2020 生物醫學大數據淘金工作坊 Part 3
Web portal for Single cell RNA-seq

Presenter:
中研院資訊所　林仲彥老師實驗室
余柏毅

2020.04

Outline

01 **Introduction**
single-cell RNA sequencing

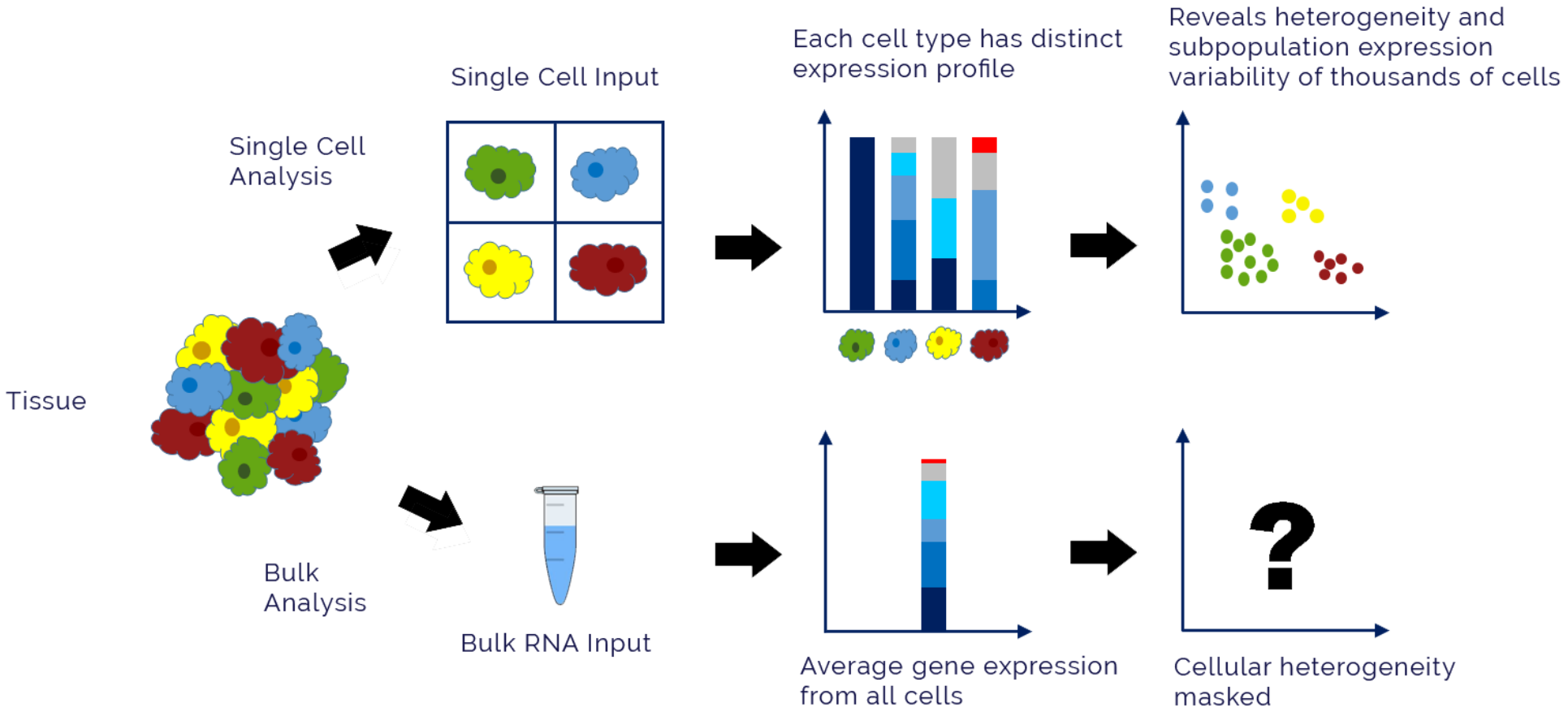
02 **Galaxy-based pre-processing system**

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

Introduction

Single-cell RNA sequencing (scRNA-seq)

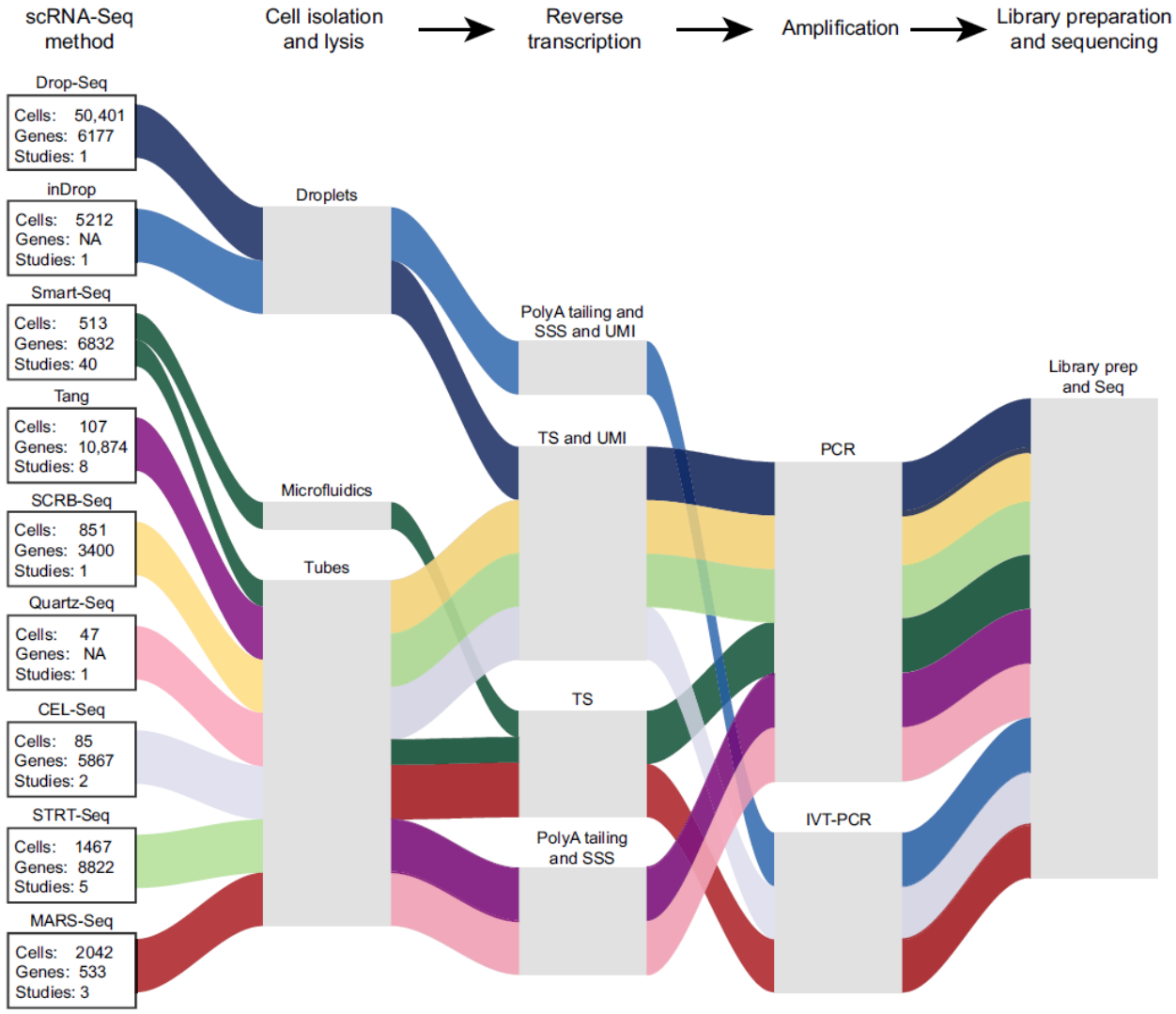
scRNA-seq technologies have combined effective single-cell isolation strategies with highly sensitive molecule detection approaches, showing promise in unravelling the heterogeneity of complex tissues or organs.



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

Single-cell RNA sequencing

A variety of scRNA-seq protocols

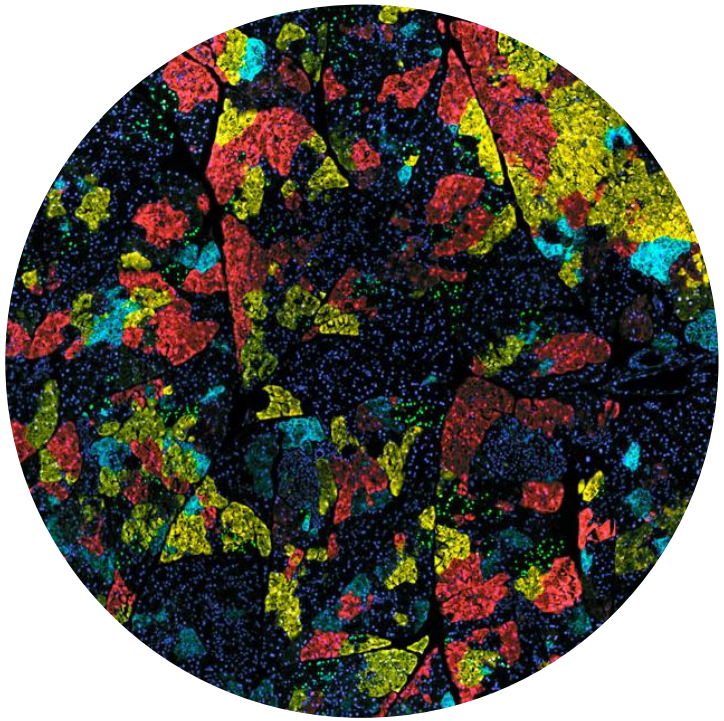


Kumar, Pavithra, Yuqi Tan, and Patrick Cahan. "Understanding development and stem cells using single cell-based analyses of gene expression." *Development* 144.1 (2017): 17-32.

Single-cell RNA sequencing

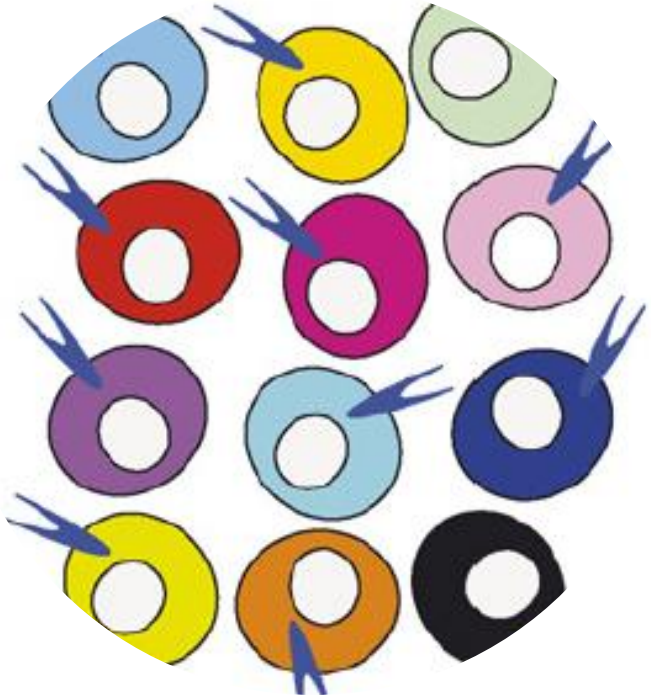
A growing number of application

Heterogenity of cancer cells



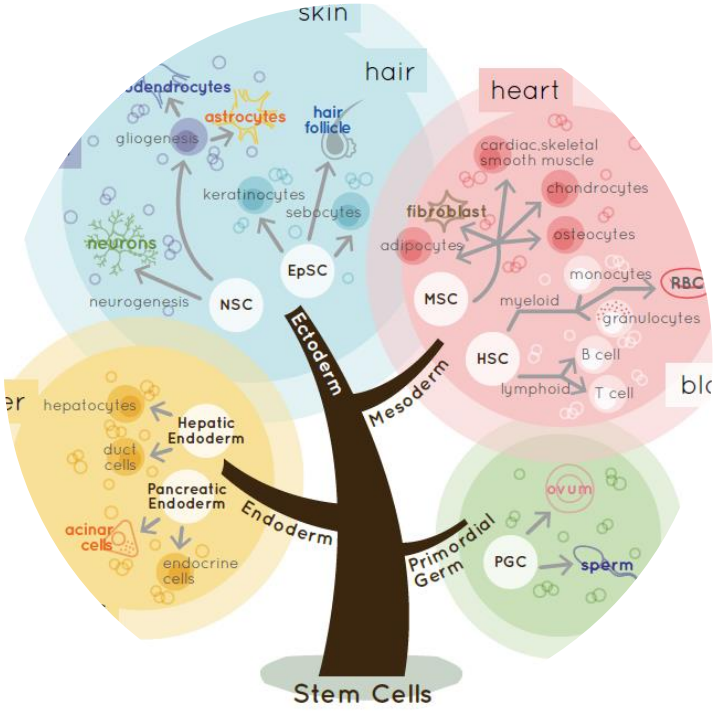
https://www.rna-seqblog.com/new-single-cell-rna-sequencing-methods-could-lead-to-better-regenerative-therapies/

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

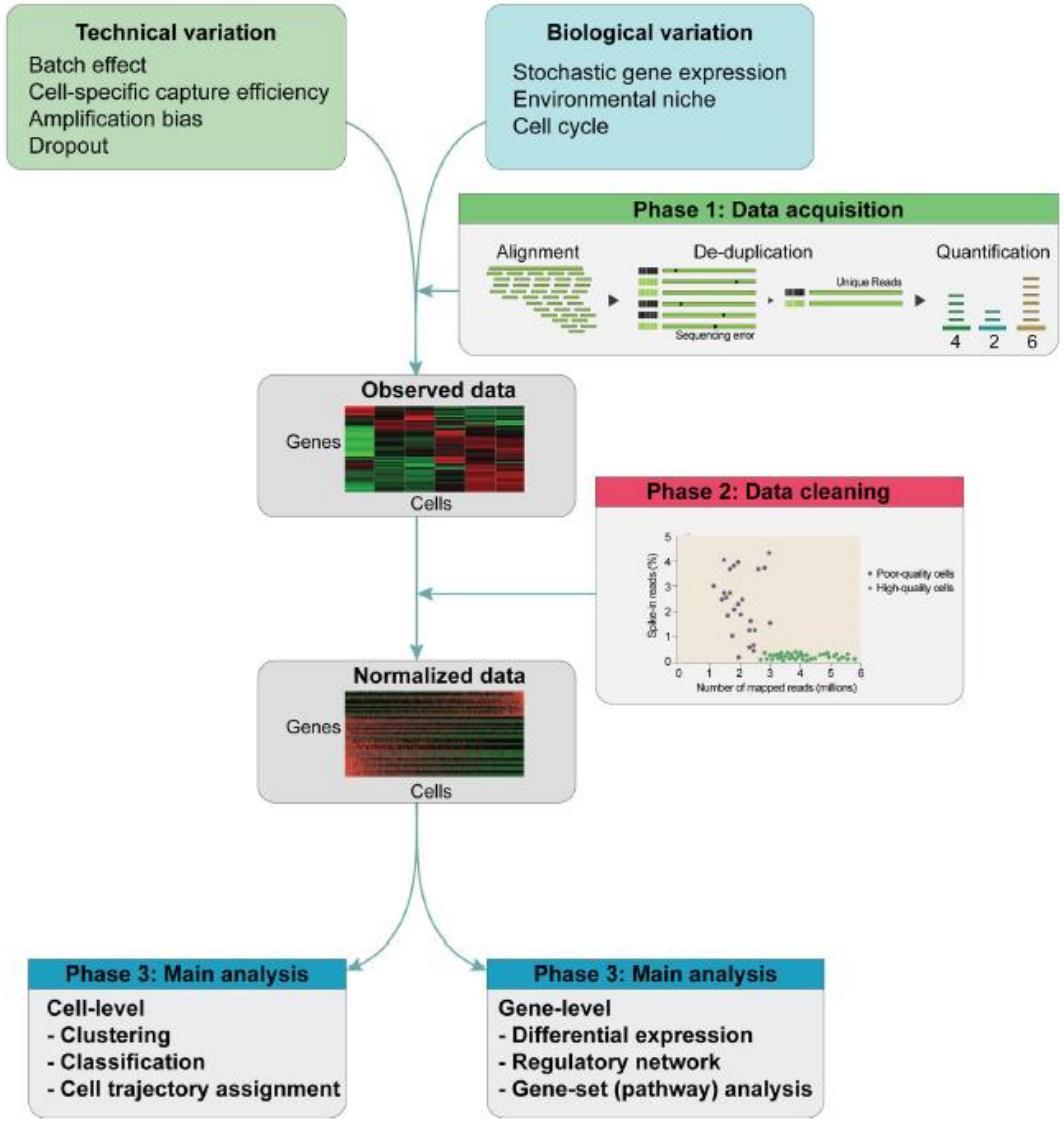


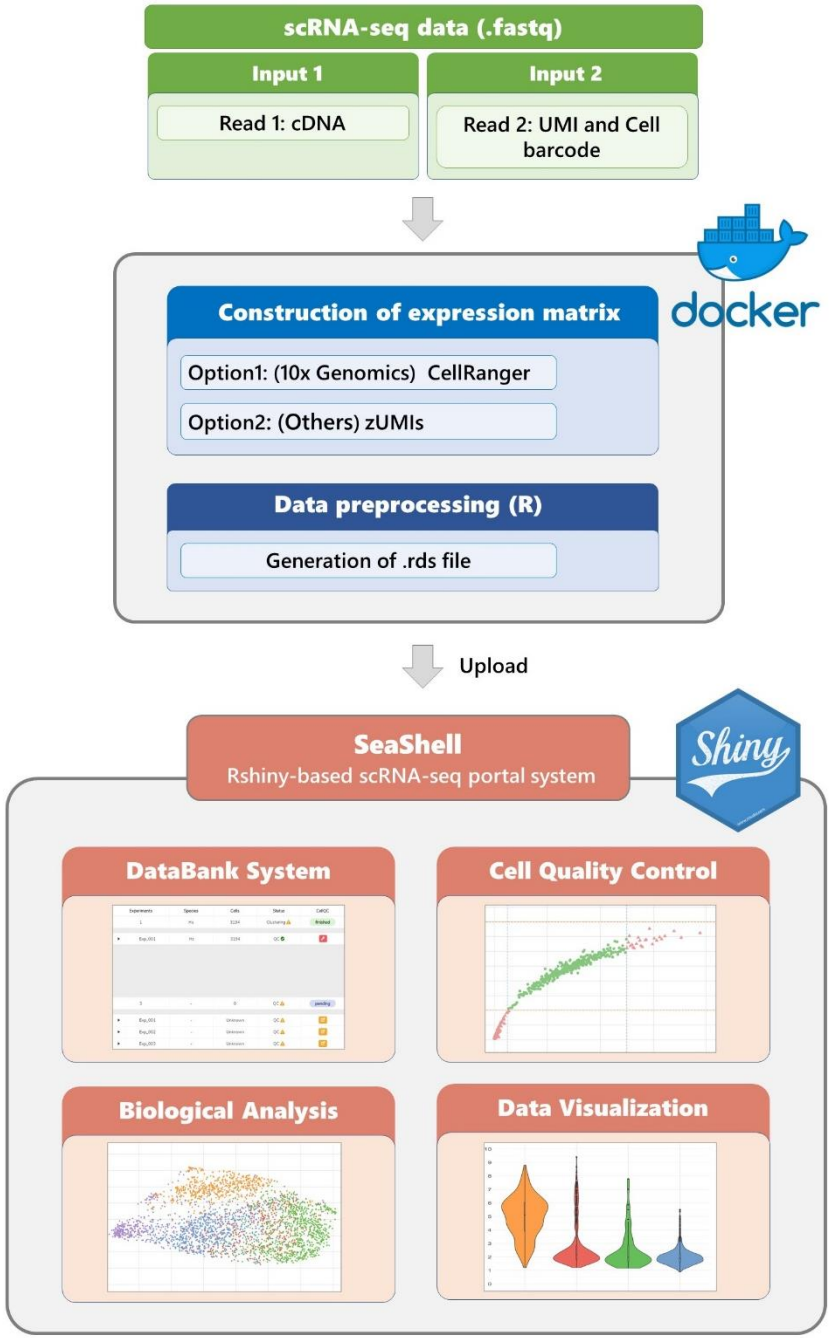
https://www.bio-connect.nl/stem-cell-and-the-regenerative-medicine-ready-for-the-patients/cnt/page/5050

Single-cell RNA sequencing

Increasing difficulties of processing sequencing data

- Serious batch effect
- Demanding computing power
- etc





Workflow of analysis

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**

Galaxy-based pre-processing system

1. Start the galaxy web server

01

Download docker image and Create a directory first for data acquisition

$ docker pull lsbnb/galaxy_sc
$ mkdir YOUR_DIR

02

Run the docker

# port can be changed
# “YOUR_DIR” will be the directory shared with docker container
$ docker run -d -t -i -p 8081:80 –v \
('pwd')/YOUR_DIR/:/root/galaxy/database/ftp/rnaanalysis@galaxy.org/ \
lsbnb/galaxy_sc /bin/bash

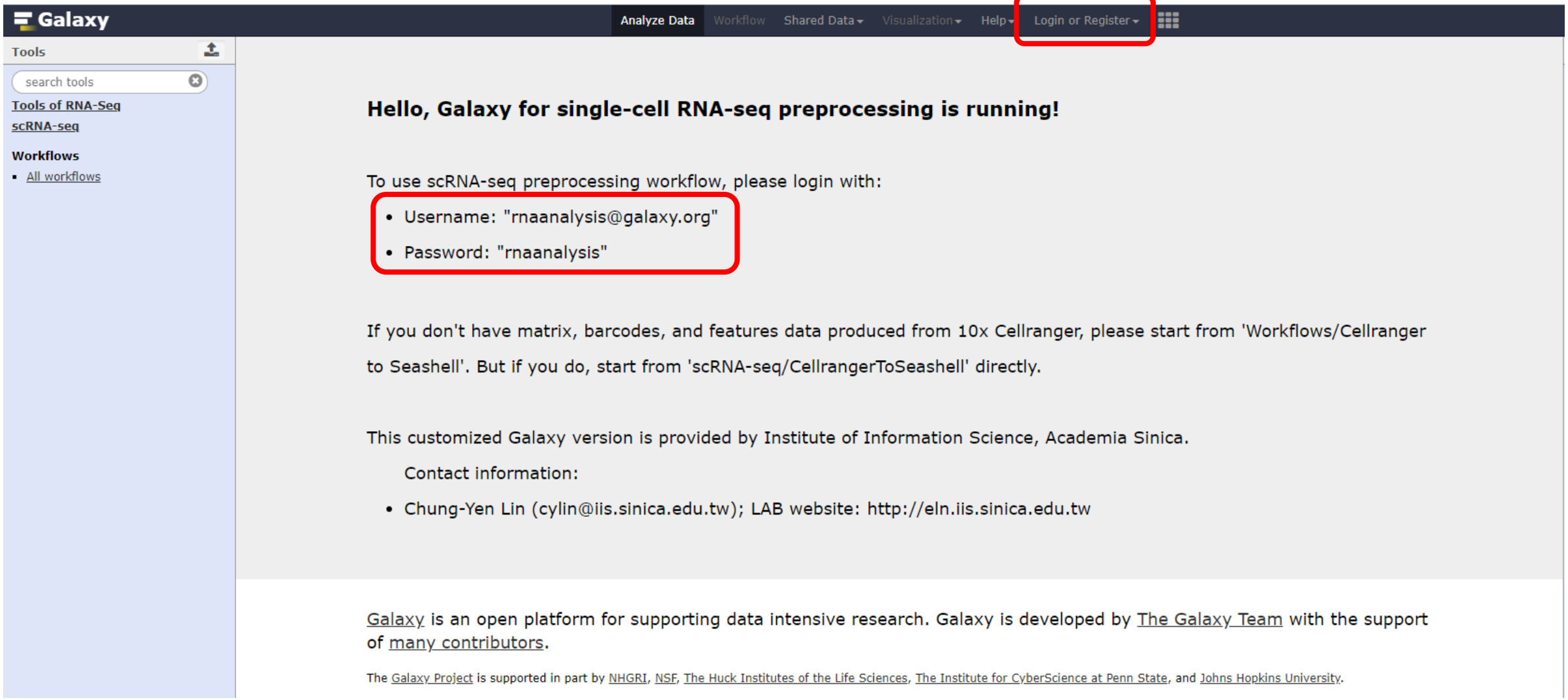
03

Access to the galaxy web server

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

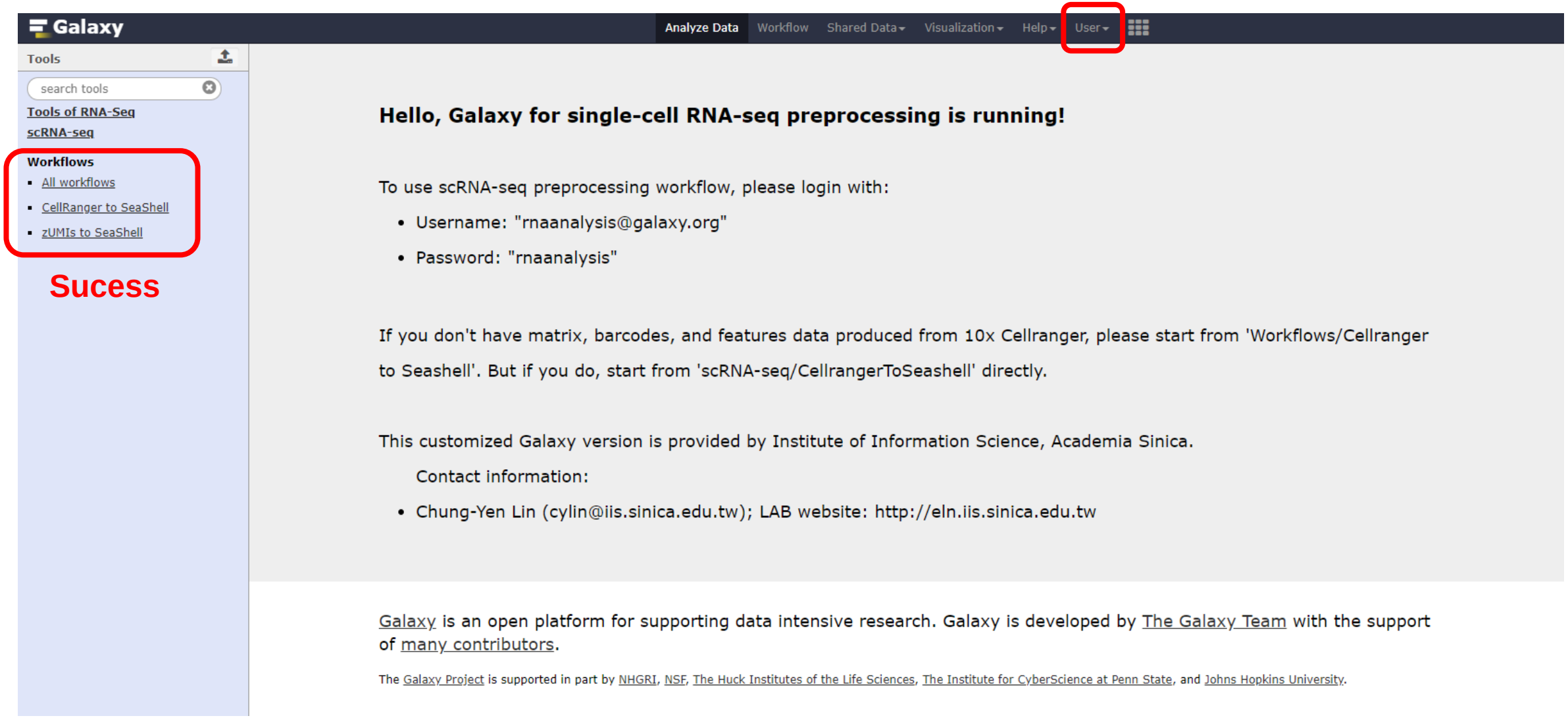
1. Start the galaxy web server

Login

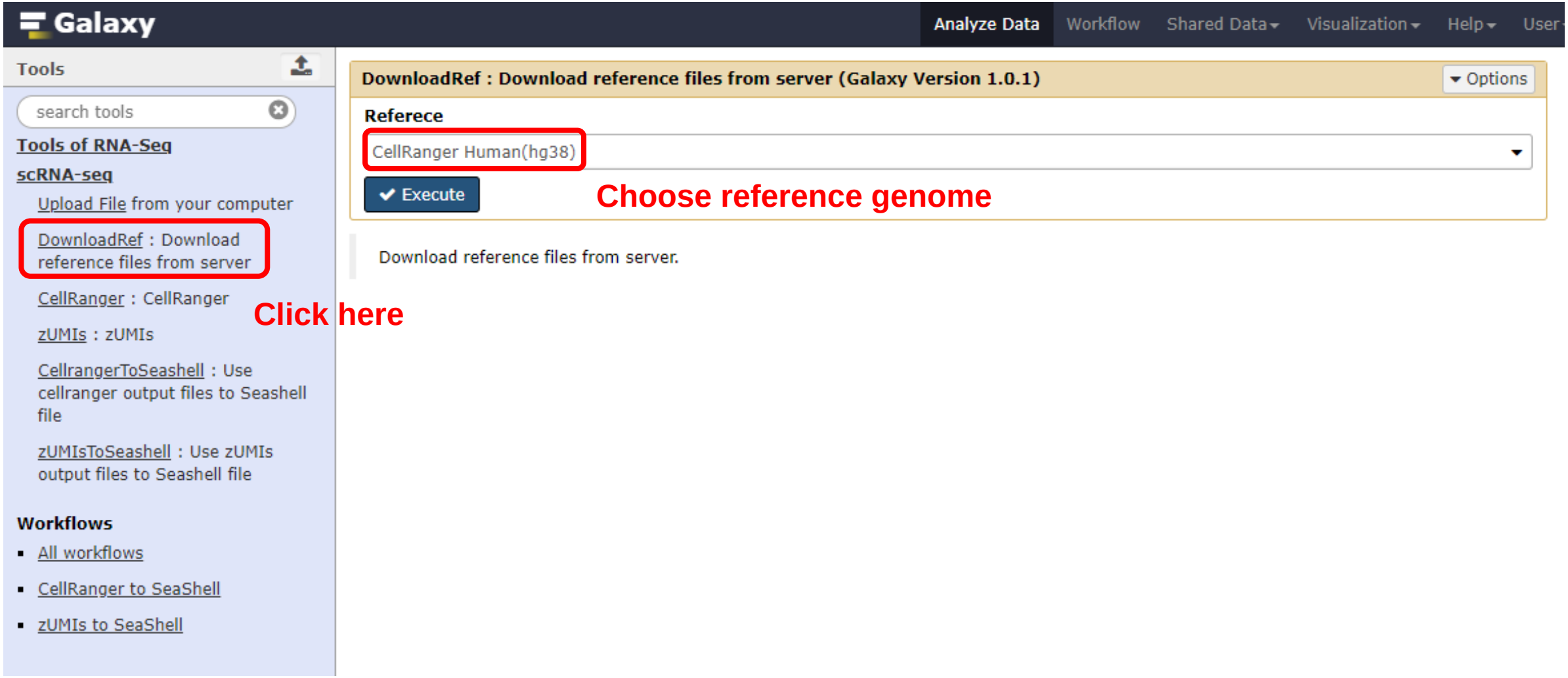


1. Start the galaxy web server

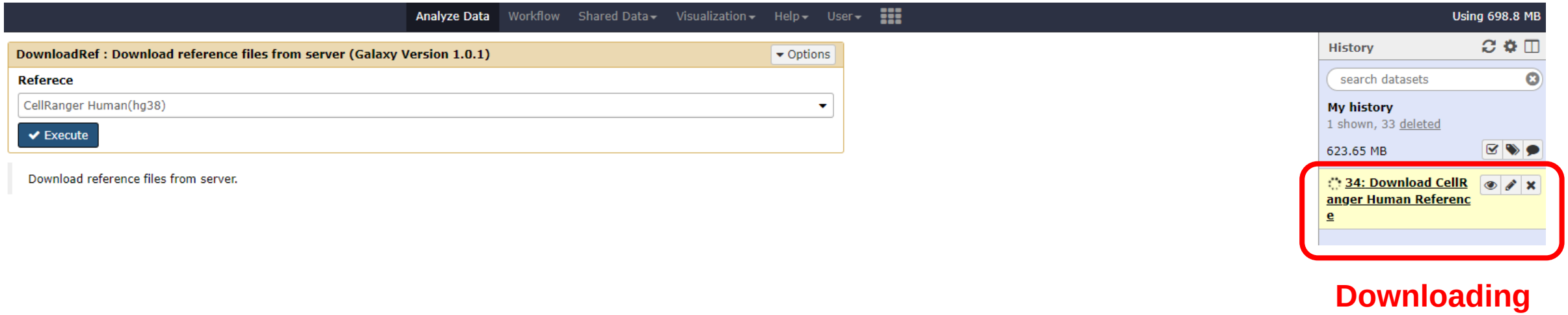
- If successfully login



2. Download your reference genome



2. Download your reference genome



Downloading

3. Prepare your sequencing data

01

Create “cellranger_input” and “YOUR_SAMPLE” folder in the directory shared with docker

./YOUR_DIR/cellranger_input/YOUR_SAMPLE/

02

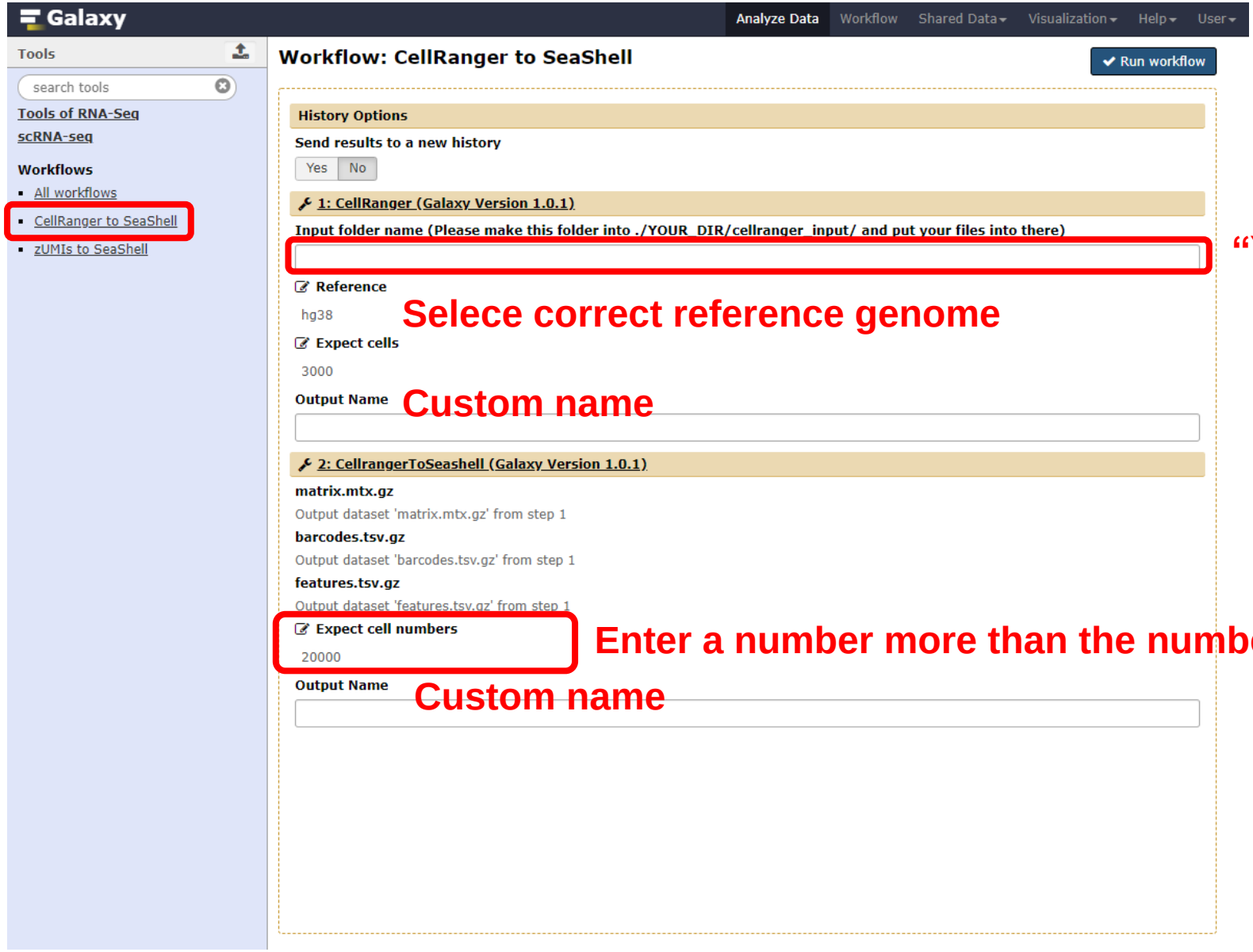
Put your fastq files in “cellranger_input”

./YOUR_DIR/cellranger_input/YOUR_SAMPLE /[Sample Name]_S1_L00[Lane Number]_R1_001.fastq.gz
./YOUR_DIR/cellranger_input/YOUR_SAMPLE /[Sample Name]_S1_L00[Lane Number]_R2_001.fastq.gz
./YOUR_DIR/cellranger_input/YOUR_SAMPLE /[Sample Name]_S1_L00[Lane Number]_I1_001.fastq.gz
If you don’t know how to name your fastq files, please refer to 10x Genomics:
https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/fastq-input

03

Enter name of “YOUR_SAMPLE” and other parameters to galaxy workflow

3. Prepare your sequencing data



Click here

Selece correct reference genome

Custom name

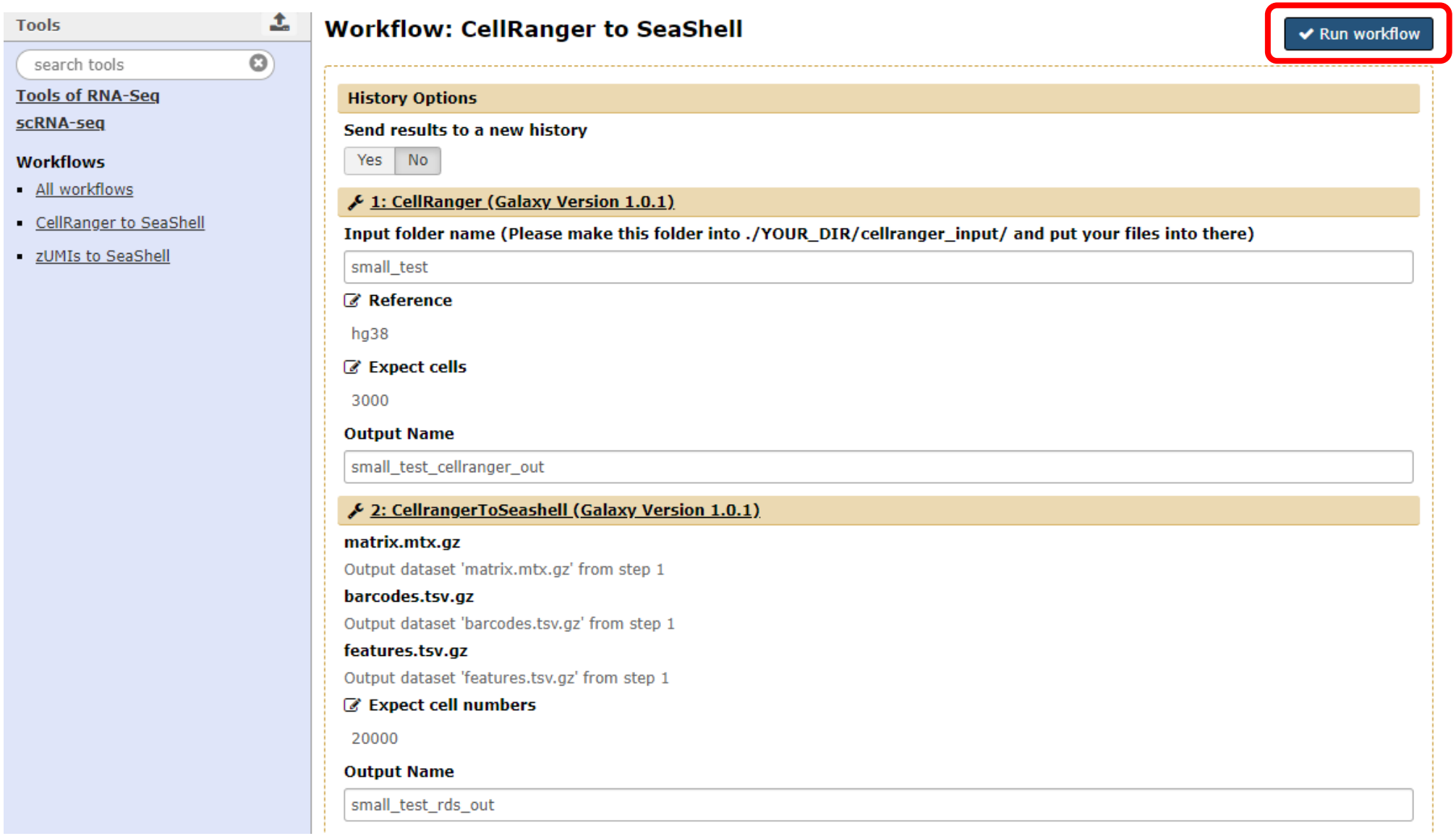
“YOUR_SAMPLE”

Enter a number more than the number of cells you expect

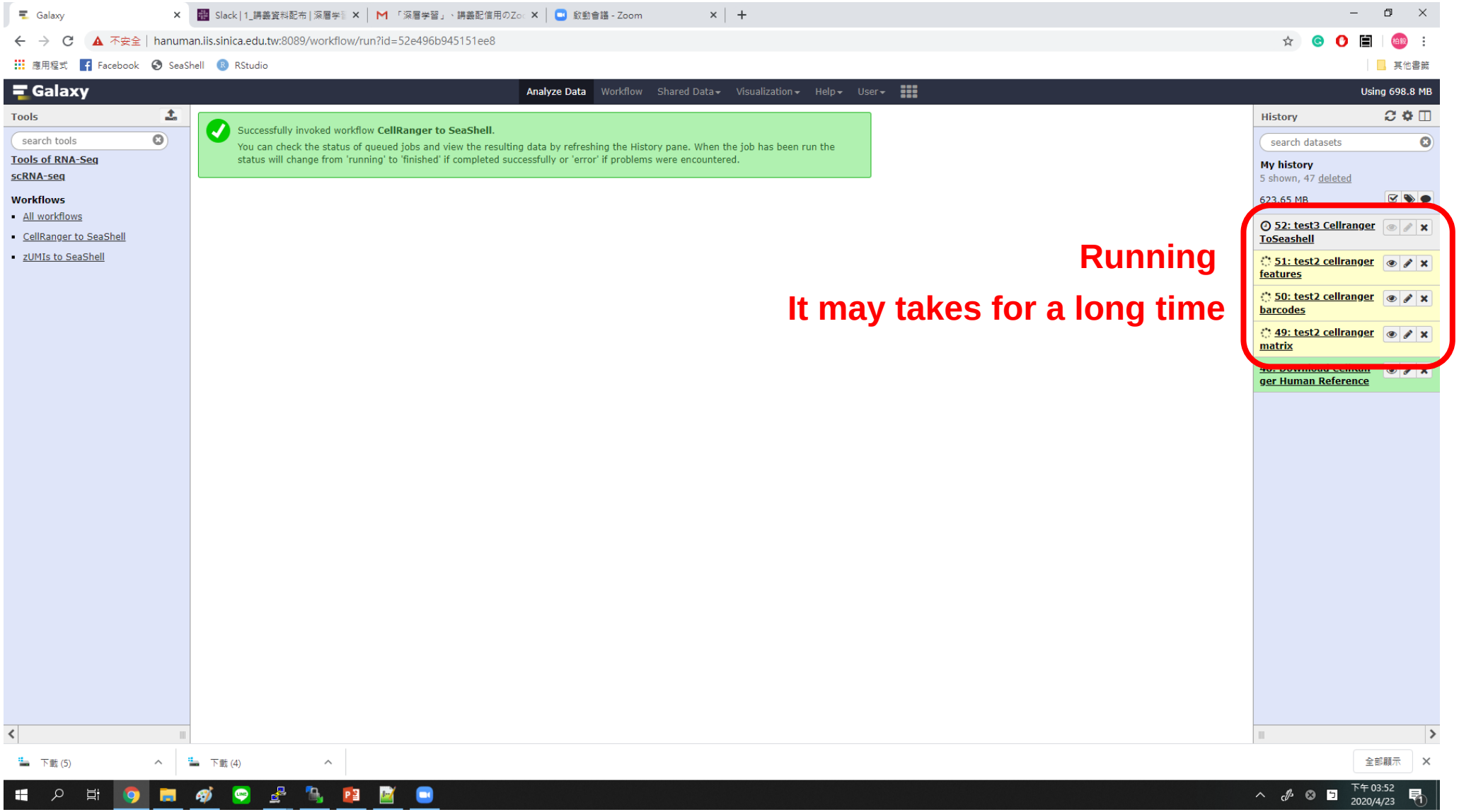
Custom name

4. Run the galaxy pipeline

Run

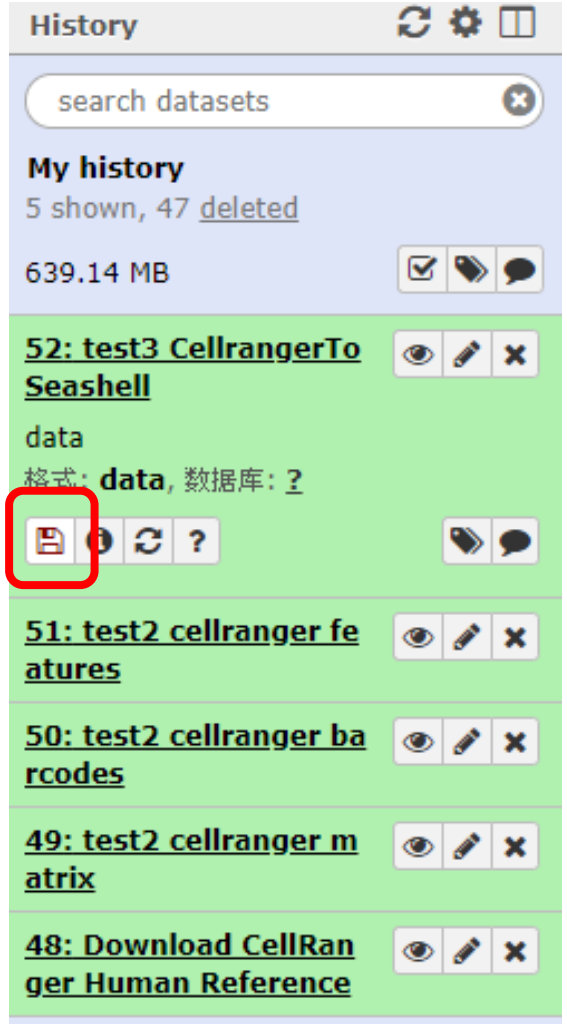


4. Run the galaxy pipeline



5. Download the .rds file

Download



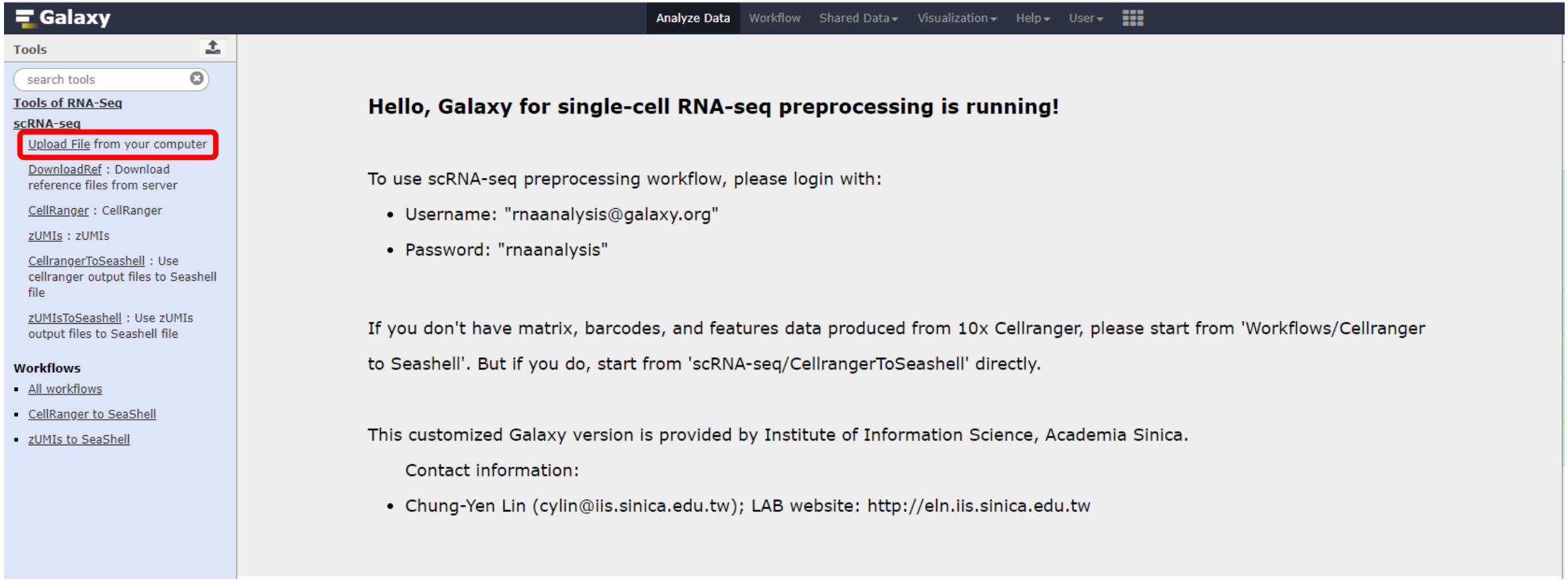
From pre-processed 10x data

01 Put (1) matrix.mtx.gz (2) barcodes.tsv.gz (3) features.tsv.gz in ”YOUR_DIR”

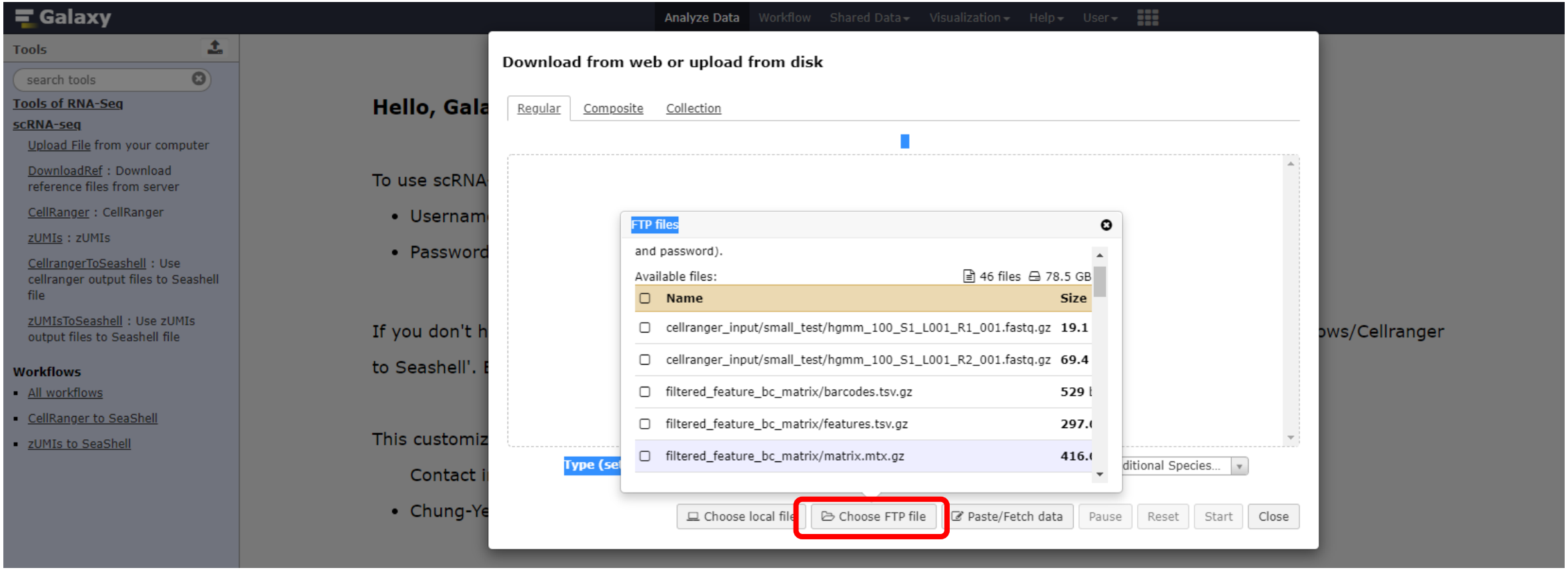
02 Upload these files to galaxy server by ftp

03 Run galaxy workflow “scRNA-seq/CellrangerToSeashell”

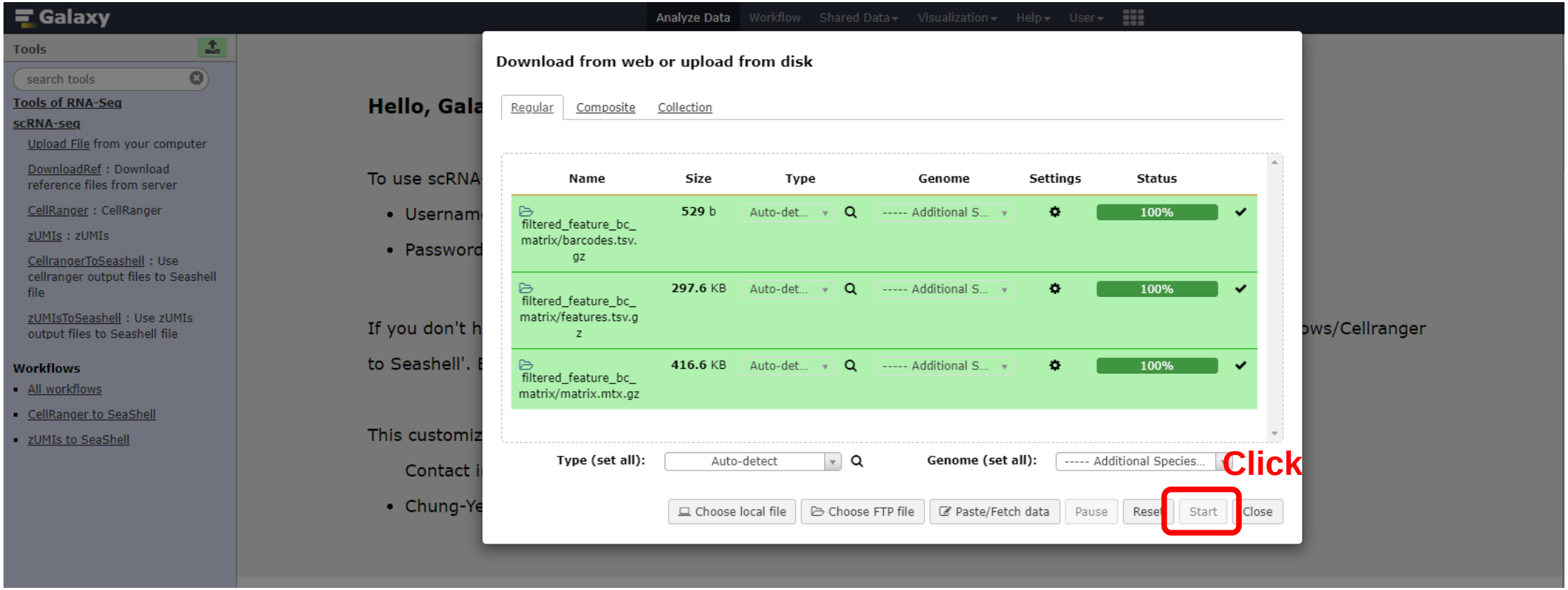
Upload files to galaxy server by ftp



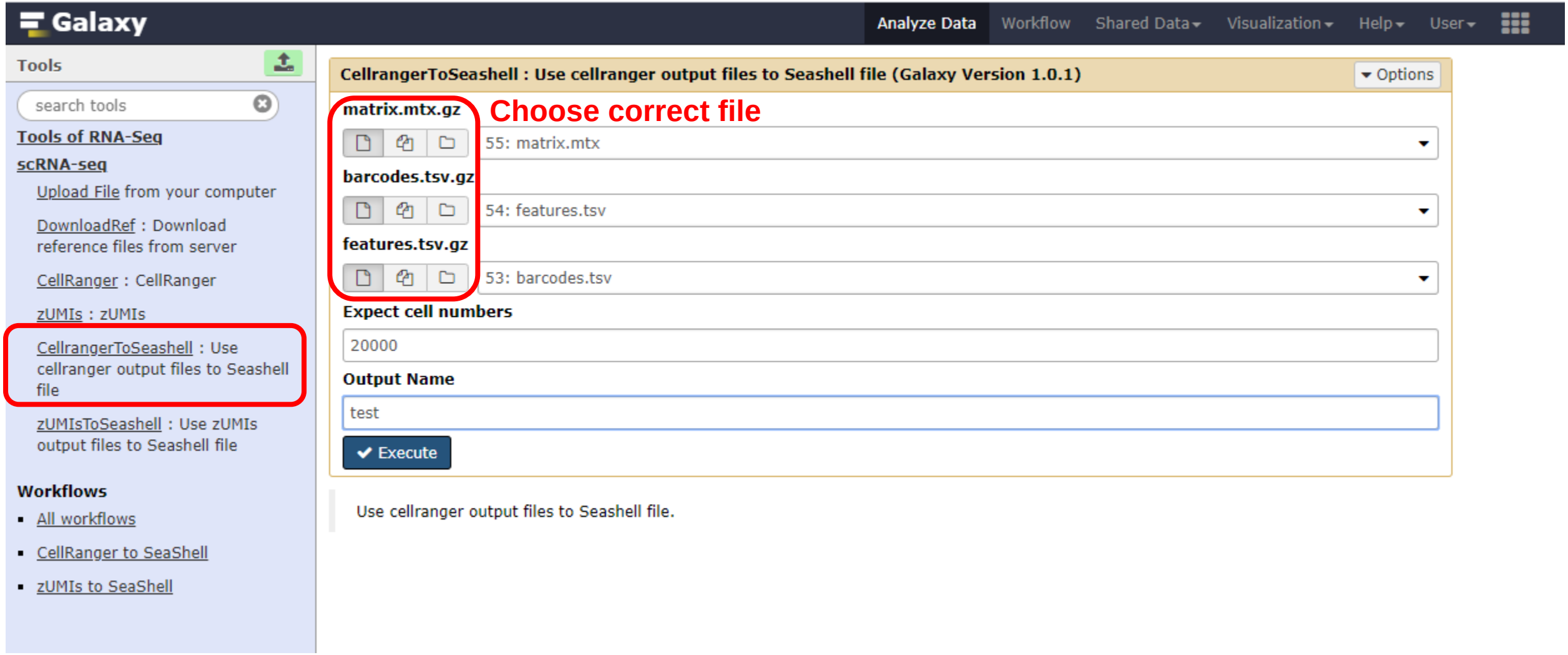
Upload files to galaxy server by ftp



Upload files to galaxy server by ftp

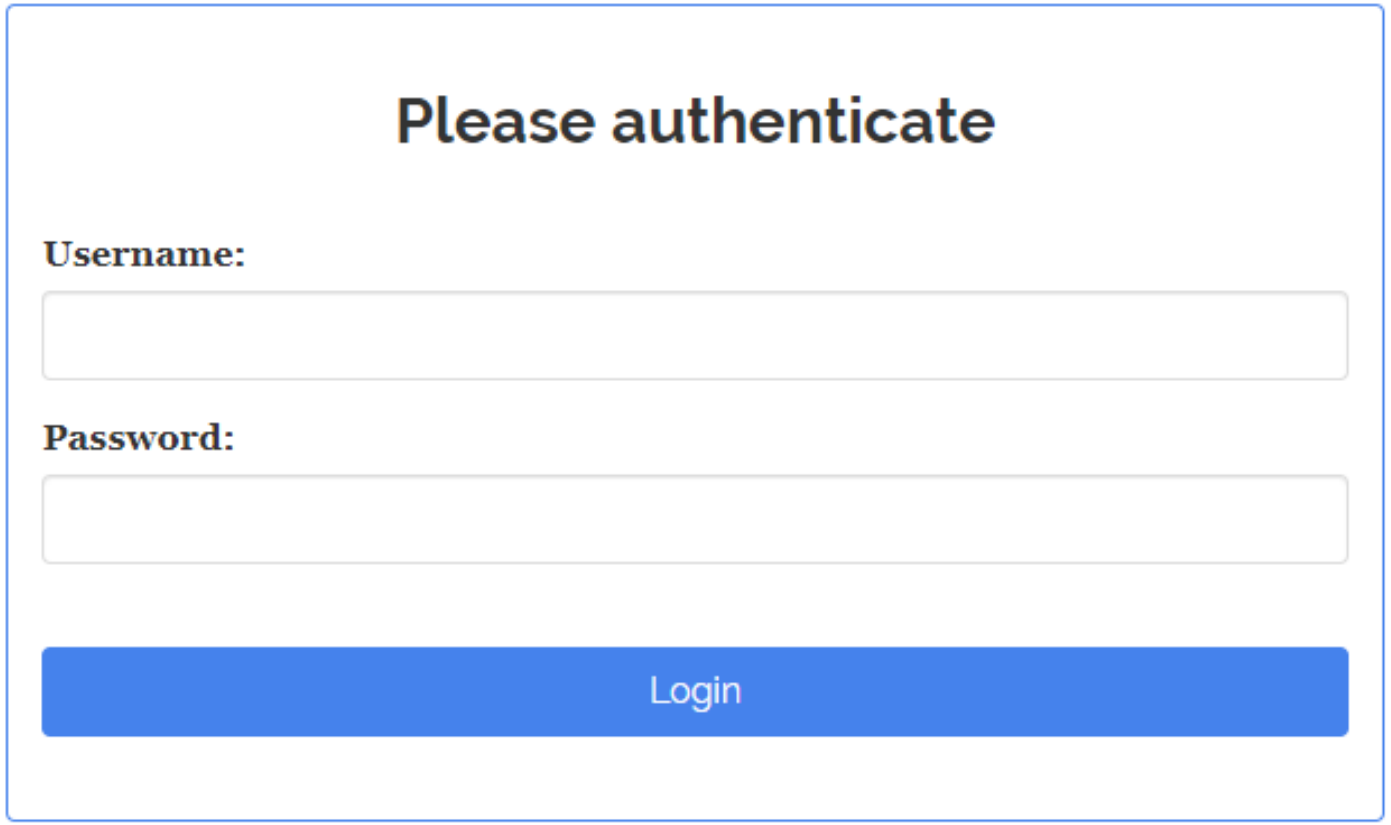


Run galaxy workflow

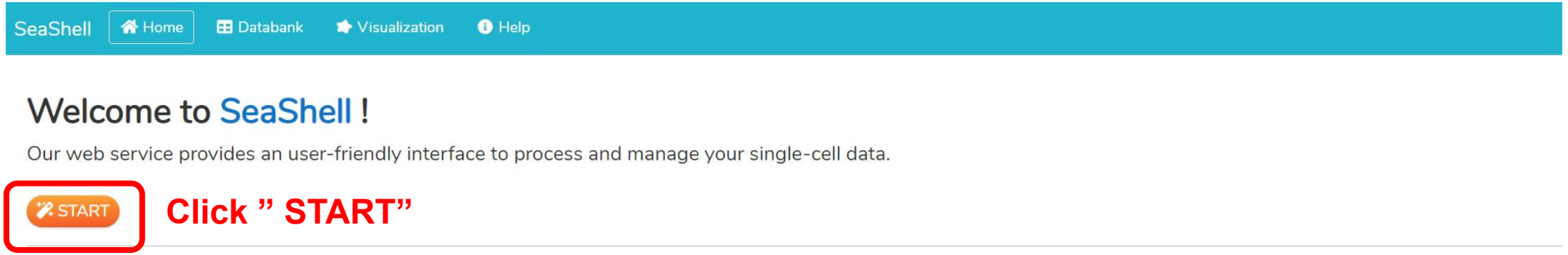


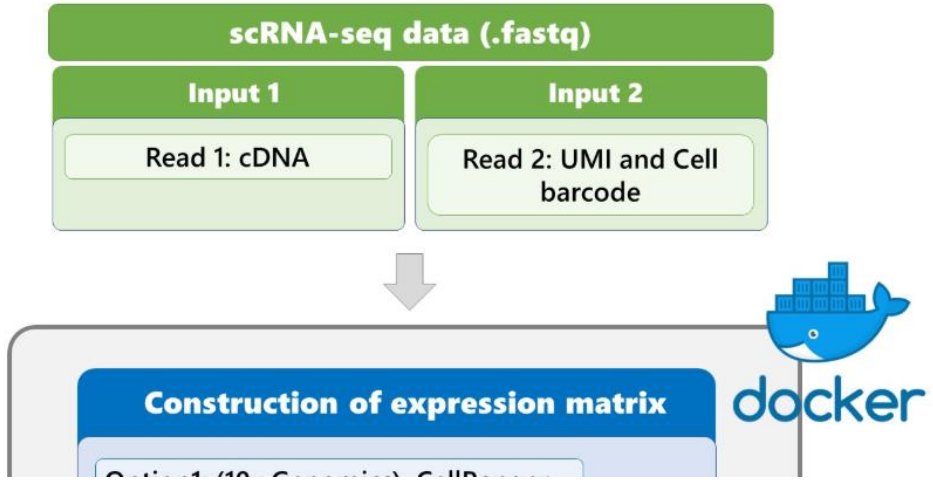
Seashell

Authentication



Homepage


SeaShell
Home
Databank
Visualization
Help
Welcome to SeaShell !
Our web service provides an user-friendly interface to process and manage your single-cell data.
START
Click ” START”


scRNA-seq data (.fastq)
Input 1
Input 2
Read 1: cDNA
Read 2: UMI and Cell barcode
Construction of expression matrix
docker

Workflow

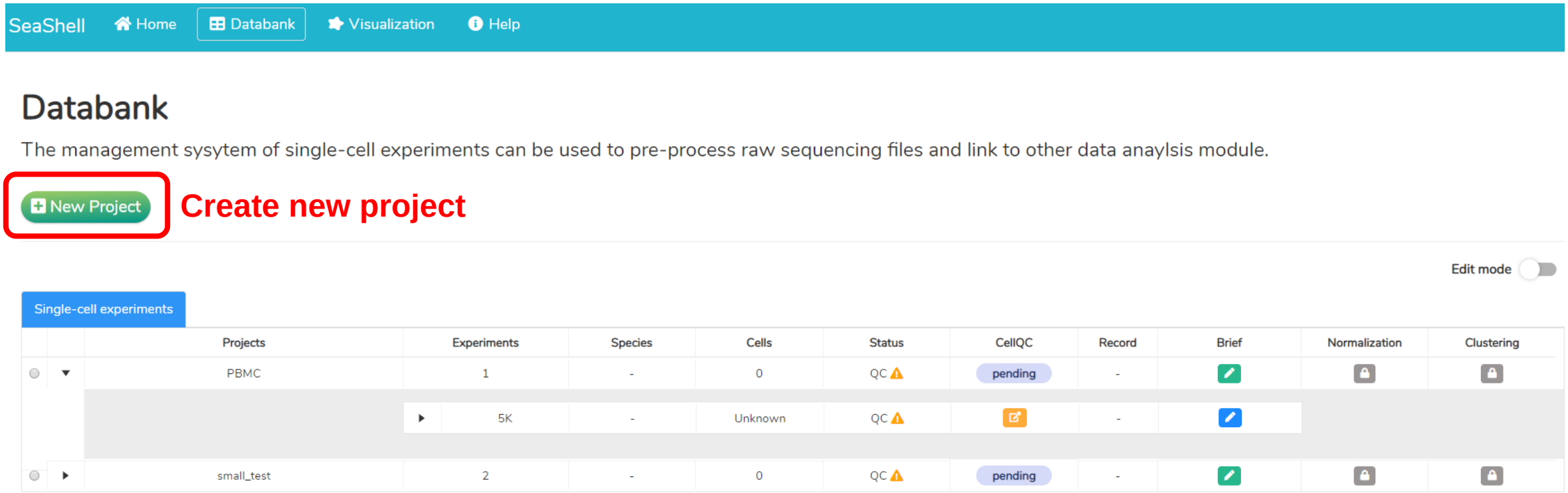
01 Upload rds file to SeaShell

02 Cell QC, Normalization, and Clustering

03 Visualization and Gene analysis

Databank

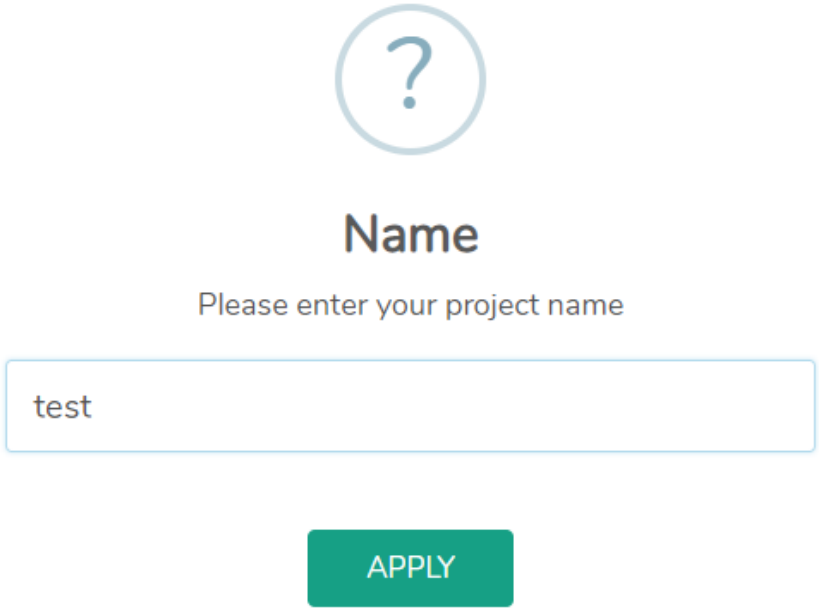
Create new project

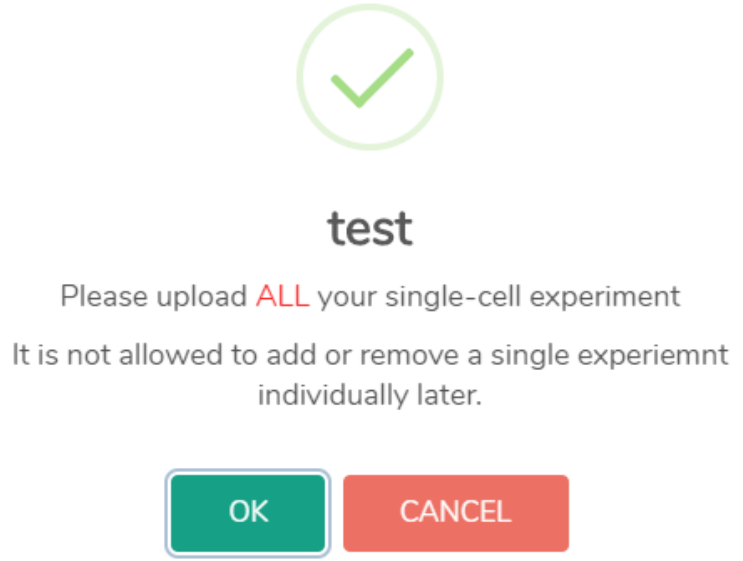


Databank

Create new project

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

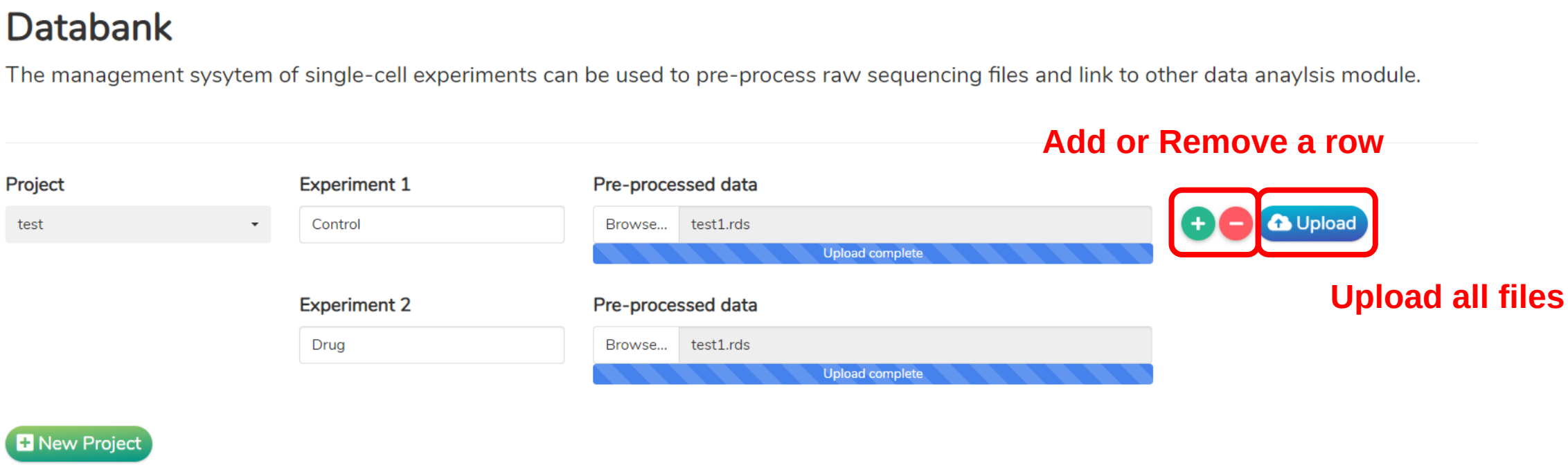




Databank

Create new project

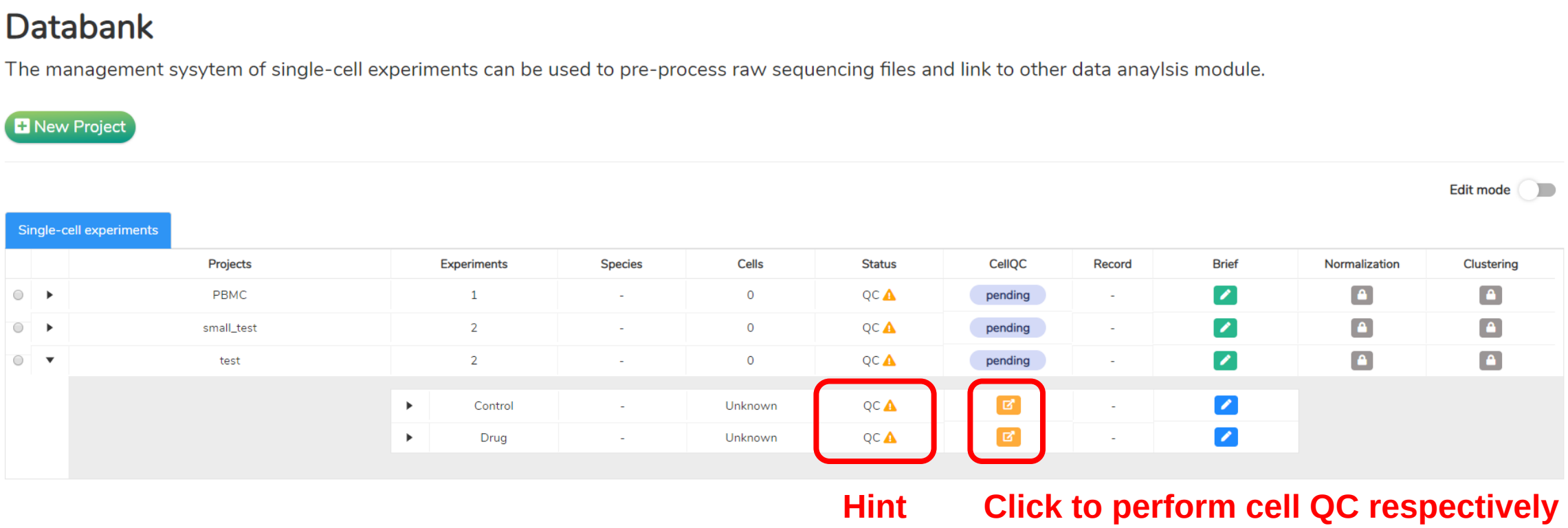
Upload **All single-cell experiments**(rds files) to one project. (For example: Control and Drug experiments)



Databank

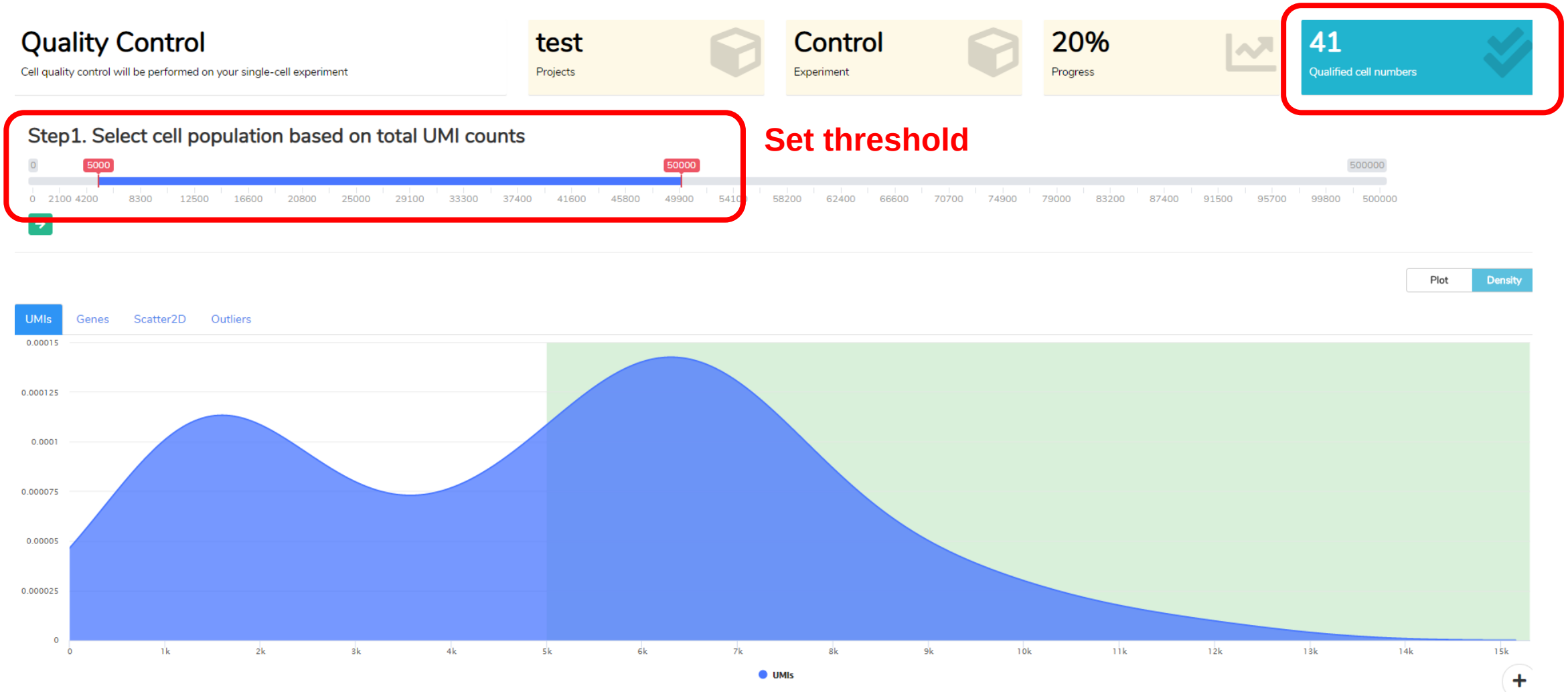
Create new project

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

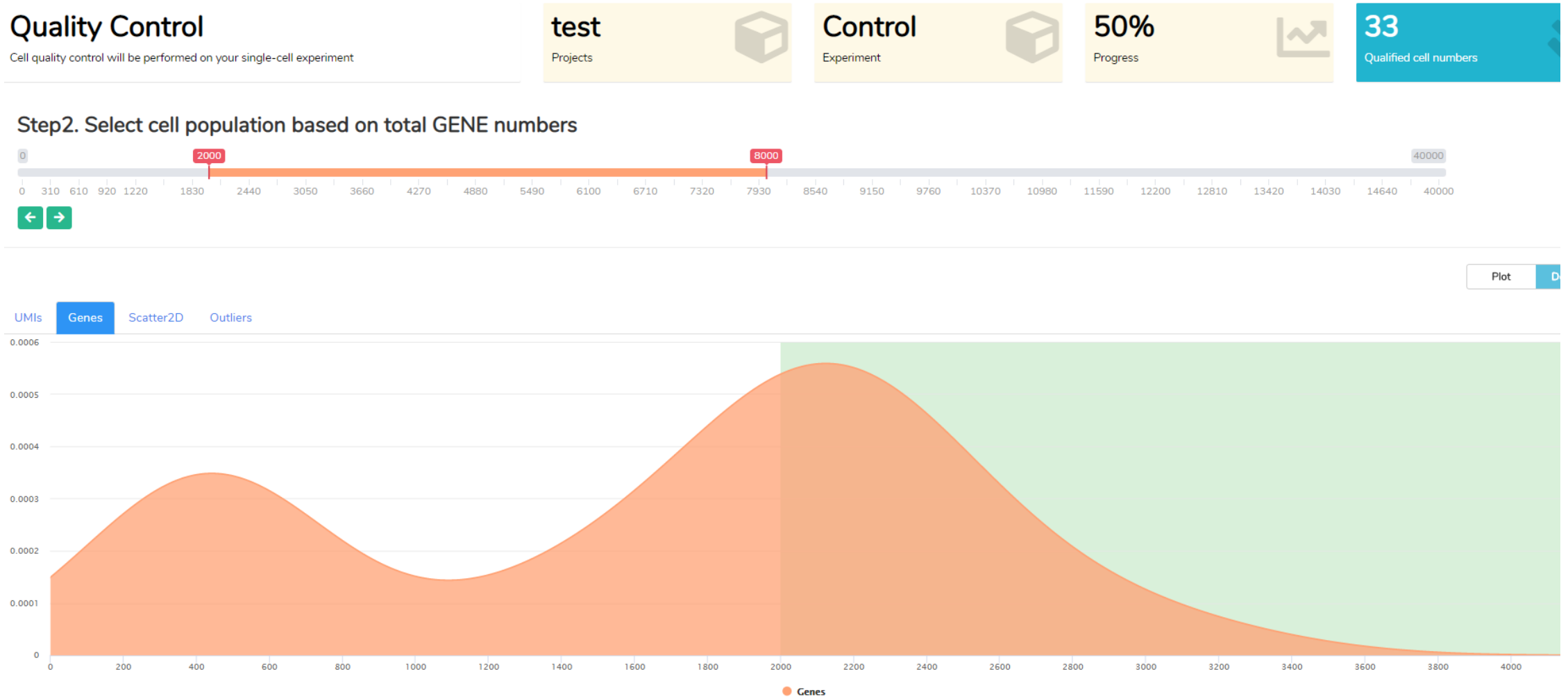


Cell QC

Real time qualified cell numbers



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
0
2000
8000
40000
0 310 610 920 1220 1830 2440 3050 3660 4270 4880 5490 6100 6710 7320 7930 8540 9150 9760 10370 10980 11590 12200 12810 13420 14030 14640 40000
Plot
UMIs
Genes
Scatter2D
Outliers
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0
0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3400 3600 3800 4000
Genes

Cell QC

Check by 2D scatter plot

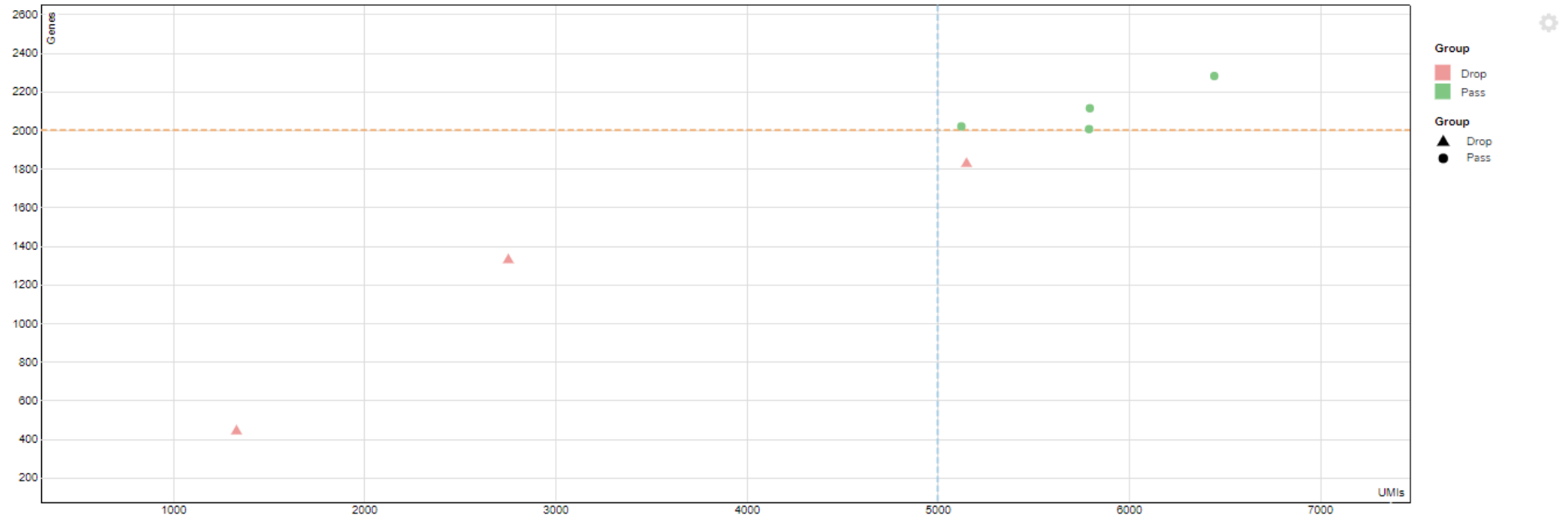
Step3. Check selected cell population

Main population are defined by UMI counts AND gene numbers

Number of observations

1% 10% 100%

UMIs Genes Scatter2D Outliers

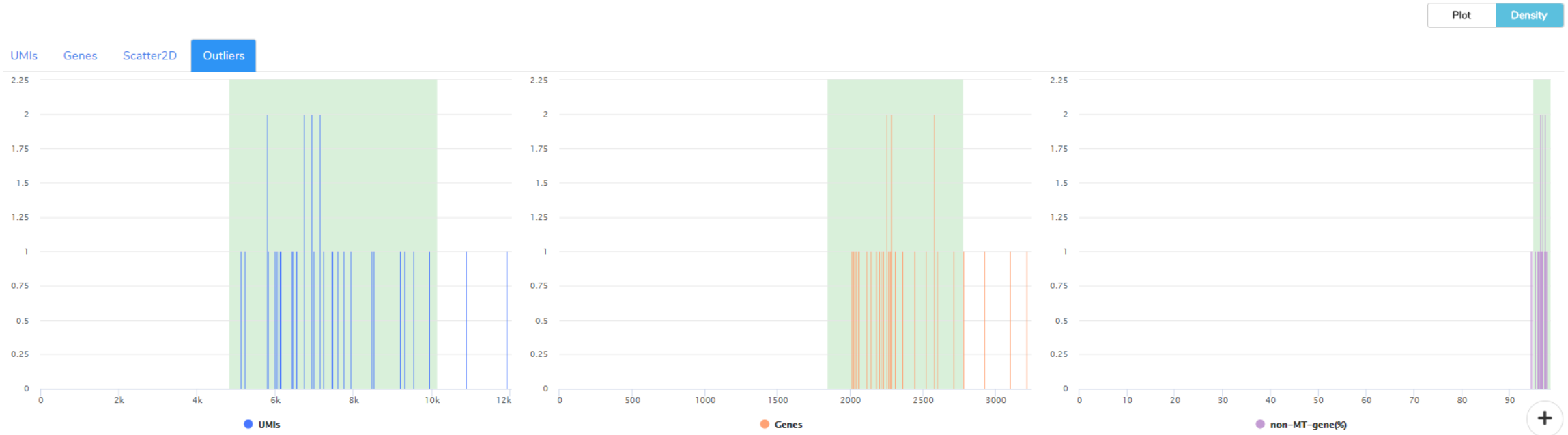


Cell QC

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.



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.

New Project

Edit mode

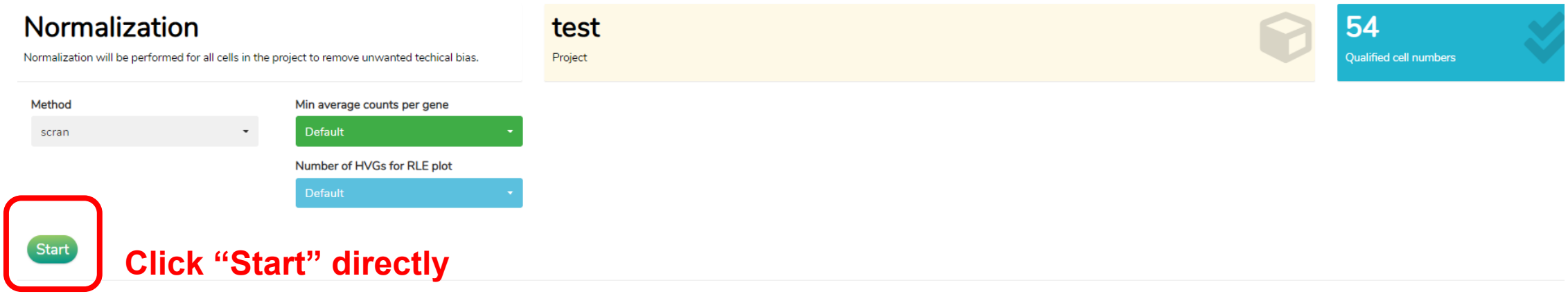
Single-cell experiments

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						
Drug	Hs	27	QC						

Perform normalization

Success and new tips

Normalization



Click “Start” directly

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

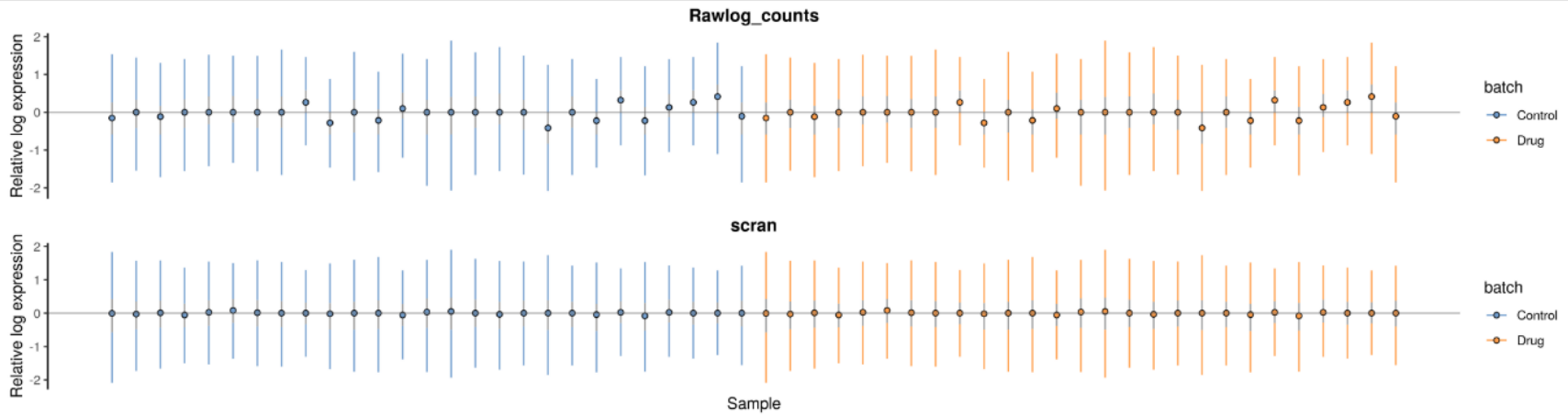
Default

SAVE CANCEL

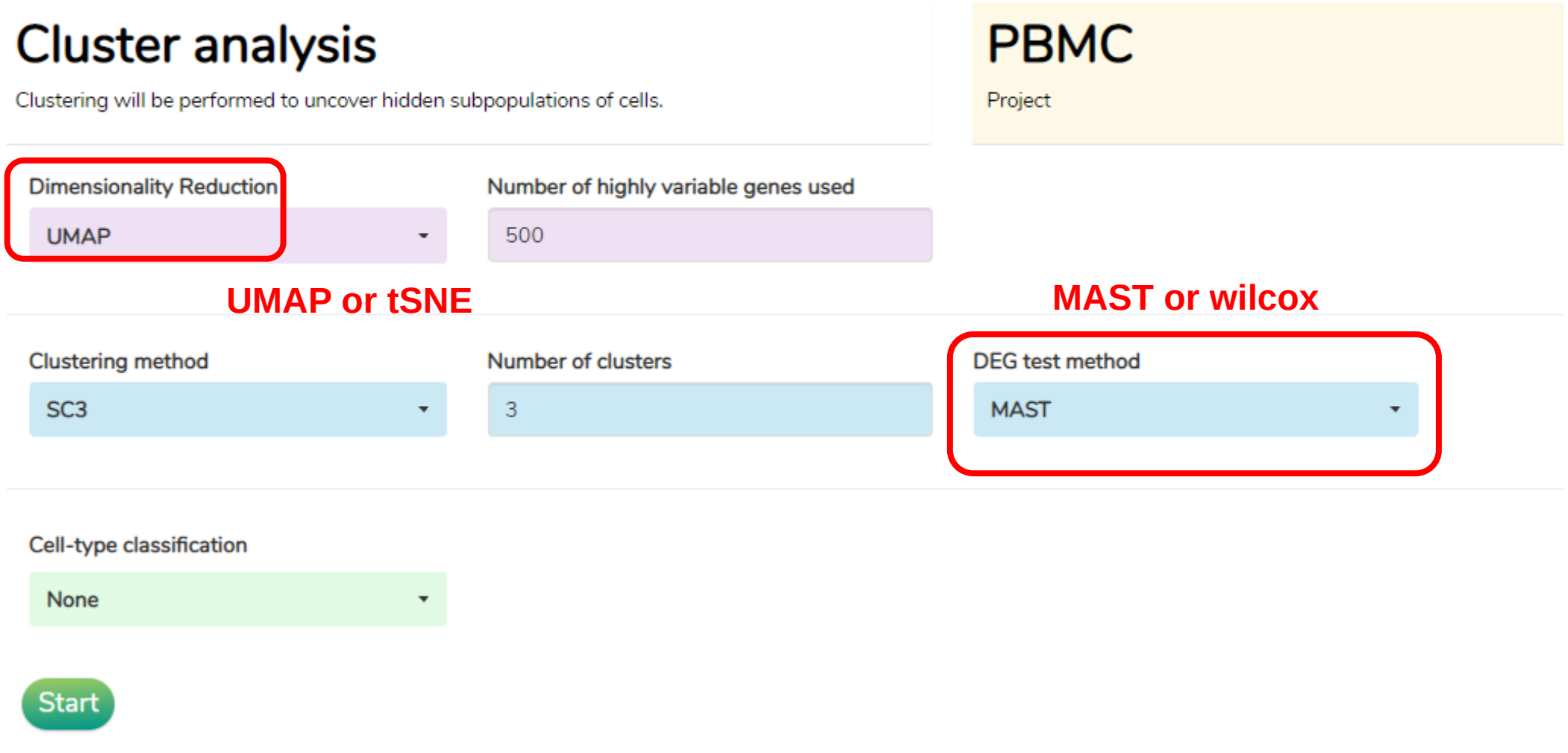
RLE plot

RLEplot

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

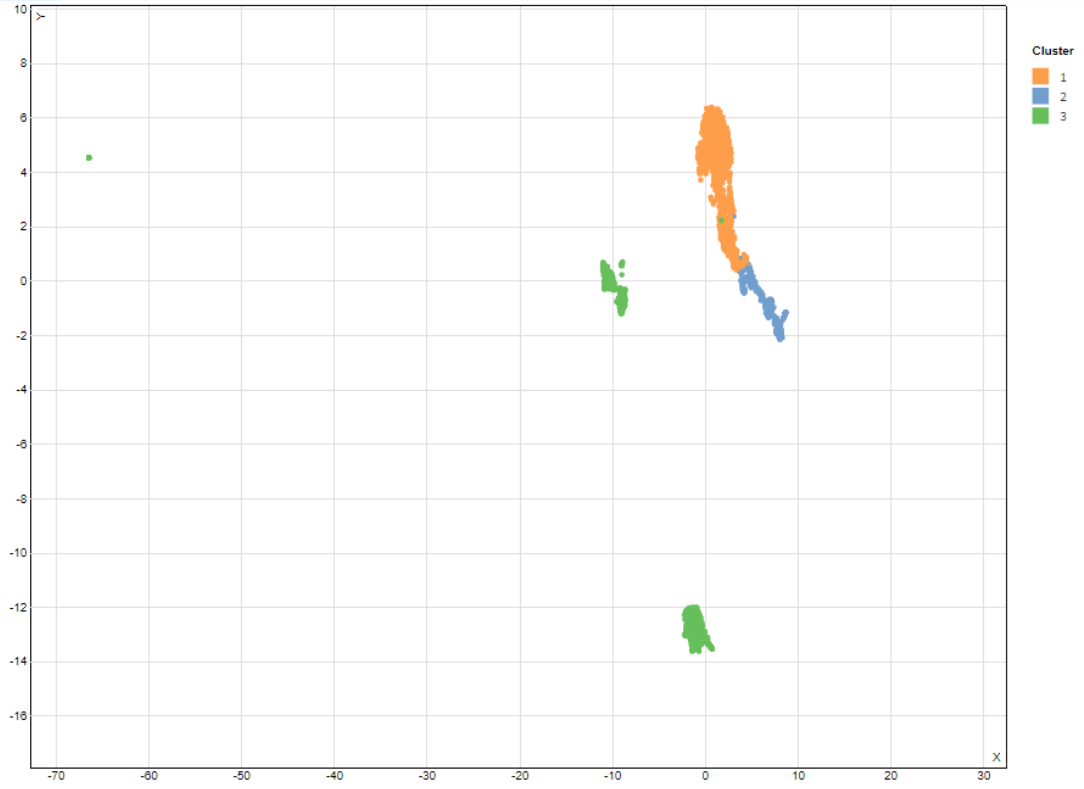


Clustering


Cluster analysis
Clustering will be performed to uncover hidden subpopulations of cells.
PBMC
Project
Dimensionality Reduction
UMAP
Number of highly variable genes used
500
UMAP or tSNE
MAST or wilcox
Clustering method
SC3
Number of clusters
3
DEG test method
MAST
Cell-type classification
None
Start

Clustering

Cluster analysis



DEG preview

1

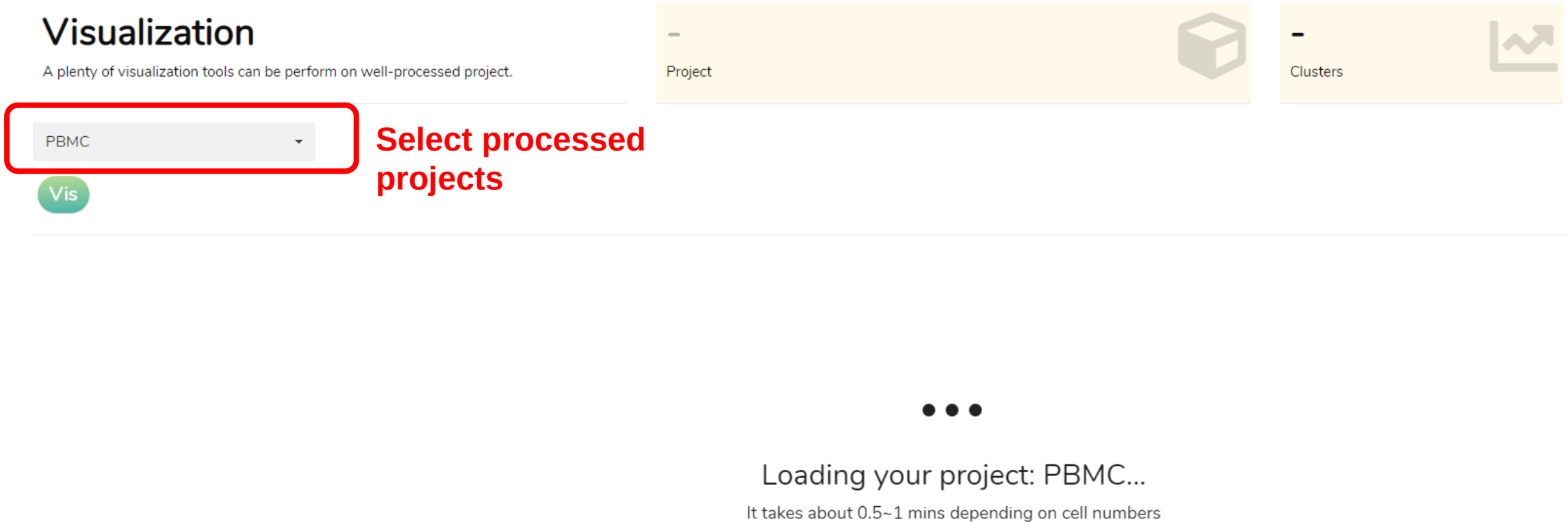
Positive only

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

Previous 1 2 3 4 5 ... 61

Visualization



Visualization

Visualization

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

PBMC
Project

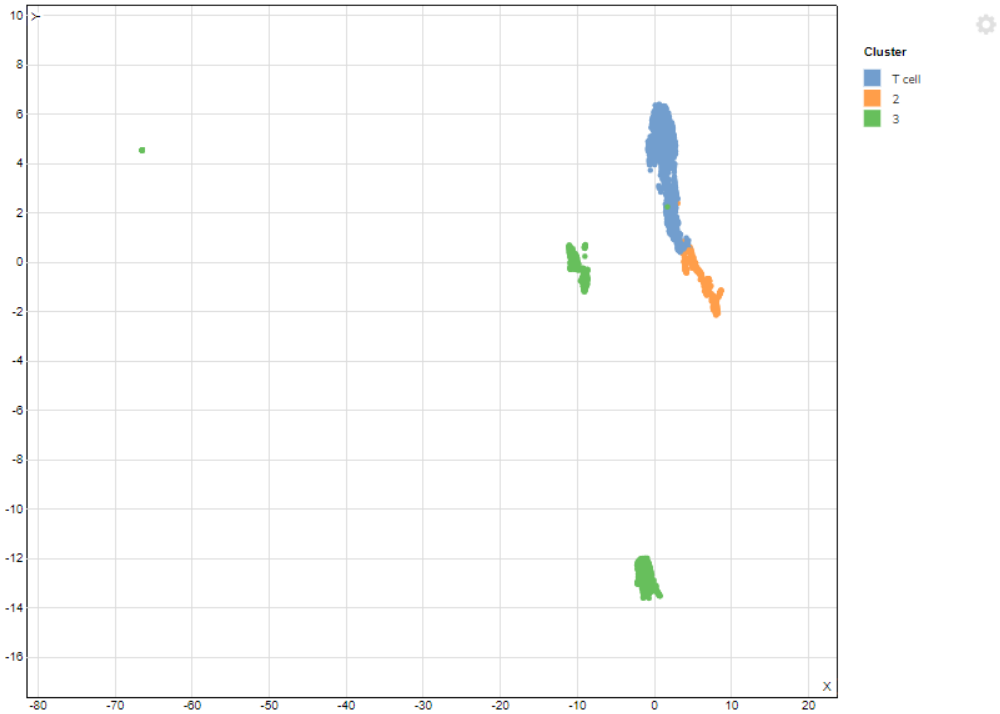
3
Clusters

3,456
Qualified cell numbers

PBMC

Vis

Rename your cluster by DEGs



DEGselector | Expression plot | Violin plot | GO analysis

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.[illegible]	Positive

1-10 of 462 rows

Previous 1 2 3 4 5 ... 47 Next

DEG filter

01-Cluster

T cell

Rename cluster

02-FoldChange

1.5

03-adjPvalue

0.1

04-Type

Positive only

Send to GO

Use these DEGs for GO analysis

Visualization

Visualization

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

PBMC

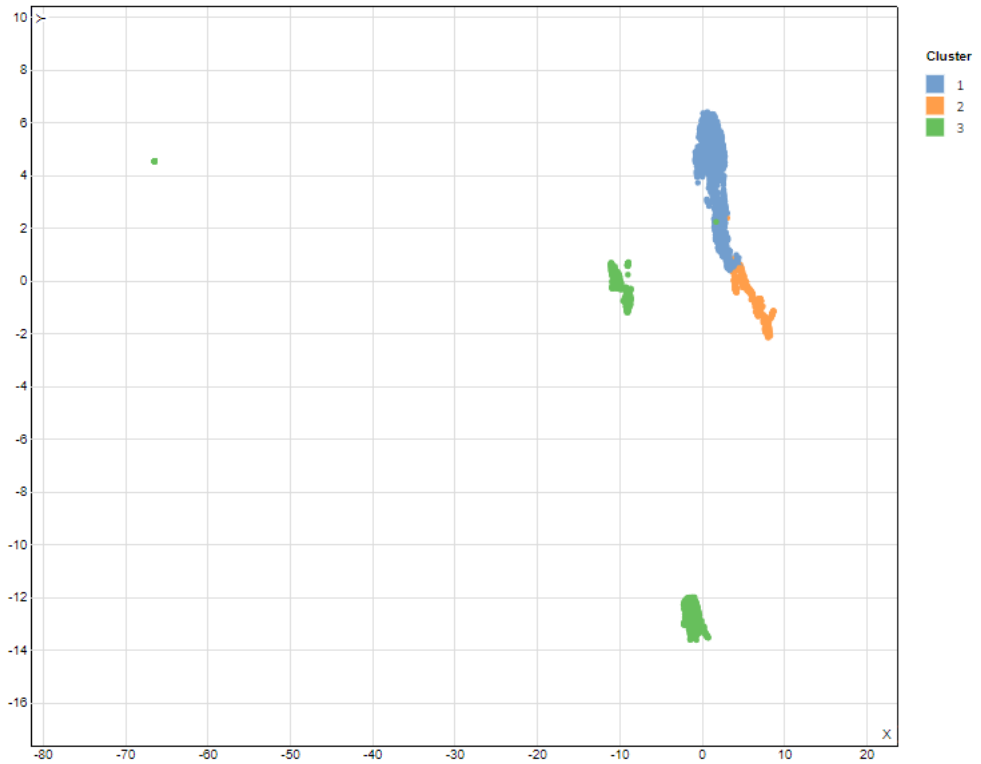
Project

3

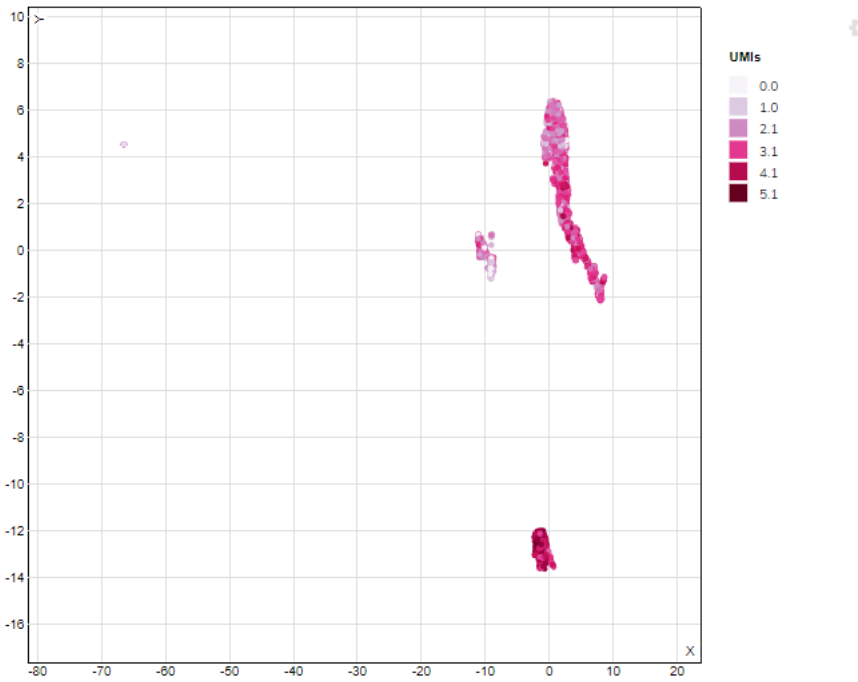
Clusters

PBMC

Vis Save



DEGselector | Expression plot | Violin plot | GO analysis



Visualization

Visualization

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

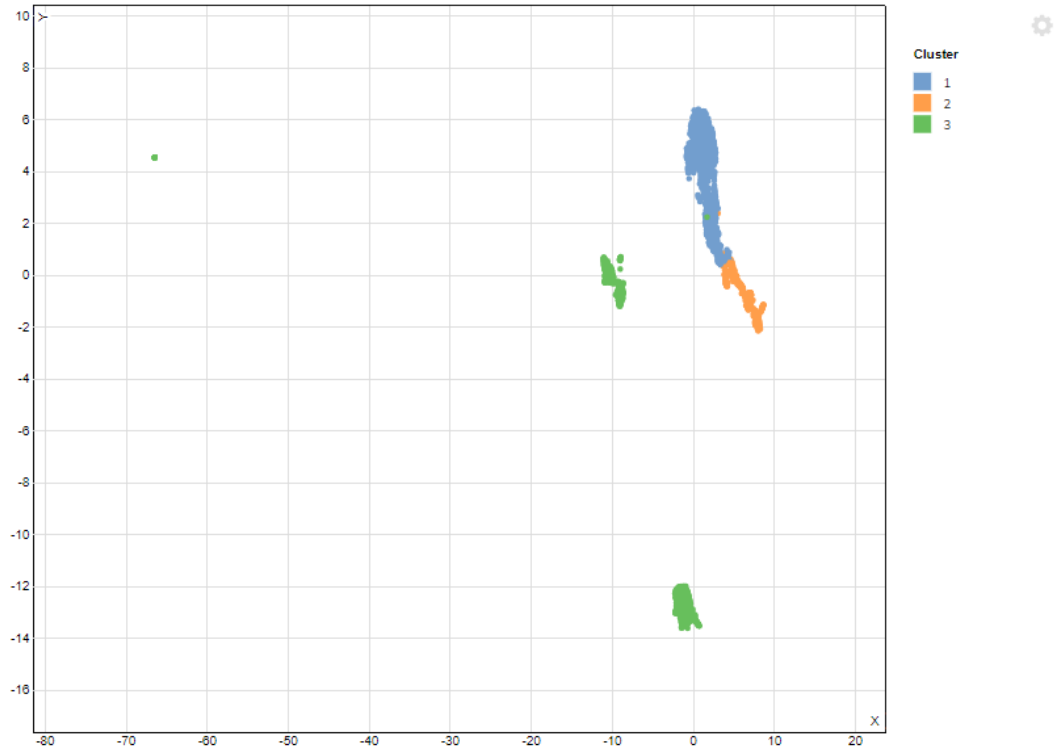
PBMC
Project

3
Clusters

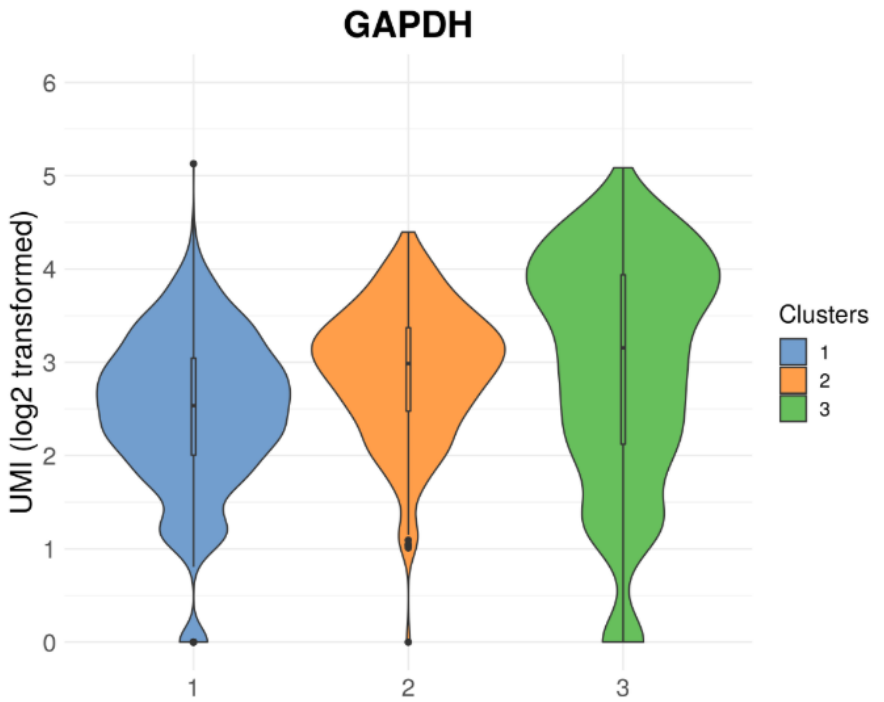
3,45
Qualified c

PBMC

Vis Save



DEGselector | Expression plot | Violin plot | GO analysis



Visualization

Visualization

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

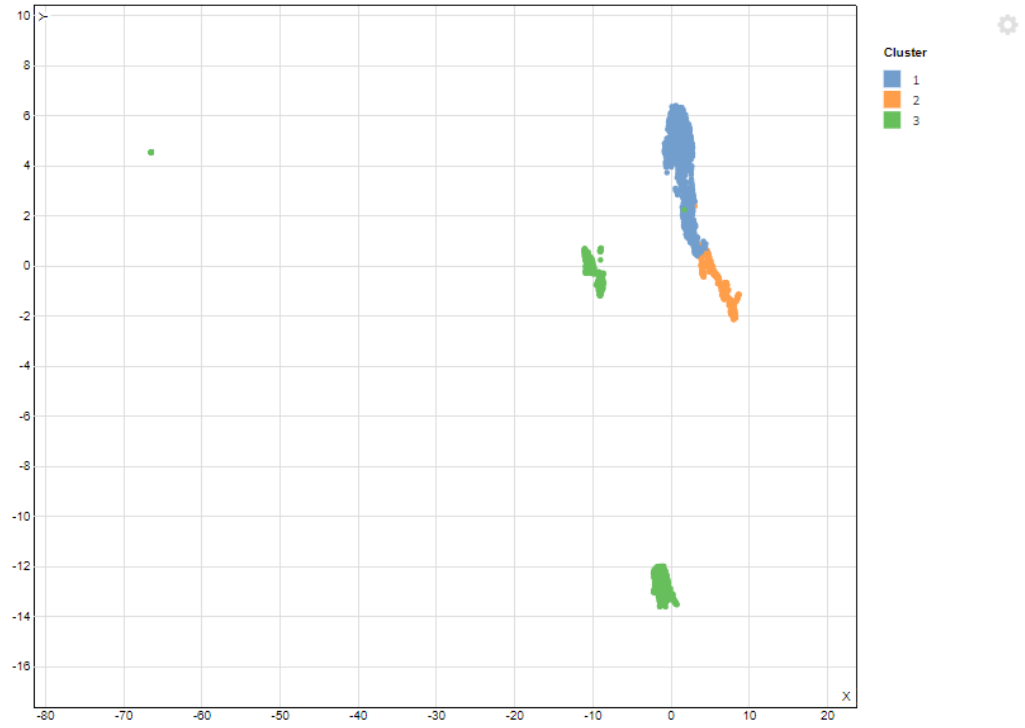
PBMC
Project

3
Clusters

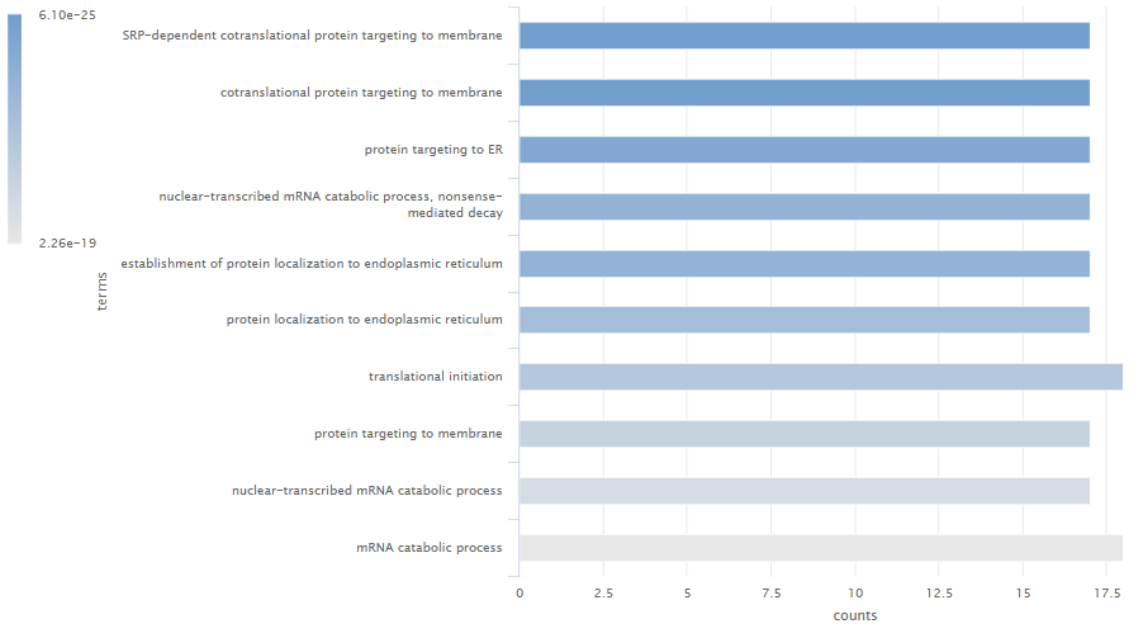
3,456
Qualified cell numbers

PBMC

Vis Save



DEGselector | Expression plot | Violin plot | GO analysis



Contact

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