

## Help document

There are four sections about how to use analysis tools implemented in EAP.

[Section 1. Requirements:](#) User registration and application for storage space.

[Section 2. Upload raw sequencing data and metadata:](#) Data transferring client installation and data transferring.

[Section 3. ChIP/ATAC-seq data preprocessing:](#) Transform raw sequencing data into analysis ready data.

[Section 4. Down-stream analysis tools:](#) Transform analysis ready data into interpretable and biological meaningful results.

### Section 1. Requirements

To run EAP, user need to register an account (url: <https://www.biosino.org/epigenetics/#/user/register>). After registration, user should send an application form (application form: [File-1](#)) to administrator (e-mail addresses are available on **Home page** and [this document](#)) to get approval to login and create storage space. Once user have an account and storage space, user can upload raw sequencing data to the storage space and run analysis workflows and analysis tools implemented in EAP. [\[back\]](#)

To register a user account as below:

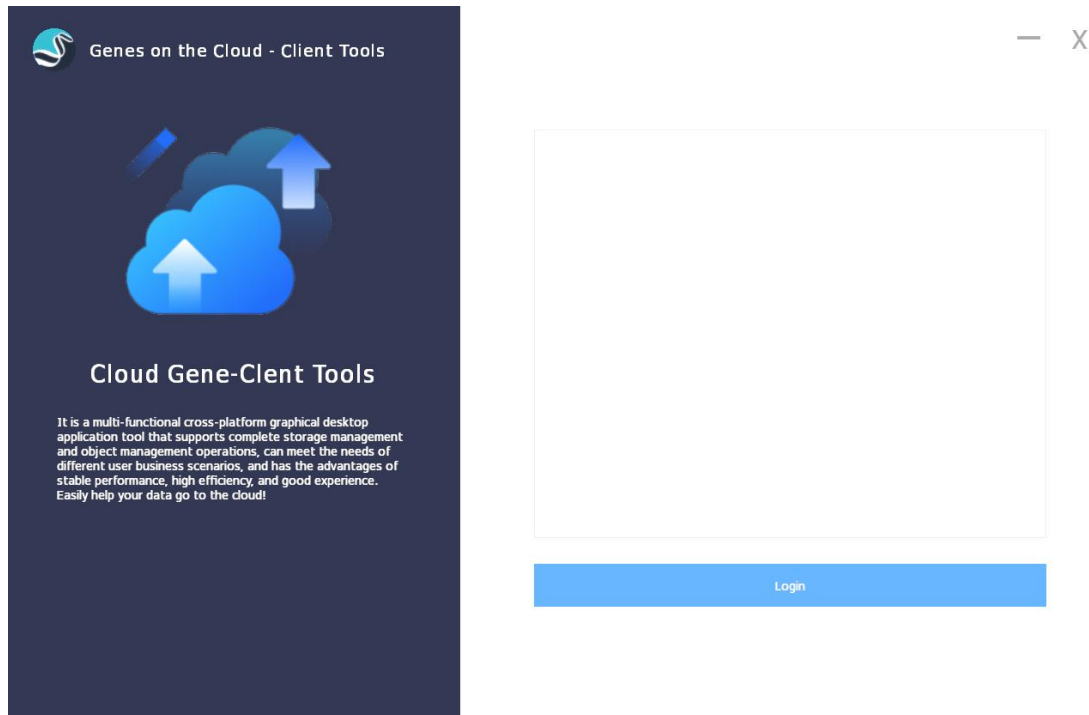


### Section 2. Upload raw sequencing data and metadata

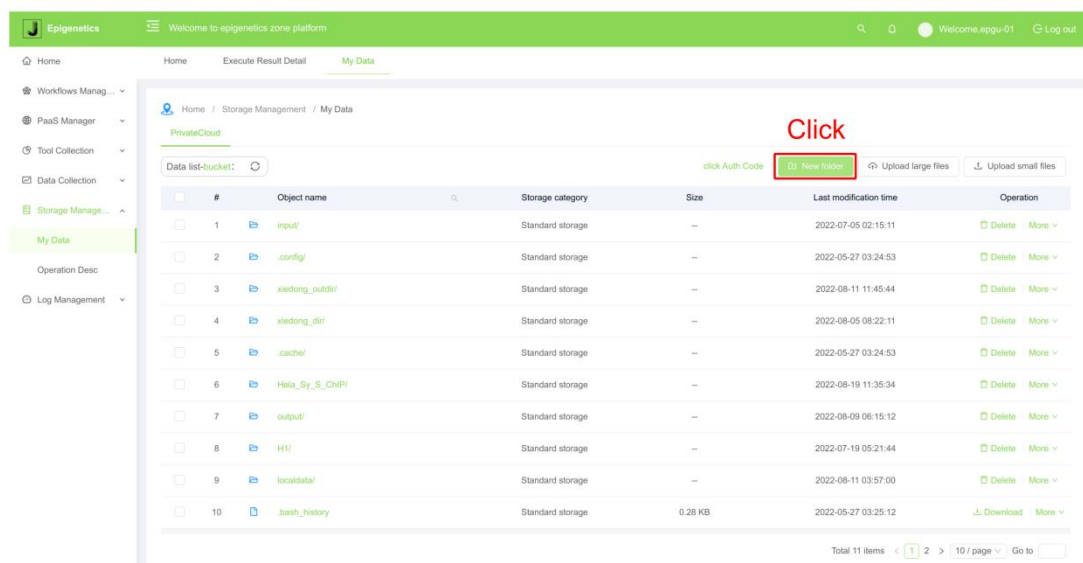
User can perform the analysis tools implemented in EAP from either raw sequencing data (i.e. FASTQ format files) or analysis ready data (following the ChIP/ATAC-seq data processing guideline in MAnorm2-utils: <https://pypi.org/project/MAnorm2-utils/>). Before running analysis tools, user need to prepare the metadata file (study design information), which describes the raw

sequencing data and the phenotypic information for each sample (metadata template could be found on **Home** page: [File-2](#)).

If user starts from analysis ready data, skips the following steps and continues to [Section 4](#). Other users should download the client compressed package from the platform (**Menu bar: Storage Management > Operation Desc**), choose an appropriate version, unzip and install it, then run the client. The running client is shown below:



Then creates a new folder (e.g. LUAD\_H3K27ac\_ChIP\_seq) in user's storage space (**Menu bar: Storage Management > My data**), click on **New folder** and give it an appropriate name. Click on **OK** to create a new folder, as a folder to store raw sequencing data and metadata file. And user can create another folder as output folder.



## New folder



Enter an appropriate name

\* Folder name: LUAD\_H3K27ac\_ChIP\_seq

Naming rules:

- Supports the creation of single folders and multi-level folders.
- A single folder name cannot contain the following characters : \:\*?"<>|.
- Folder name cannot be in English period (.) Or the beginning or end of a slash (/).
- A single slash (/) separates and creates multiple levels of folders.
- The total absolute path length of a folder cannot exceed 1023 characters.
- Cannot contain more than two adjacent slashes (/).

关闭

OK

Click on [More](#) and [Copy authorization code](#) (for the destination folder, **LUAD\_H3K27ac\_ChIP\_seq in this example**) from web page and paste it into the input box of client, click on [Login](#).

Home / Storage Management / My Data

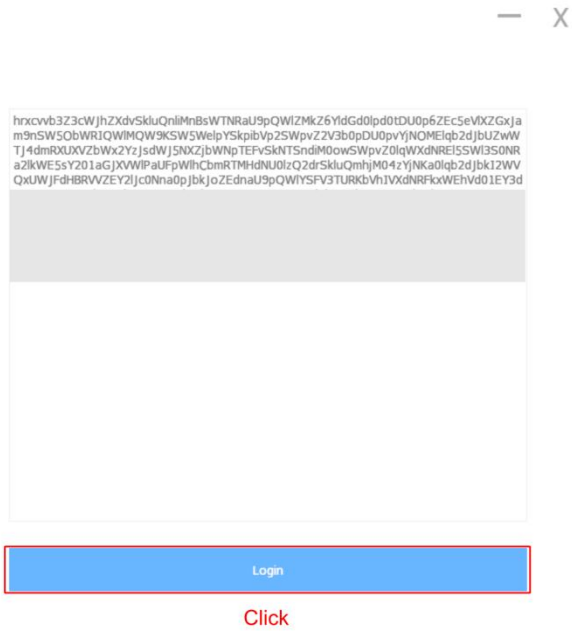
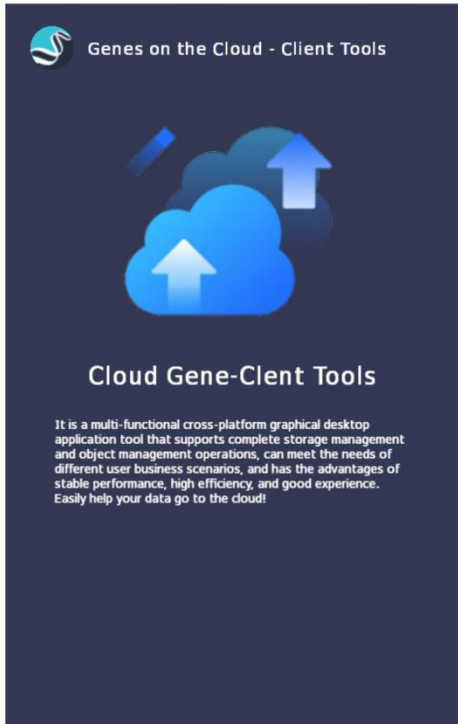
PrivateCloud

Data list-bucket:

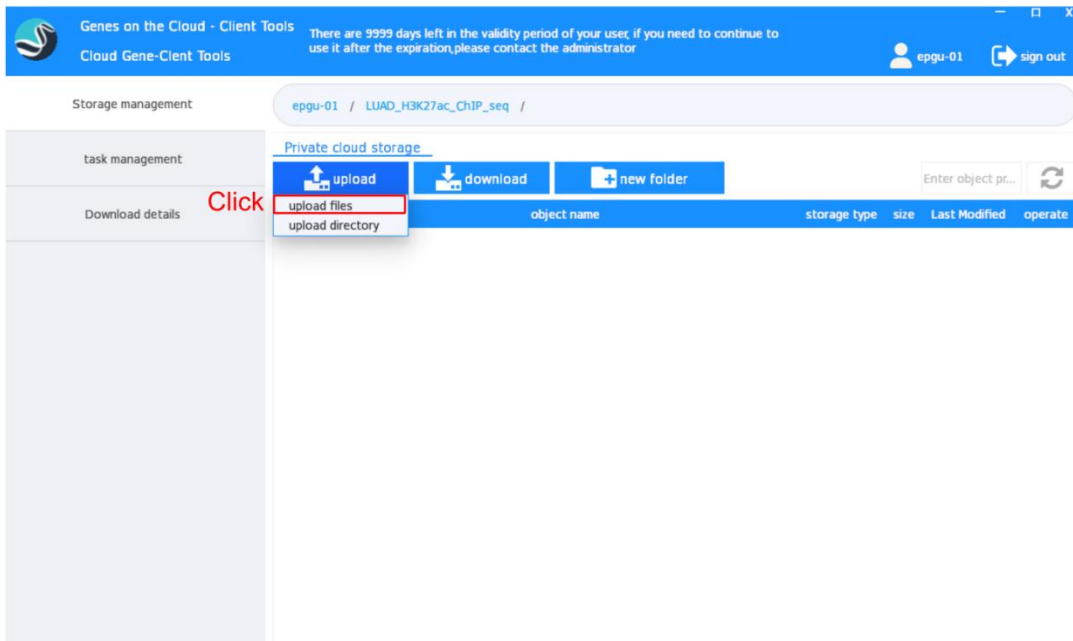
click Auth Code New Folder Upload large files Upload small files

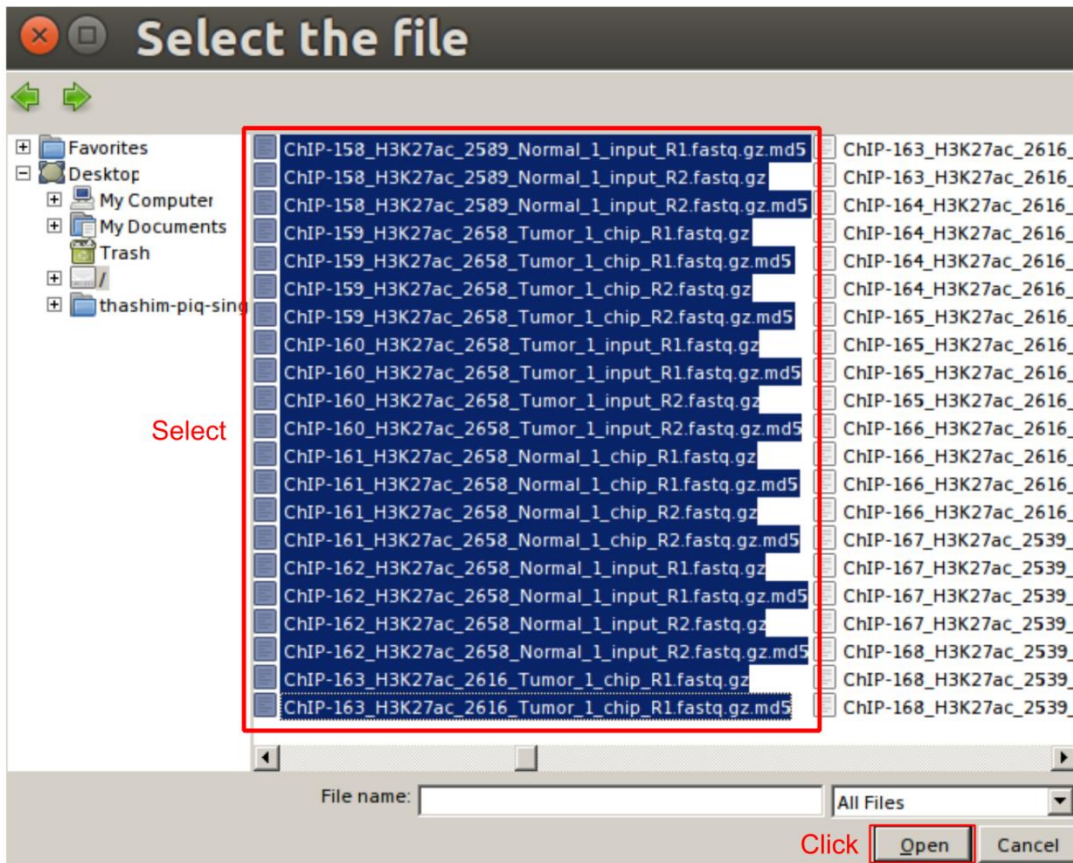
#	Object name	Storage category	Size	Last modification time	Operation
1	input/	Standard storage	--	2022-07-05 02:15:11	Delete More
2	config/	Standard storage	--	2022-05-27 03:24:53	Delete More
3	xiedong_outdir/	Standard storage	--	2022-08-11 11:45:44	Delete More
4	xiedong_dir/	Standard storage	--	2022-08-05 08:22:11	Delete More
5	cache/	Standard storage	--	2022-05-27 03:24:53	Delete More
6	HeLa_Sy_S_ChIP/	Standard storage	--	2022-08-19 11:35:34	Delete More
7	output/	Standard storage	--	2022-08-09 06:15:12	Delete More
8	bash_history	Standard storage	0.28 KB	2022-05-27 03:25:12	Download More
9	LUAD_H3K27ac_ChIP_seq/	Standard storage	--	2022-08-20 09:21:40	Delete More
10	opt/	Standard storage	--	2022-08-20 09:21:40	Delete More

Total 10 items < 1 > 50 / page

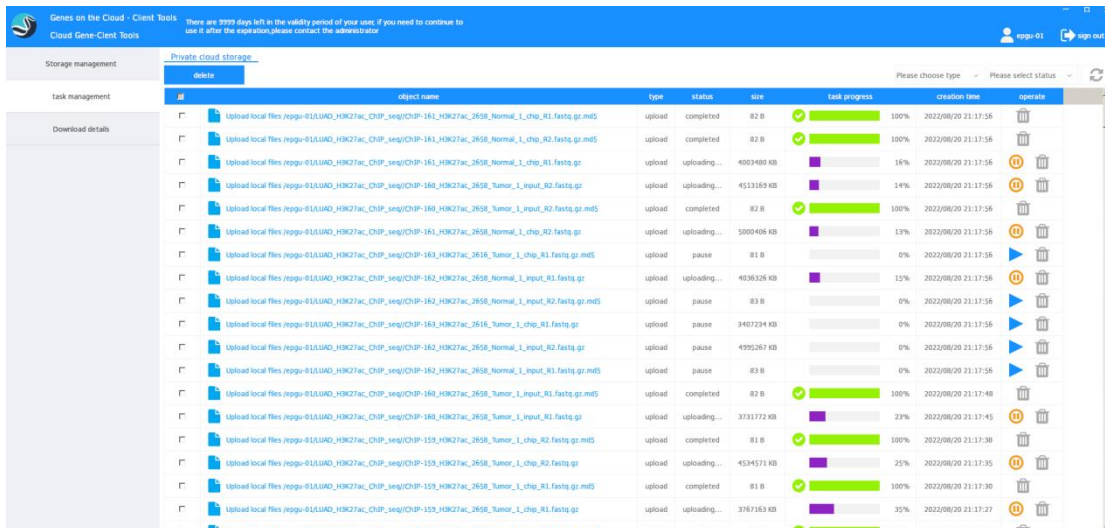


Select the raw sequencing data and corresponding metadata files to upload. While EAP does not impose a strict file size limit, we recommend keeping each upload below 50 GB in total. For larger datasets, we suggest uploading the files in separate batches. **Please note that this is a recommended threshold, not a hard limit.**





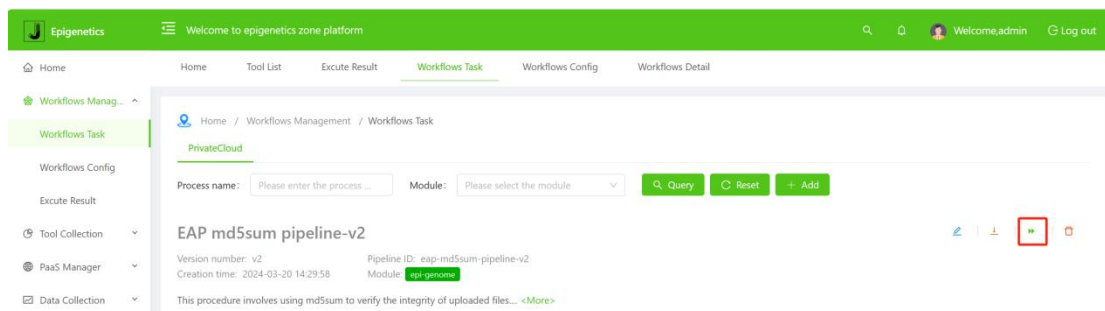
Once the raw sequencing data and metadata are under transferring, the upload task progress could be checked in the **task management** section in the client. If the transferring task is completed, user can continue to next step. [\[back\]](#)



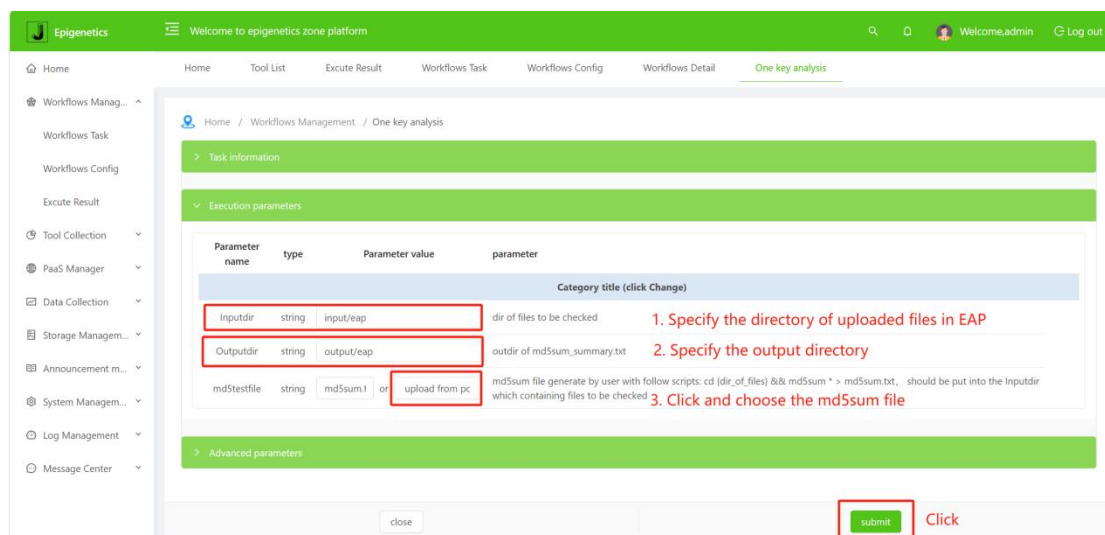
### Section 3. ChIP/ATAC-seq data preprocessing

Before running the data preprocessing analysis, we recommend that users apply the md5sum checking procedure to verify the integrity of the uploaded files. First, running command “cd [files directory] && md5sum \* > md5sum.txt” in your local system to create the md5sum for

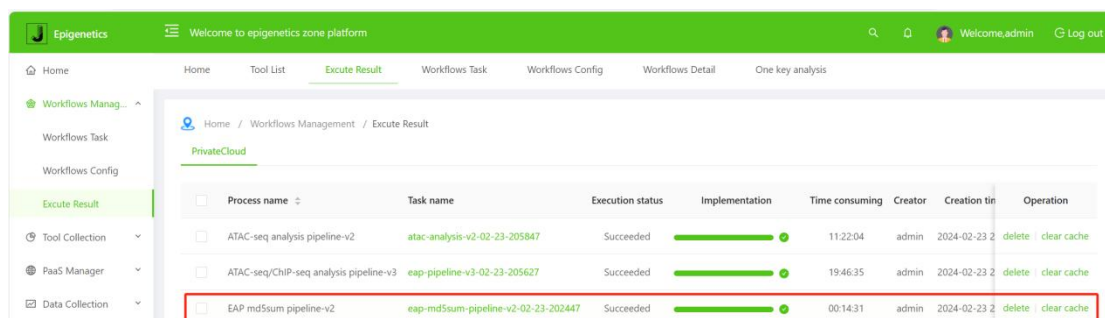
each file. Then navigate to **Workflows Task (Menu bar: Workflows management > Workflows Task)** and choose the EAP md5sum pipeline, click on the **running button** (Rectangle in red in the image below).



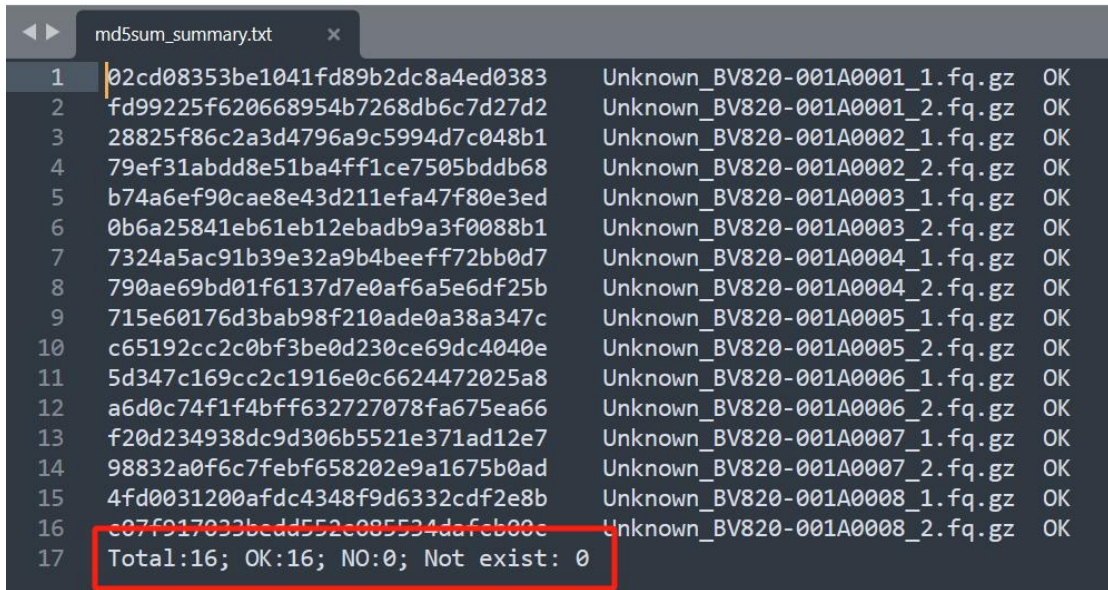
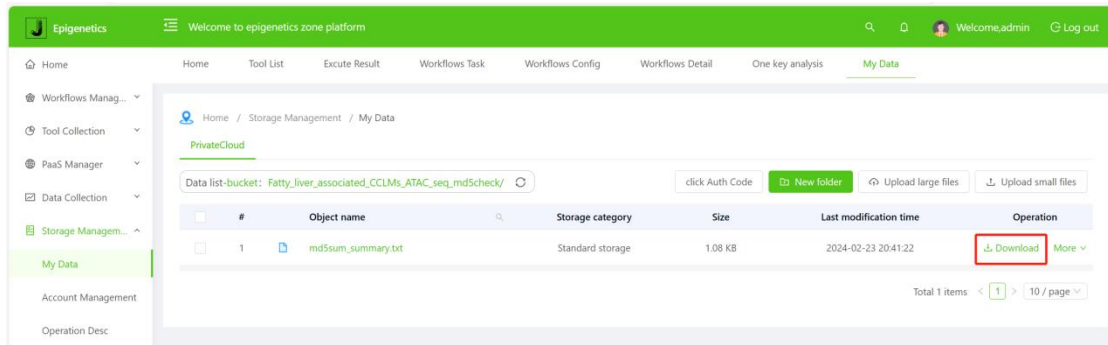
A screen like the image below will appear. Then specify the input and output directories and choose the md5sum file for upload. Click on **submit** to run the workflow (Rectangle in red in the image below).



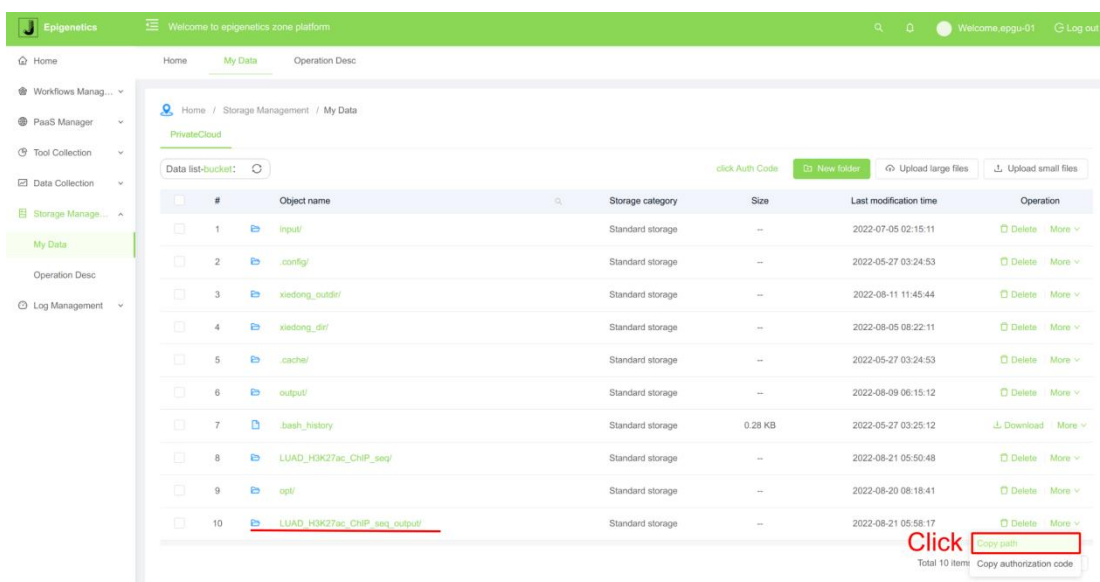
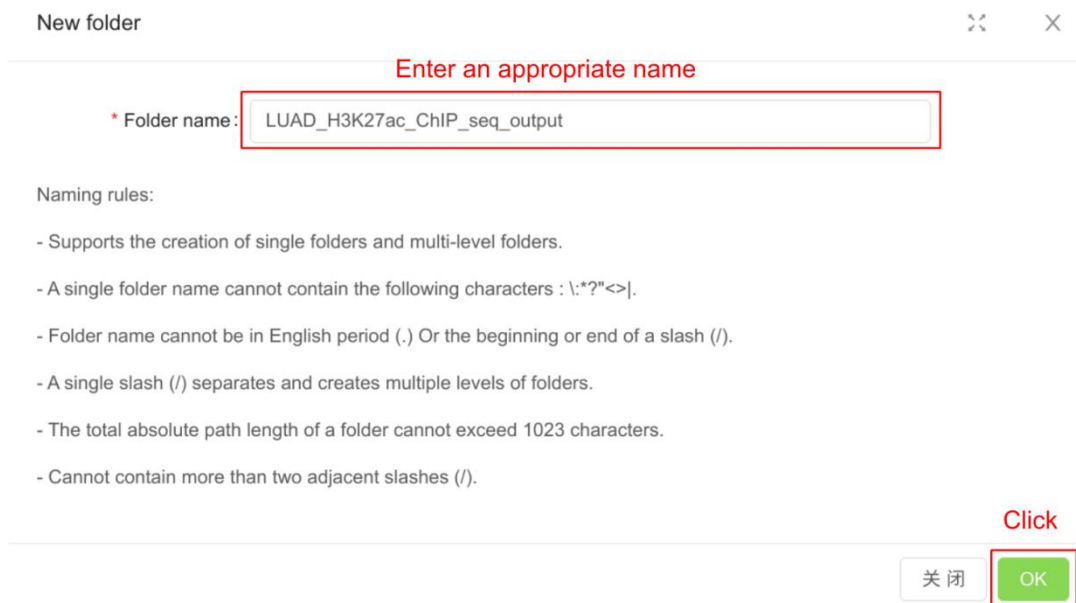
The progress of running task could be monitored in the **Excute Result from Workflows Management (Menu bar: Workflows Management > Excute Result)**.



Upon completion of this pipeline, user can navigate to output directory in **Storage Management (Menu bar: Storage Management > My Data > output directory)** and download and check the summary result.



For ChIP/ATAC-seq data preprocessing, EAP provides a standardized analysis pipeline to quality control, read alignment, peak calling and read counting, and also creates a PDF report including quality control plots and summary statistics to facilitate filtering poor quality samples. User should make sure the raw sequencing data and metadata have been uploaded to user's storage space before running the preprocessing workflow (Data preprocessing module). First, create a new folder for output result files (**LUAD\_H3K27ac\_ChIP\_seq\_output** in this example, refer to *Section 2*).



Then navigate to **Workflows Task (Menu bar: Workflows management > Workflows Task)** and choose the EAP data preprocessing pipeline, click on the **running button** (Rectangle in red in the image below).



A screen like the image below will appear. Then fill out the task identifier **Task name** with an appropriate name (**luad-h3k27ac-chip-seq-20220821 in this example**). All required parameters must be filled (Details see the parameter description).

Task information

\* Task name:

Timeout:  day  hour  minute

priority:  Priority range: [0-999]  
reminder: High value priority

Image pull policy:  Use last pulled image(recommend)  Each re pull

retry count:  Retry count range: [0-9]

Click on **submit** to run the workflow (Rectangle in red in the image below).

Execution parameters

Parameter name	type	Parameter value	parameter
Category title (click Change)			
Inputdir	string	<input type="text" value="LUAD_H3K27ac_ChIP_seq/"/>	data dir for your input files
Outputdir	string	<input type="text" value="LUAD_H3K27ac_ChIP_seq_output/"/>	result dir for all samples
sequencing_type	string	<input type="text" value="ChIPPE"/>	Single end CHIP or Paired end CHIP or ATAC,choose between ATAC, ChIPPE and CHIPSE
information_file	string	<input type="text" value="LUAD_H3K27ac_ChIP_seq.txt"/> or <input type="button" value="upload from pc"/>	information text,should be deposited in input_dir. <a href="#">template</a>
REF_INDEX	string	<input type="text" value="hg19"/>	Genome index,choose among :hg19,hg38,mm9,mm10
typical_bin_size	number	<input type="text" value="2000"/>	typical bin size for profile-bins
variable_of_interest	string	<input type="text" value="tissue_type"/>	interested variable for differetial analysis
project_name	string	<input type="text" value="LUAD_H3K27ac_ChIP_seq"/>	name of project

> Advanced parameters

The progress of running task could be monitored in the **Excute Result** from **Workflows Management (Menu bar: Workflows Management > Excute Result)**

Home / Workflows Management / Excute Result

PrivateCloud

Process name	Task name	Execution status	Implementation	Time consuming	Creator	Creation time	Operation
<input type="checkbox"/> ATAC-seq/ChIP-seq analysis pipeline-v3	<a href="#">luad-h3k27ac-chip-seq-20220821</a>	Running	<div style="width: 18%;"><div style="width: 18%;"></div></div> 18%	00:03:06	epgu-01	2022-08-21 21:59:23	<a href="#">delete</a> <a href="#">clear cache</a>
<input type="checkbox"/> ATAC-seq/ChIP-seq analysis pipeline-v3	<a href="#">eap-pipeline-v3-08-19-133854</a>	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div> 100%	01:14:05	epgu-01	2022-08-19 13:40:48	<a href="#">delete</a> <a href="#">clear cache</a>
<input type="checkbox"/> ATAC-seq/ChIP-seq analysis pipeline-v3	<a href="#">eap-pipeline-v3-08-18-152545</a>	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div> 100%	03:23:35	epgu-01	2022-08-18 15:27:00	<a href="#">delete</a> <a href="#">clear cache</a>
<input type="checkbox"/> ChIP-seq analysis pipeline-v2	<a href="#">chip-analysis-v2-08-02-124031</a>	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div> 100%	01:28:33	epgu-01	2022-08-02 12:41:50	<a href="#">delete</a> <a href="#">clear cache</a>
<input type="checkbox"/> ATAC-seq analysis pipeline-v2	<a href="#">atac-analysis-v2-08-02-110556</a>	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div> 100%	00:57:33	epgu-01	2022-08-02 11:06:37	<a href="#">delete</a> <a href="#">clear cache</a>

Total 5 items < 1 > 10 / page

Details could be accessed by clicking on **Task name**, then the running status of the task as shown below:

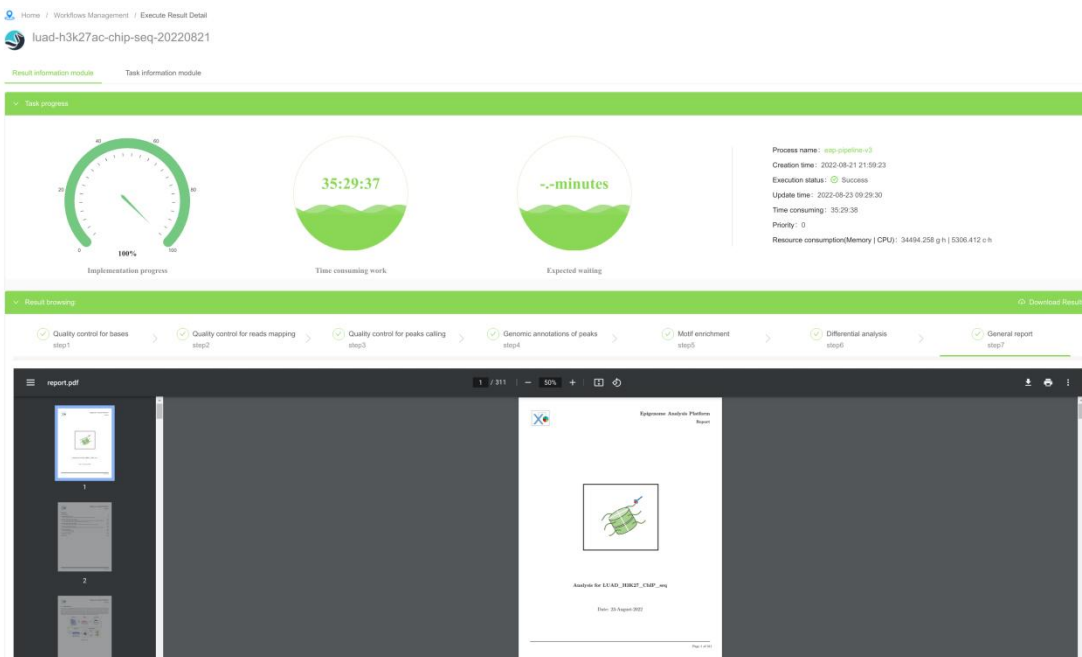


After data preprocessing, a PDF report will be generated, including quality control plots and summary statistics (**Menu bar: Workflows Management > Excute Result**). User can filter poor quality samples based on this summary report.

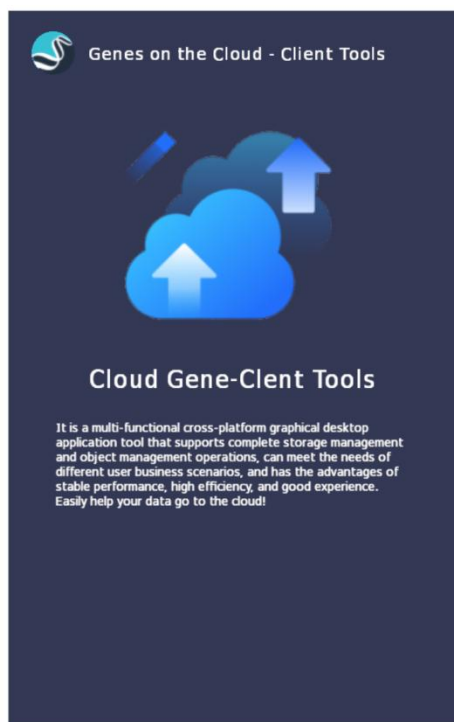
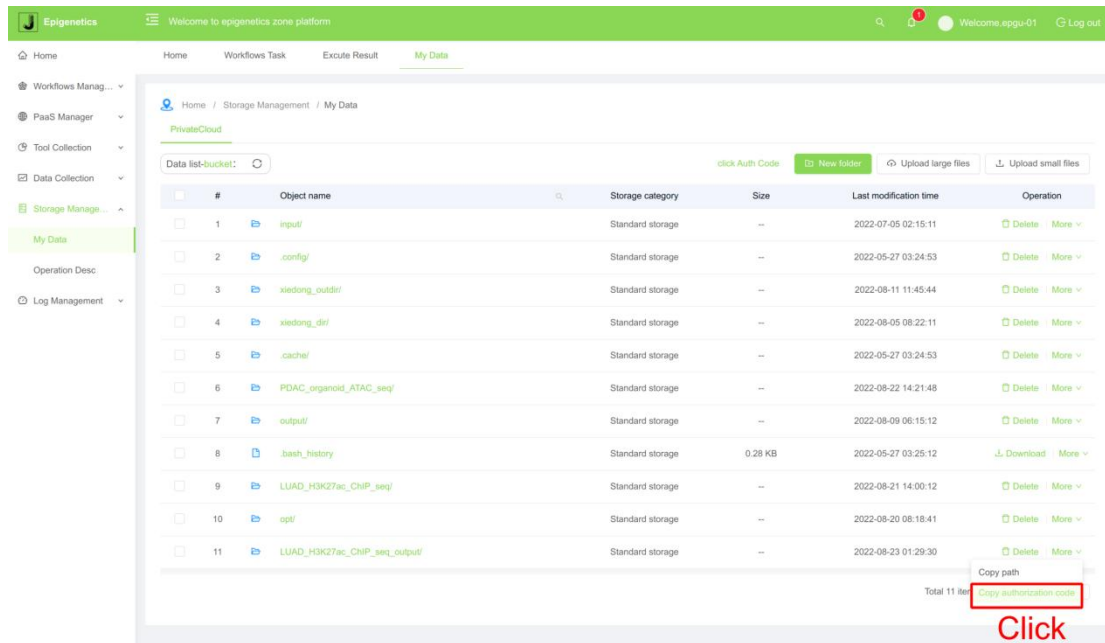
Home / Workflows Management / Execute Result  
PrivateCloud

Process name	Task name	Execution status	Implementation	Time consuming	Creator	Creation time	Operation
<input type="checkbox"/> ATAC-seq/CHIP-seq analysis pipeline-v3	luad-h3k27ac-chip-seq-20220821	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div>	35:29:37	epgu-01	2022-08-21 21:59:23	<a href="#">delete</a> <a href="#">clear cache</a>
<input type="checkbox"/> ATAC-seq/CHIP-seq analysis pipeline-v3	eap-pipeline-v3-08-18-152545	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div>	03:23:35	epgu-01	2022-08-18 15:27:00	<a href="#">delete</a> <a href="#">clear cache</a>
<input type="checkbox"/> CHIP-seq analysis pipeline-v2	chip-analysis-v2-08-02-124031	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div>	01:28:33	epgu-01	2022-08-02 12:41:50	<a href="#">delete</a> <a href="#">clear cache</a>
<input type="checkbox"/> ATAC-seq analysis pipeline-v2	atac-analysis-v2-08-02-110556	Succeeded	<div style="width: 100%;"><div style="width: 100%;"></div></div>	00:57:33	epgu-01	2022-08-02 11:08:37	<a href="#">delete</a> <a href="#">clear cache</a>

Total 4 items < 1 > 10 / page v

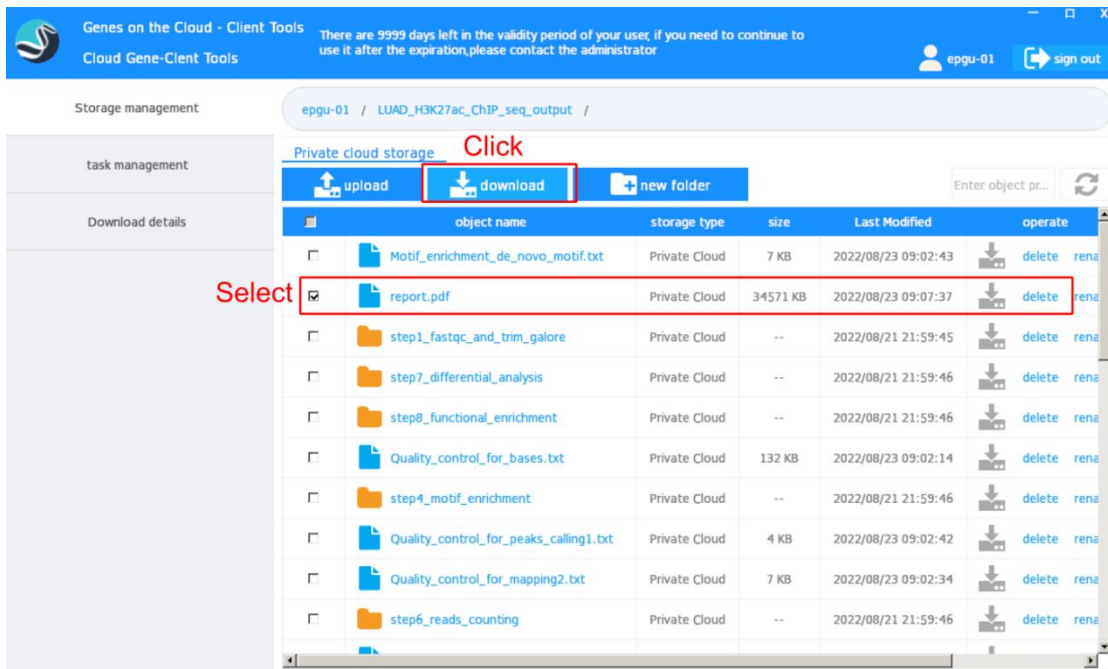


The complete results including all output tables and figures could be found in the output folder and all files could be downloaded through the client tool. Here we demonstrate how to download summary report and count tables. Click on **More** and **Copy authorization code** (for the output folder, **LUAD\_H3K27ac\_ChIP\_seq\_output** in this example) from web page and paste it into the input box of client, click on **Login**.

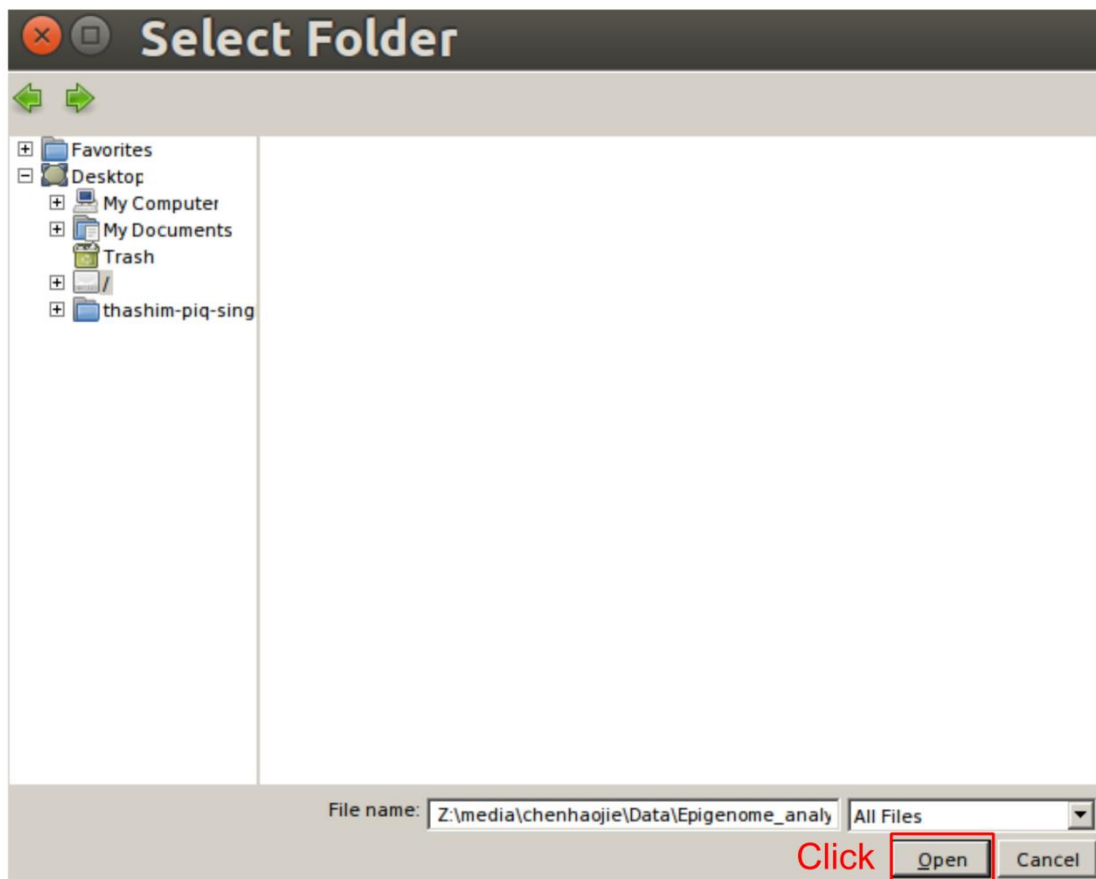


Select files (or folder) and click on **download**.

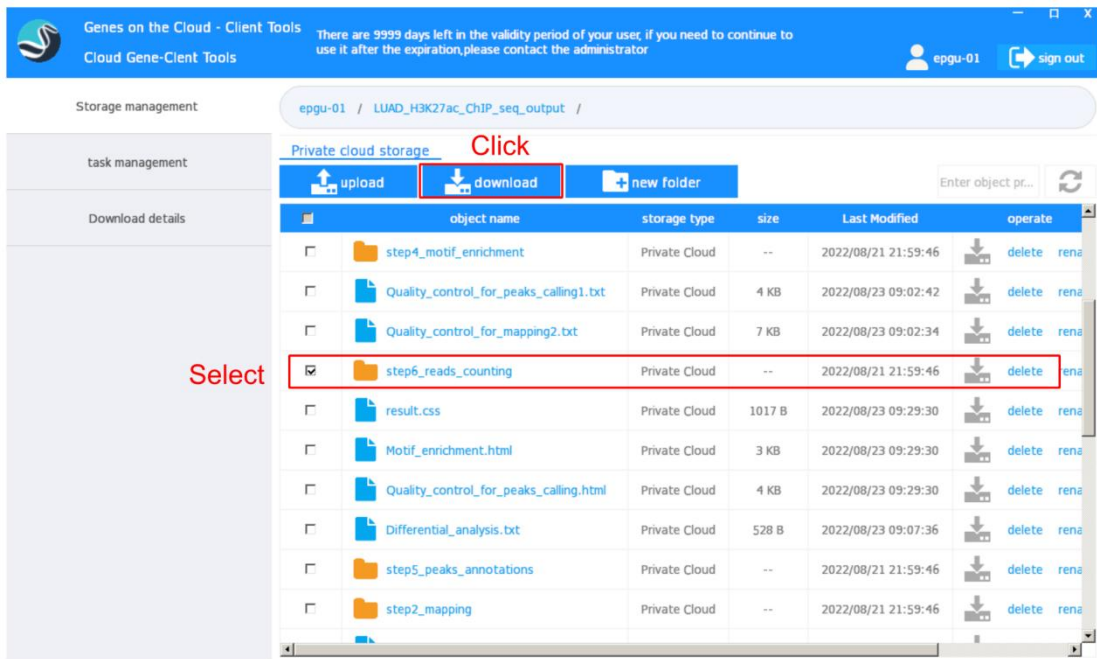
**Example for download files (report.pdf):**



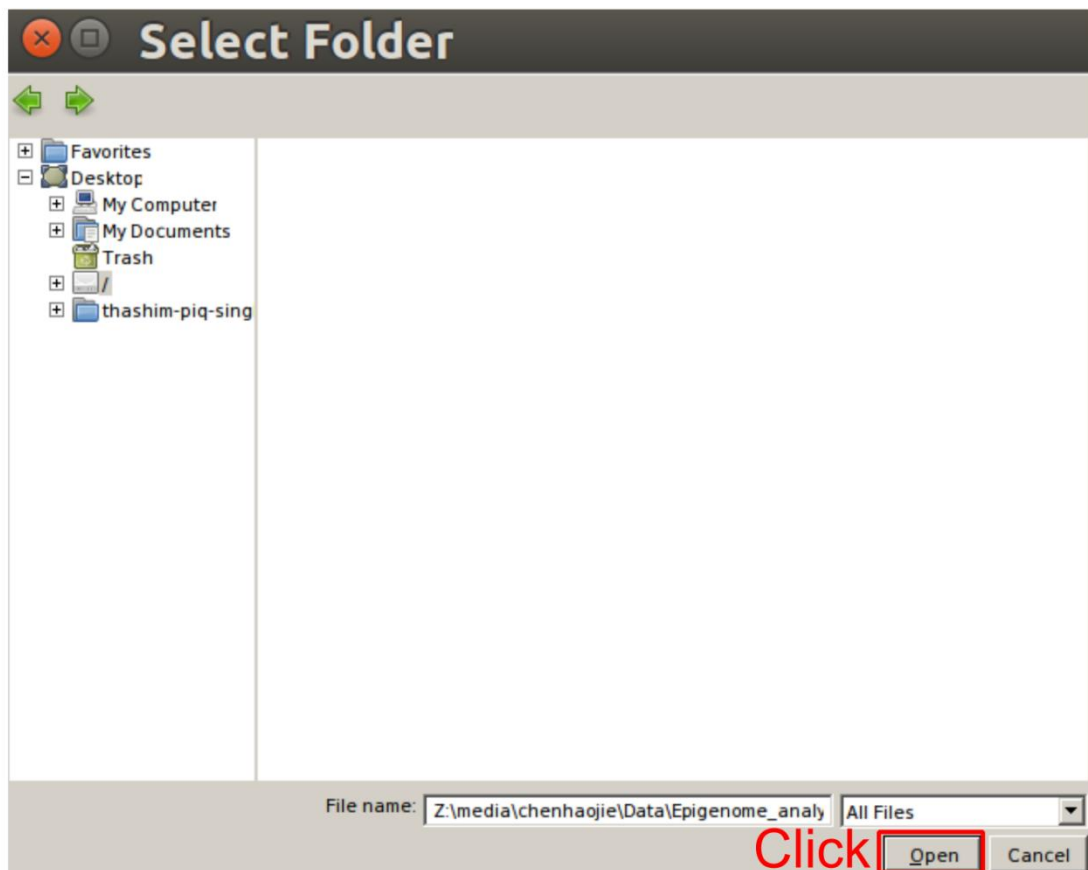
Select a local folder and click on Open to download the files.



Example for download a folder (step6\_reads\_counting):



Select a local folder and click on Open to download the files.



The download task progress could be monitored in the **task management** section in the client and details are available in **Download details**.

Genes on the Cloud - Client Tools There are 9999 days left in the validity period of your user, if you need to continue to use it after the expiration, please contact the administrator

Cloud Gene-Client Tools eggu-01 [sign out](#)

Storage management  [Refresh](#)

Numbering	object name
1740	Download server files:/eggu-01/LUAD_H3K27ac_Chip_seq_output/step6_reads_counting/distal_peak_regions_2000bp.txt
1741	Download server files:/eggu-01/LUAD_H3K27ac_Chip_seq_output/step6_reads_counting/proximal_peak_regions_2000bp.txt
1742	Download server files:/eggu-01/LUAD_H3K27ac_Chip_seq_output/step6_reads_counting/NA_profile_bins.xls
1743	Download server files:/eggu-01/LUAD_H3K27ac_Chip_seq_output/report.pdf

task management

Download details

If you encounter any errors, you can click on **Task information module** to check the details or click on the **green arrow** to send message to administrators. Administrators will help your resolve the issue as soon as possible. [\[back\]](#)

Home / Workflows Management / Execute Result Detail

luad-h3k27ac-chip-seq-20220821 **Click**

Result information module [Task information module](#)


The following is a reference to Huawei cloud data!

Process events Task event Input

Task name	Task type	Process events	Completion time	Occurrences	Status reason
job10-0	CCE.Job	创建成功 ##SuccessfulCreate	2022-08-23 08:55:58	1	创建*8s Job对象 job10-0-c32e0104ac1043c2 成功. ## Create Job Instance(job10-0-c32e0104ac1043c2) Successfully
job10-0	CCE.Job	等待执行完毕 ##WaitJobFinished	2022-08-23 08:55:59	1	等待任务 job10-0-c32e0104ac1043c2 执行完成 ## Waiting job job10-0-c32e0104ac1043c2 finished <b>Click</b>
job11-0	CCE.Job	创建成功 ##SuccessfulCreate	2022-08-23 08:57:59	1	创建*8s Job对象 job11-0-c32e0104ac1043c2 成功. ## Create Job Instance(job11-0-c32e0104ac1043c2) Successfully
job11-0	CCE.Job	等待执行完毕 ##WaitJobFinished	2022-08-23 08:57:59	1	等待任务 job11-0-c32e0104ac1043c2 执行完成 ## Waiting job job11-0-c32e0104ac1043c2 finished
luad-h3k27ac-chip-seq-20220821166109036216745	GCS.Action	流程执行成功 ##ExecutionSuccess	2022-08-23 09:29:00	1	执行create成功 ## Action(create) Success.


< 1 >

Chat dialogX



No Data

Total 0 items < 0 >



An error occurred when I run this task and job2 was failed, then the task was abruptly closed. I need help.

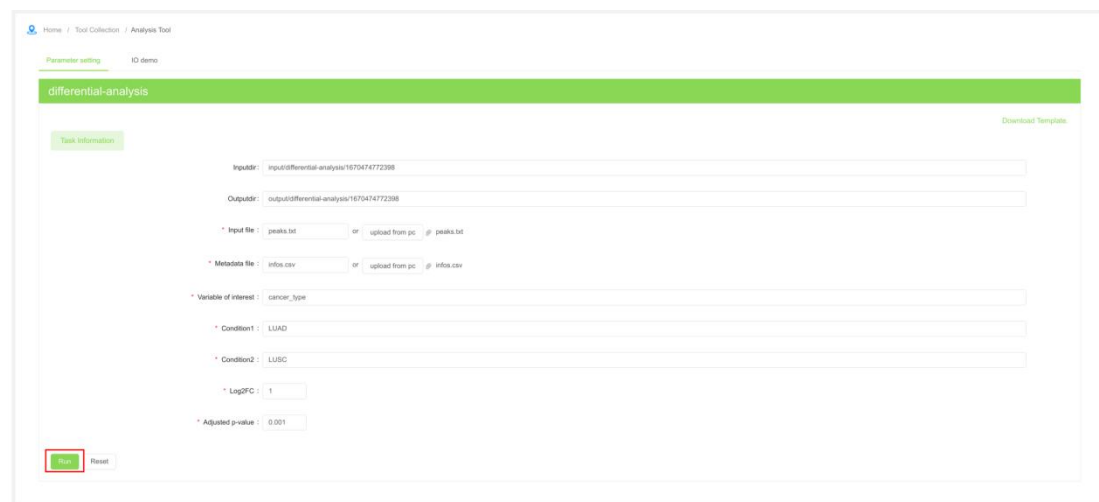
Click  
close send

#### Section 4. Down-stream analysis tools

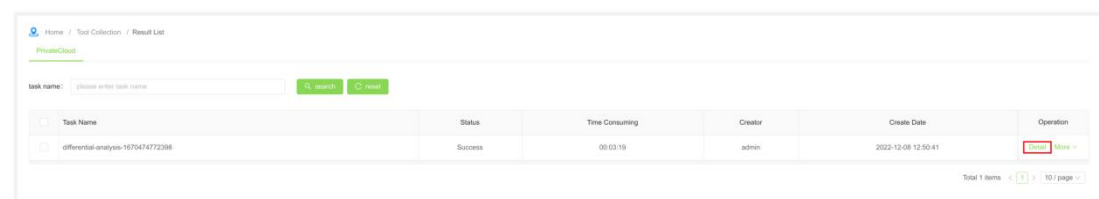
After data preprocessing, the output files from step6 are the input files for downstream analysis tools implemented in EAP. EAP uses statistical algorithms developed by our teams to transform the CHIP/ATAC-seq data into interpretable and biological meaningful results. At present, EAP provides a comprehensive of CHIP/ATAC-seq data down-stream analysis tools including [differential analysis](#), [differential TF motif enrichment analysis](#), [hypervariable analysis](#), [differential inferred TF regulatory potential analysis](#), [clustering analysis](#) and [signature genes score analysis](#) (**Menu bar: Tool Collection > Tool List**). All these tools will generate publication-ready figures and tables for users. Users can choose an appropriate analysis tool for their research problem by clicking on **Execute**. Input files could be uploaded from local PC (personal computer) or output directory of preprocessing module (i.e. output directory in step6) or deposited in the users' storage space (all input files should be deposited in the same directory). Output directory could be specified by users or created by EAP automatically. If output directory already exists and is not empty, the original results in the output directory may be overwritten. [\[back\]](#)



**Differential analysis:** One of the most common analyses of CHIP/ATAC-seq data is identification of differentially enriched peaks (DEPs)/differentially accessible peaks (DAPs). This enables us to elucidate the alterations of gene expression regulation related to phenotypic changes and it is important for exploration of therapeutic targets and biomarkers. To perform this analysis, users can upload the reads count table from the output result of Data preprocessing module and the metadata file contained variables of interest or choose the corresponding files in user's storage space, select the variable of interest, choose two conditions to perform the comparison and set the adjusted p-value cutoff (e.g. 0.001) and log2 transformed fold change cutoff (e.g. 1) used for identifying significant DEPs or DAPs, then perform differential analysis based on the user specified variable of interest (e.g. cancer\_type) by clicking on button **Run**. Details of input and output files are available in IO demo.

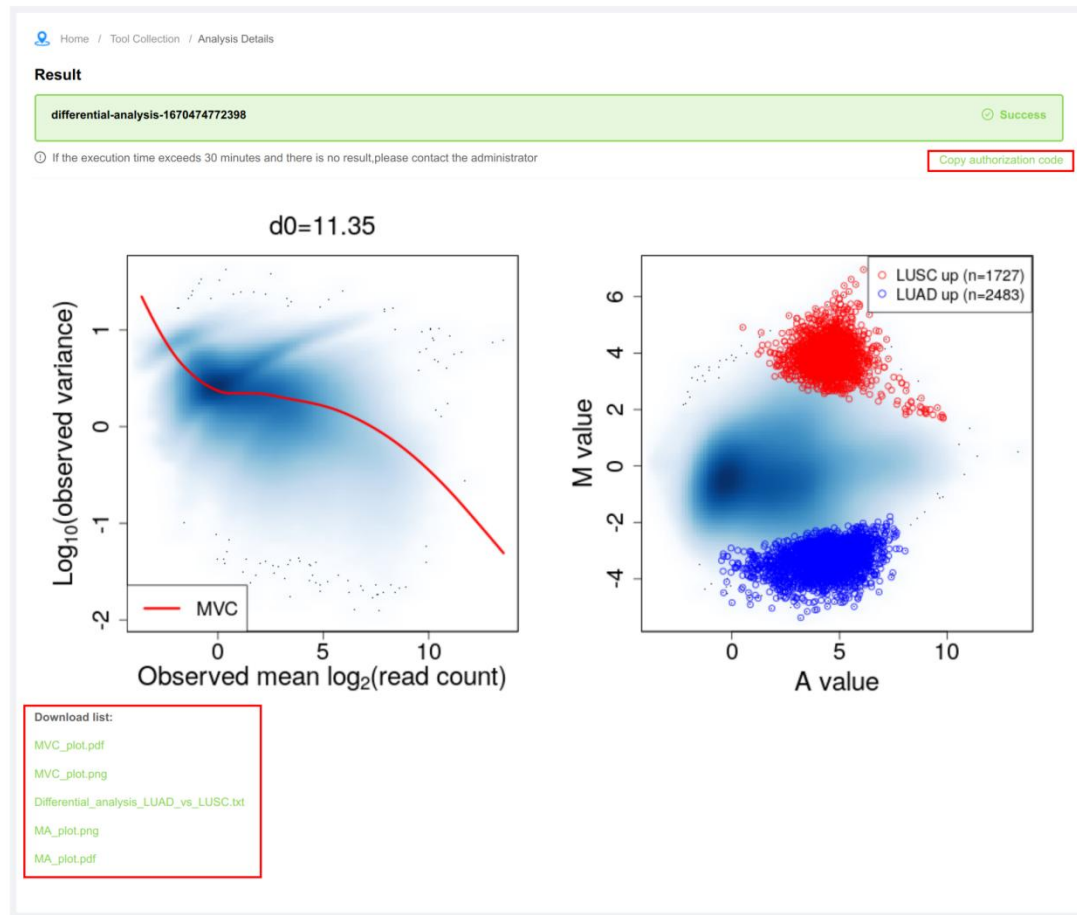


The progress of running task could be monitored in the Result List from Tool Collection (**Menu bar: Tool Collection > Result List**).



After successful completion of the submitted task. A mean-variance curve (MVC) plot, a MA plot and a result table will be generated and can be accessed by clicking on **Detail** (Rectangle

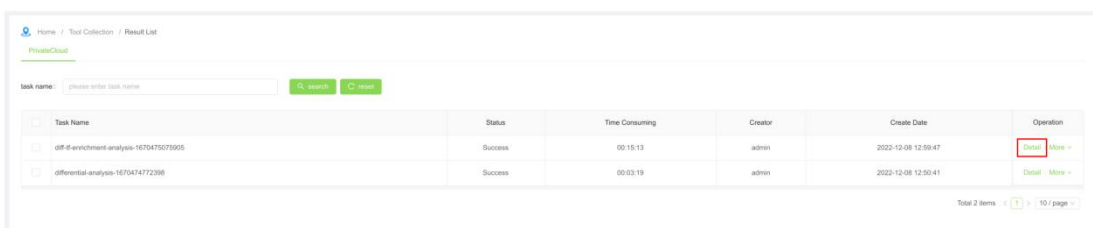
in Red in the image above). A screen like the image below will appear. Results with file size below 40MB could be downloaded directly in the web page by clicking the links from the [Download list](#). Other results should be downloaded through the client by copying the [authorization code](#) and pasting it in client. [\[back\]](#)



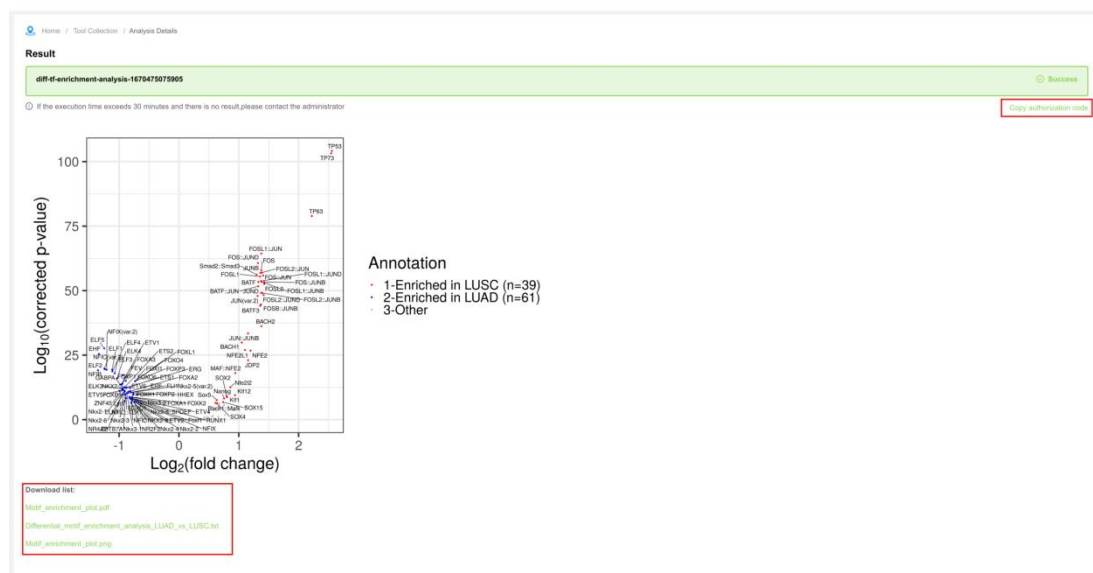
**Differential TF motif enrichment analysis:** Differential TF motif enrichment analysis can potentially identify key transcription factors associated with the changes of regulatory element activity between different biological conditions or developmental stages. Users can submit the result from Differential analysis, choose an appropriate reference genome version (e.g. hg19) and set the adjusted p-value cutoff (e.g. 0.001) and log2 transformed fold change cutoff (e.g. 1) used for identifying significant DEPs or DAPs, set the number of most significantly differentially enriched TF motifs highlighted in the volcano plot, then run the analysis tool by clicking on button [Run](#) will lead to a table of differential enriched TF motifs.



The progress of running task could be monitored in the Result List from Tool Collection (**Menu bar: Tool Collection > Result List**).



After successful completion of the submitted task. A volcano plot and a result table will be generated and can be accessed by clicking on **Detail** (Rectangle in Red in the image above). A screen like the image below will appear. Results with file size below 40MB could be downloaded directly in the web page by clicking the links from the **Download list**. Other results should be downloaded through the client by copying the **authorization code** and pasting it in client. [\[back\]](#)



**Hypervariable analysis:** Hypervariable analysis allows users to detect peaks that contribute to sample-to-sample differences within a population. Therefore, using these variable signals to investigate the similar structure in the population by clustering samples into groups with similar patterns. Users can specify the adjusted p-value cutoff (e.g. 0.001) to determine significantly

variable peaks then EAP performs principal component analysis (PCA) based on these signal. Samples are visualized in the two-dimensional PCA space and assigns color to the sample based on the user-specified categorical variable (e.g. cancer\_type). However, in some complex data set, two-dimensional PCA space will not be able to interpret the complex similar relationship between samples, this tool provides t-SNE visualization based on more than two principal components. Users can set the number of principal components (e.g. 3) used for performing t-SNE dimension reduction and test several different perplexity values (0 represent default value, range from 1 to  $(\text{number of samples}-1)/3$ , default value is  $(\text{number of samples}-1)/3$ ), then making some assessments with t-SNE visualization plots. Variance-stabilizing transformation can be enabled by setting this parameter to TRUE. Running the analysis tool by clicking on button **Run**. [\[back\]](#)

hypervariable-analysis-prd Download Template.

**Task Information**

Inputdir:

Outputdir:

\* Proximal peak regions file:  or

\* Distal peak regions file:  or

\* Metadata file:  or

\* Categorical variable:

\* The number of PCs:

\* Perplexity:

\* Adjusted p-value:

\* Variance stabilizing transformation:

The progress of running task could be monitored in the Result List from Tool Collection (**Menu bar: Tool Collection > Result List**).

Home / Tool Collection / Result List

task name:

Task Name	Status	Time Consuming	Creator	Create Date	Operation
hypervariable-analysis-1670494889625	Success	00:02:28	admin	2022-12-08 15:52:34	<input type="button" value="Detail"/> <input type="button" value="More"/>
differential-enrichment-analysis-1670475075905	Success	00:15:13	admin	2022-12-08 12:59:47	<input type="button" value="Detail"/> <input type="button" value="More"/>
differential-analysis-1570474722386	Success	00:03:19	admin	2022-12-08 12:56:41	<input type="button" value="Detail"/> <input type="button" value="More"/>

Total 3 items  / 10 / page

After successful completion of the submitted task. Two MVC plots, two scatter plot and two result tables will be generated and can be accessed by clicking on Detail (Rectangle in Red in the image above). A screen like the image below will appear. Results with file size below 40MB could be downloaded directly in the web page by clicking the links from the **Download list**. Other results should be downloaded through the client by copying the **authorization code** and pasting it in client. [\[back\]](#)

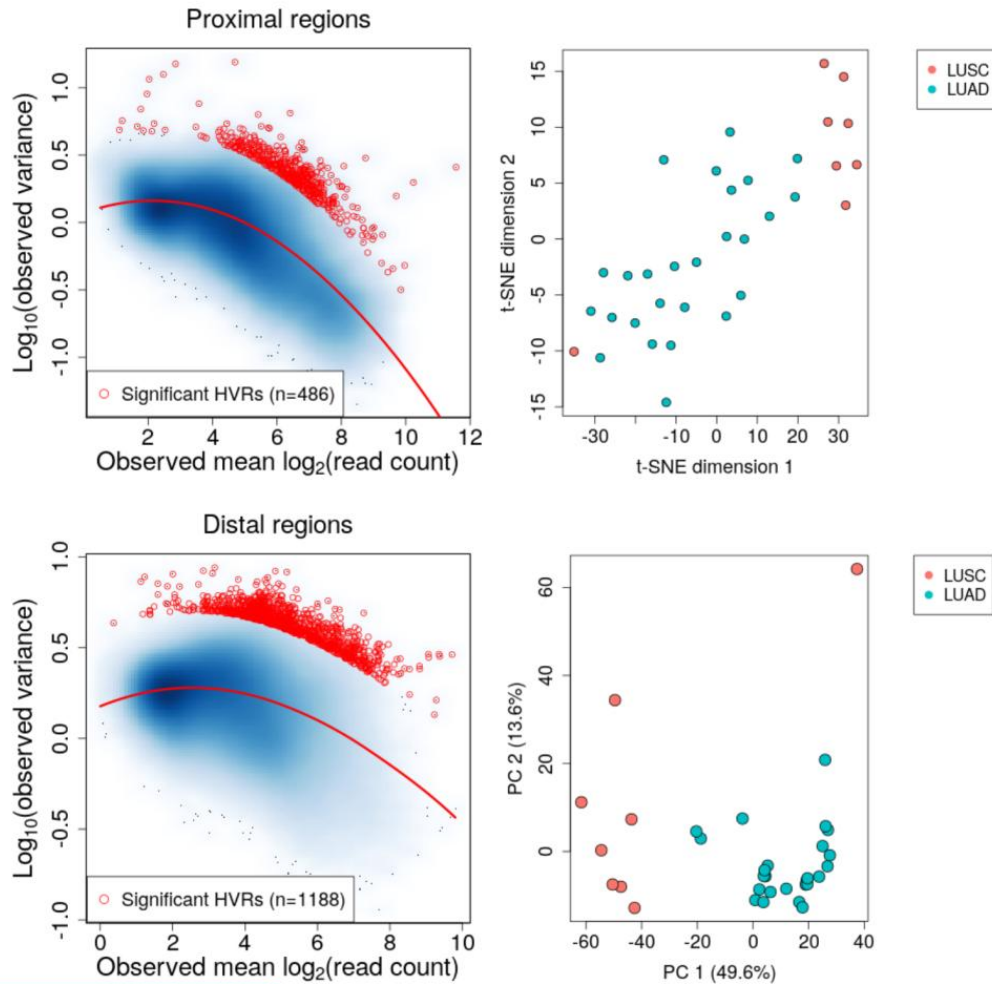
### Result

hypervariable-analysis-1670484980625

Success

If the execution time exceeds 30 minutes and there is no result, please contact the administrator

Copy authorization code



- Download list:
- [Distal\\_MVC\\_and\\_HVRs\\_plot.pdf](#)
  - [TSNE\\_plot.pdf](#)
  - [Proximal\\_MVC\\_and\\_HVRs\\_plot.png](#)
  - [PCA\\_scatter\\_plot.pdf](#)
  - [Distal\\_hypervariable\\_analysis.txt](#)
  - [TSNE\\_plot.png](#)
  - [Distal\\_MVC\\_and\\_HVRs\\_plot.png](#)
  - [PCA\\_scatter\\_plot.png](#)
  - [Proximal\\_MVC\\_and\\_HVRs\\_plot.pdf](#)
  - [Proximal\\_hypervariable\\_analysis.txt](#)

**Differential inferred TF regulatory potential analysis:** Differential inferred TF regulatory potential analysis can be used to detect TF motif associated with variable CHIP/ATAC-seq

signal and characterized different biological samples. Users can submit the result of Hypervariable analysis and then select the variable of interest from metadata used for analysis. Users choose an appropriate reference genome version (e.g. hg19) and set the adjusted p-value cutoff (e.g. 0.001) to determine significantly variable peaks. Then EAP performs motif scanning on these genomic regions and aggregates TF motif associated signals in each sample into a score, represents the TF regulatory activity. Samples are visualized in the two-dimensional PCA space or the two-dimensional t-SNE space and assigns color to the sample based on the TF activity of user specified TF (e.g. TP63 or Nkx2-1). Finally, EAP identifies TFs associated with the user specified variable of interest. The outputs in this analysis including a table of TF activities in each sample, a table of t- statistic of the association test and plots for dimension reduction visualization of samples and the activities of user specified TFs. After setting the required analysis parameters, user runs the analysis tool by clicking on button **Run**.

The progress of running task could be monitored in the Result List from Tool Collection (**Menu bar: Tool Collection > Result List**).

Task Name	Status	Time Consuming	Creator	Create Date	Operation
differential-tf-activity-analysis-1670503735285	Success	00:08:53	admin	2022-12-08 21:14:06	<b>Detail</b> More >
hypervariable-analysis-1670484680625	Success	00:02:28	admin	2022-12-08 15:52:34	Detail More >
gdf-if-enrichment-analysis-1670475075905	Success	00:15:13	admin	2022-12-08 12:59:47	Detail More >
differential-analysis-1670474772398	Success	00:03:19	admin	2022-12-08 12:50:41	Detail More >

After successful completion of the submitted task. Scatter plots, rank plots, a heatmap and two result tables will be generated and can be accessed by clicking on Detail (Rectangle in Red in the image above). Rank plot showed the ranks and t-statistics of top ranked TFs. Scatter plot showed the activity scores of TF of interest in the two-dimension reduction space. Heatmap

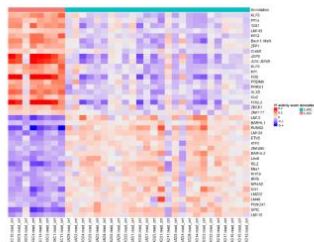
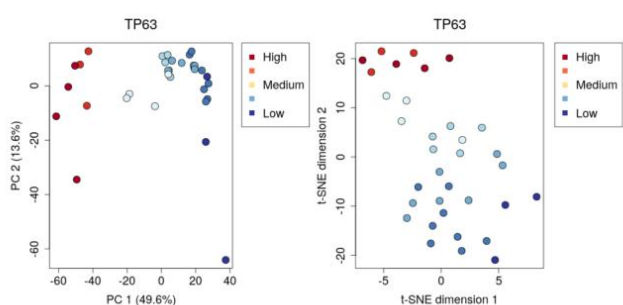
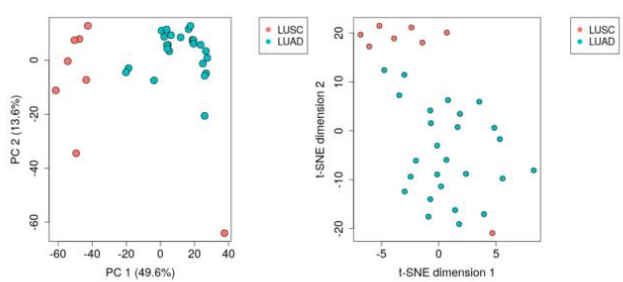
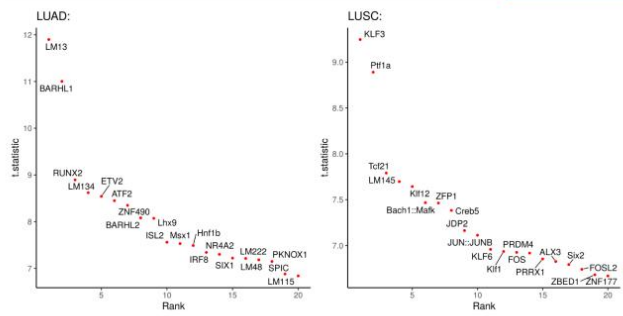
showed the activity scores of top ranked differential TFs. A screen like the image below will appear. Results with file size below 40MB could be downloaded directly in the web page by clicking the links from the [Download list](#). Other results should be downloaded through the client by copying the [authorization code](#) and pasting it in client. [\[back\]](#)

### Result

differential-tf-activity-analysis-1672730260794 Success

If the execution time exceeds 30 minutes and there is no result, please contact the administrator

[Copy authorization code](#)



- Download list:
- [PCA\\_plot.txt](#)
  - [TF\\_activity\\_score\\_heatmap.pdf](#)
  - [Differential\\_TF\\_activity\\_analysis.txt](#)
  - [TP63\\_TF\\_activity\\_score\\_TSNE\\_plot.pdf](#)
  - [LUAD\\_TF\\_activity\\_rank\\_plot.png](#)
  - [TSNE\\_plot1.txt](#)
  - [TSNE\\_plot\\_1.pdf](#)
  - [LUSC\\_TF\\_activity\\_rank\\_plot.png](#)
  - [TP63\\_TF\\_activity\\_scatter\\_plot.pdf](#)
  - [PCA\\_scatter\\_plot\\_1.png](#)
  - [LUAD\\_TF\\_activity\\_rank\\_plot.pdf](#)
  - [LUSC\\_TF\\_activity\\_rank\\_plot.pdf](#)
  - [TSNE\\_plot\\_1.png](#)
  - [TP63\\_TF\\_activity\\_scatter\\_plot.png](#)
  - [TP63\\_TF\\_activity\\_score\\_TSNE\\_plot.png](#)
  - [TF\\_activity\\_scores.txt](#)
  - [TF\\_activity\\_score\\_heatmap.png](#)
  - [PCA\\_scatter\\_plot\\_1.pdf](#)

**Clustering analysis:** Clustering of ChIP/ATAC-seq data is widely used to identify novel subtypes of cancer in cancer epigenomic studies. It is an unsupervised strategy for finding subgroups with similar patterns. A common practice for clustering analysis is to preprocess the data by detecting hypervariable signal and performing unsupervised hierarchical clustering based on these variable signals. Users can submit the result of Hypervariable analysis and metadata file, then choose the number of clusters (e.g. 2), select an appropriate adjusted p-value cutoff (e.g. 0.001) for defining hypervariable peak regions and choose the number of principal components (e.g. 0, 0 represents choosing the recommended optimal number of principal components) used for hierarchical clustering. Click on button **Run** to perform the analysis.

The progress of running task could be monitored in the Result List from Tool Collection (**Menu bar: Tool Collection > Result List**).

Task Name	Status	Time Consuming	Creator	Create Date	Operation
clustering-analysis-167050999976	Success	00:02:03	admin	2022-12-08 22:48:57	<b>Detail</b> More
differential-4f-activity-analysis-167050373225	Success	00:08:53	admin	2022-12-08 21:14:06	Detail More
hypervariable-analysis-167048480625	Success	00:02:28	admin	2022-12-08 15:52:34	Detail More
qRf-enrichment-analysis-167047507980	Success	00:15:13	admin	2022-12-08 12:58:47	Detail More
differential-analysis-167047477238	Success	00:03:19	admin	2022-12-08 12:50:41	Detail More

After successful completion of the submitted task. A rank plot, a dendrogram plot and a result table will be generated and can be accessed by clicking on Detail (Rectangle in Red in the image above). The rank plot indicates the recommended optimal number of principal components used for hierarchical clustering (However, user could specify the number of principal components used for hierarchical clustering). A screen like the image below will appear. Results with file size below 40MB could be downloaded directly in the web page by clicking the links from the **Download list**. Other results should be downloaded through the client by copying the **authorization code** and pasting it in client. [\[back\]](#)

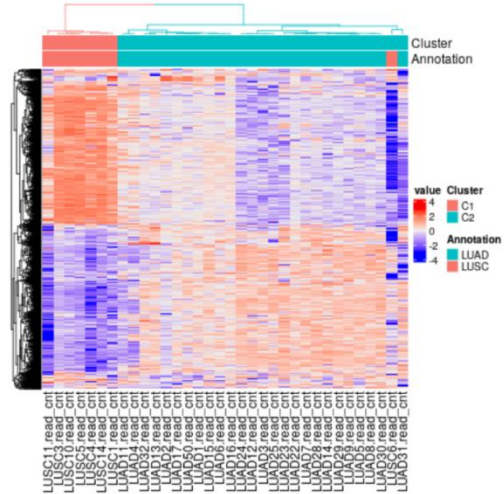
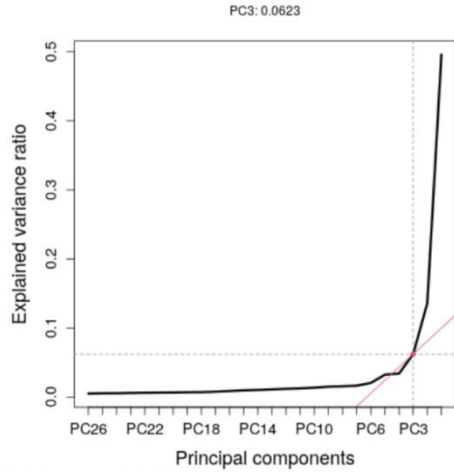
## Result

clustering-analysis-1672742532828

Success

If the execution time exceeds 30 minutes and there is no result, please contact the administrator

Copy authorization code



- Download list:
- Test\_clustering\_plot.pdf
  - Test\_clustering\_reuslt.txt
  - Explained\_variance\_ratio\_in\_each\_PC.png
  - Test\_clustering\_plot.png
  - Explained\_variance\_ratio\_in\_each\_PC.pdf

**Signature genes score analysis:** Given a gene set of interest, it is usually more desirable to summarize the expression level of that gene set using a single integrated score. This tool standardizes the ChIP/ATAC-seq signals in the proximal regions within a given dataset by z-score transformation. Then summarizes resulting scores of those proximal regions linked to the genes of interest, minus the mean of z-scores of all proximal regions as negative control. This analysis provides visualization to improve interpretation of the clustering results. For example, users can annotate the cluster based on these signature genes scores. Users can submit the result of Hypervariable analysis (i.e. Proximal\_hypervariable\_analysis.RData) , metadata file and a set of signature genes in GMT format. Click on button **Run** to perform the analysis.

Home / Tool Collection / Analysis Tool

Parameter setting ID: demo

signature-genes-score-estimation [Download Template](#)

Task Information

InputId: input/signature-genes-score-estimation/1672752431156

OutputId: output/signature-genes-score-estimation/1672752431156

\* Input file: Proximal\_hypervariable\_ or upload from pc # Proximal\_hypervariable\_analysis\_RData

\* Peak to genes file: proximal\_regions\_peaks or upload from pc # proximal\_regions\_peaks\_to\_genes\_links.txt

\* Metadate file: Test\_clustering\_result.txt or upload from pc # Test\_clustering\_result.txt

\* Signature genes: NPC1.gmt or upload from pc # NPC1.gmt

\* Variable\_of\_interest: cluster

**Run** Reset

The progress of running task could be monitored in the Result List from Tool Collection (**Menu bar: Tool Collection > Result List**).

Home / Tool Collection / Result List

ProximalCloud

task name: please enter task name

Task Name	Status	Time Consuming	Creator	Create Date	Operation
signature-genes-score-estimation-1672752431156	Success	00:01:39	admin	2023-01-03 21:30:22	<b>Detail</b> More >
clustering-analysis-1672742532826	Success	00:01:57	admin	2023-01-03 18:44:04	Detail More >
differential-activity-analysis-167273260794	Success	00:07:40	admin	2023-01-03 15:19:21	Detail More >
hypervariable-analysis-167273067237	Success	00:02:40	admin	2023-01-03 14:44:20	Detail More >
differential-enrichment-analysis-1672727176269	Success	00:15:12	admin	2023-01-03 14:28:49	Detail More >
differential-analysis-1672726708304	Success	00:02:22	admin	2023-01-03 14:22:40	Detail More >

Total 6 items < 1 > 13 / page >

After successful completion of the submitted task. A Box plot will be generated and can be accessed by clicking on Detail (Rectangle in Red in the image above). A screen like the image below will appear. Box plot showed the signature genes score distribution in each group. Groups were defined based on user specified variable of interest. Results with file size below 40MB could be downloaded directly in the web page by clicking the links from the [Download list](#). Other results should be downloaded through the client by copying the [authorization code](#) and pasting it in client. [\[back\]](#)

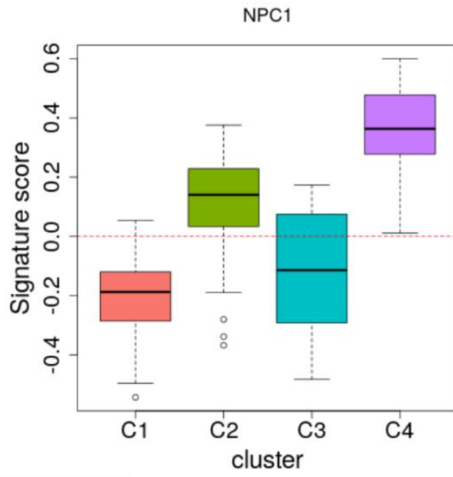
## Result

signature-genes-score-estimation-1672752431156

Success

If the execution time exceeds 30 minutes and there is no result, please contact the administrator

Copy authorization code



Download list:

[NPC1.png](#)

[NPC1.pdf](#)

## Supplementary Materials

File-1: [Storage space application form.xlsx](#)

File-2: [metadata.csv](#)

Administrators' email address:

Haojie Chen     [chenhaojie2017@sinh.ac.cn](mailto:chenhaojie2017@sinh.ac.cn)

Zhen Shao     [shaozhen@sinh.ac.cn](mailto:shaozhen@sinh.ac.cn)

Zhijie Guo     [guozhijie2018@sinh.ac.cn](mailto:guozhijie2018@sinh.ac.cn)

These files and information are also available on **Home page** by clicking on

**Tutorial.** [\[back\]](#)

