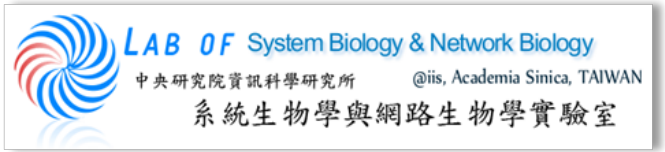
Docexpress –

A galaxy docker for estimation of Expression profiling in RNA-seq

SPEAKER: Ping-Heng Hsieh
DATE: 2018/12/07



Outline

- Introduction for Docexpress
- Walkthrough – Estimate Expression Profiling
- Introduction for Docmethyl

1.

INTRODUCTION FOR DOCEXPRESS

Docexpress

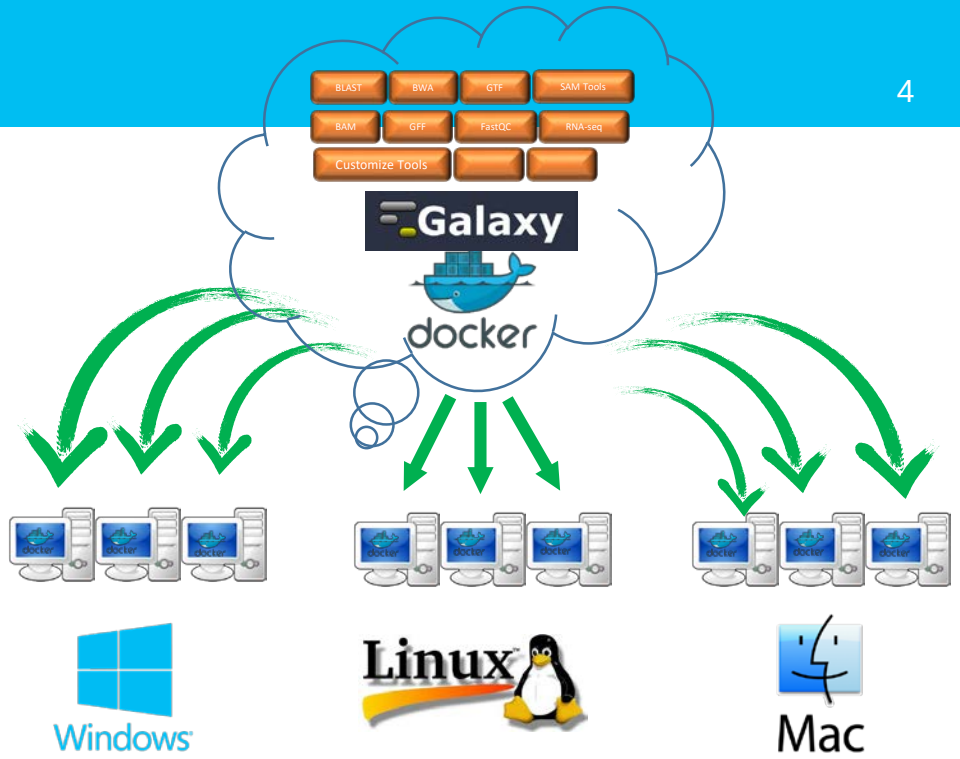
- Galaxy + Docker

Galaxy

- “Galaxy is an open source, web-based platform for data intensive biomedical research.” – by Galaxy

Docker

- “Docker is an open platform for developing, shipping, and running applications.” – by Docker docs



Docexpress

- For alleviating the burden on processing massive NGS data, we create the workflows based on galaxy/ Docker to estimate expression profiling in RNA-seq.

- After the data pre-processing done, the expression profiling can be submitted to MOLAS, is a robust web application that can take gene expression data (FPKM/TPM) from different libraries as inputs, map these expressed genes with build-in annotations for further analyses and reveal biological meaning of the complex data in the intuitive interface.

1. Install Docker on your local system [Ref.]
2. Open a terminal or Use command mode
3. Following the “Install & Usage” on our docker hub [link]

How to lunch Docexpress

- Create data store directory “galaxy_guest”

  md galaxy_guest #for windows command line

- Pull the docexpress images

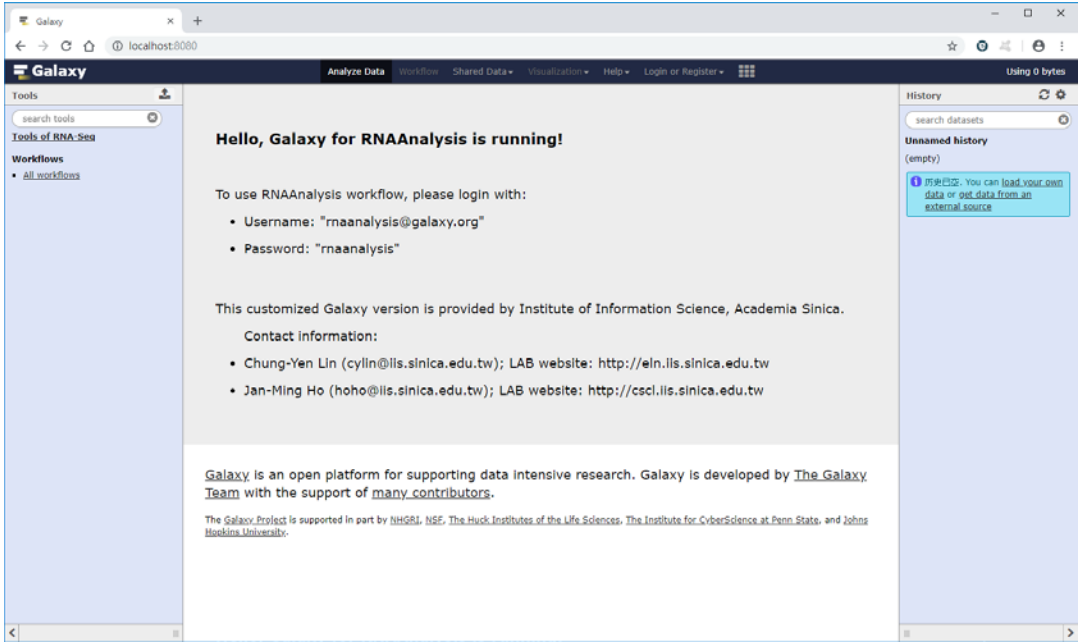
  docker pull lsbnb/docexpress_fastqc

- Run docexpress

  docker run -d -t -i -p 8080:80 -p 8021:21 -p 8022:22 -v %cd%/galaxy_guest/:/root/galaxy/database/ftp/rnaanalysis@galaxy.org/ lsbnb/docexpress_fastqc /bin/bash

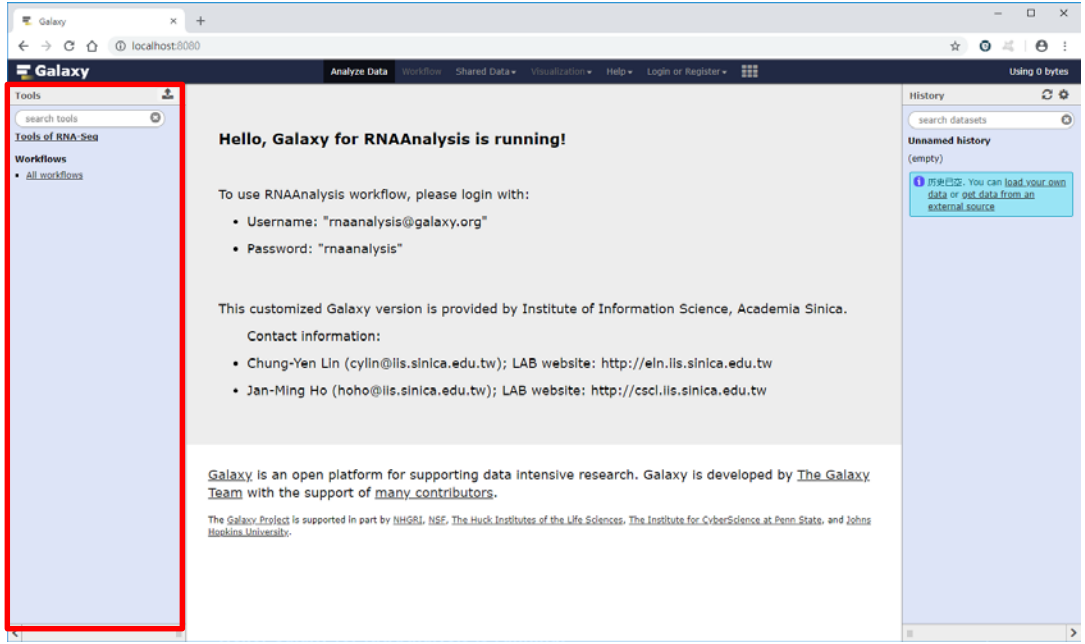
- Open the browser and input the address “http://localhost:8080/”

Lunch succeeded!



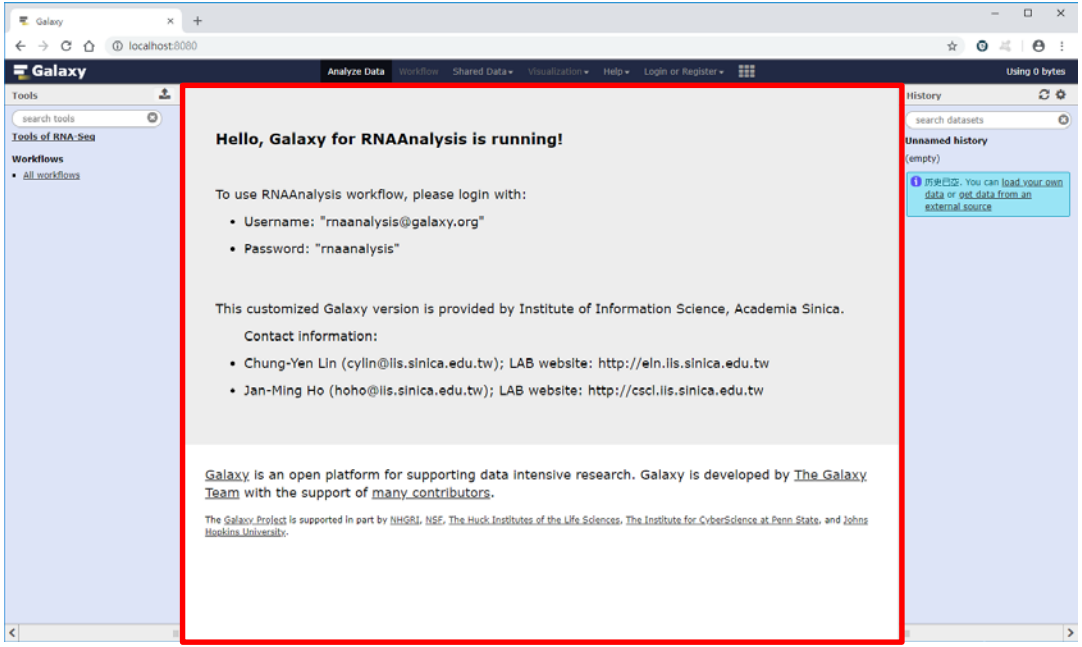
Lunch succeeded!

Tool panel: which contains all tools and workflows



Lunch succeeded!

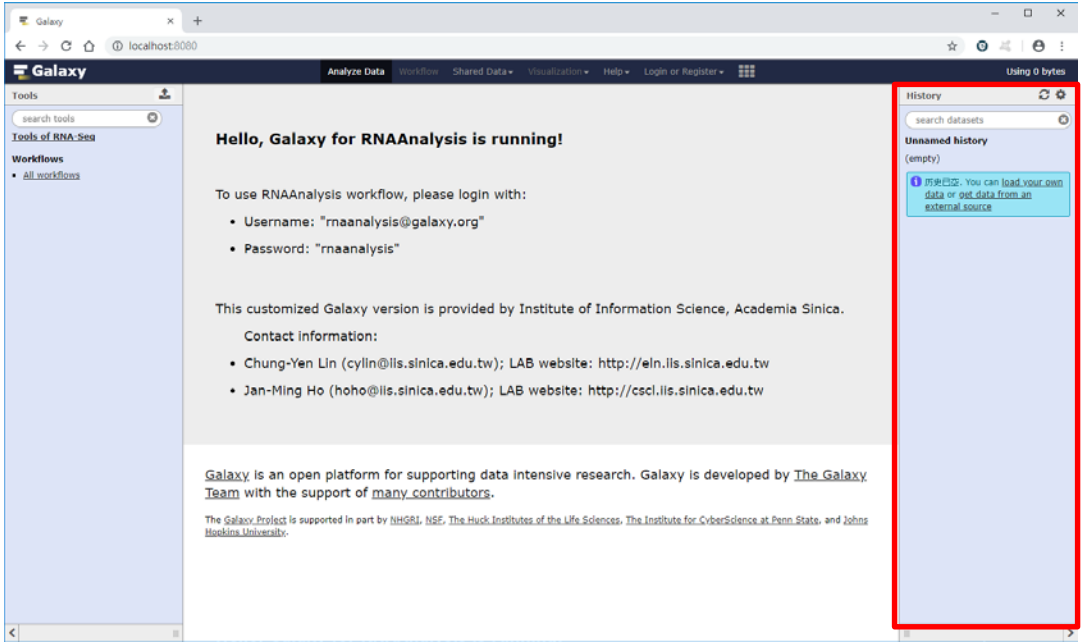
Main : show the homepage and execution result…etc



Lunch succeeded!

History: contains all input and output data and result file.

Please get on Galaxy 101 for more details [link]



2.

WALKTHROUGH

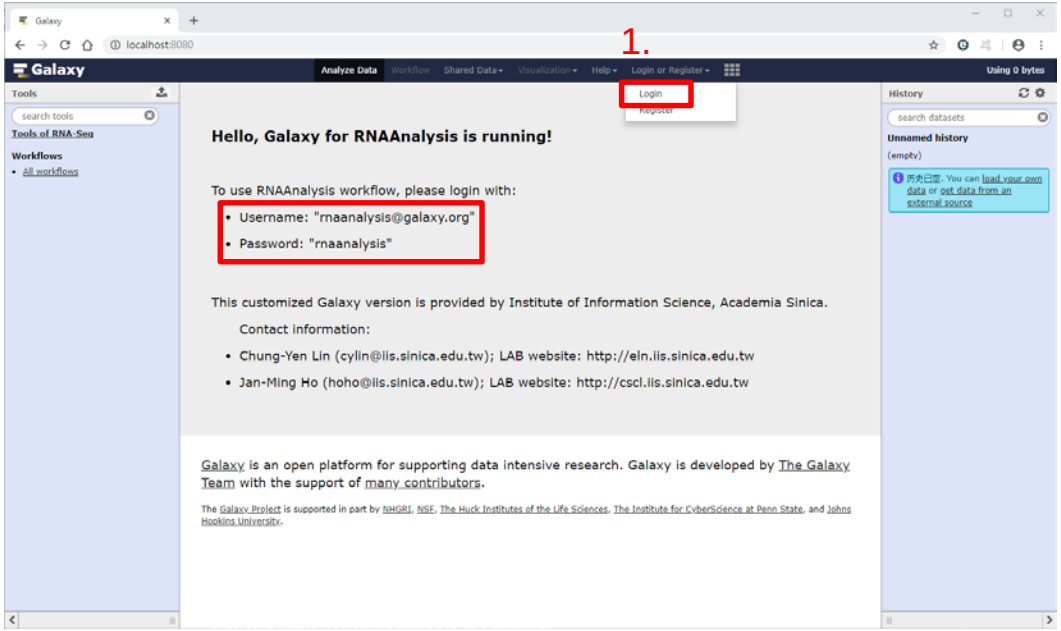
– Estimate Expression Profiling

Download test data

- Download [link]
- Put all test data into directory “galaxy_guest”

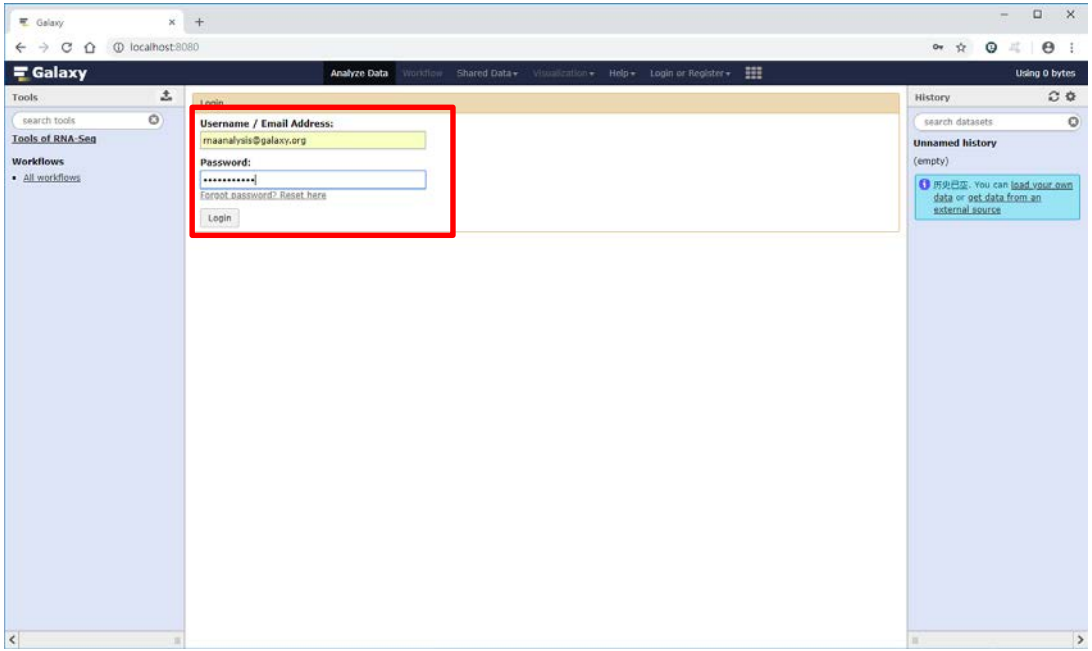
Login Galaxy

Login for full accessing all functions of Docexpress



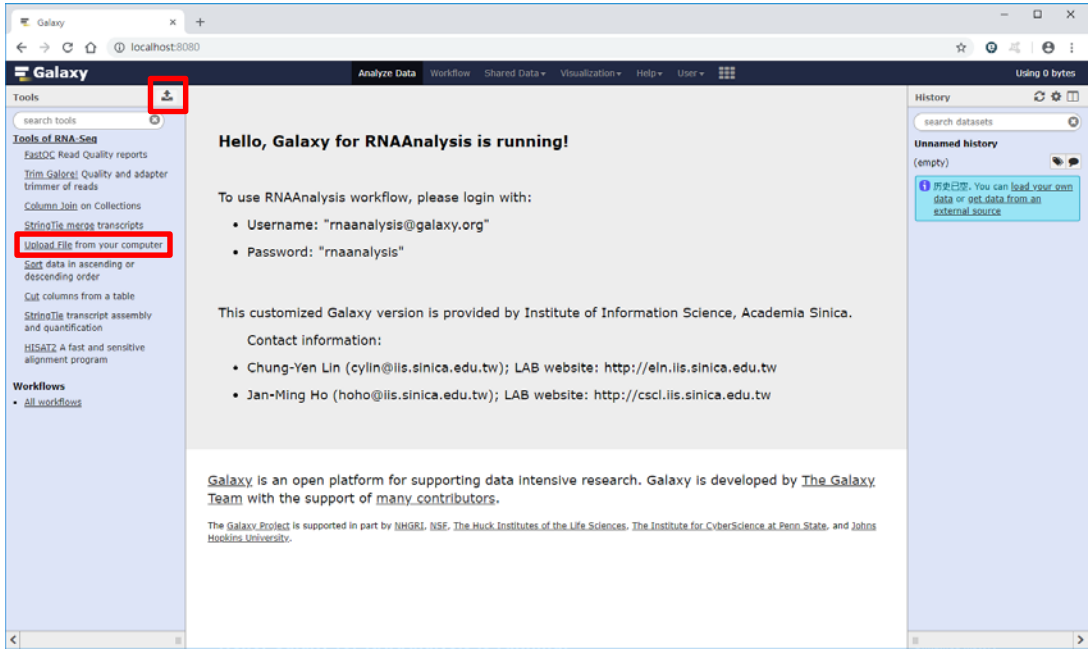
Login Galaxy

Login for full accessing all functions of Docexpress



Upload files

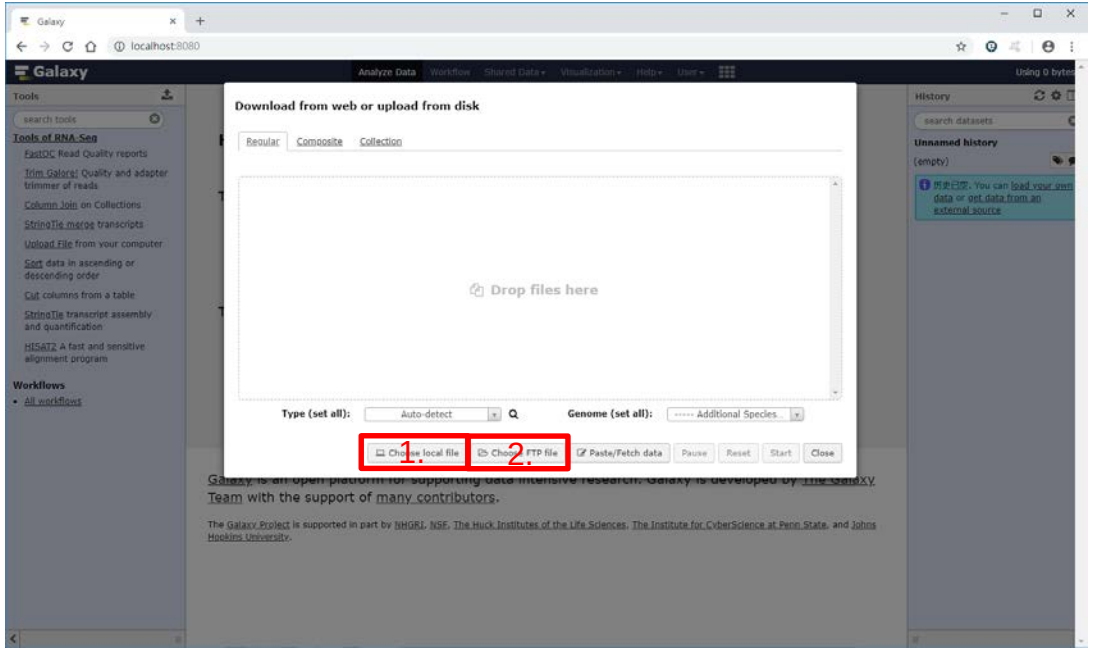
Two ways for calling the upload function



Upload files

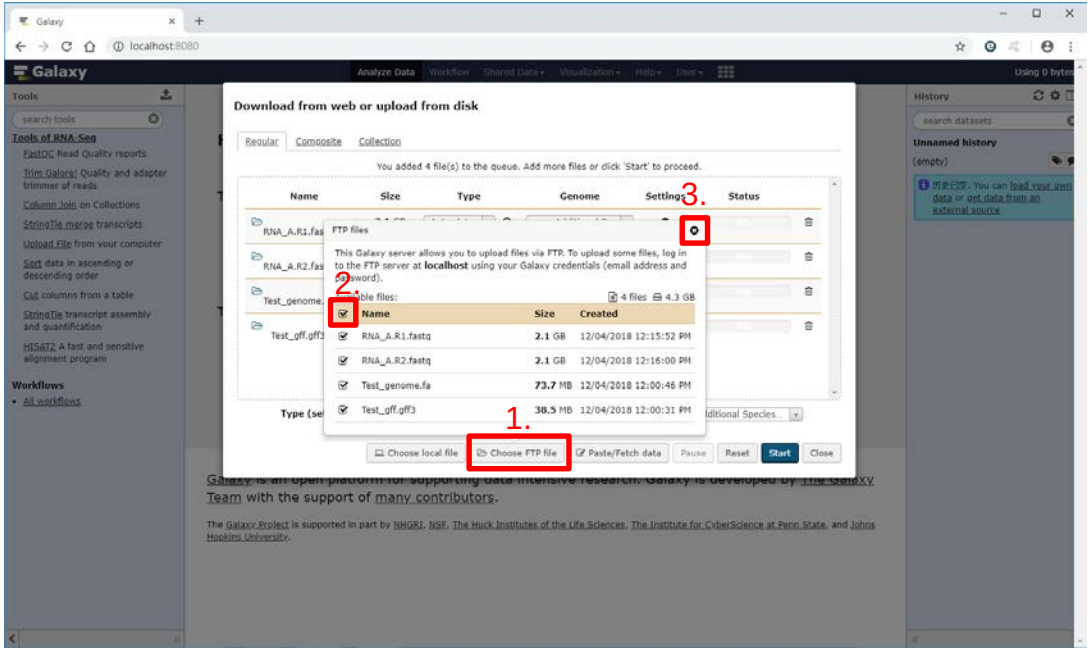
Two ways for uploading data,

1. Single file < 2 GB
2. No limitation, depend on the capability of local system



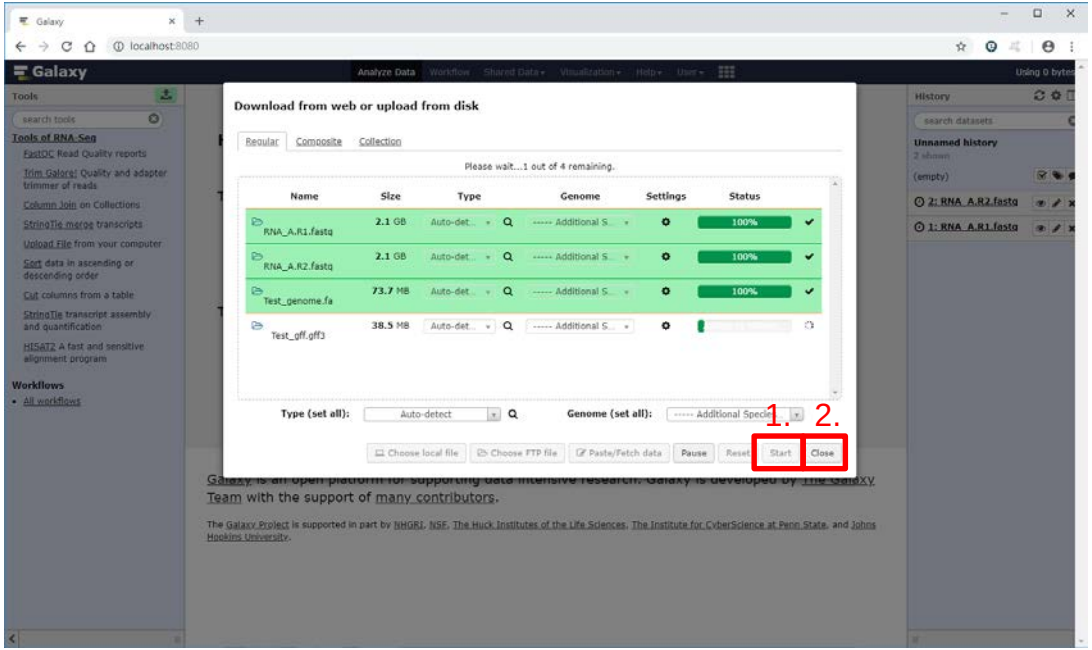
Upload files

1. Choose FTP file
2. Check all files
3. Close selection window



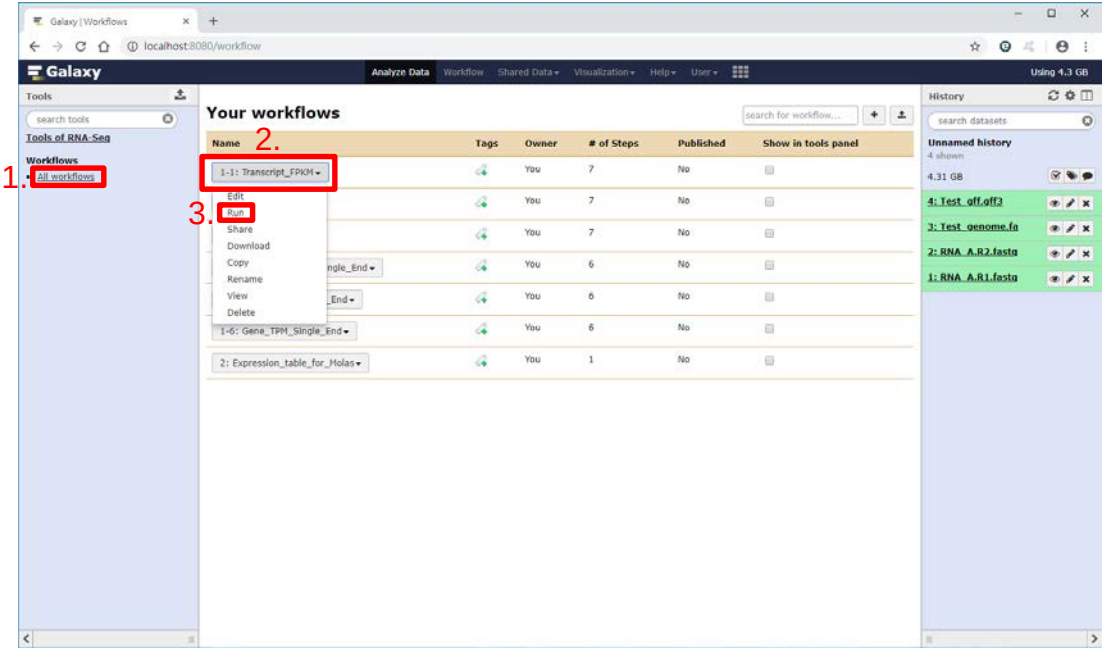
Upload files

1. Upload files
2. Close upload function



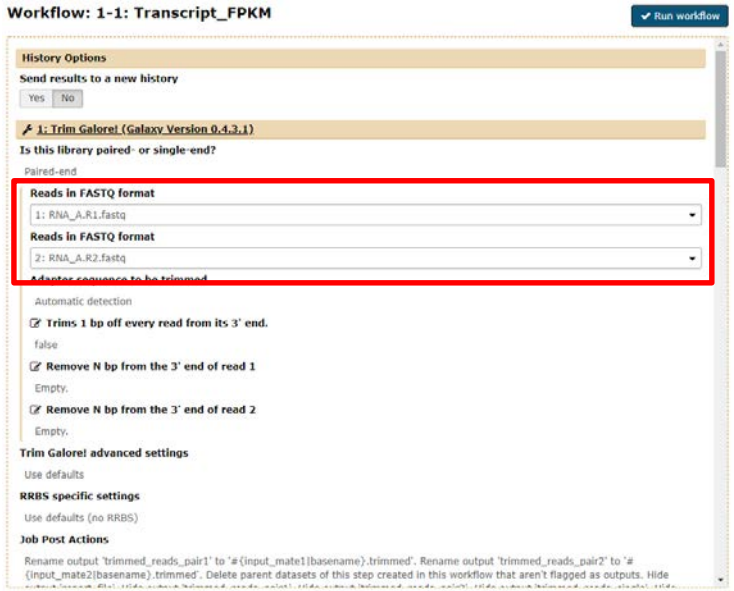
Run Workflow : 1-1: Transcript_FFKM

1. Choose “All workflows”
2. Choose “1-1: Transcription_FPKM”
3. Run the workflow



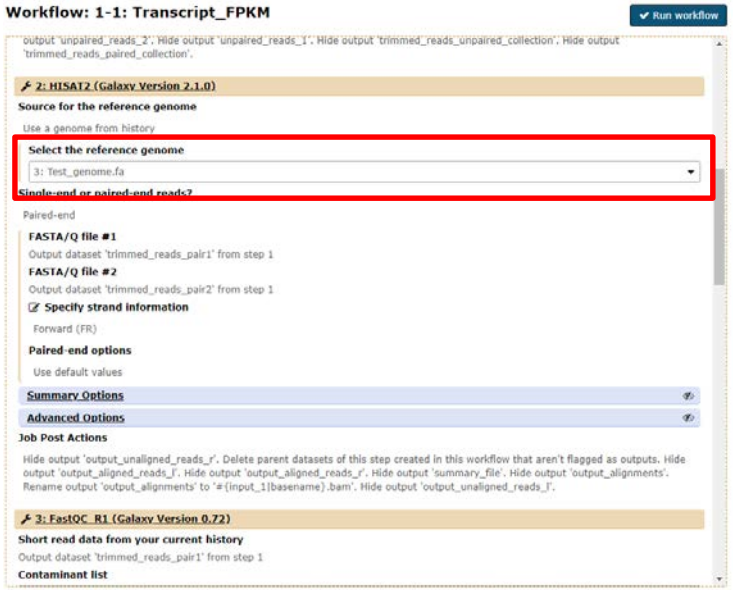
Run Workflow : 1-1: Transcript_FFKM

1. Choose RAN-seq data as the input



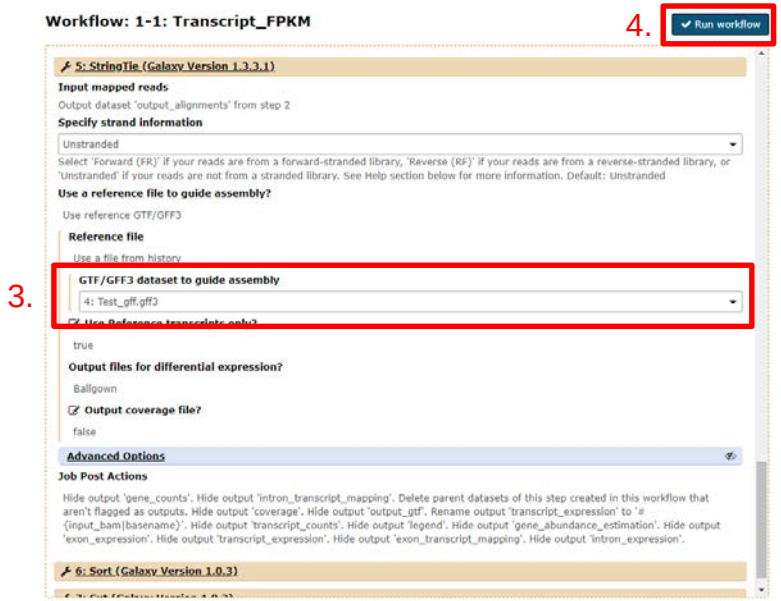
Run Workflow : 1-1: Transcript_FFKM

1. Choose RAN-seq data as the input
2. Select genome



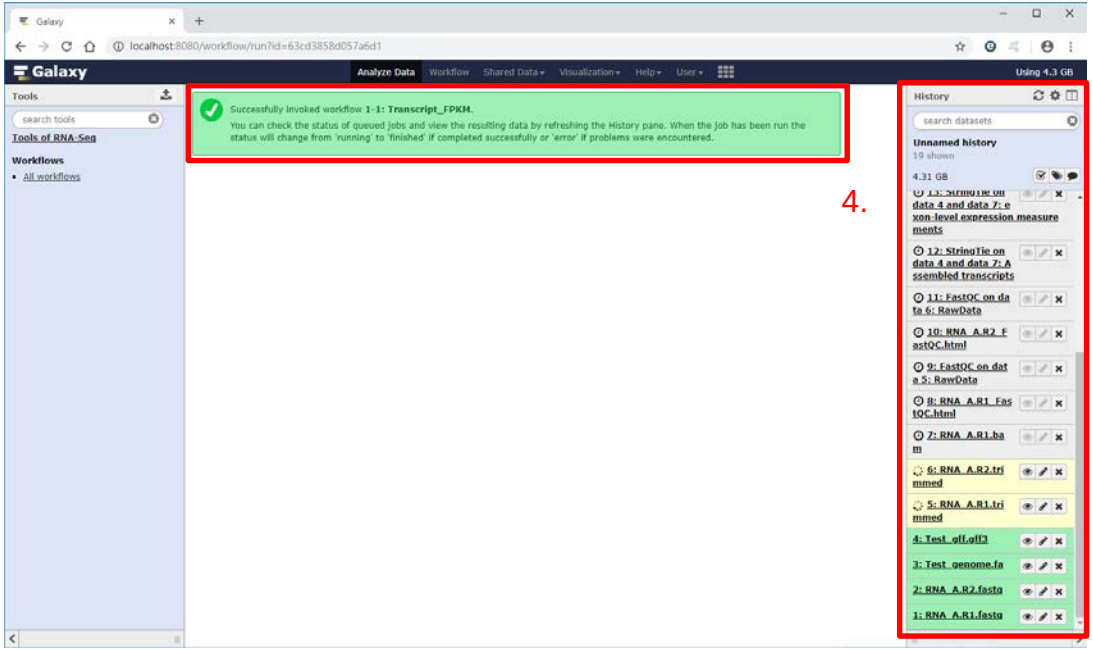
Run Workflow : 1-1: Transcript_FFKM

1. Choose RAN-seq data as the input
2. Select genome
3. Select GFF/GTF file
4. Run workflow



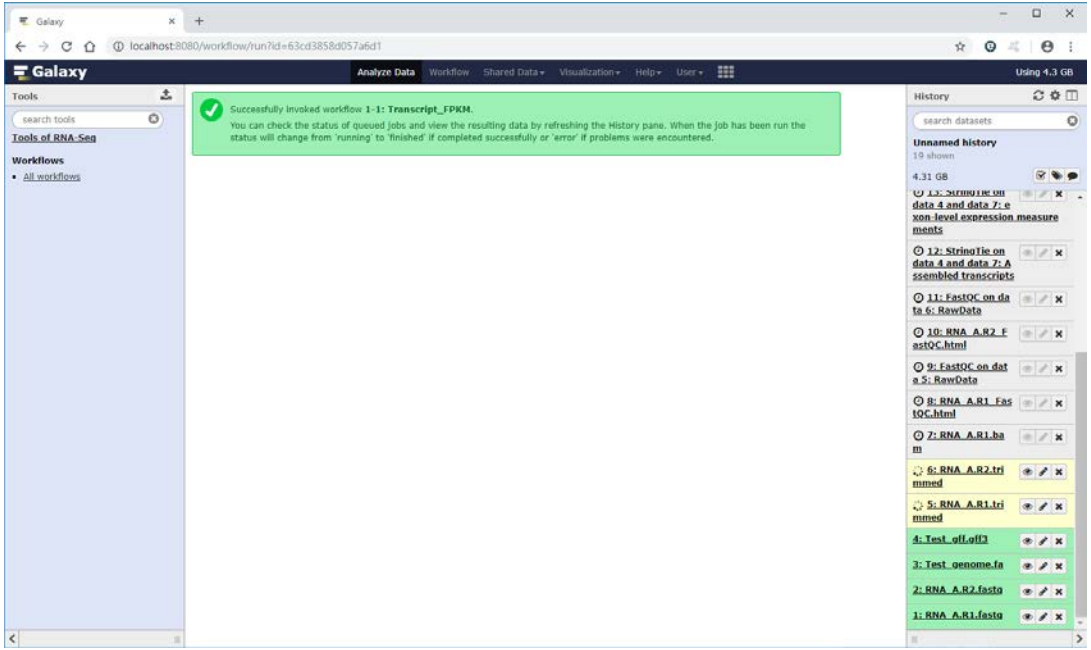
Run Workflow : 1-1: Transcript_FFKM

1. Choose RAN-seq data as the input
2. Select genome
3. Select GFF/GTF file
4. Run workflow



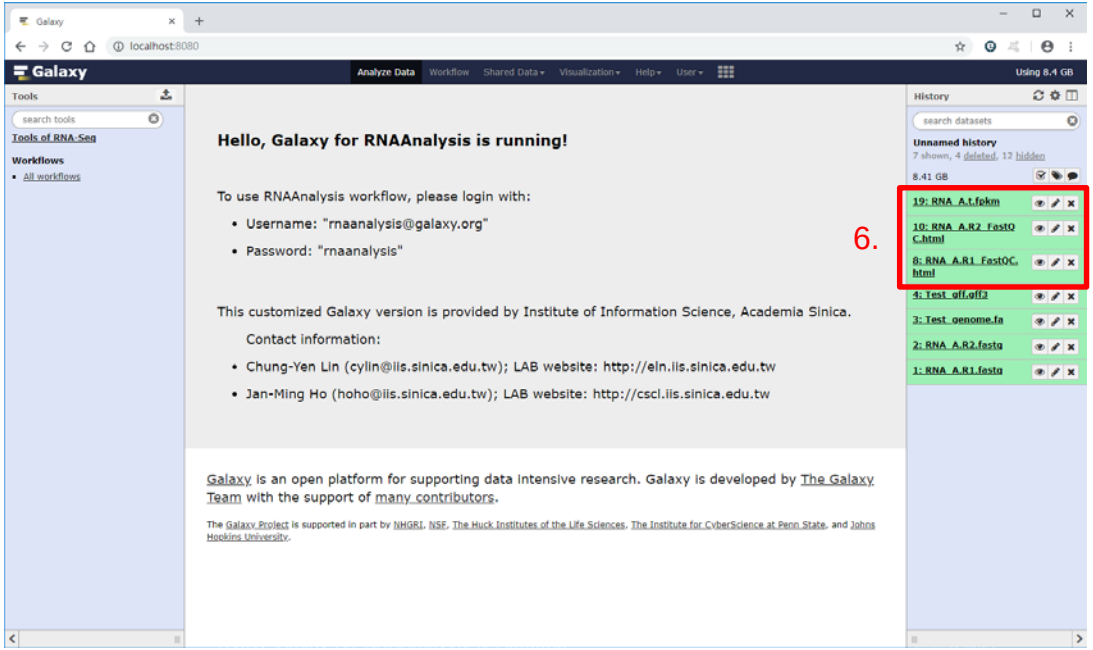
Run Workflow : 1-1: Transcript_FFKM

1. Choose RAN-seq data as the input
2. Select genome
3. Select GFF/GTF file
4. Run workflow
5. Repeat 1. - 4. until all RNA-seq data selected and execute the workflow



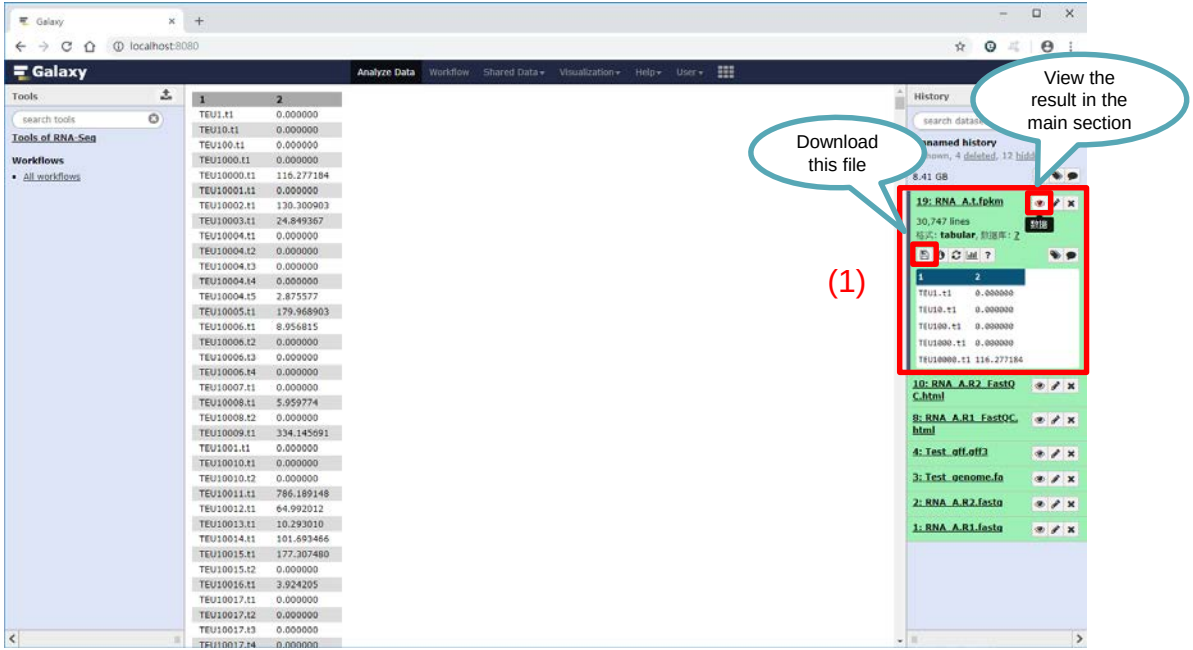
Run Workflow : 1-1: Transcript_FFKM

1. Choose RAN-seq data as the input
2. Select genome
3. Select GFF/GTF file
4. Run workflow
5. Repeat 1. - 4. until all RNA-seq data selected and execute the workflow
6. Outputs of workflow,
   (1) FastQC.html
   (2) Samlpe.t.fpkm



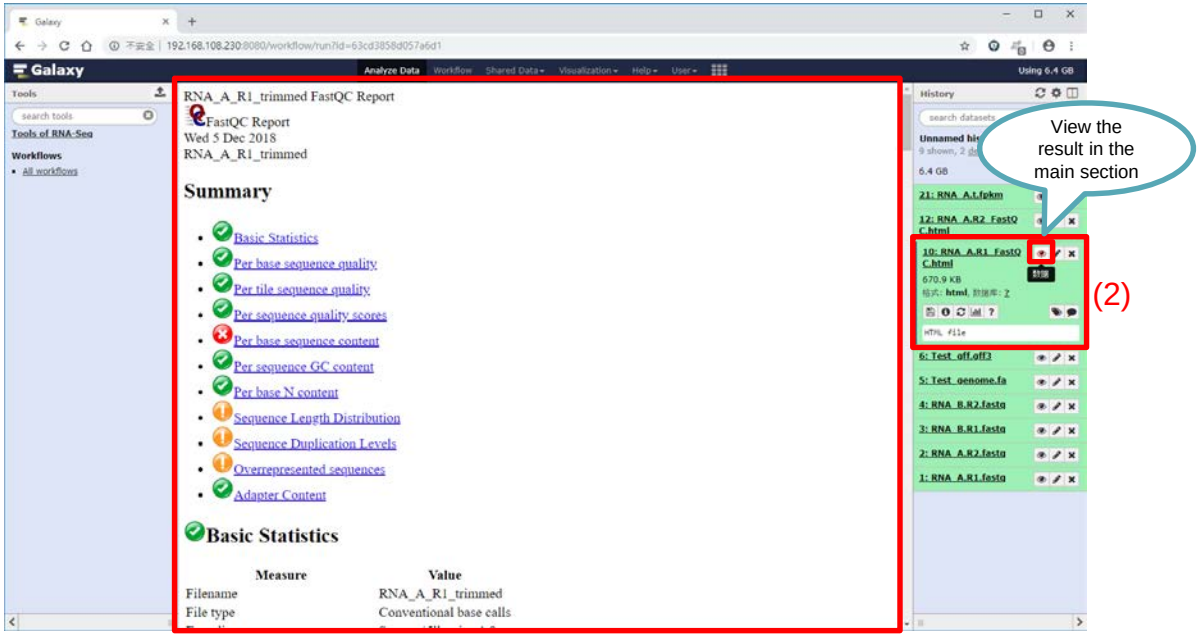
Output of Workflow : 1-1: Transcript_FFKM

- Outputs of workflow,
  (1) Samlpe.t.fpkm
  (2) FastQC.html



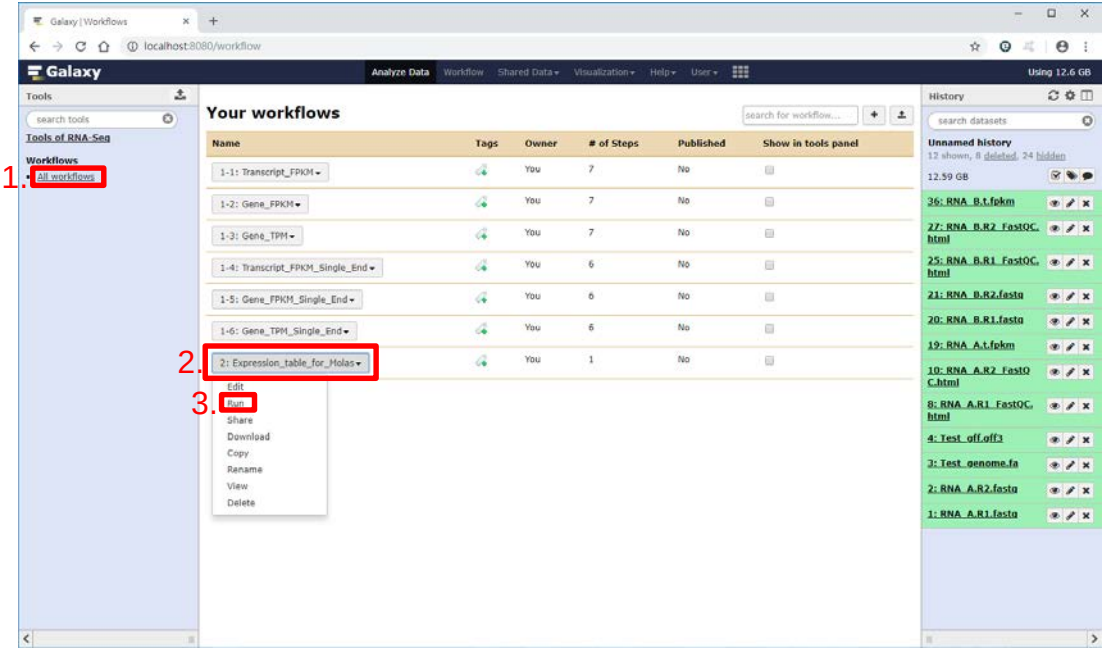
Output of Workflow : 1-1: Transcript_FFKM

- Outputs of workflow,
  (1) Samlpe.t.fpkm
  (2) FastQC.html



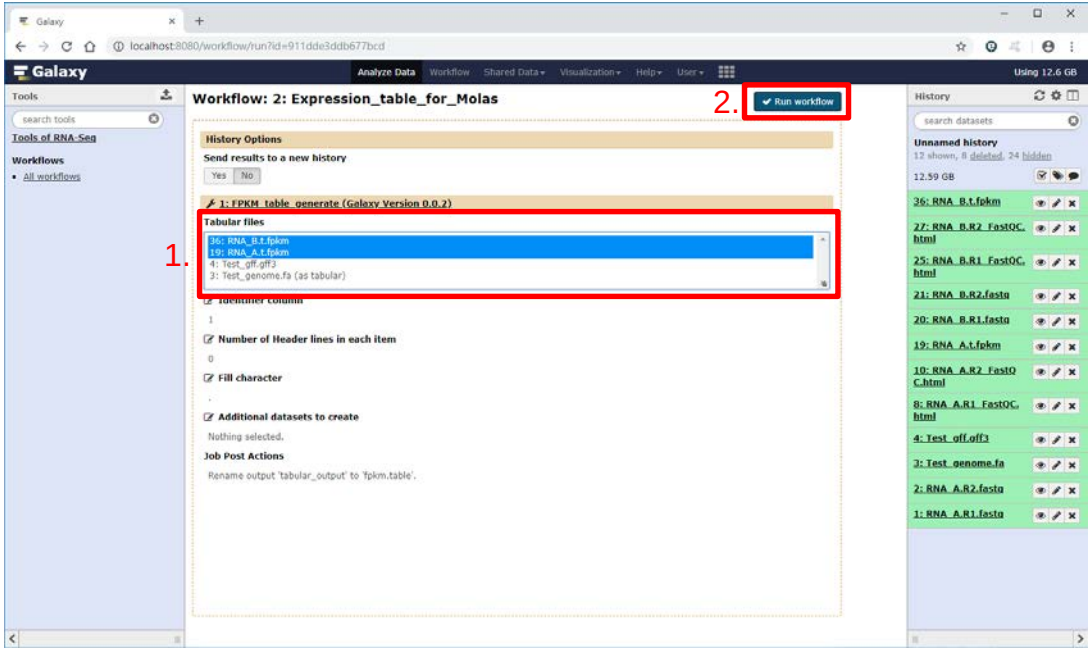
Run Workflow : 2: Expression_table_for_Molas

1. Choose “All workflows”
2. Choose “2:Expression_table_for_Molas”
3. Run the workflow



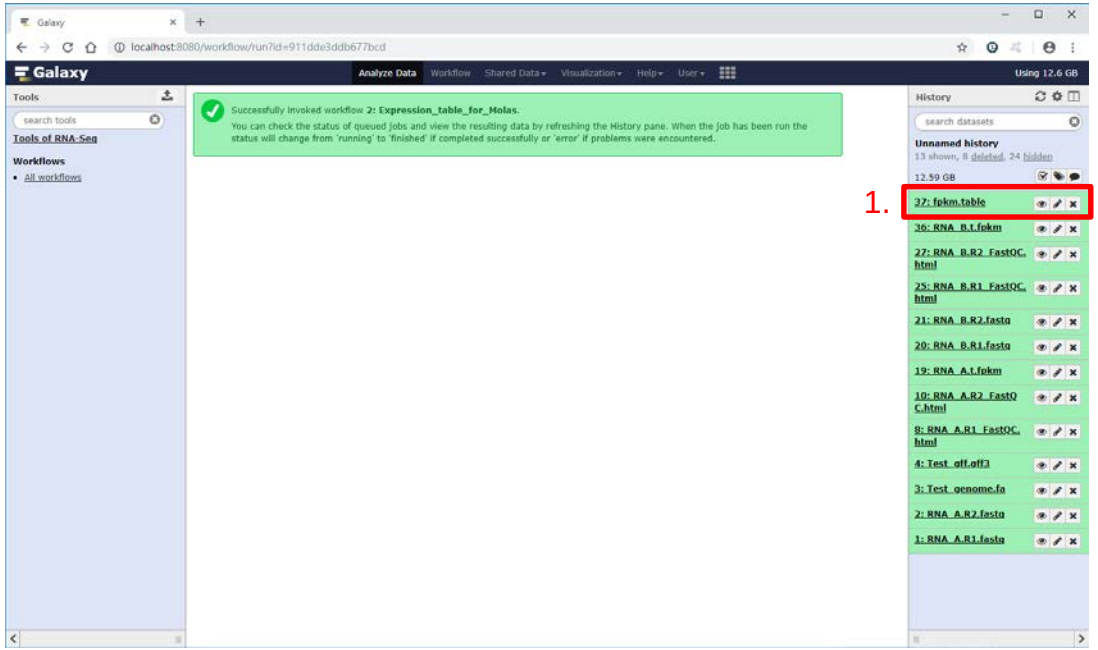
Run Workflow : 2: Expression_table_for_Molas

1. Choose all .t.fpkm files with pressing the “Ctrl”
2. Run workflow



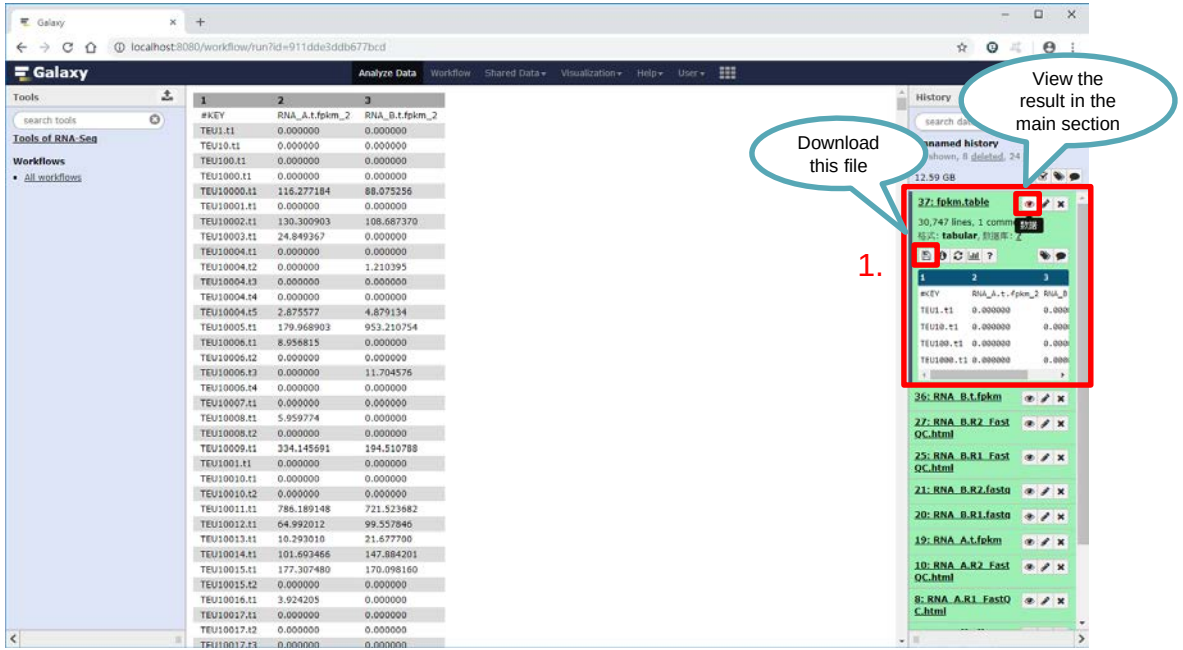
Run Workflow : 2: Expression_table_for_Molas

1. Result: fpkm.table



Run Workflow : 2: Expression_table_for_Molas

1. View and download the fpkm.table for MOLAS system



3.

INTRODUCTION FOR DOCMETHYL

DocMethyl

» We packed a Docker container DocMethyl to deal with raw data processing, mapping, and methylation calling/ scoring to give the summary, mtable, of the whole genome methylation status by the gene.

» Mtables are uploaded to the web server EpiMOLAS_web for linking with gene annotation databases that enable rapid data retrieval and analyses.

1. Install Docker on your local system [Ref.]
2. Open a terminal or Use command mode
3. Following the “Install & Usage” on our docker hub [link]

Any questions?

LAB OF System Biology & Network Biology
中央研究院資訊科學研究所 @iis, Academia Sinica, TAIWAN
系統生物學與網路生物學實驗室

References

1. Docker - https://www.docker.com/
2. Docker Install - https://docs.docker.com/install/linux/docker-ce/ubuntu/
3. Galaxy - https://usegalaxy.org/
4. Docexpress - https://hub.docker.com/r/lsbnb/docexpress/
5. DocMethyl - https://hub.docker.com/r/lsbnb/docmethyl/
6. Galaxy basic usage https://galaxyproject.github.io/training-material/topics/introduction/tutorials/galaxy-intro-101/tutorial.html