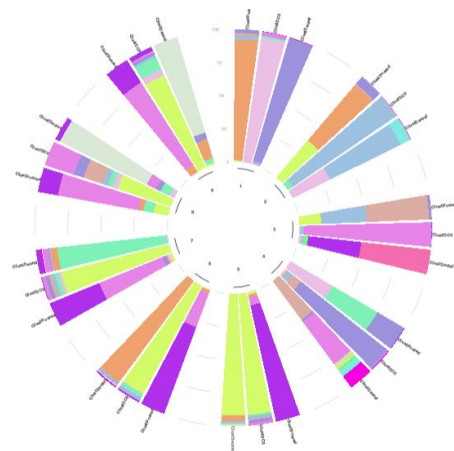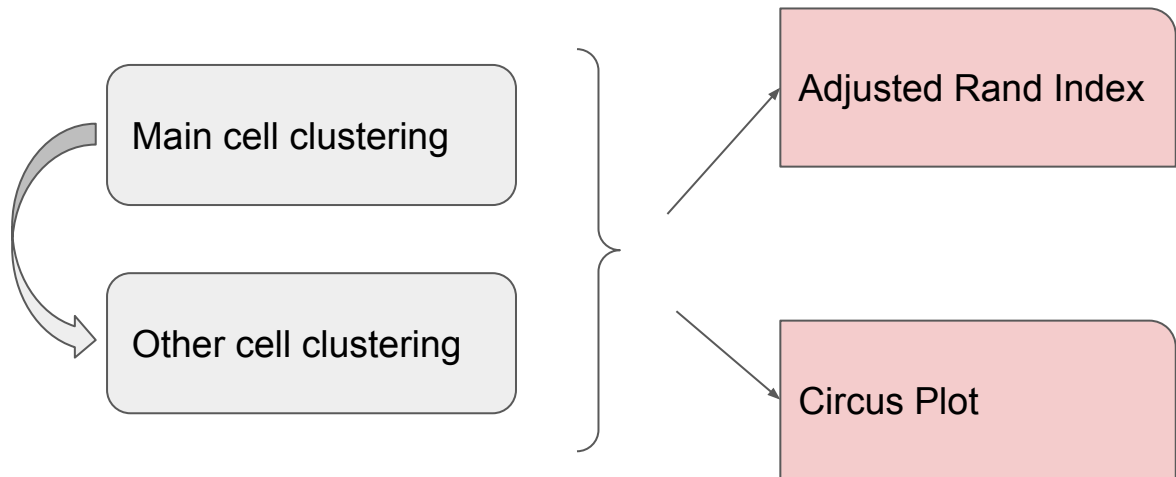
level1class   ▾

Choose the number of dimensions:

2

Graph of the cells in a low-dimensional space obtained with the chosen dimensionality reduction method.

level1class

- astrocytes_ependymal
- endothelial-mural
- interneurons
- microglia
- oligodendrocytes
- pyramidal CA1
- pyramidal SS

⬇ Download Plot

Comparison of cell clusterings

Main cell clustering

Other cell clustering

Adjusted Rand Index

Circus Plot

sc-ASTEC

Introduction   Upload Data   Descriptive statistics   Visualization of data   **Comparison of several cell clusterings**   Chosen cell clustering   Detection of marker genes

Study of selected genes

Choose the cell clusterings to compare
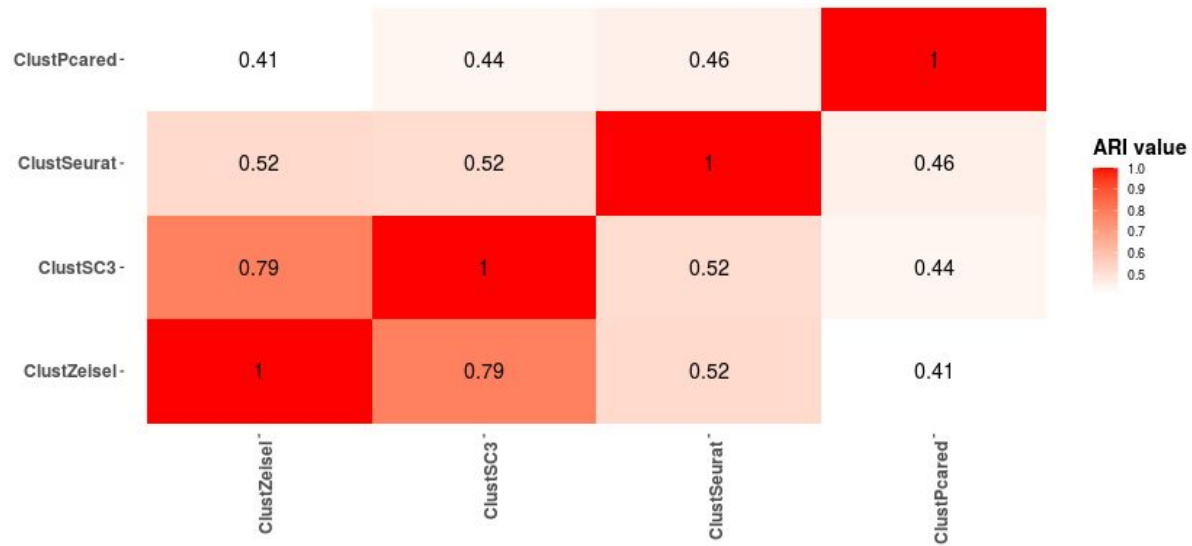
The main cell clustering:

ClustZeisel   ▾

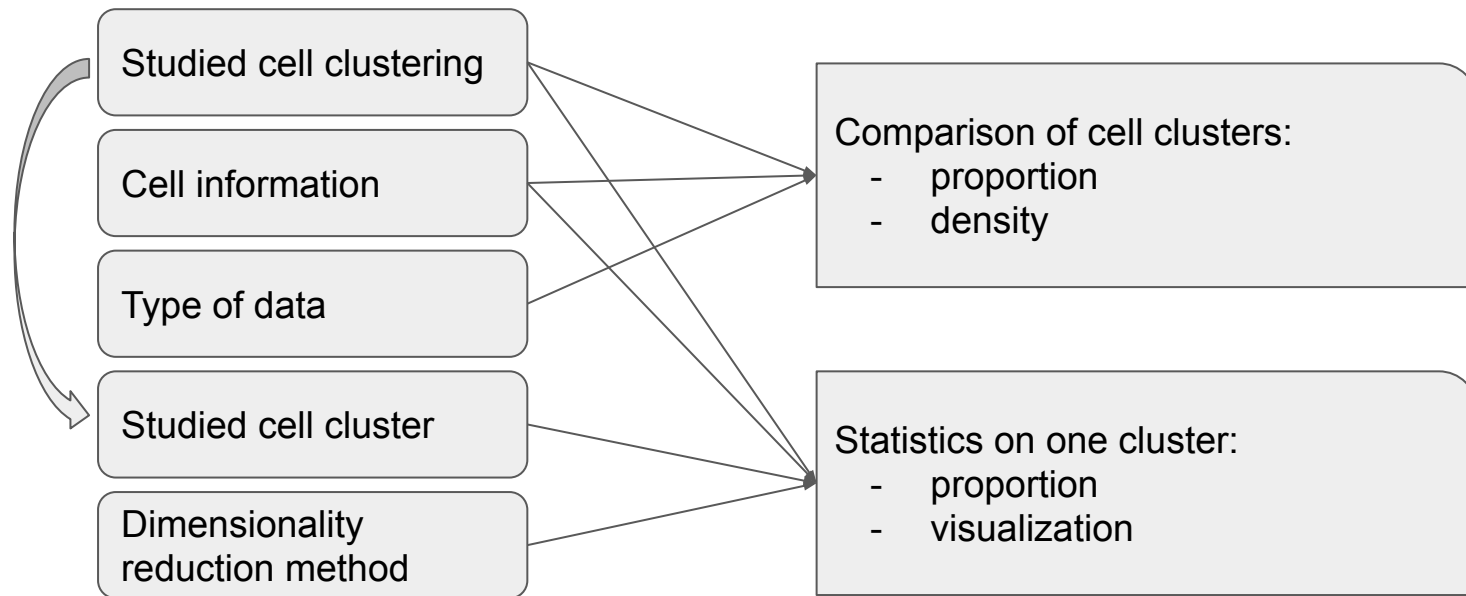**The other cell clusterings:**

ClustSC3   ClustSeurat   ClustPcared

ARI   CircusPlot

Table of adjusted rand index (ARI) between each cell clusterings.

⬇ Download Plot

Analysis of cell clustering

Introduction   Upload Data   Descriptive statistics   Visualization of data   Comparison of several cell clusterings   Chosen cell clustering   Detection of marker genes

Study of selected genes

Choose the studied cell clustering:
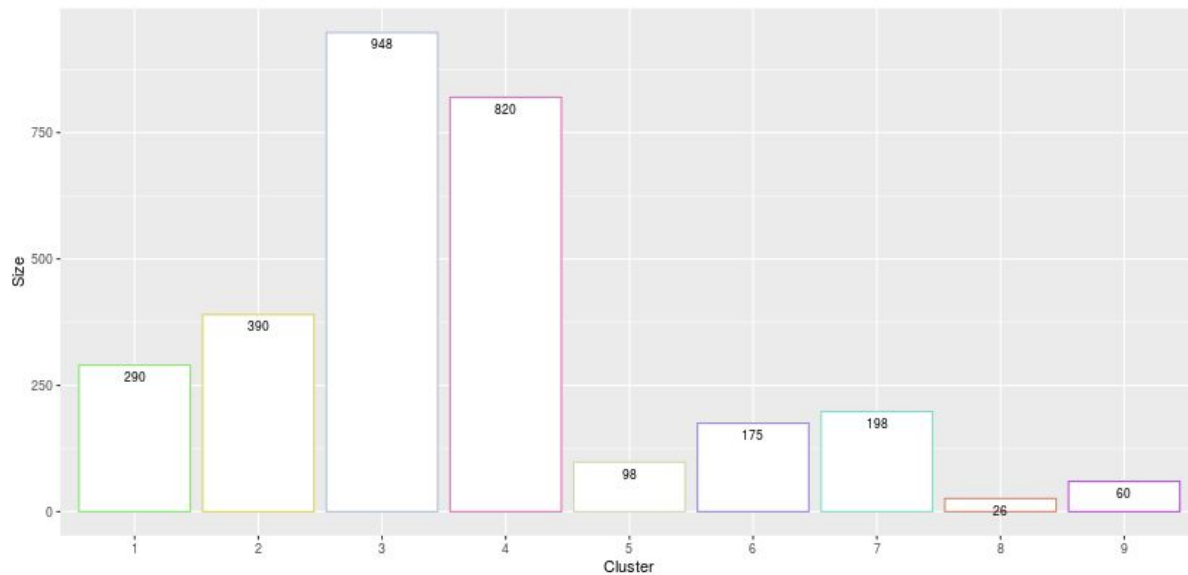
ClustZeisel

Choose cell information:
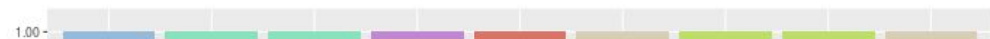
level1class

Choose a type of data:

counts

Description   Specific cluster

Number of cells per cluster:

Download Plot

Proportion of cell information per cluster:

Detection of marker genes

One to one
One to many
Many to one
Many to many

Choose two cells groups (based on clusters) [C1, C2]

Calculation of indicators by gene:

- **pct.1 (pct.2)**: % of express cells in C1 (resp. C2)
- **avg_logFC**: log fold-change of the average expression between the two groups
- **AUC**: the area under the ROC curve
- **p_val_adj**: the adjusted p-value on *Wilcoxon test*, based on *Bonferroni correction* using all genes in the dataset

Marker genes := Selection with threshold on theses indicators

Detection of marker genes

Introduction   Upload Data   Descriptive statistics   Visualization of data   Comparison of several cell clusterings   Chosen cell clustering   Detection of marker genes
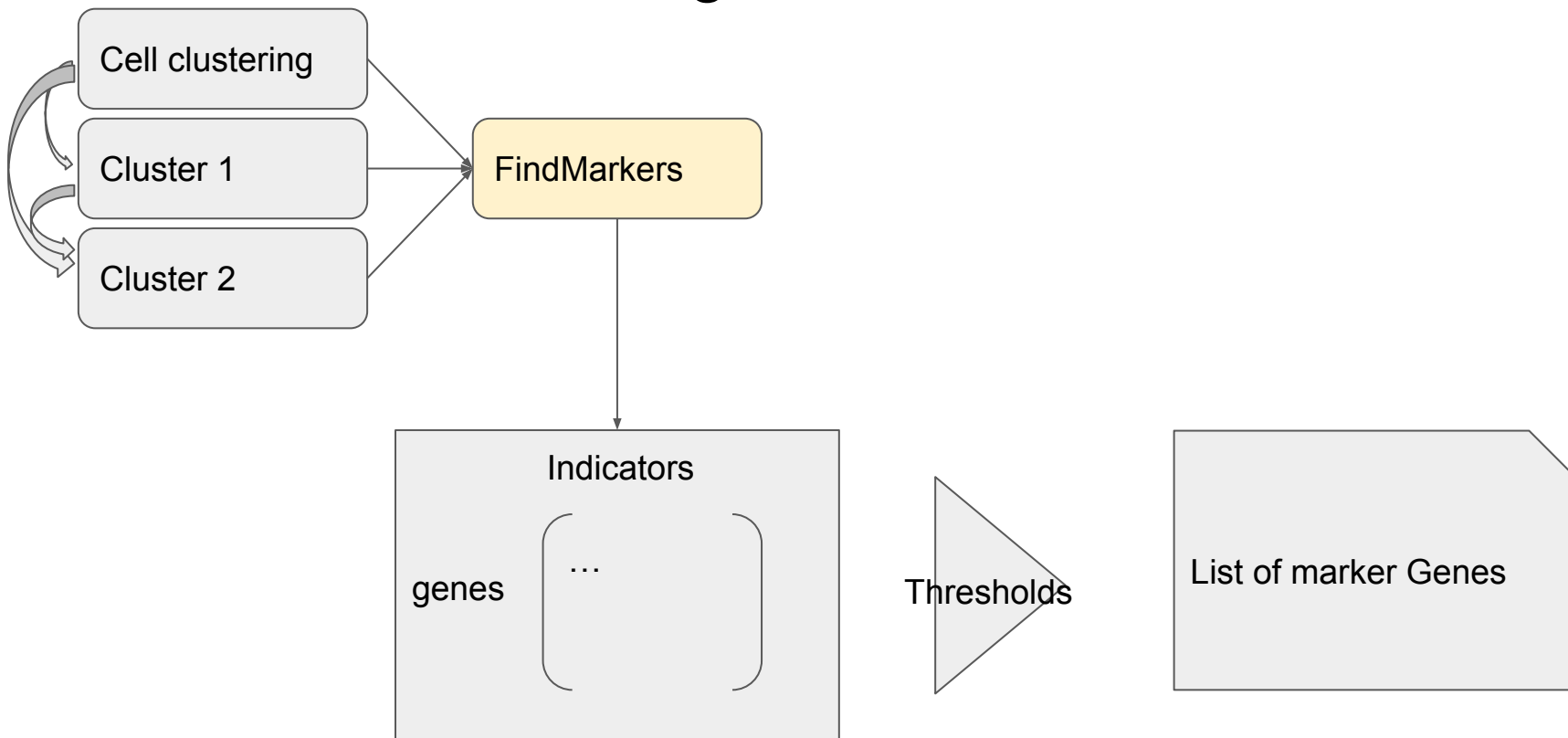
Study of selected genes

Choose the studied cell clustering:

ClustZeisel ▾

Choose one (some) studied cluster(s):

6

Choose the comparing group of clusters (one, some or all others):

9

Go - Compute indicators.

Calibration of indicators

Maximum threshold for the adjusted p-val:

0.05 ●━━━━━━━━━━━━━━━━━━ 1

0   0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Minimum threshold for the logfoldchange:

0 ━━━━━━━━2.44●━━━━━━━ 5

| Find Marker Genes | Indicators |

With the chosen thresholds, 10 marker genes are detected.

Show [10 ▾] entries   Search: [   ]

|	gene	avg_logFC	pct.1	pct.2	p_val_adj	AUC	delta	func
1	Arap3	2.913	0.754	0.133	0.000	0.833	0.621	
2	Fam101b	2.998	0.560	0.033	0.000	0.767	0.527	
3	Flt1	2.742	0.863	0.133	0.000	0.886	0.730	
4	Gimap6	3.057	0.783	0.100	0.000	0.860	0.683	
5	Itm2a	2.736	0.994	0.333	0.000	0.933	0.661	
6	Ly6c1	2.976	0.914	0.217	0.000	0.893	0.697	
7	mt-Tc	3.026	0.891	0.550	0.000	0.817	0.341	
8	Slc38a3	3.236	0.640	0.033	0.000	0.805	0.607	
9	Slc7a5	2.848	0.749	0.167	0.000	0.821	0.582	
10	Slco1a4	2.755	0.846	0.150	0.000	0.873	0.696	

Showing 1 to 10 of 10 entries   Previous [1] Next

Study of selected genes
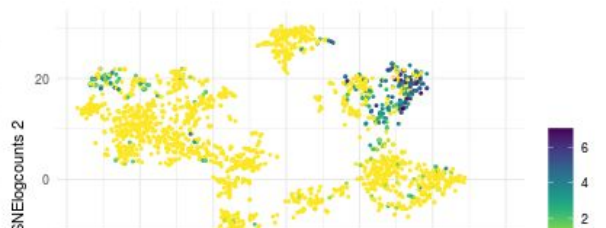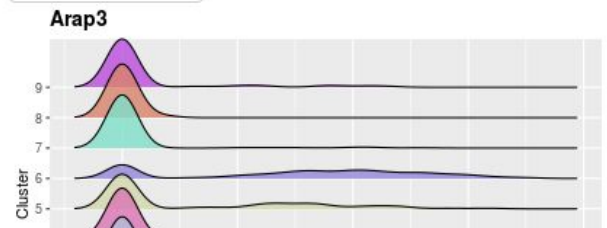
Study of selected genes
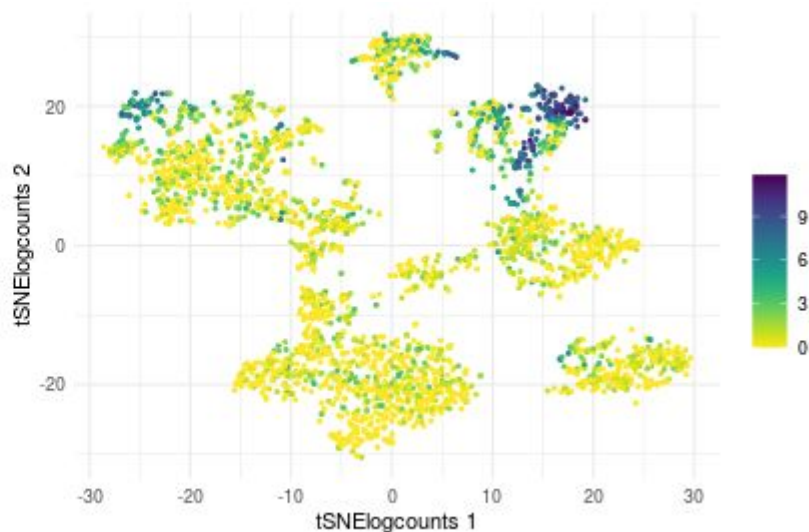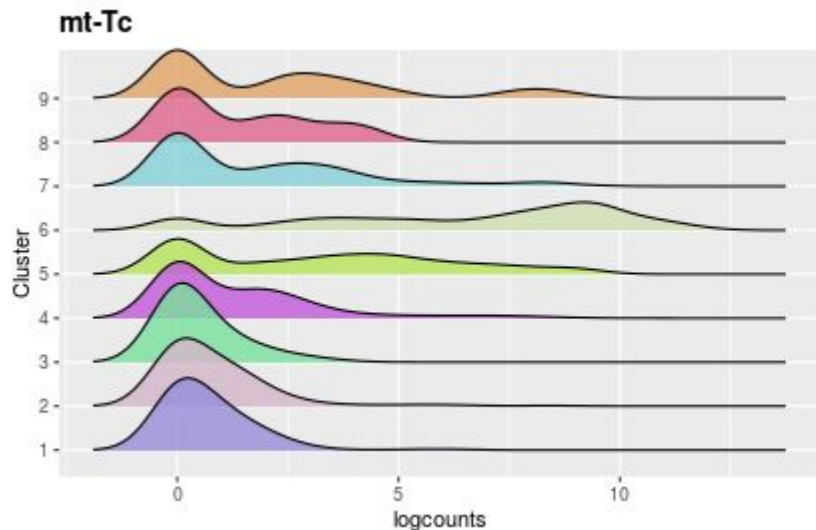
Contents

1. Context
2. Overview of the app ASTEC-sc
3. What's next

Advantages

User-friendliness

Many options for analysis

Autonomy

Comparison of cell clusterings

Improvement

More comparison of cell clusterings

Open structure of sce object

Deployment

Mathrice : the CNRS server

Access only with internet

The way to use ASTEC-sc without R and packages