2021 生物醫學大數據淘金工作坊
Seashell: Web portal for Single cell RNA-seq

Presenter:
中研院資訊所 林仲彥老師實驗室
莊佳憲

2022.01

Outline

01 Introduction

single-cell RNA sequencing

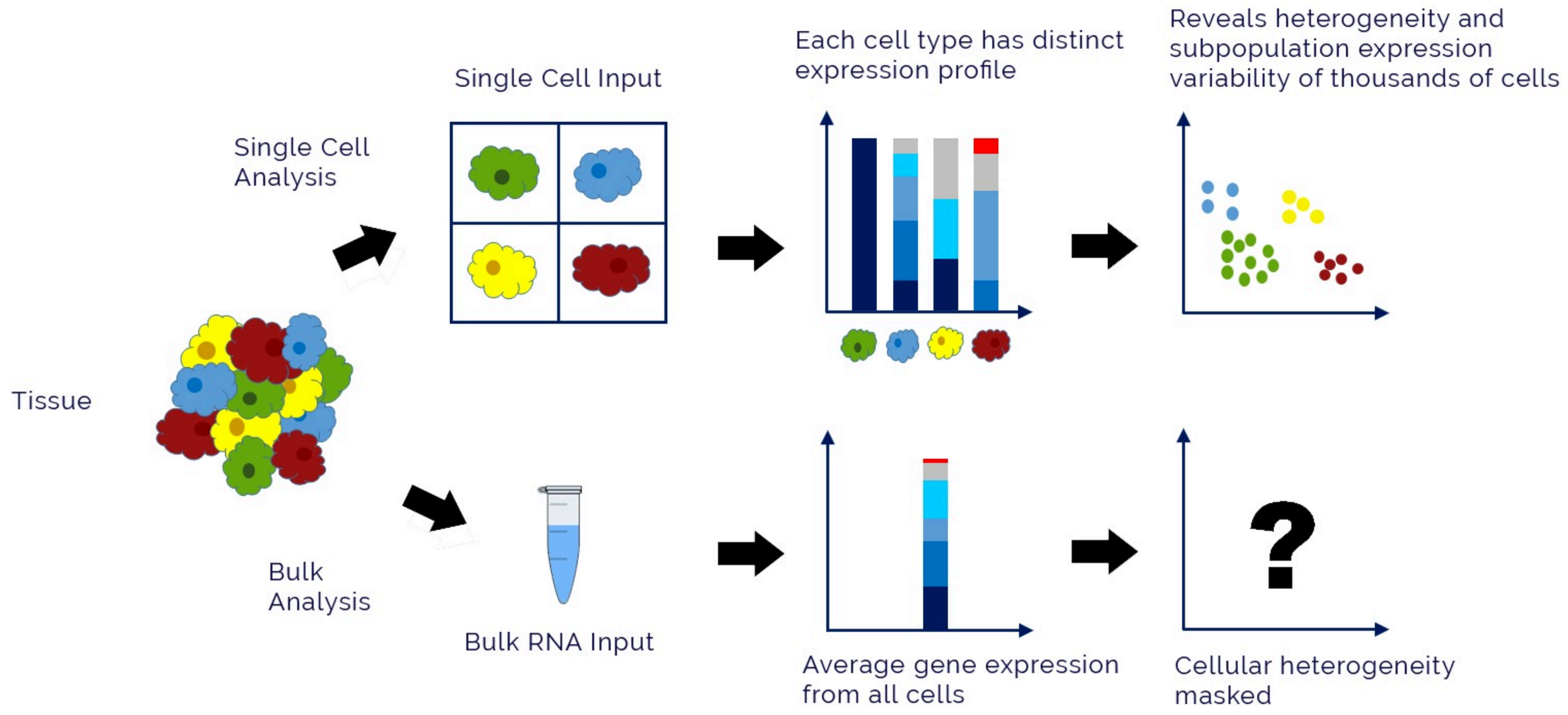
02 Seashell: web portal for single cell RNA-seq analysis and visualization

- SeaShell_Data
- SeaShell_Analyzer

Introduction

Single-cell RNA sequencing (scRNA-seq)

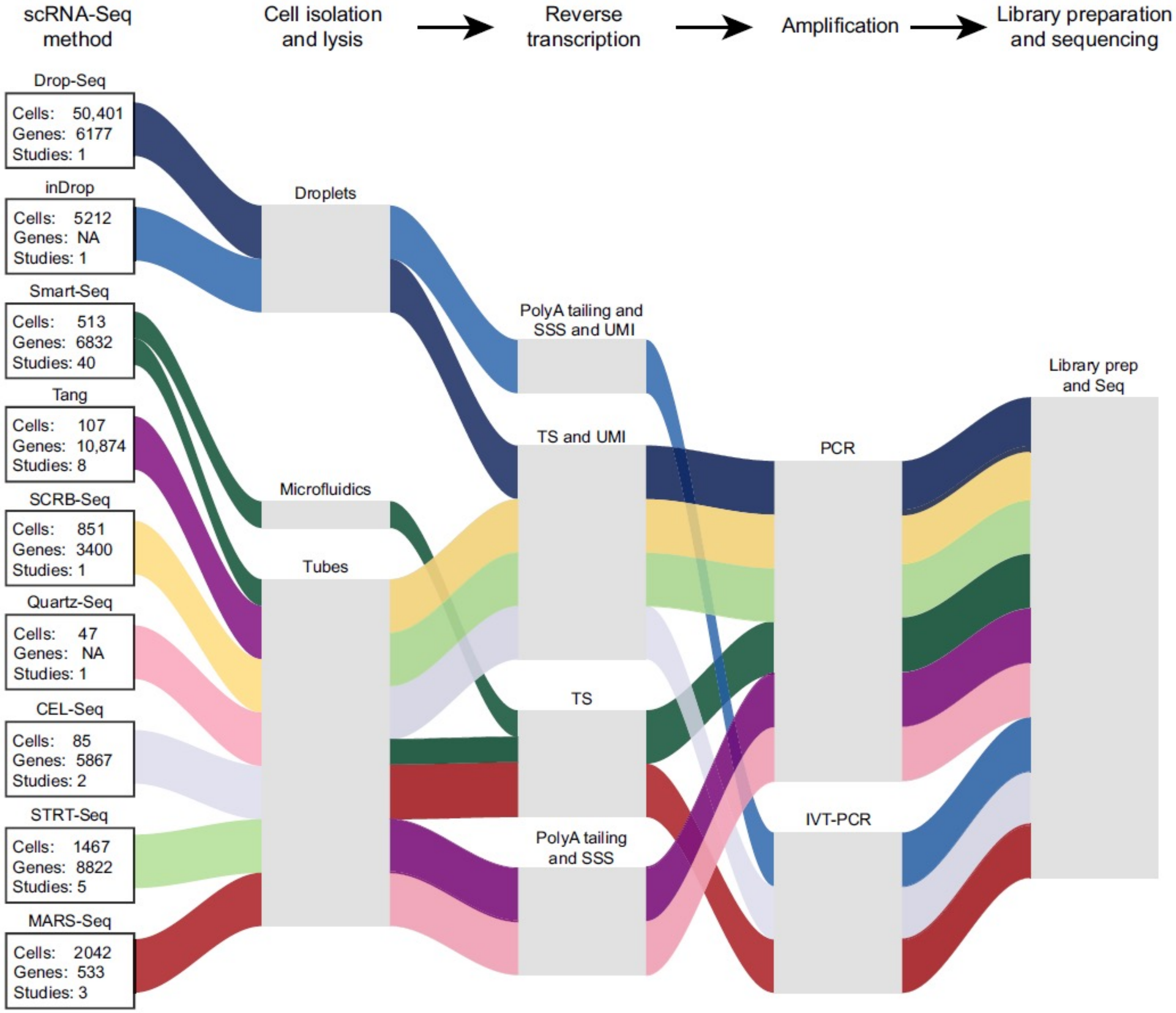
scRNA-seq technologies have combined effective single-cell isolation strategies with highly sensitive molecule detection approaches, showing promise in unravelling the heterogeneity of complex tissues or organs.



https://www.lcsciences.com/discovery/applications/transcriptomics/single-cell-rna-seq-sequencing-service/

Single-cell RNA sequencing

- **A variety of scRNA-seq protocols**



scRNA-Seq method

Cell isolation and lysis → Reverse transcription → Amplification → Library preparation and sequencing

Drop-Seq
Cells: 50,401
Genes: 6177
Studies: 1

inDrop
Cells: 5212
Genes: NA
Studies: 1

Smart-Seq
Cells: 513
Genes: 6832
Studies: 40

Tang
Cells: 107
Genes: 10,874
Studies: 8

SCRB-Seq
Cells: 851
Genes: 3400
Studies: 1

Quartz-Seq
Cells: 47
Genes: NA
Studies: 1

CEL-Seq
Cells: 85
Genes: 5867
Studies: 2

STRT-Seq
Cells: 1467
Genes: 8822
Studies: 5

MARS-Seq
Cells: 2042
Genes: 533
Studies: 3

Droplets

Microfluidics

Tubes

PolyA tailing and SSS and UMI

TS and UMI

TS

PolyA tailing and SSS

PCR

IVT-PCR

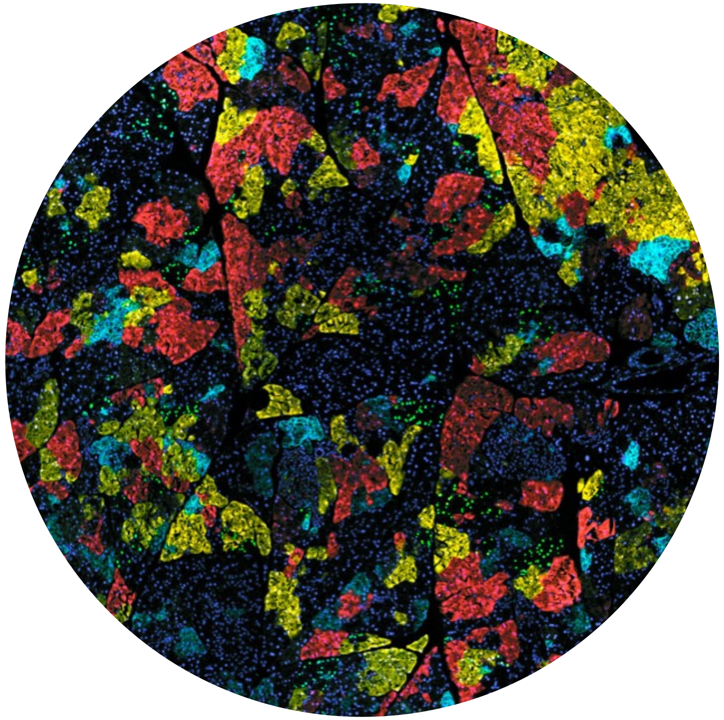
Library prep and Seq

Kumar, Pavithra, Yuqi Tan, and Patrick Cahan. "Understanding development and stem cells using single cell-based analyses of gene expression." *Development* 144.1 (2017): 17-32.

Single-cell RNA sequencing

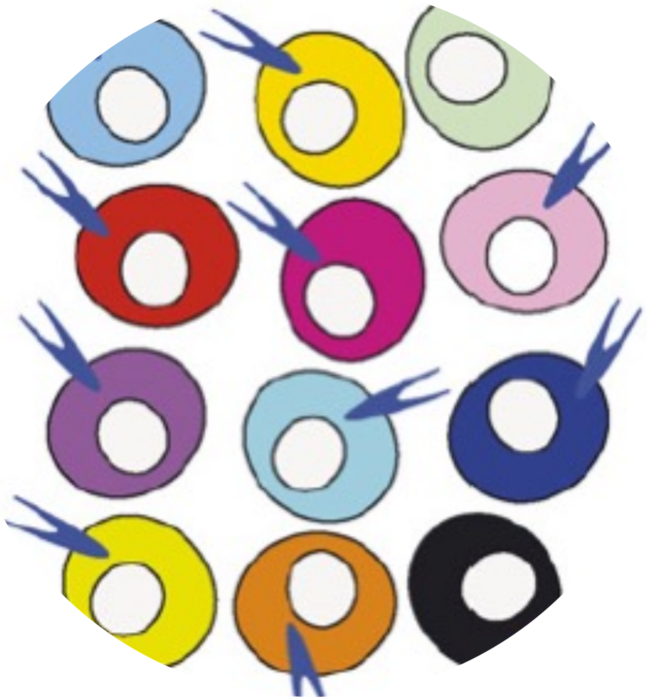
- **A growing number of application**

Heterogenity of cancer cells



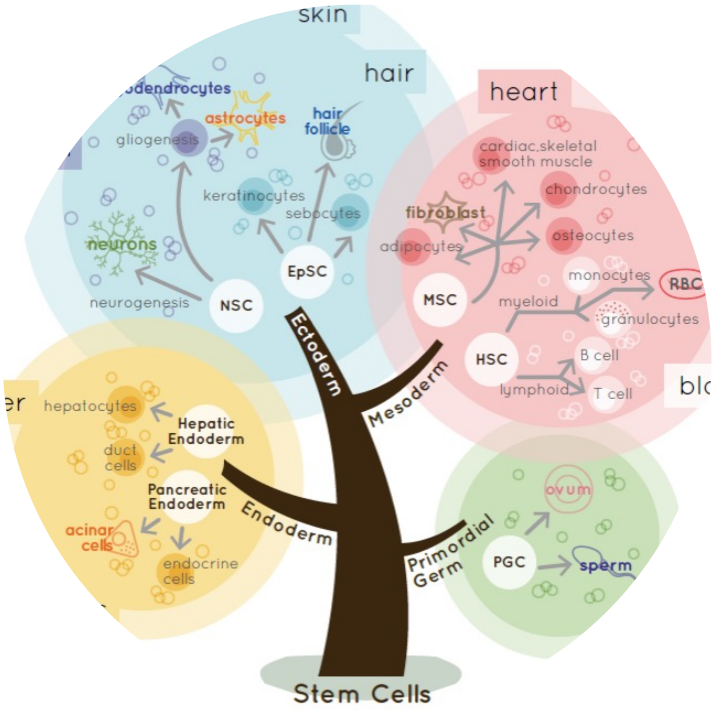
https://www.rna-seqblog.com/new-single-cell-rna-sequencing-methods-could-lead-to-better-regenerative-therapies/

Diversity of T cell receptor



Vallejo, Abbe N. "Immune remodeling: lessons from repertoire alterations during chronological aging and in immune-mediated disease." *Trends in molecular medicine* 13.3 (2007): 94-102.

Differentiation of stem cells

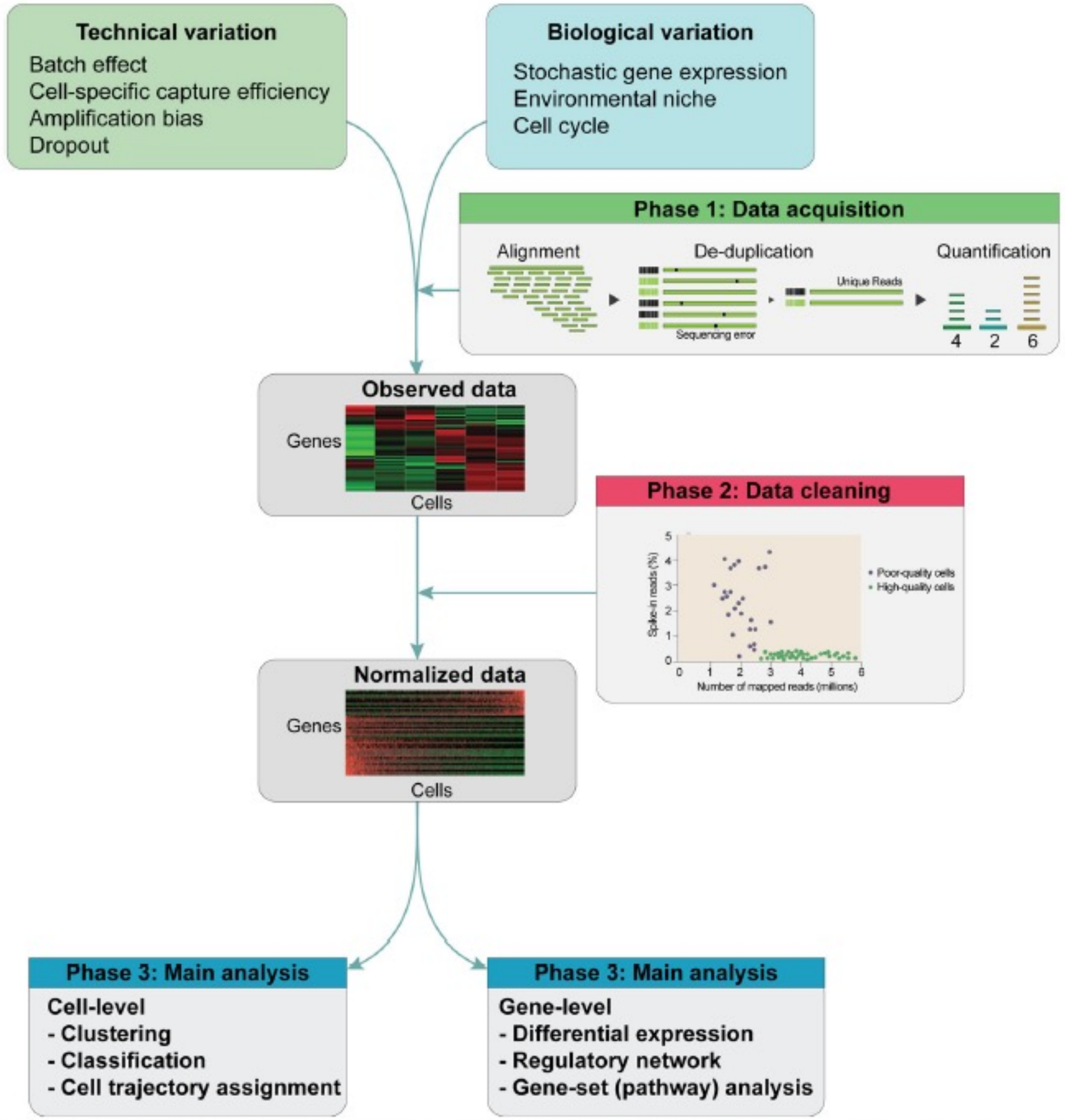


https://www.bio-connect.nl/stem-cell-and-the-regenerative-medicine-ready-for-the-patients/cnt/page/5050

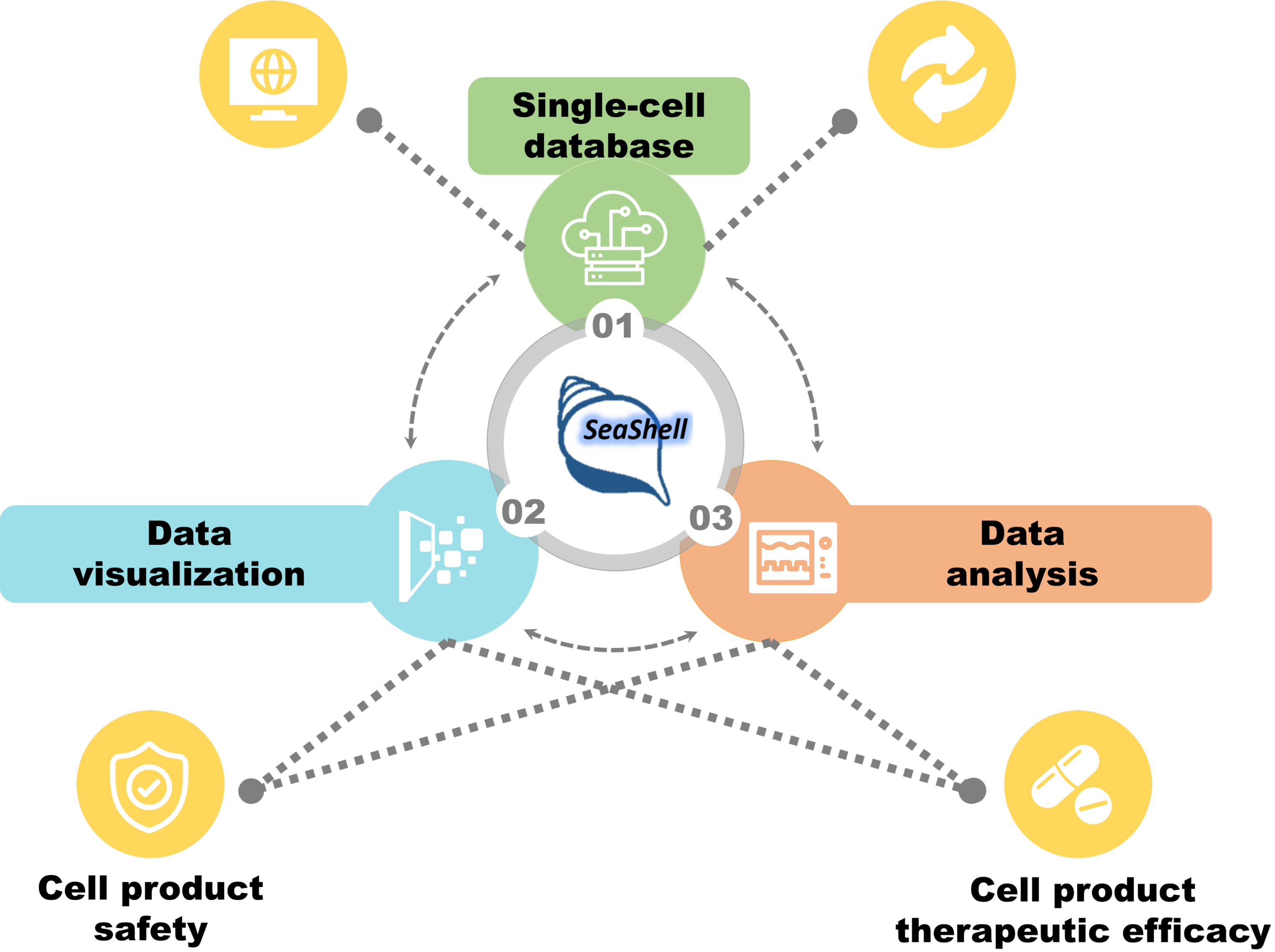
Single-cell RNA sequencing

Increasing difficulties of processing sequencing data

- Serious batch effect
- Demanding computing power
- etc



Seashell: an automatic web portal system for single-cell RNA sequencing (scRNA-seq) analysis



Single-cell database

01

SeaShell

02

Data visualization

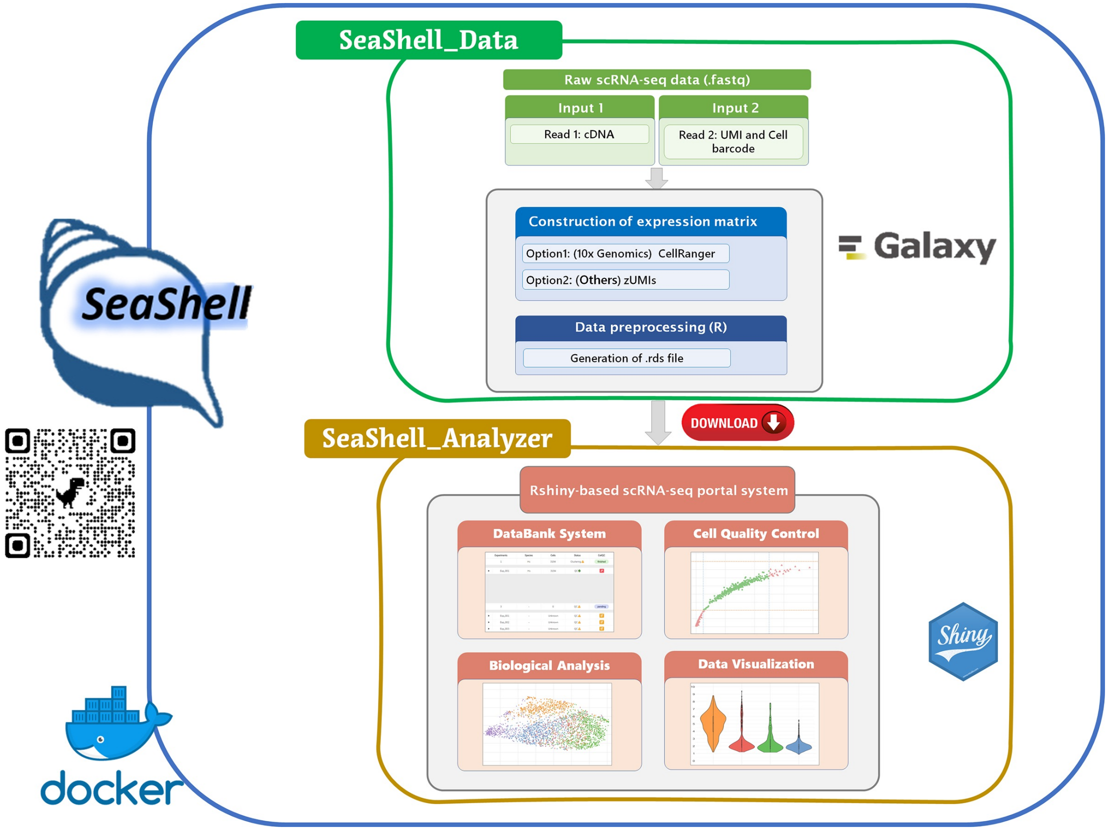
03

Data analysis

Cell product safety

Cell product therapeutic efficacy

Workflow of analysis



SeaShell_Data

Raw scRNA-seq data (.fastq)

Input 1

Read 1: cDNA

Input 2

Read 2: UMI and Cell barcode

Construction of expression matrix

Option1: (10x Genomics) CellRanger

Option2: (Others) zUMIs

Data preprocessing (R)

Generation of .rds file

DOWNLOAD

SeaShell_Analyzer

Rshiny-based scRNA-seq portal system

DataBank System

Cell Quality Control

Biological Analysis

Data Visualization

Docker Hub: https://hub.docker.com/r/lsbnb/seashell

The portal system can be divided into two parts:

Galaxy-based docker image for data preprocessing

Users can run this docker image by a Galaxy-based GUI, and then easily obtain expression matrices for further analysis.

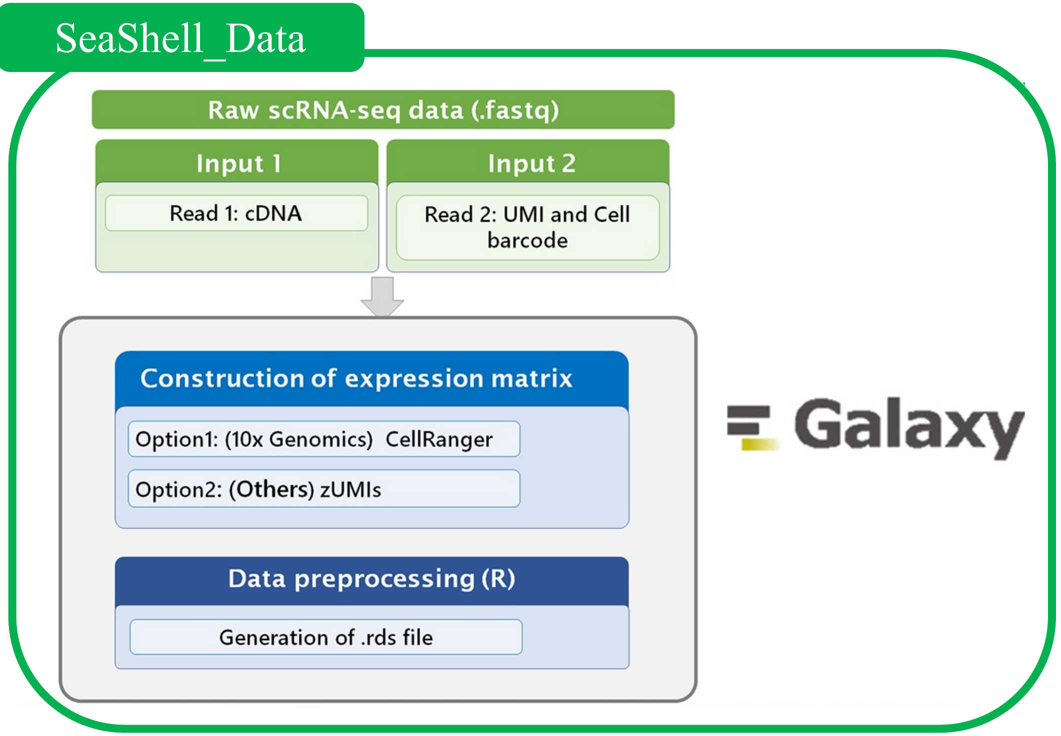
RShiny-based web portal system

The web-portal system was well-developed by R language, JavaScript D3 library, and other visualization tools, allowing researchers to automatically perform a user-friendly and up-to-date scRNA-seq analysis pipeline on scRNA-seq experiments. The analysis pipeline contains: cell quality control, normalization, cluster analysis, differentially expressed genes analysis, marker identification, gene ontology analysis, and other popular tools.

*** Now only available for 10x drop-seq raw data**

SeaShell_Data

1. Start the galaxy web server



01

Set up docker environment

1-1. Create required folders for input and output files
1-2. Data hierarchy
1-3. Download reference files from docker image
1-4. Run docker

02

Run SeaShell_Data by command mode or galaxy mode (http://[Your IP]:8080)

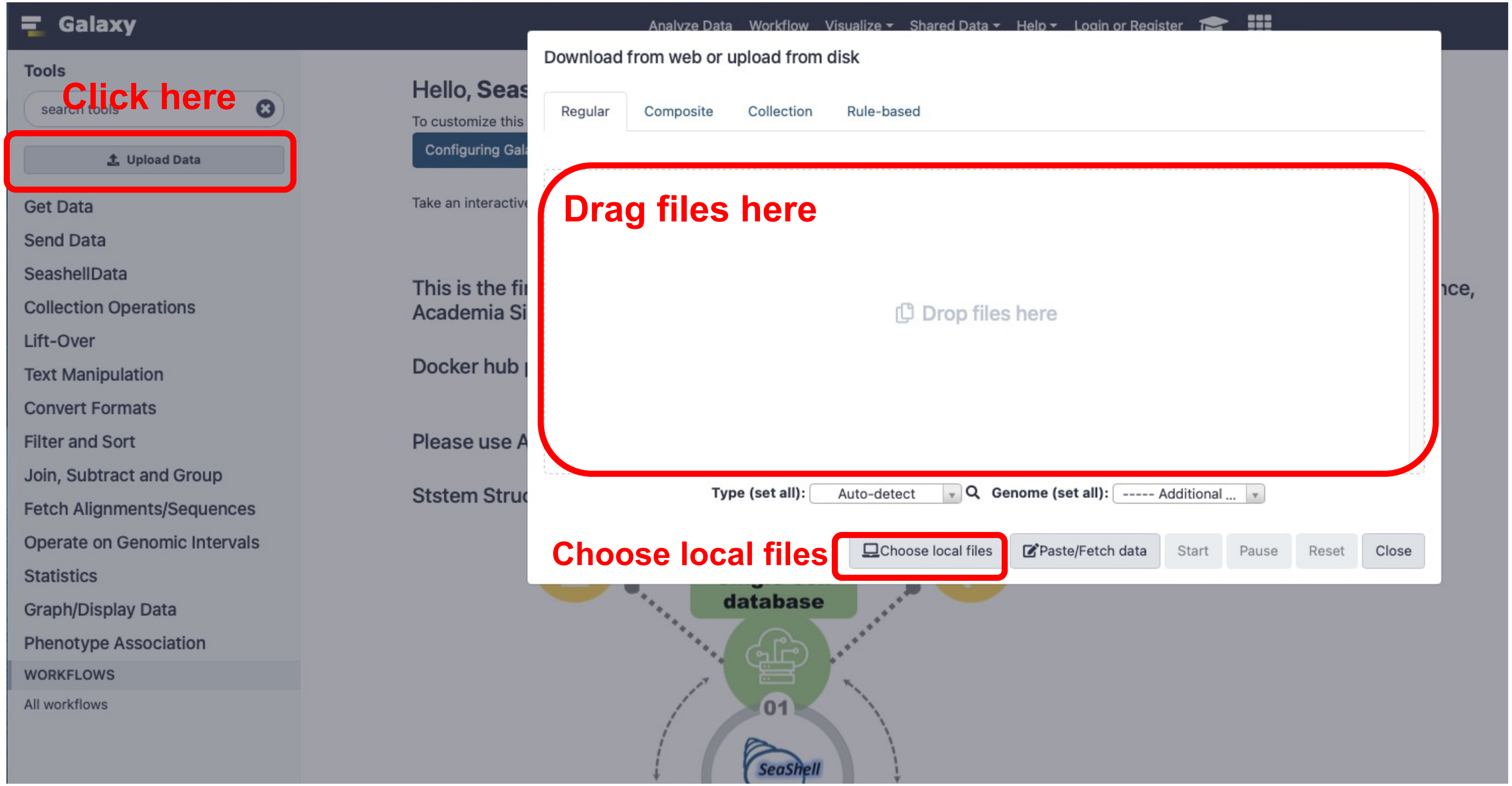
1. CellRanger 3.0.2 -> CellrangerToSeashell
2. zUMIs 2.9.4 -> zUMIsToSeashell
3. CsvLoomToSeashell

03

Use Seashell_Analyzer

http://[Your IP]:8081/seashell/ will be the access to galaxy web server

0. Upload Data


Click here

Drag files here

Choose local files

Download from web or upload from disk

Hello, Seas

To customize this

Take an interactive

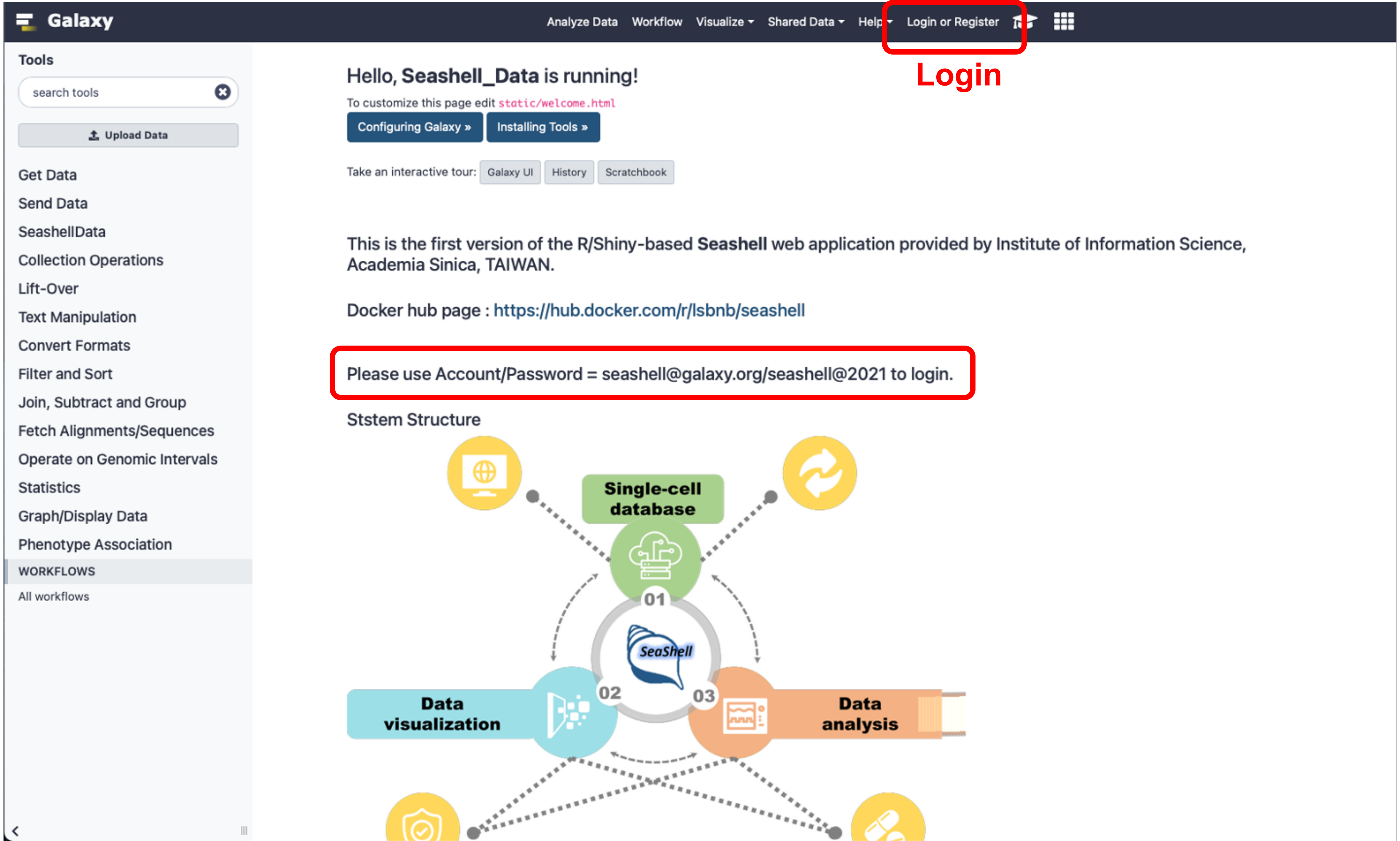
This is the fi
Academia Si

Docker hub

Please use A

Ststem Struc

1. Start the galaxy web server



Hello, **Seashell_Data** is running!

To customize this page edit `static/welcome.html`

Take an interactive tour:

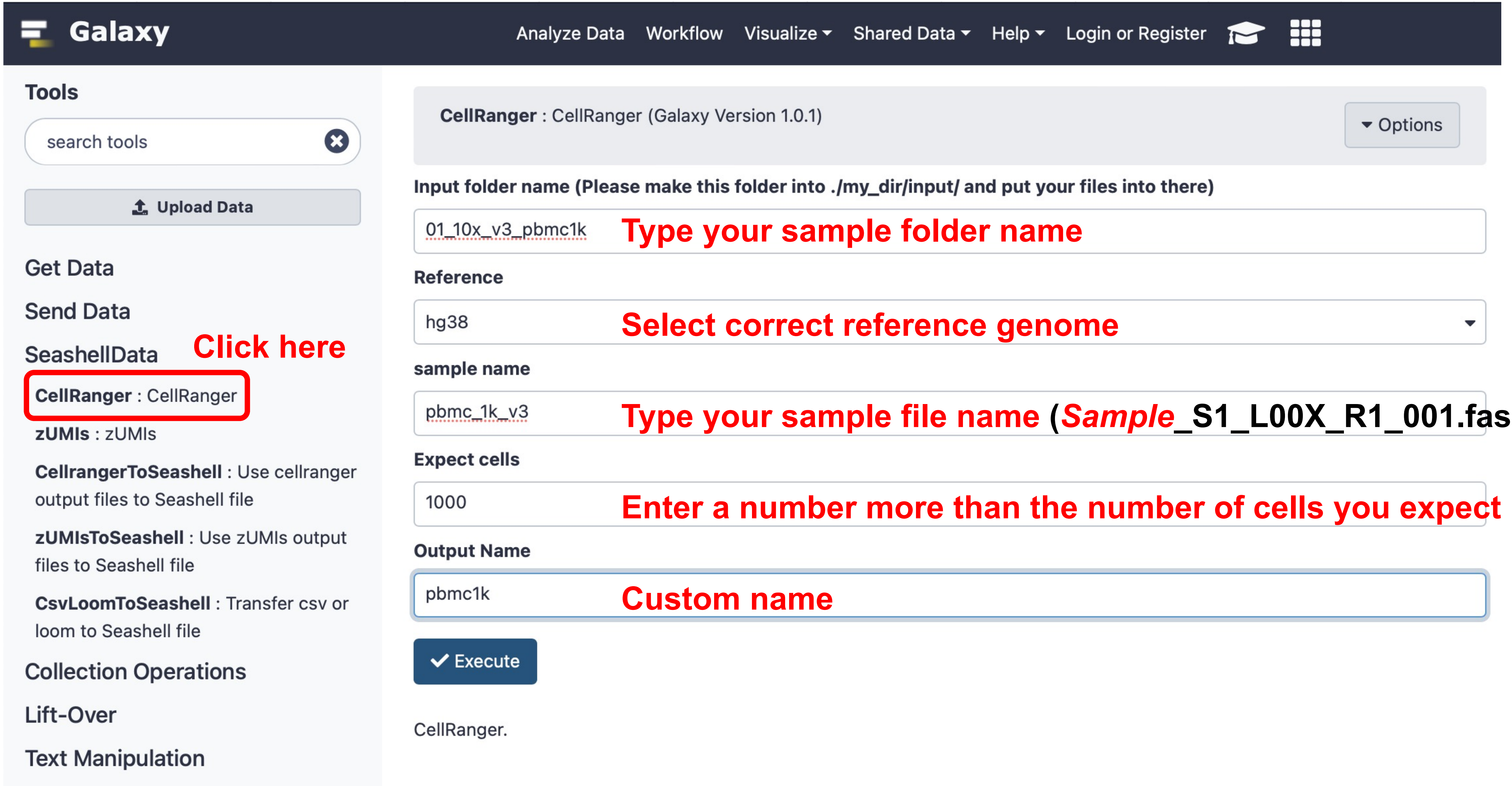
This is the first version of the R/Shiny-based **Seashell** web application provided by Institute of Information Science, Academia Sinica, TAIWAN.

Docker hub page : https://hub.docker.com/r/lsbnb/seashell

Please use Account/Password = seashell@galaxy.org/seashell@2021 to login.

Ststem Structure

3. Run CellRanger



Click here

CellRanger : CellRanger (Galaxy Version 1.0.1)

Input folder name (Please make this folder into ./my_dir/input/ and put your files into there)
01_10x_v3_pbmc1k

Type your sample folder name

Reference
hg38

Select correct reference genome

sample name
pbmc_1k_v3

Type your sample file name (*Sample*_S1_L00X_R1_001.fastq.gz)

Expect cells
1000

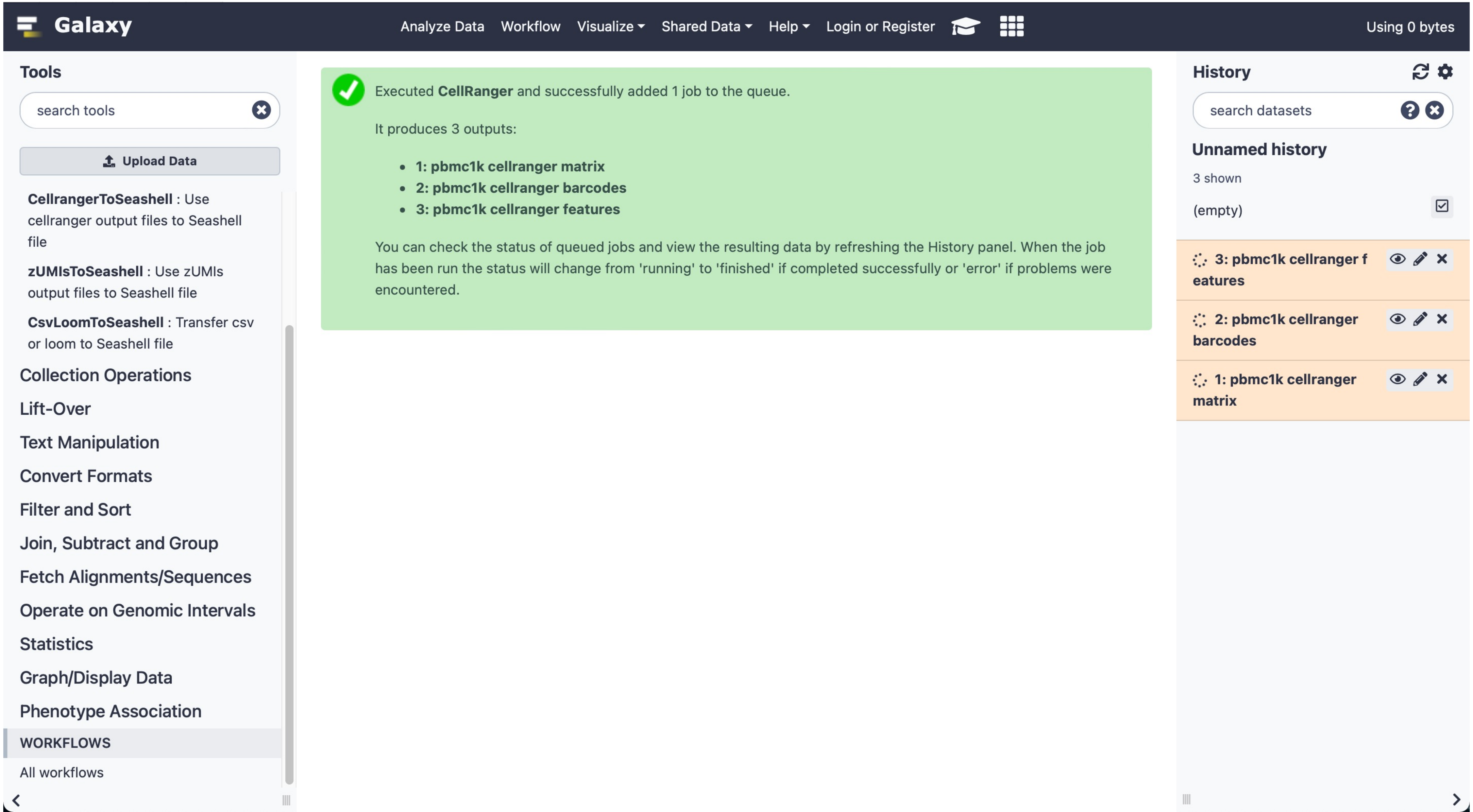
Enter a number more than the number of cells you expect

Output Name
pbmc1k

Custom name

CellRanger.

3. Run CellRanger



CellrangerToSeashell : Use cellranger output files to Seashell file

zUMIsToSeashell : Use zUMIs output files to Seashell file

CsvLoomToSeashell : Transfer csv or loom to Seashell file

Executed **CellRanger** and successfully added 1 job to the queue.

It produces 3 outputs:

- **1: pbmc1k cellranger matrix**
- **2: pbmc1k cellranger barcodes**
- **3: pbmc1k cellranger features**

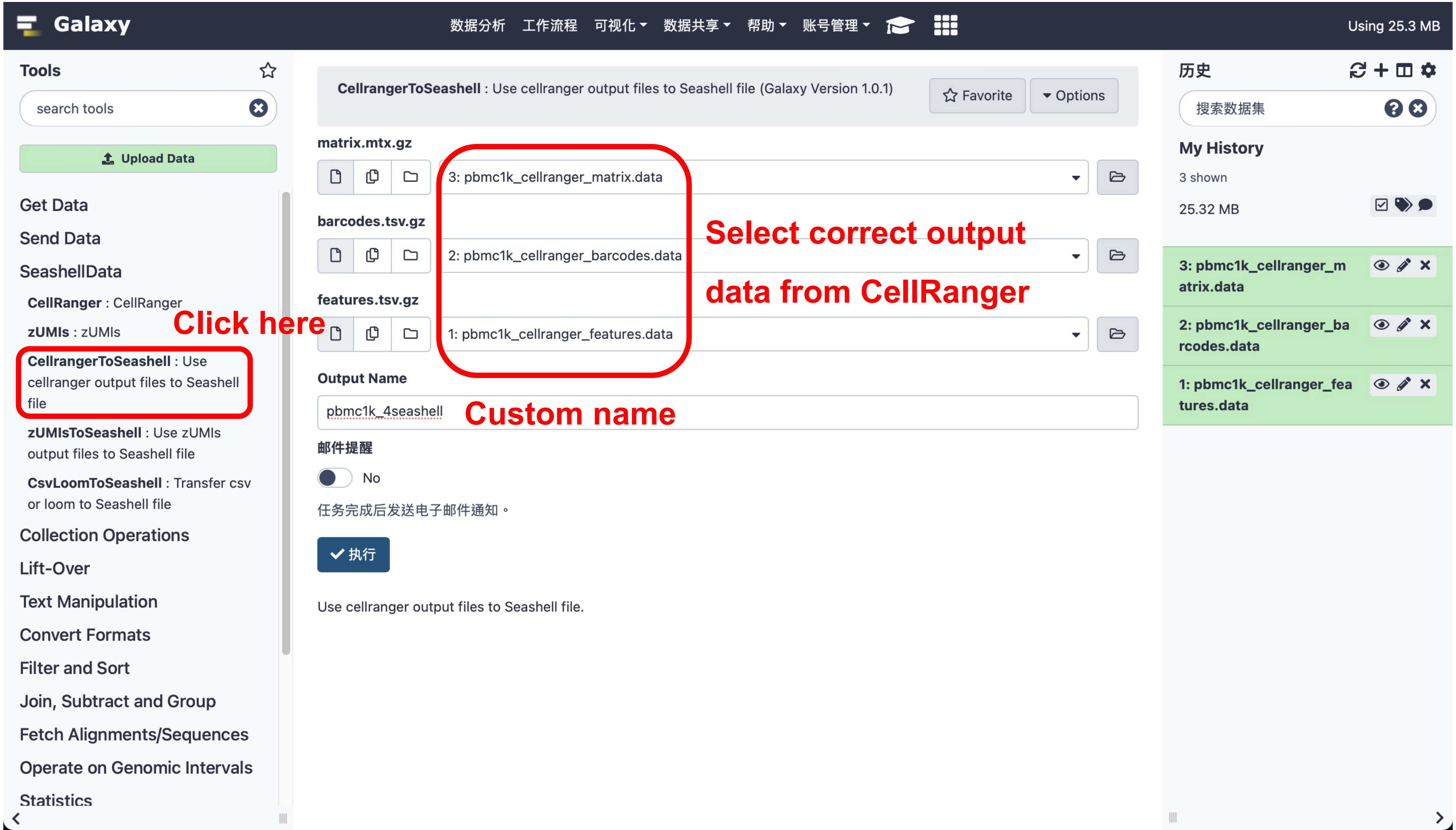
You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

3: pbmc1k cellranger features

2: pbmc1k cellranger barcodes

1: pbmc1k cellranger matrix

4. Run CellRanger to Seashell



CellRanger : CellRanger

zUMIs : zUMIs

CellrangerToSeashell : Use cellranger output files to Seashell file

zUMIsToSeashell : Use zUMIs output files to Seashell file

CsvLoomToSeashell : Transfer csv or loom to Seashell file

Click here

CellrangerToSeashell : Use cellranger output files to Seashell file (Galaxy Version 1.0.1)

matrix.mtx.gz
3: pbmc1k_cellranger_matrix.data

barcodes.tsv.gz
2: pbmc1k_cellranger_barcodes.data

features.tsv.gz
1: pbmc1k_cellranger_features.data

Select correct output data from CellRanger

Output Name
pbmc1k_4seashell

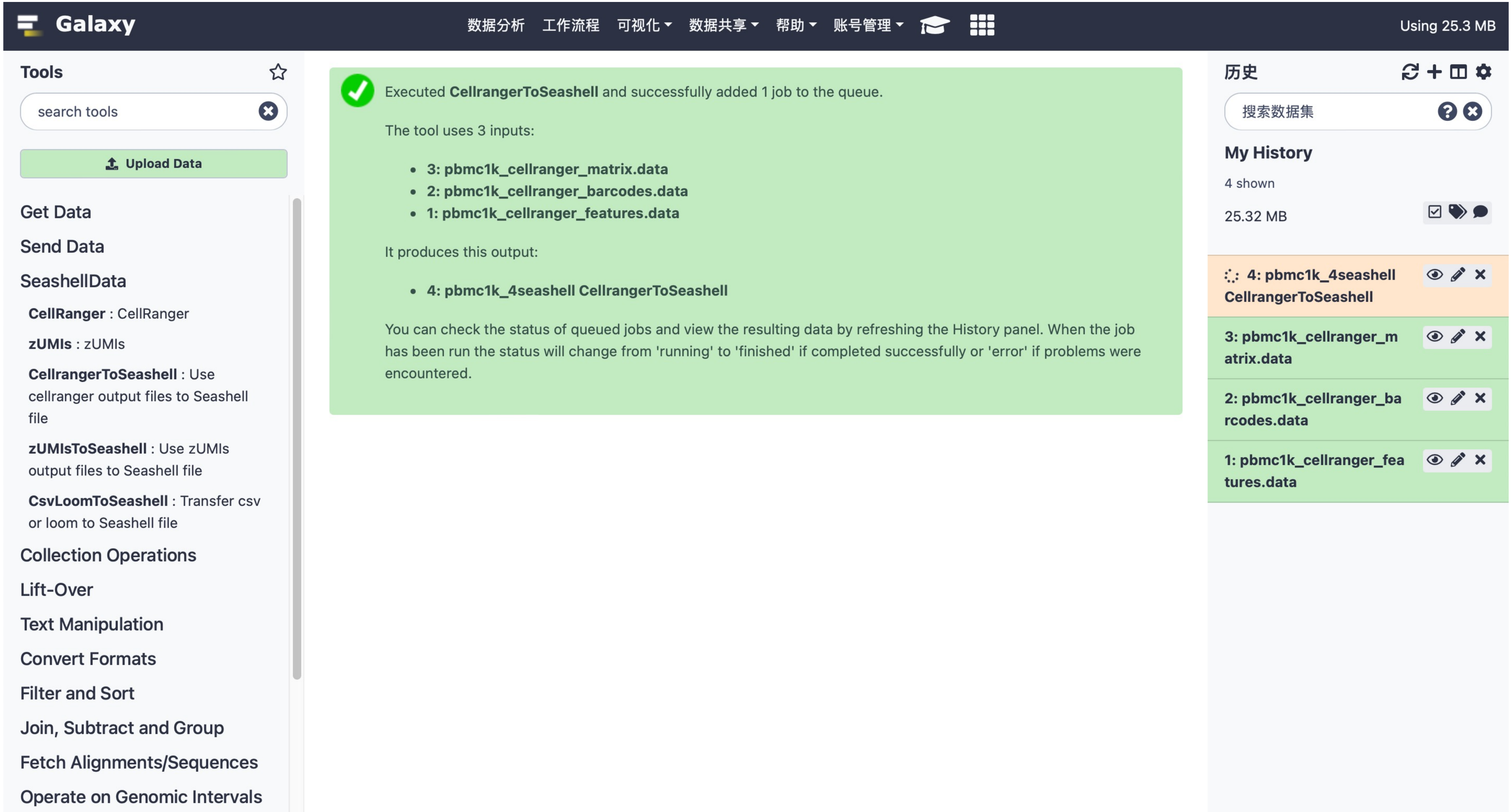
Custom name

邮件提醒
No
任务完成后发送电子邮件通知。

Use cellranger output files to Seashell file.

3: pbmc1k_cellranger_matrix.data
2: pbmc1k_cellranger_barcodes.data
1: pbmc1k_cellranger_features.data

4. Run CellRanger to Seashell



Executed **CellrangerToSeashell** and successfully added 1 job to the queue.

The tool uses 3 inputs:

- **3: pbmc1k_cellranger_matrix.data**
- **2: pbmc1k_cellranger_barcodes.data**
- **1: pbmc1k_cellranger_features.data**

It produces this output:

- **4: pbmc1k_4seashell CellrangerToSeashell**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

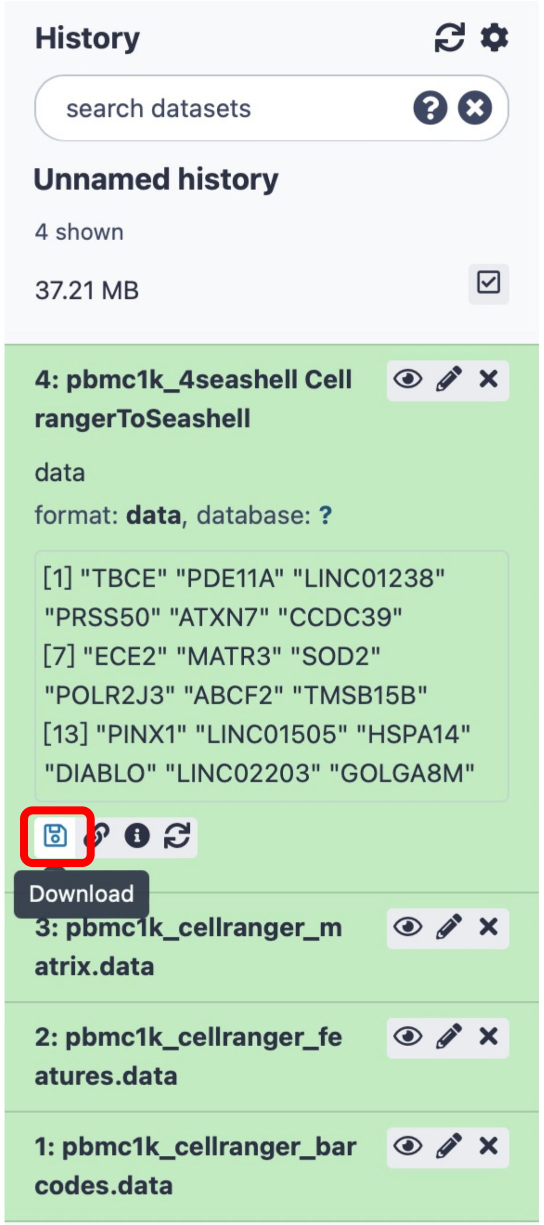
4: pbmc1k_4seashell CellrangerToSeashell

3: pbmc1k_cellranger_matrix.data

2: pbmc1k_cellranger_barcodes.data

1: pbmc1k_cellranger_features.data

5. Download the .rds file for SeaShell_Analyzer



Unnamed history

4 shown

37.21 MB

4: pbmc1k_4seashell CellrangerToSeashell

data

format: **data**, database: ?

```
[1] "TBCE" "PDE11A" "LINC01238"
"PRSS50" "ATXN7" "CCDC39"
[7] "ECE2" "MATR3" "SOD2"
"POLR2J3" "ABCF2" "TMSB15B"
[13] "PINX1" "LINC01505" "HSPA14"
"DIABLO" "LINC02203" "GOLGA8M"
```

3: pbmc1k_cellranger_matrix.data

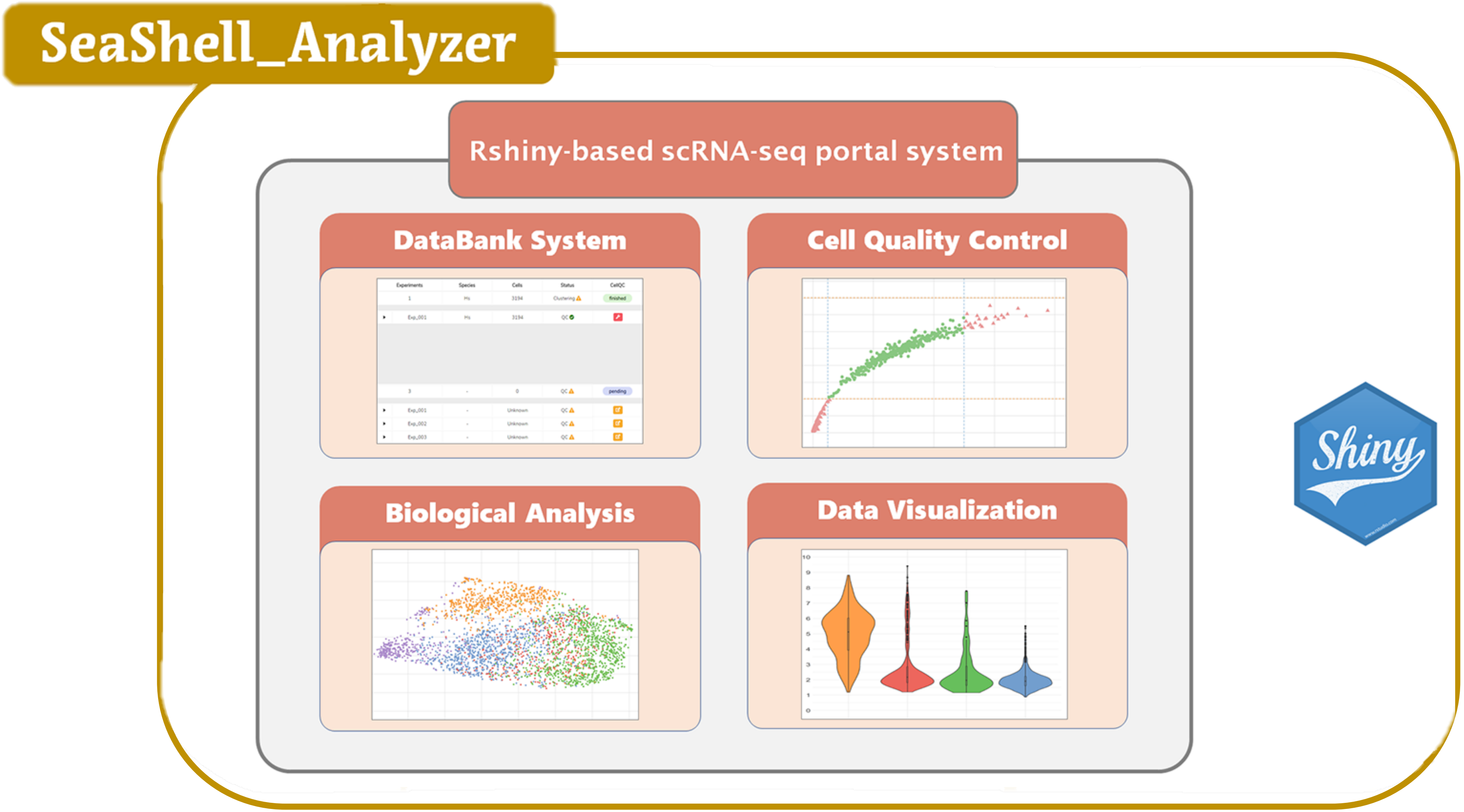
2: pbmc1k_cellranger_features.data

1: pbmc1k_cellranger_barcodes.data

Download

SeaShell_Analyzer

SeaShell_Analyzer



SeaShell_Analyzer

Rshiny-based scRNA-seq portal system

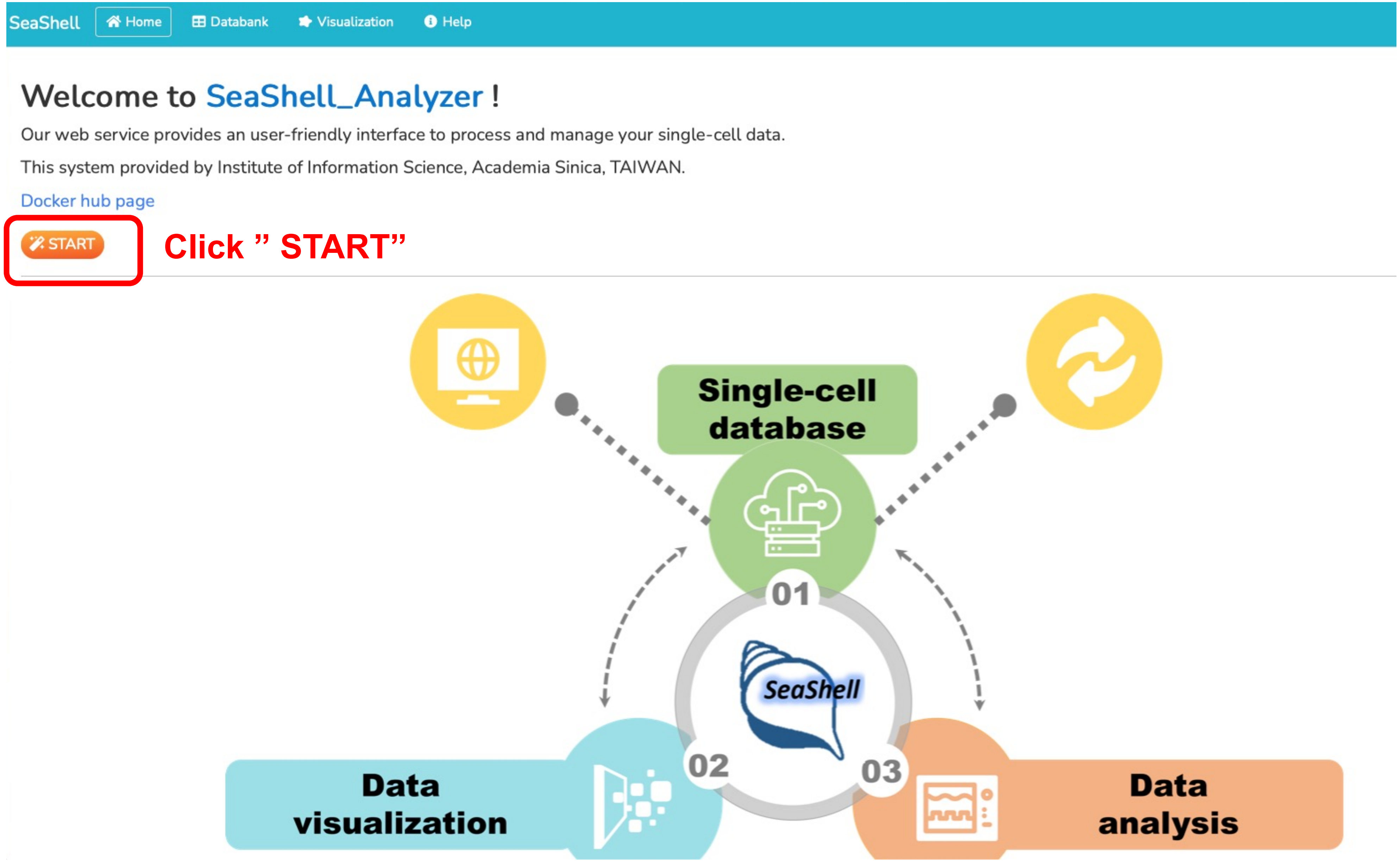
DataBank System

Cell Quality Control

Biological Analysis

Data Visualization

Homepage



Welcome to SeaShell_Analyzer !

Our web service provides an user-friendly interface to process and manage your single-cell data.

This system provided by Institute of Information Science, Academia Sinica, TAIWAN.

Docker hub page

Click " START"

Workflow

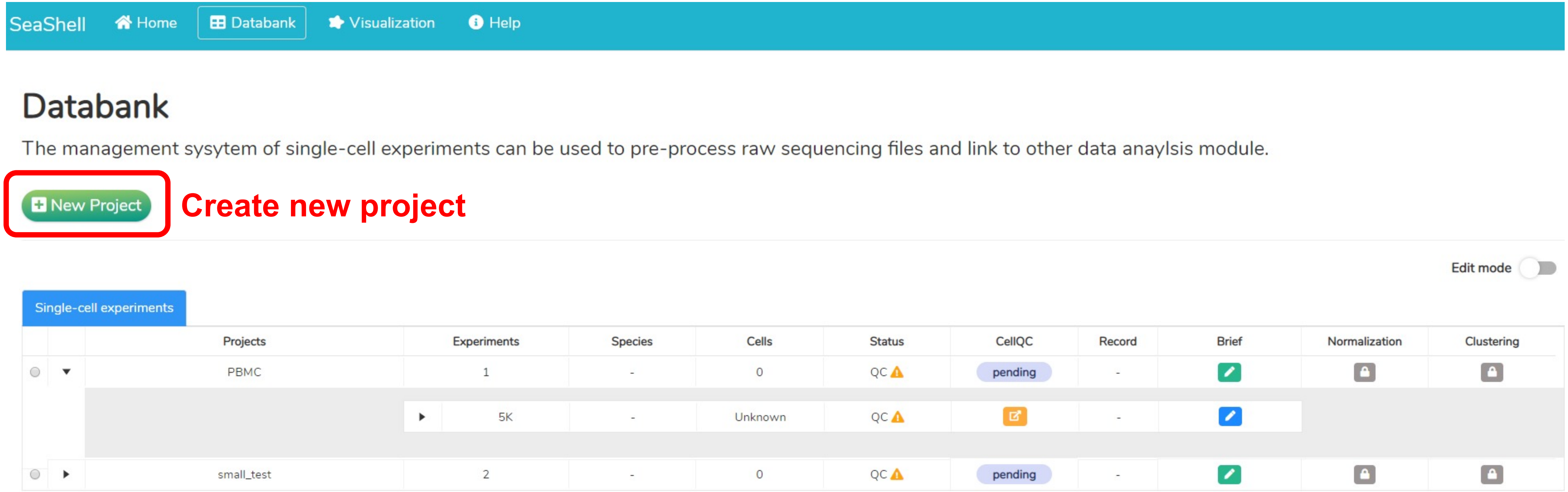
01 Upload rds file to SeaShell

02 Cell QC, Normalization, and Clustering

03 Visualization and Gene analysis

Databank

Create new project



Databank

The management sysytem of single-cell experiments can be used to pre-process raw sequencing files and link to other data anaylsis module.

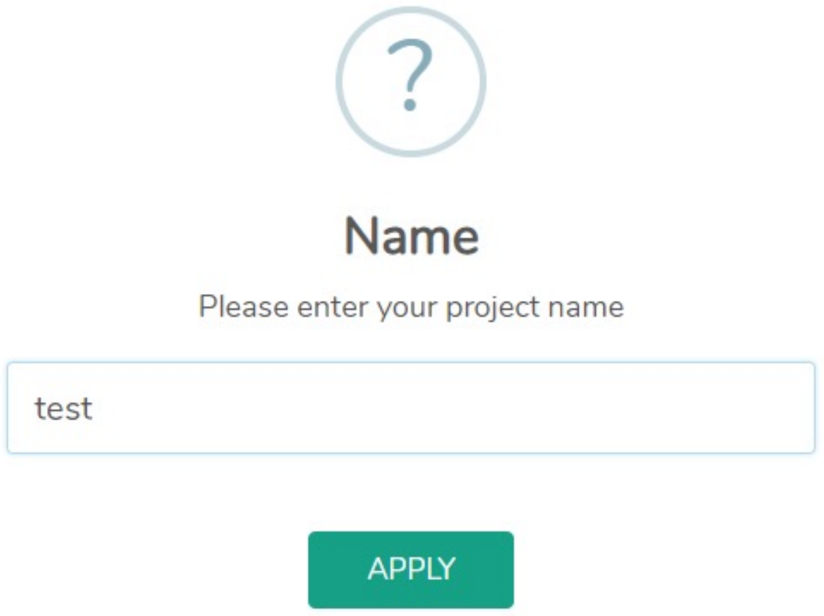
Create new project

		Projects	Experiments	Species	Cells	Status	CellQC	Record	Brief	Normalization	Clustering
○	▾	PBMC	1	-	0	QC ⚠	pending	-	✎	🔒	🔒
		▸	5K	-	Unknown	QC ⚠	↗	-	✎		
○	▸	small_test	2	-	0	QC ⚠	pending	-	✎	🔒	🔒

Databank

Create new project

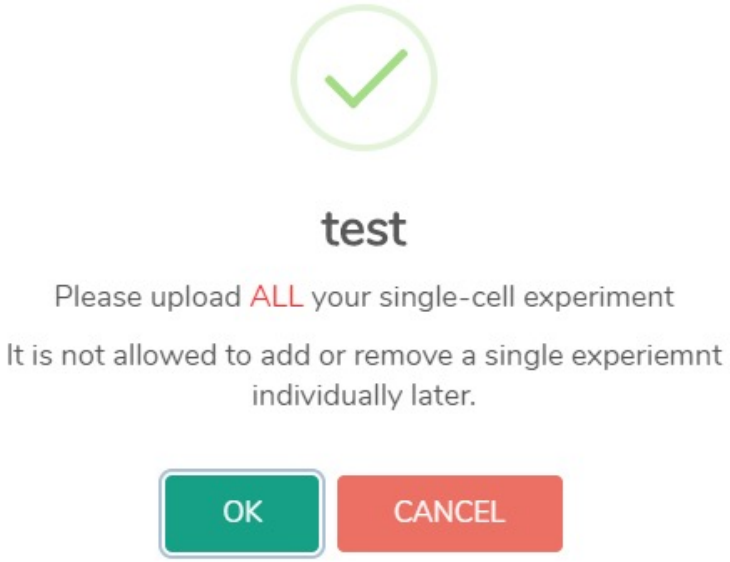
Upload **All single-cell experiments**(rds files) to one project. (For example: Control and Drug experiments)



Name

Please enter your project name

test



test

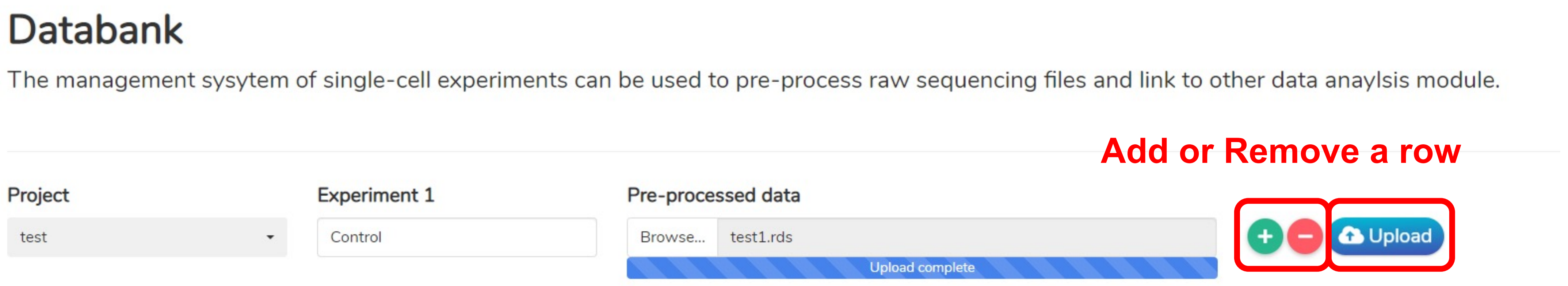
Please upload ALL your single-cell experiment

It is not allowed to add or remove a single experiemnt individually later.

Databank

Create new project

Upload **All single-cell experiments**(rds files) to one project. (For example: Control and Drug experiments)



Databank

The management sysytem of single-cell experiments can be used to pre-process raw sequencing files and link to other data anaylsis module.

Project: test

Experiment 1: Control

Pre-processed data: Browse... test1.rds

Upload complete

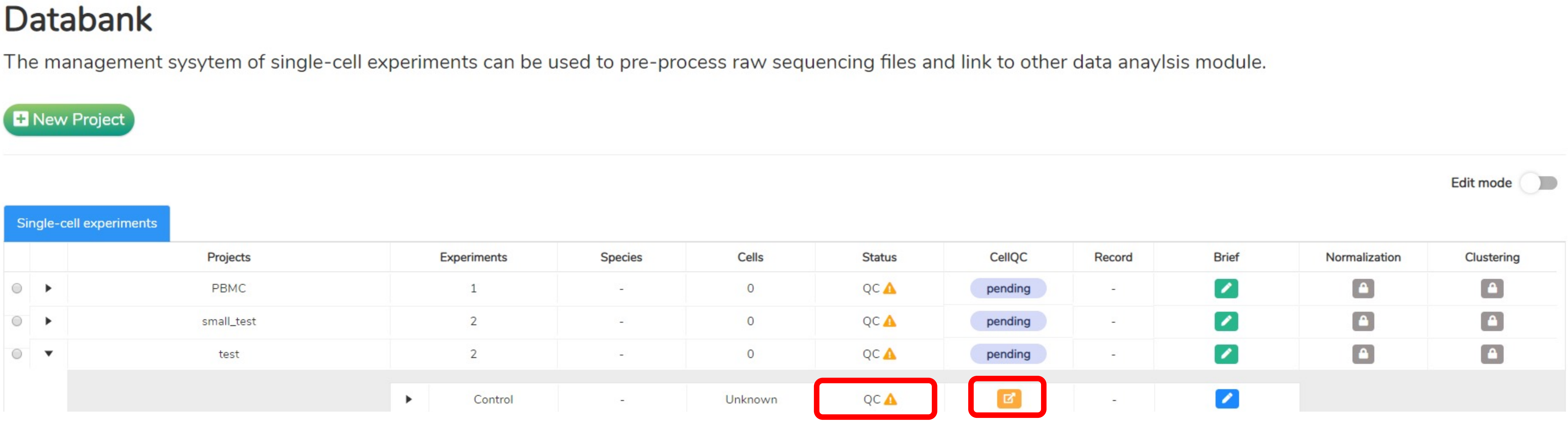
Add or Remove a row

Upload all files

Databank

Create new project

Upload **All single-cell experiments**(rds files) to one project. (For example: Control and Drug experiments)



Databank

The management sysytem of single-cell experiments can be used to pre-process raw sequencing files and link to other data anaylsis module.

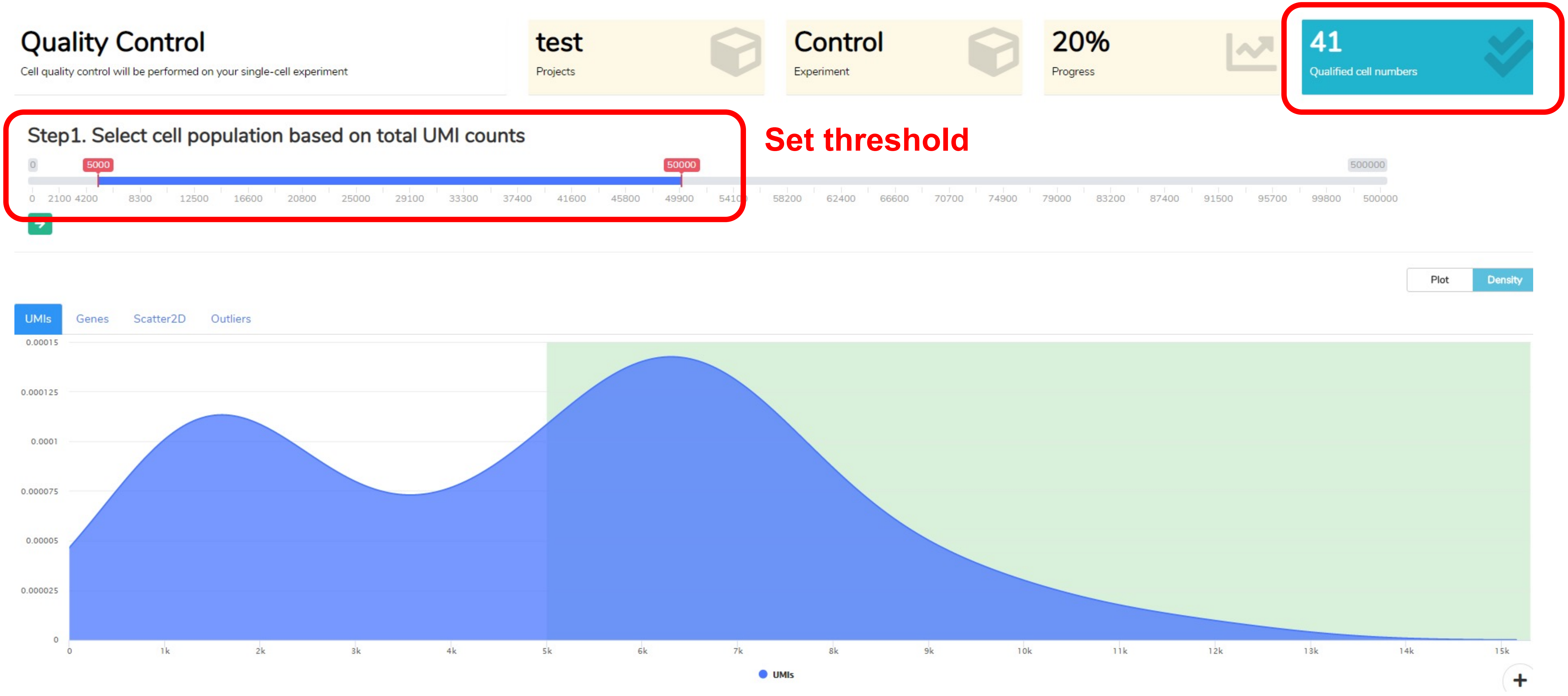
		Projects	Experiments	Species	Cells	Status	CellQC	Record	Brief	Normalization	Clustering
○	▸	PBMC	1	-	0	QC ⚠	pending	-	✎	🔒	🔒
○	▸	small_test	2	-	0	QC ⚠	pending	-	✎	🔒	🔒
○	▾	test	2	-	0	QC ⚠	pending	-	✎	🔒	🔒
			▸ Control	-	Unknown	QC ⚠	✎	-	✎		

Hint

Click to perform cell QC respectively

Cell QC

Real time qualified cell numbers



Quality Control

Cell quality control will be performed on your single-cell experiment

test

Projects

Control

Experiment

20%

Progress

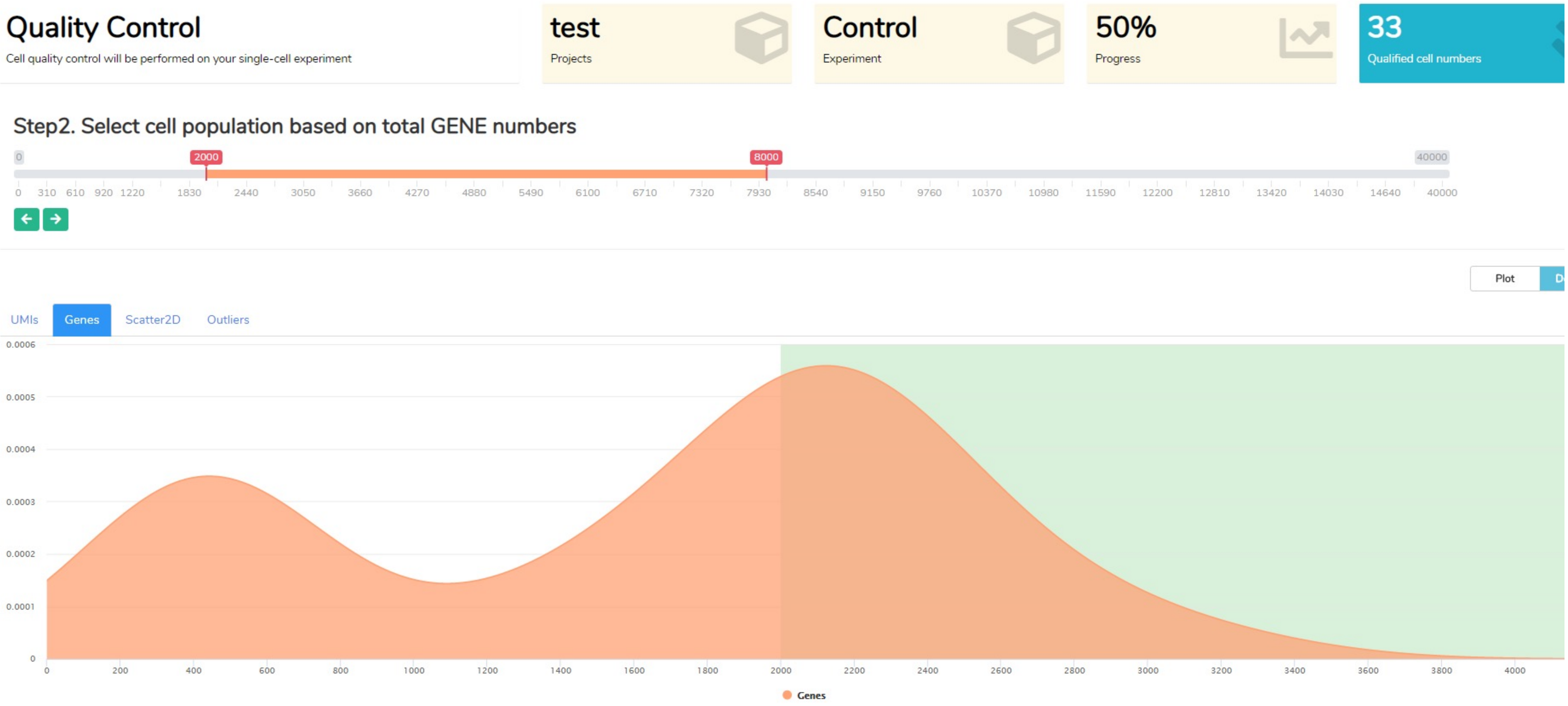
41

Qualified cell numbers

Step1. Select cell population based on total UMI counts

Set threshold

Cell QC



Quality Control

Cell quality control will be performed on your single-cell experiment

test

Projects

Control

Experiment

50%

Progress

33

Qualified cell numbers

Step2. Select cell population based on total GENE numbers

Cell QC

Check by 2D scatter plot

Step3. Check selected cell population

Main population are defined by UMI counts AND gene numbers

Number of observations

Cell QC

Secondary cell QC

Step4. Define outliers by statistical distribution

Outliers are defined as cells certain median absolute deviation (MAD) away from the median in the following distribution.

Save QC result

Cell QC

Databank

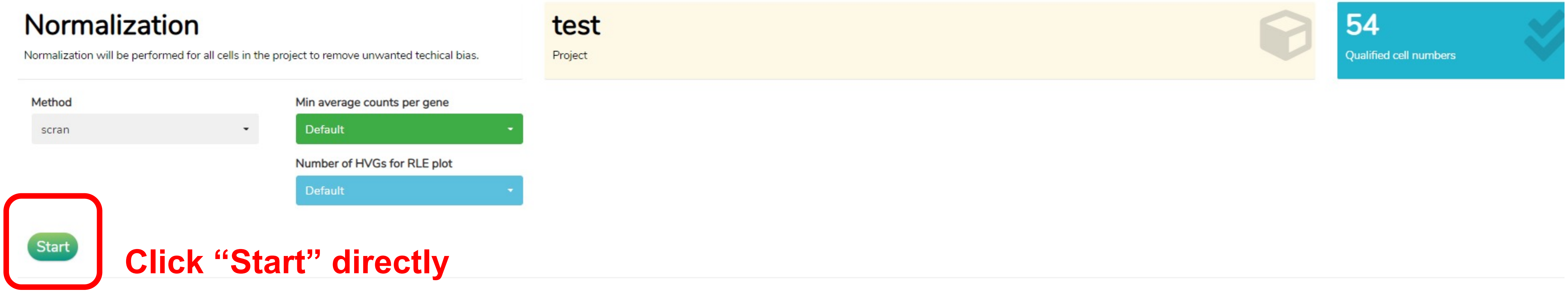
The management sysytem of single-cell experiments can be used to pre-process raw sequencing files and link to other data anaylsis module.

		Projects	Experiments	Species	Cells	Status	CellQC	Record	Brief	Normalization	Clustering
○	▸	PBMC	1	-	0	QC ⚠	pending	-	✎	🔒	🔒
○	▸	small_test	2	-	0	QC ⚠	pending	-	✎	🔒	🔒
○	▾	test	2	Hs	54	Normalization ⚠	finished	-	✎	↗	🔒
			▸ Control	Hs	27	QC ✓	🔧	⤓	✎		

Success and new tips

Perform normalization

Normalization



Normalization

Normalization will be performed for all cells in the project to remove unwanted techical bias.

test

Project

54

Qualified cell numbers

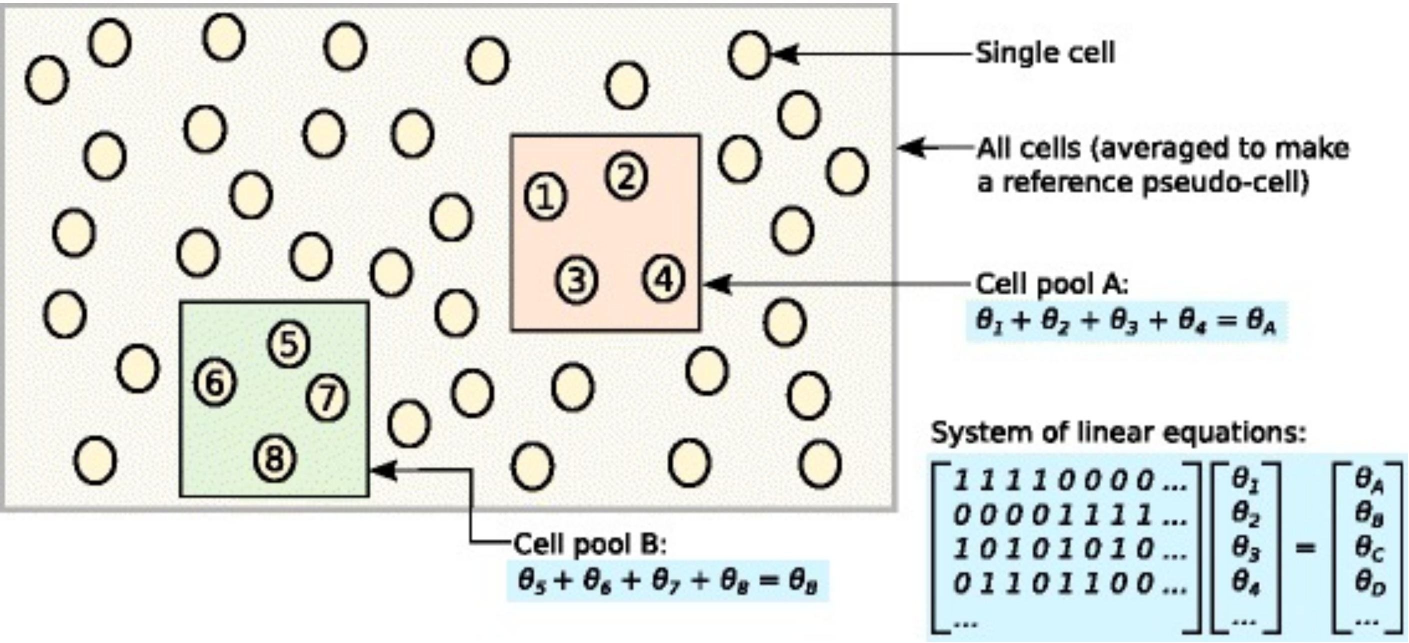
Method: scran

Min average counts per gene: Default

Number of HVGs for RLE plot: Default

Click “Start” directly

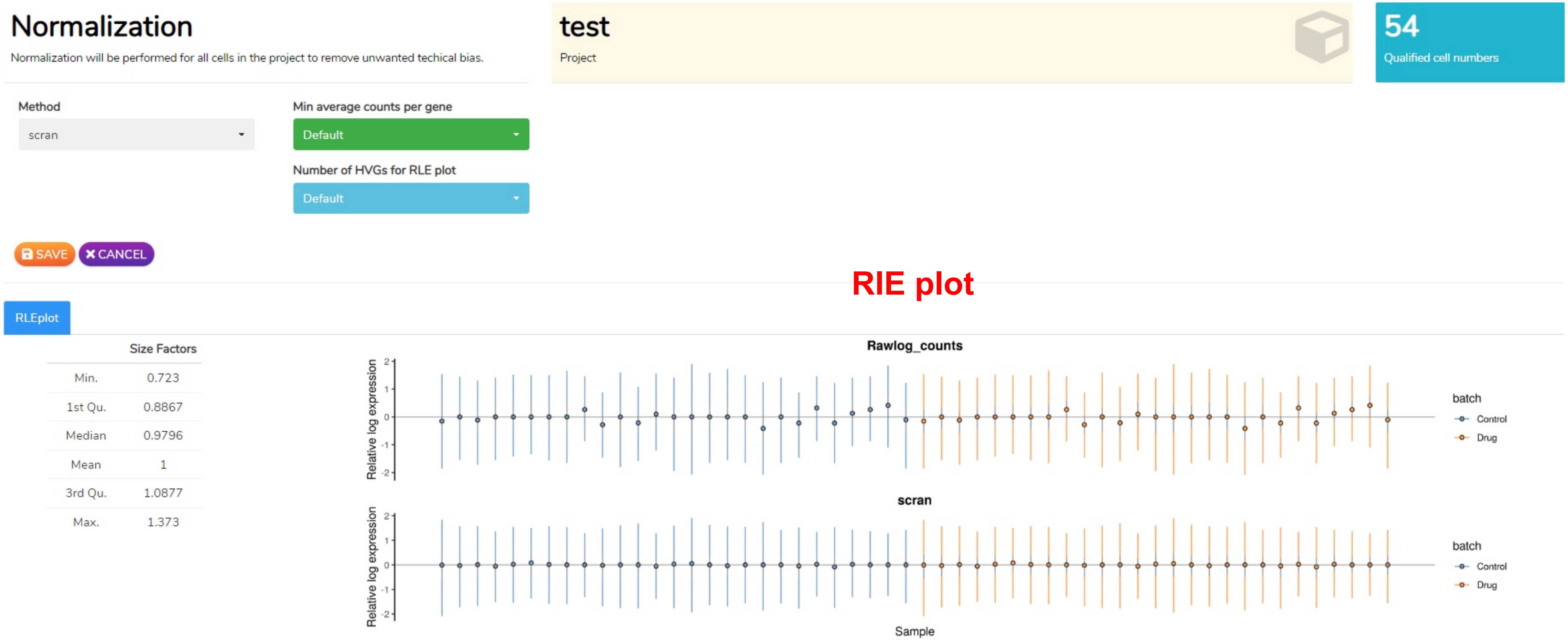
Normalization (scran)



- Defining a pool of cells
- Summing expression values across all cells in the pool
- Normalizing the cell pool against an average reference, using the summed expression values
- Repeating this for many different pools of cells to construct a linear system
- Deconvolving the pool-based size factors to their cell-based counterparts

Aaron T L Lun, *et al. Genome Biol* 2016;17:75. doi: 10.1186/s13059-016-0947-7.

Normalization



Normalization

Normalization will be performed for all cells in the project to remove unwanted techical bias.

test
Project

54
Qualified cell numbers

Method: scran

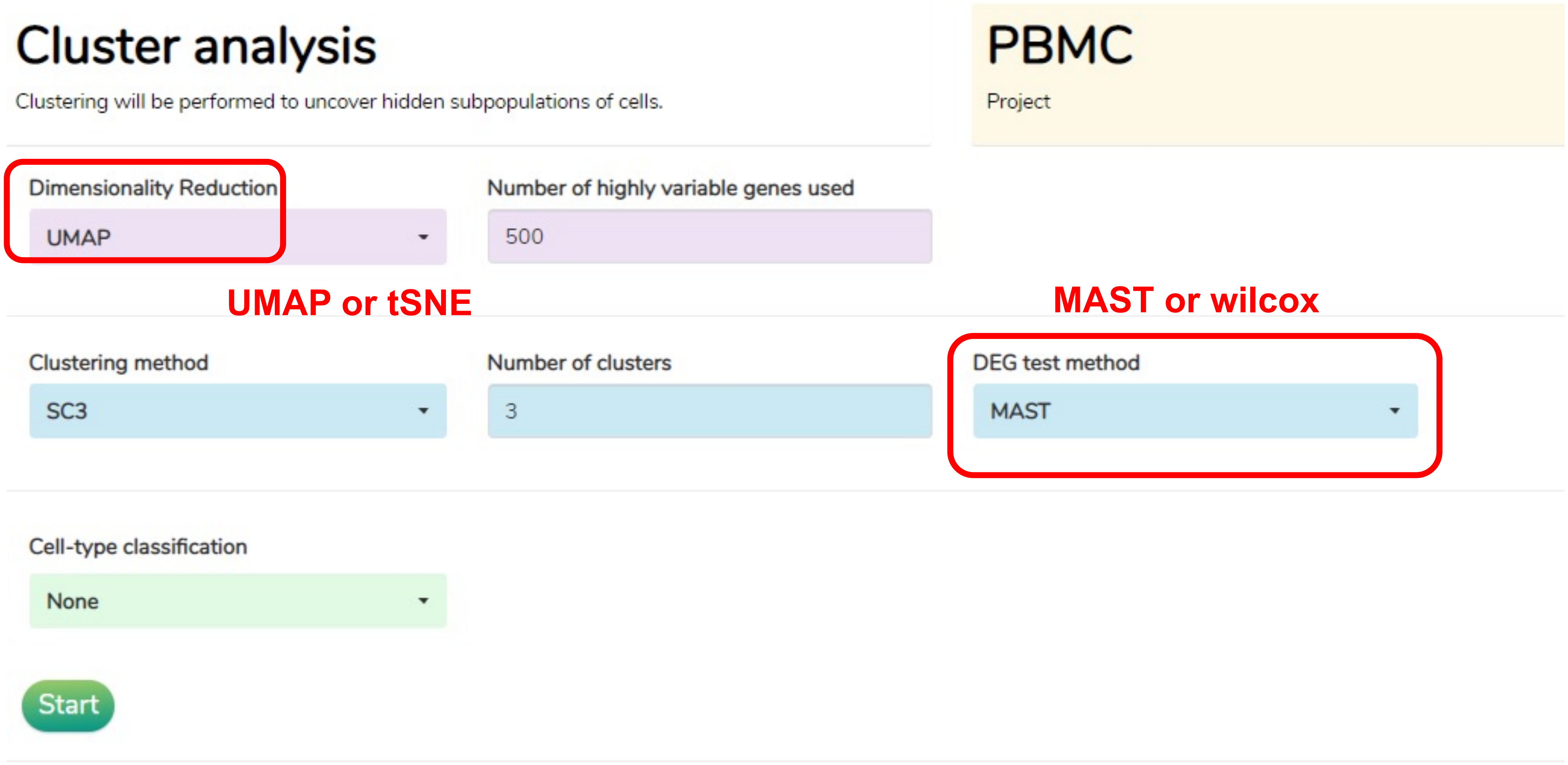
Min average counts per gene: Default

Number of HVGs for RLE plot: Default

RIE plot

	Size Factors
Min.	0.723
1st Qu.	0.8867
Median	0.9796
Mean	1
3rd Qu.	1.0877
Max.	1.373

Clustering



Cluster analysis

Clustering will be performed to uncover hidden subpopulations of cells.

PBMC

Project

Dimensionality Reduction: UMAP

Number of highly variable genes used: 500

Clustering method: SC3

Number of clusters: 3

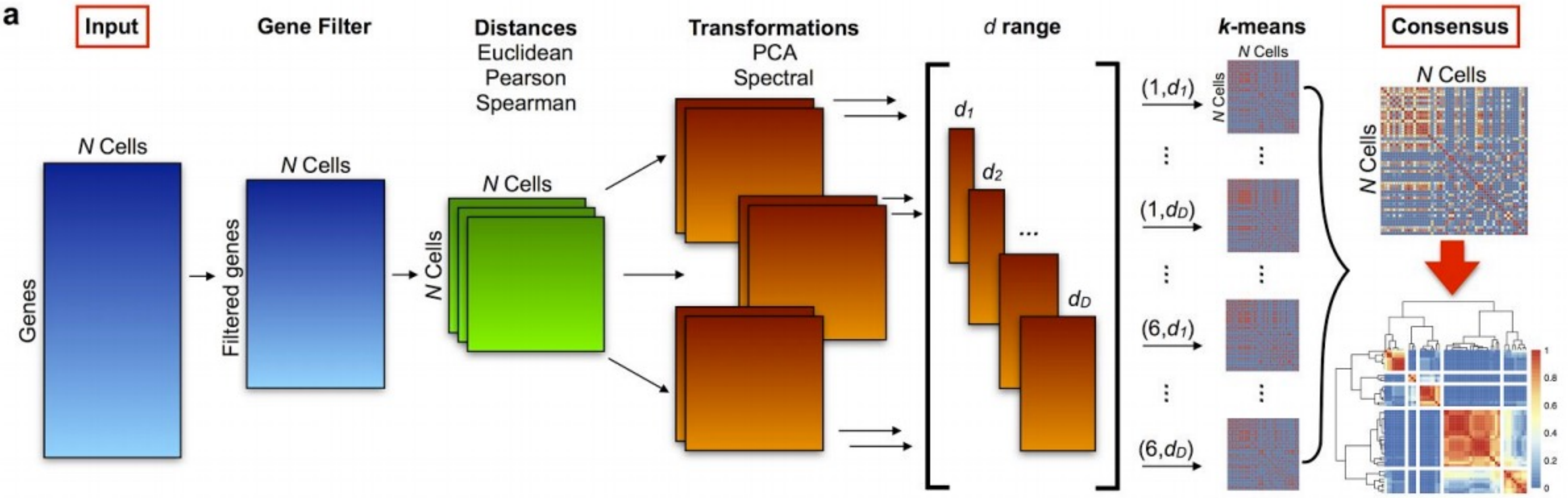
DEG test method: MAST

Cell-type classification: None

UMAP or tSNE

MAST or wilcox

Clustering (SC3)

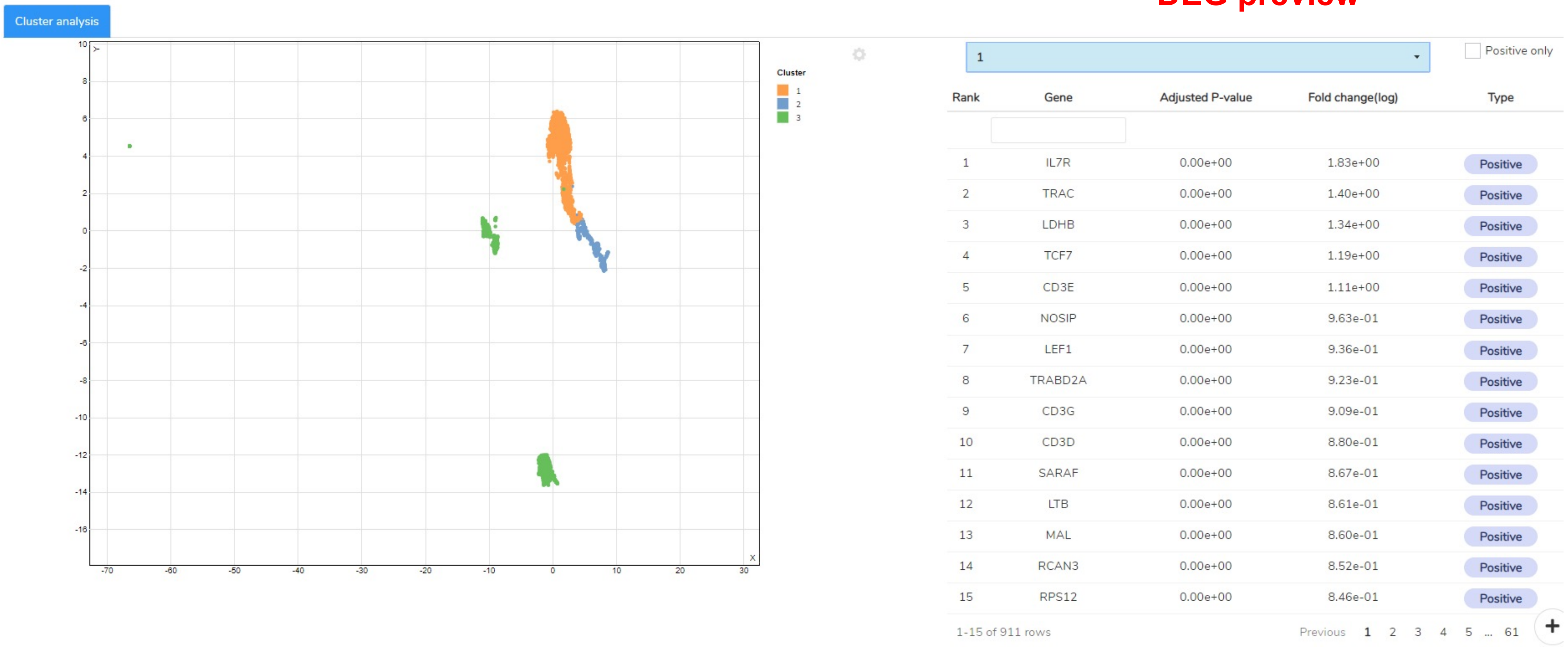


- PCA and spectral dimensionality reductions
- Use k-means
- Consensus clustering

Vladimir Yu Kiselev, *et al. Nat Methods* 2017;14:483-486. doi:10.1038/nmeth.4236.

Clustering

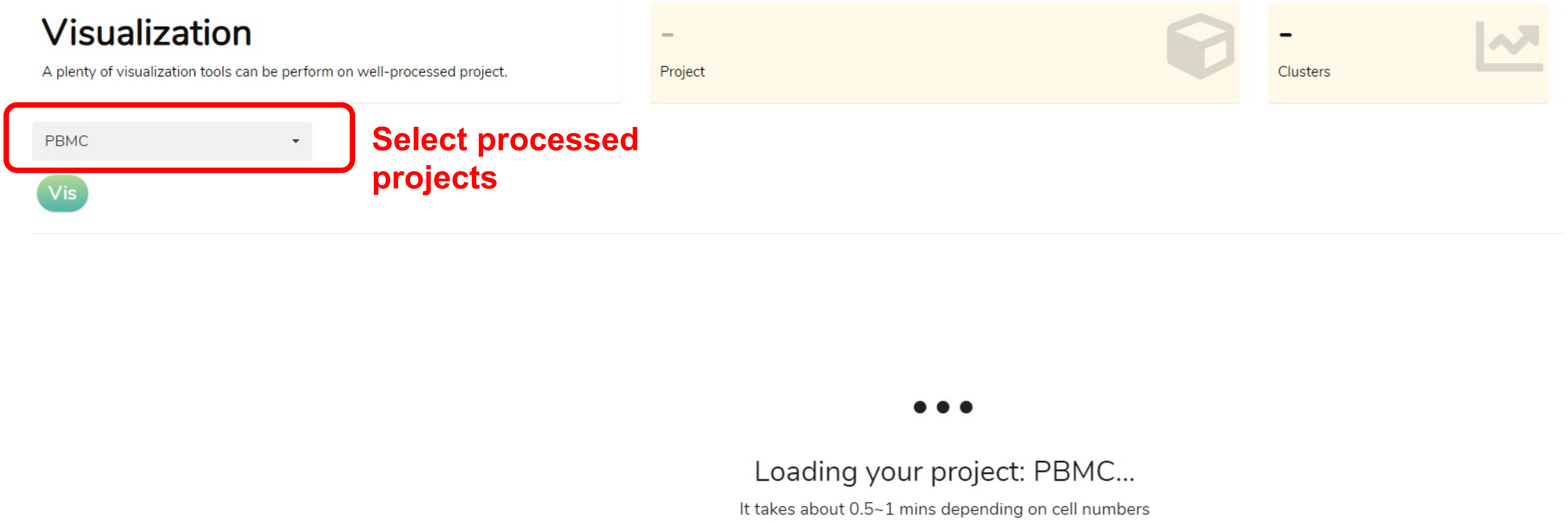
DEG preview



Rank	Gene	Adjusted P-value	Fold change(log)	Type
1	IL7R	0.00e+00	1.83e+00	Positive
2	TRAC	0.00e+00	1.40e+00	Positive
3	LDHB	0.00e+00	1.34e+00	Positive
4	TCF7	0.00e+00	1.19e+00	Positive
5	CD3E	0.00e+00	1.11e+00	Positive
6	NOSIP	0.00e+00	9.63e-01	Positive
7	LEF1	0.00e+00	9.36e-01	Positive
8	TRABD2A	0.00e+00	9.23e-01	Positive
9	CD3G	0.00e+00	9.09e-01	Positive
10	CD3D	0.00e+00	8.80e-01	Positive
11	SARAF	0.00e+00	8.67e-01	Positive
12	LTB	0.00e+00	8.61e-01	Positive
13	MAL	0.00e+00	8.60e-01	Positive
14	RCAN3	0.00e+00	8.52e-01	Positive
15	RPS12	0.00e+00	8.46e-01	Positive

1-15 of 911 rows

Visualization



Visualization

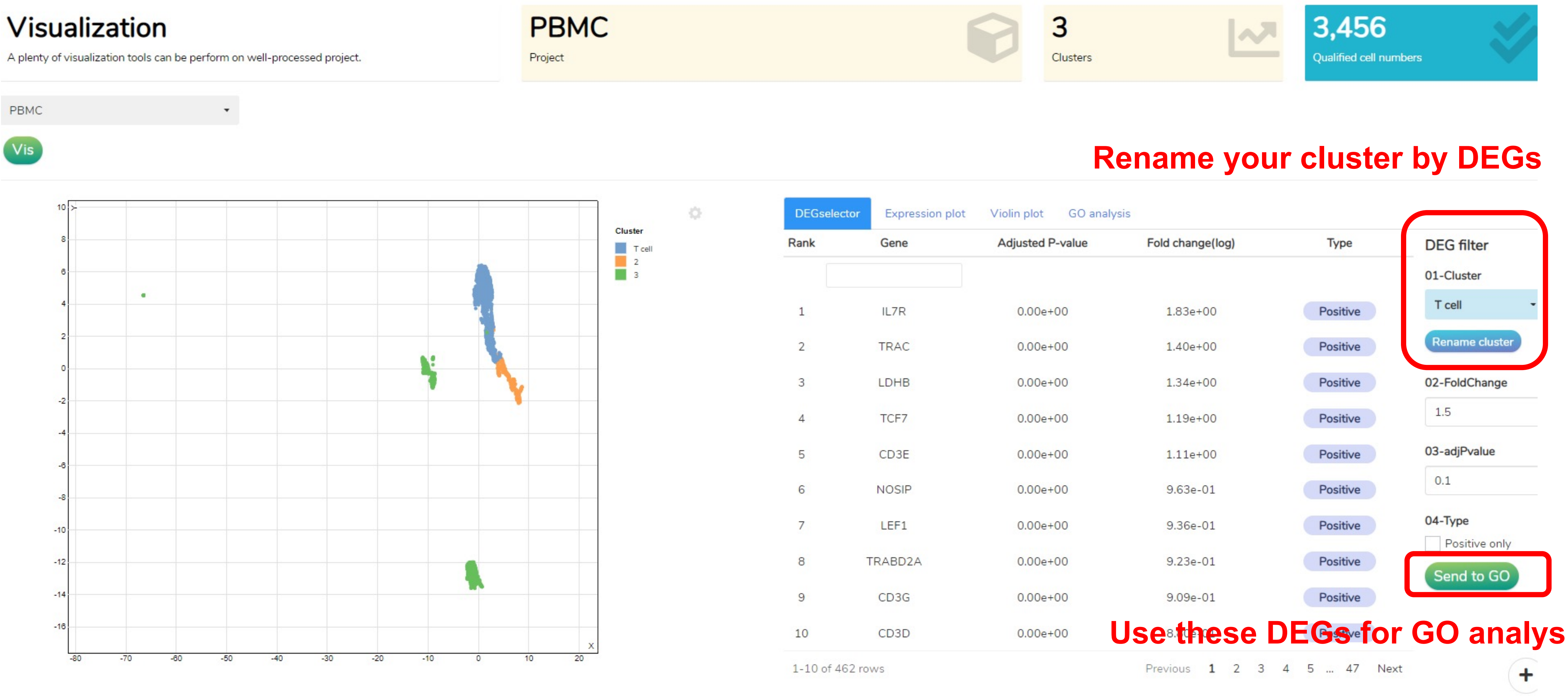
A plenty of visualization tools can be perform on well-processed project.

Loading your project: PBMC...

It takes about 0.5~1 mins depending on cell numbers

Select processed projects

Visualization



Visualization

A plenty of visualization tools can be perform on well-processed project.

PBMC
Project

3
Clusters

3,456
Qualified cell numbers

PBMC

Rename your cluster by DEGs

Rank	Gene	Adjusted P-value	Fold change(log)	Type
1	IL7R	0.00e+00	1.83e+00	Positive
2	TRAC	0.00e+00	1.40e+00	Positive
3	LDHB	0.00e+00	1.34e+00	Positive
4	TCF7	0.00e+00	1.19e+00	Positive
5	CD3E	0.00e+00	1.11e+00	Positive
6	NOSIP	0.00e+00	9.63e-01	Positive
7	LEF1	0.00e+00	9.36e-01	Positive
8	TRABD2A	0.00e+00	9.23e-01	Positive
9	CD3G	0.00e+00	9.09e-01	Positive
10	CD3D	0.00e+00	8.80e-01	Positive

1-10 of 462 rows

DEG filter

01-Cluster: T cell

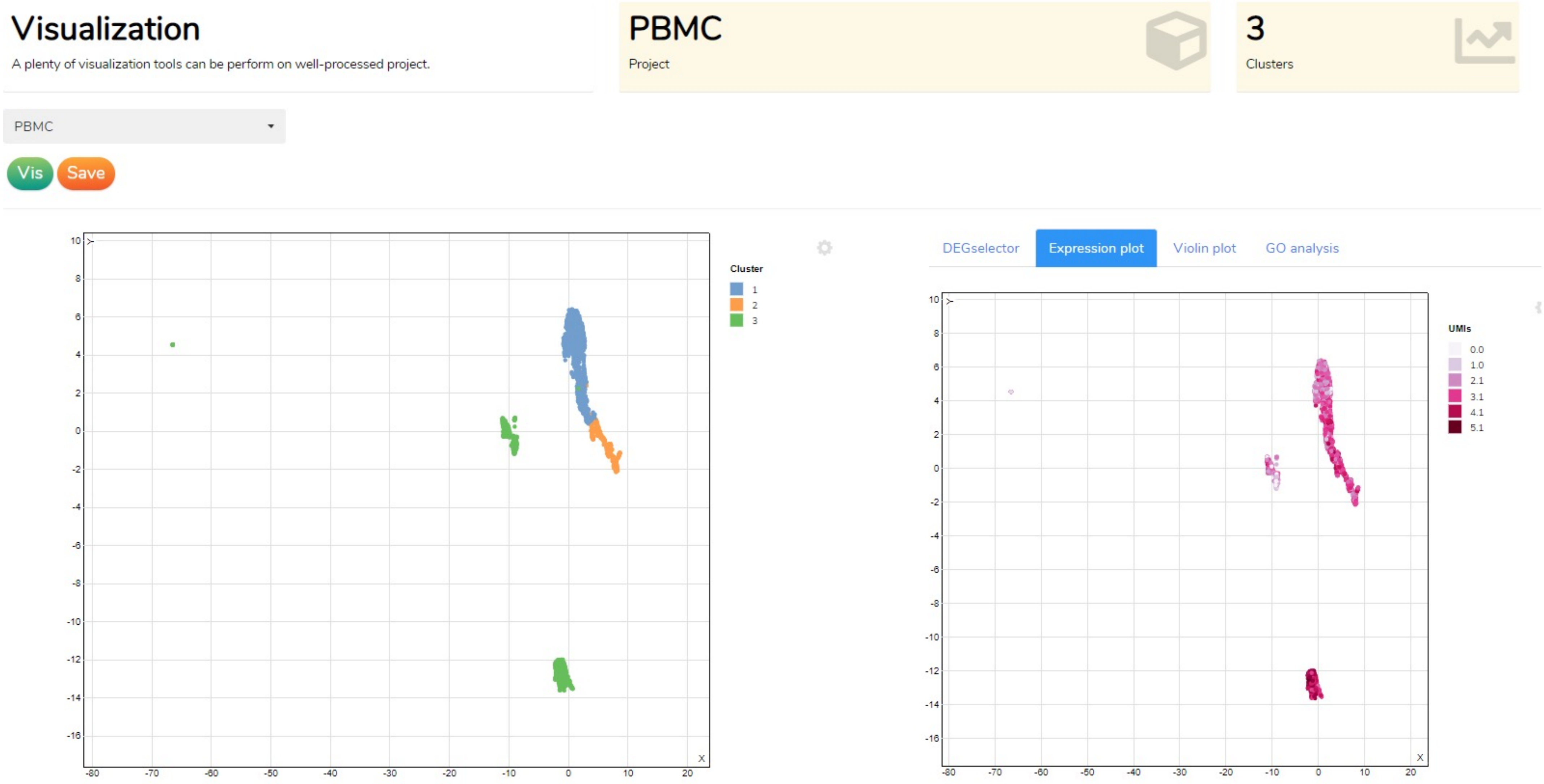
02-FoldChange: 1.5

03-adjPvalue: 0.1

04-Type: ☐ Positive only

Use these DEGs for GO analysis

Visualization



Visualization

A plenty of visualization tools can be perform on well-processed project.

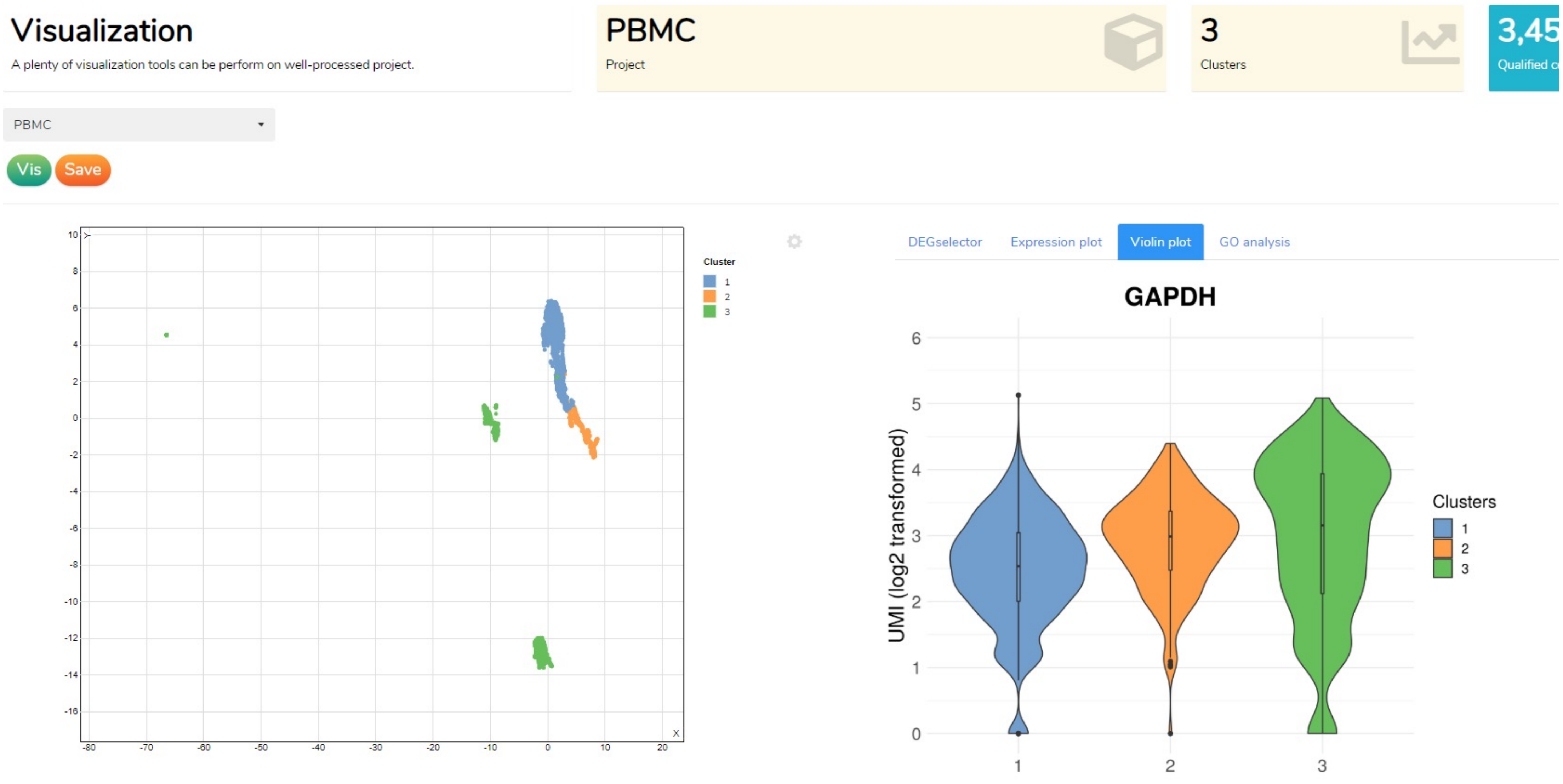
PBMC

Project

3

Clusters

Visualization



Visualization

A plenty of visualization tools can be perform on well-processed project.

PBMC

Project

3

Clusters

3,45

Qualified c

Visualization

Visualization

A plenty of visualization tools can be perform on well-processed project.

PBMC
Project

3
Clusters

3,456
Qualified cell numbers

Contact

Lab of Systems and Network Biology

中央研究院 資訊科學研究所 系統生物學暨網路生物學實驗室
主持人　林仲彥　研究員