

## 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.

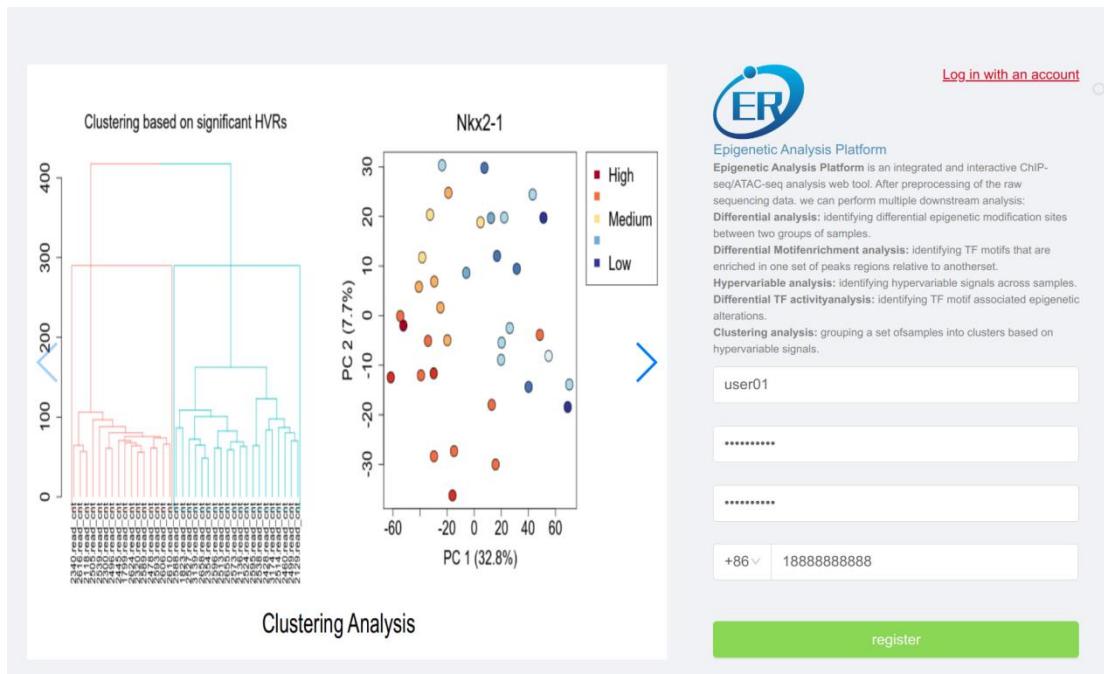
[Section 3. ChIP/ATAC-seq data preprocessing:](#) Transform raw sequencing data into analysis ready data.

[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://sdap.biosino.org/epigenetics/#/user/register>) and create storage space by sending an application form (application form: [File-1](#)) to administrator (e-mail addresses are available on **Home page**). 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. [\[back\]](#)

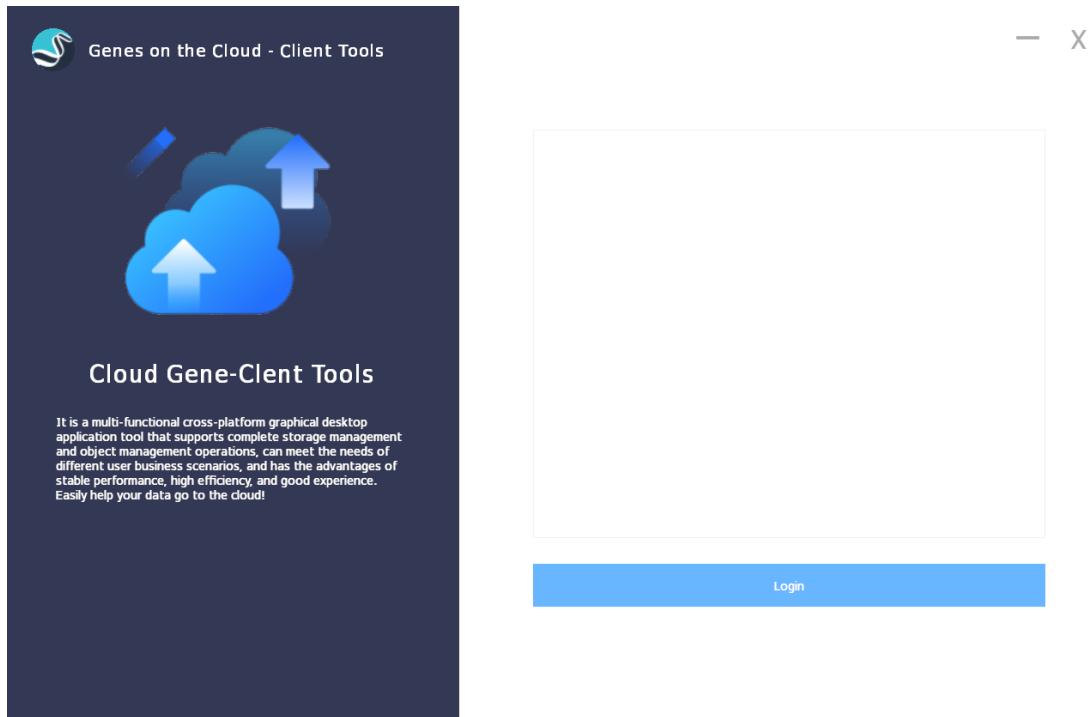
To register a user account as bellow:



### 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.

#	Object name	Storage category	Size	Last modification time	Operation
1	input/	Standard storage	--	2022-07-05 02:15:11	<input type="button"/> Delete <input type="button"/> More
2	config/	Standard storage	--	2022-05-27 03:24:53	<input type="button"/> Delete <input type="button"/> More
3	xiedong_outdir/	Standard storage	--	2022-08-11 11:45:44	<input type="button"/> Delete <input type="button"/> More
4	xiedong_dir/	Standard storage	--	2022-08-05 08:22:11	<input type="button"/> Delete <input type="button"/> More
5	cache/	Standard storage	--	2022-05-27 03:24:53	<input type="button"/> Delete <input type="button"/> More
6	Hela_Sy_S_ChIP/	Standard storage	--	2022-08-19 11:35:34	<input type="button"/> Delete <input type="button"/> More
7	output/	Standard storage	--	2022-08-09 06:15:12	<input type="button"/> Delete <input type="button"/> More
8	H1/	Standard storage	--	2022-07-19 05:21:44	<input type="button"/> Delete <input type="button"/> More
9	localdata/	Standard storage	--	2022-08-11 03:57:00	<input type="button"/> Download <input type="button"/> More
10	.bash_history	Standard storage	0.28 KB	2022-05-27 03:25:12	<input type="button"/> Download <input type="button"/> More

## 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 (/).

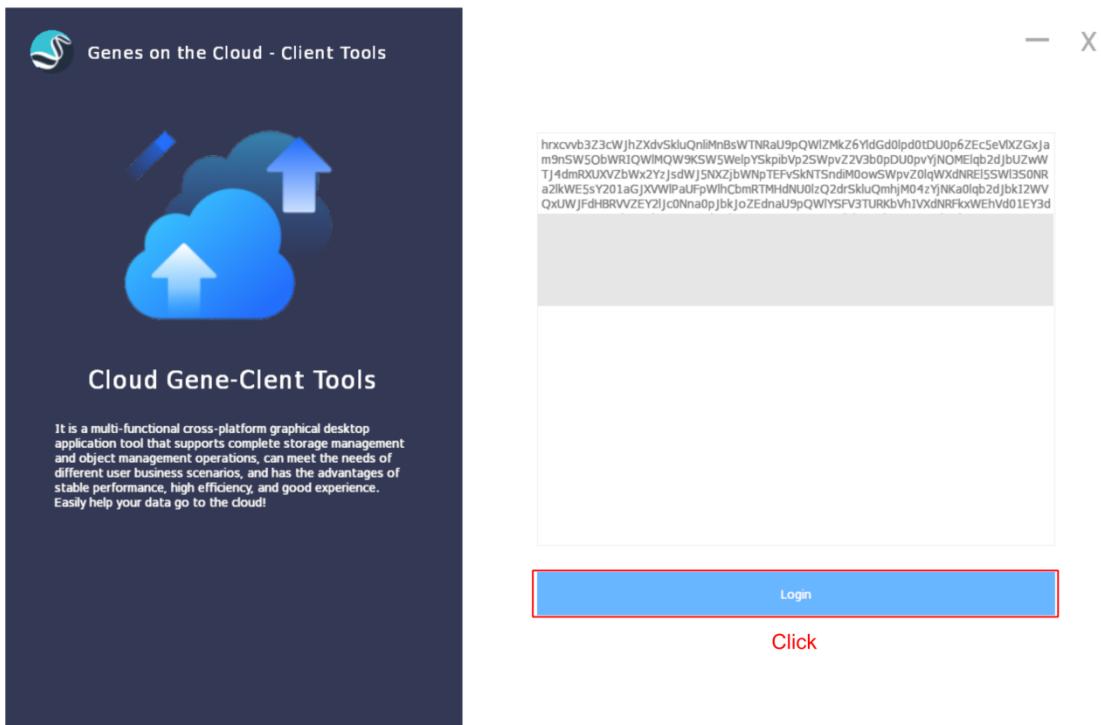
[关闭](#) [OK](#)

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**.

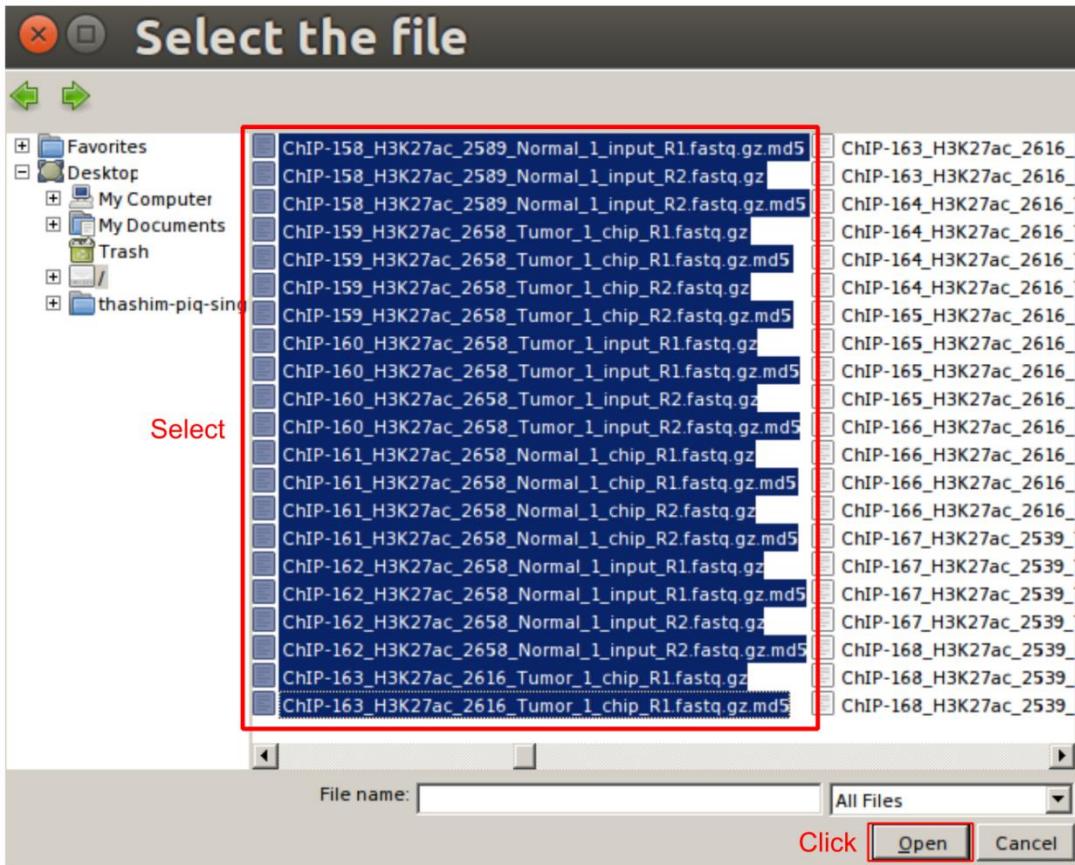
The screenshot shows a cloud storage interface with the following details:

- Header:** Home / Storage Management / My Data
- Toolbar:** Data list-bucket: (dropdown), click Auth Code, New folder, Upload large files, Upload small files
- Table:** A list of 10 items (files/folders) in a bucket named "click Auth Code".

#	Object name	Storage category	Size	Last modification time	Operation
1	input/	Standard storage	--	2022-07-05 02:15:11	<a href="#">Delete</a> <a href="#">More</a>
2	.config/	Standard storage	--	2022-05-27 03:24:53	<a href="#">Delete</a> <a href="#">More</a>
3	xiedong_outdir/	Standard storage	--	2022-08-11 11:45:44	<a href="#">Delete</a> <a href="#">More</a>
4	xiedong_dir/	Standard storage	--	2022-08-05 08:22:11	<a href="#">Delete</a> <a href="#">More</a>
5	.cache/	Standard storage	--	2022-05-27 03:24:53	<a href="#">Delete</a> <a href="#">More</a>
6	Hela_Sy_S_ChIP/	Standard storage	--	2022-08-19 11:35:34	<a href="#">Delete</a> <a href="#">More</a>
7	output/	Standard storage	--	2022-08-09 06:15:12	<a href="#">Delete</a> <a href="#">More</a>
8	.bash_history	Standard storage	0.28 KB	2022-05-27 03:25:12	<a href="#">Download</a> <a href="#">More</a>
9	LUAD_H3K27ac_ChIP_seq/	Standard storage	--	2022-08-20 09:21:40	<a href="#">Delete</a> <a href="#">More</a>
10	opt/	Standard storage	--	2022-08-20 09:21:40	<a href="#">Delete</a> <a href="#">More</a>
- Bottom Right:** Total 10 items, page navigation (1, 50 / page, etc.), and a red box highlighting the "Copy authorization code" button.



Select raw sequencing data and metadata file and upload (While there is no strict size limit for files uploaded to EAP, for each data uploading, it is recommended that you upload files below 50 GB in total size. When your upload files exceed this files size limit, you would upload these files separately).



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. [\[back\]](#)

object name	type	status	size	task progress	creation time	operate
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-161_H3K27ac_2658_Normal_1_input_R1.fastq.gz.md5	upload	completed	82 B	100%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-161_H3K27ac_2658_Normal_1_input_R2.fastq.gz.md5	upload	completed	82 B	100%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-161_H3K27ac_2658_Normal_1_chip_R1.fastq.gz	upload	uploading...	4003480 KB	16%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Normal_1_input_R1.fastq.gz	upload	uploading...	4113169 KB	14%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Tumor_1_input_R1.fastq.gz.md5	upload	completed	82 B	100%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Normal_1_input_R2.fastq.gz	upload	uploading...	5000406 KB	13%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2616_Tumor_1_chip_R1.fastq.gz.md5	upload	pause	81 B	0%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Normal_1_input_R1.fastq.gz	upload	uploading...	4936326 KB	15%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Normal_1_input_R2.fastq.gz.md5	upload	pause	83 B	0%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-161_H3K27ac_2616_Tumor_1_chip_R1.fastq.gz	upload	pause	3407234 KB	0%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Normal_1_input_R2.fastq.gz	upload	pause	4995267 KB	0%	2022/08/20 21:17:56	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Normal_1_input_R1.fastq.gz.md5	upload	completed	83 B	100%	2022/08/20 21:17:48	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Tumor_1_input_R1.fastq.gz	upload	uploading...	3731722 KB	23%	2022/08/20 21:17:45	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-152_H3K27ac_2658_Tumor_1_chip_R1.fastq.gz.md5	upload	completed	81 B	100%	2022/08/20 21:17:36	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-159_H3K27ac_2658_Tumor_1_chip_R1.fastq.gz	upload	uploading...	4534571 KB	25%	2022/08/20 21:17:35	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-159_H3K27ac_2658_Tumor_1_chip_R2.fastq.gz.md5	upload	completed	81 B	100%	2022/08/20 21:17:30	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-153_H3K27ac_2658_Tumor_1_chip_R1.fastq.gz	upload	uploading...	3167163 KB	35%	2022/08/20 21:17:27	
Upload local files /rppu-01/UAQ_H3K27ac_CHIP_seq/ChIP-163_H3K27ac_2616_Tumor_1_chip_R1.fastq.gz.md5	upload	completed	83 B	100%	2022/08/20 21:17:25	

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

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*).

New folder

Enter an appropriate name

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

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 (/).

**Click**

关闭 OK

#	Object name	Storage category	Size	Last modification time	Operation
1	input/	Standard storage	--	2022-07-05 02:15:11	<input type="checkbox"/> Delete More
2	.config/	Standard storage	--	2022-05-27 03:24:53	<input type="checkbox"/> Delete More
3	xiedong_outdir/	Standard storage	--	2022-08-11 11:45:44	<input type="checkbox"/> Delete More
4	xiedong_dir/	Standard storage	--	2022-08-05 08:22:11	<input type="checkbox"/> Delete More
5	.cache/	Standard storage	--	2022-05-27 03:24:53	<input type="checkbox"/> Delete More
6	output/	Standard storage	--	2022-08-09 06:15:12	<input type="checkbox"/> Delete More
7	.bash_history	Standard storage	0.28 KB	2022-05-27 03:25:12	<input type="checkbox"/> Download More
8	LUAD_H3K27ac_ChIP_seq/	Standard storage	--	2022-08-21 05:50:48	<input type="checkbox"/> Delete More
9	opt/	Standard storage	--	2022-08-20 08:18:41	<input type="checkbox"/> Delete More
10	<b>LUAD_H3K27ac_ChIP_seq_output/</b>	Standard storage	--	2022-08-21 05:58:17	<input type="checkbox"/> Delete More

Click Copy path  
Total 10 items Copy authorization code

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).

Epigenetics Welcome to epigenetics zone platform

Home My Data Operation Desc Workflows Task

Workflows Management Workflows Task

Process name: Please enter the process name Module: Please select the module

ATAC-seq/ChIP-seq analysis pipeline-v3

Pipeline ID: esp-pipeline-v3 Module: **esp-genome ATAC-seq ChIP-seq**

Version number: v1 Creation time: 2022-07-21 15:55:57

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) is the premier technology for profiling genome-wide localization of DNA binding proteins, including transcription factors and histones with various modifications. Besides, assay for transposase-accessible chromatin using sequencing (ATAC-seq) has been widely adopted for detection of open chromatin. Based on these data, we can identify the ChIP/ATAC-seq signal intensities.

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).

Home / Workflows Management / One key analysis

Task information

\* Task name: **luad-h3k27ac-chip-seq-20220821**

Timeout: 2 day 0 hour 0 minute

priority: 0 Priority range: [0-999]

reminder: High value priority

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

retry count: 3 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
Inputdir	string	LUAD_H3K27ac_ChIP_seq/	data dir for your input files
Outputdir	string	LUAD_H3K27ac_ChIP_seq_output/	result dir for all samples
sequencing_type	string	CHIPPE	Single end ChIP or Paired end ChIP or ATAC,choose between ATAC, CHIPPE and CHIPSE
information_file	string	LUAD_H3K27ac_ChIP_seq/ or upload from pc	information text,should be deposited in input_dir. template
REF_INDEX	string	hg19	Genome index,choose among :hg19,hg38,mm9,mm10
typical_bin_size	number	2000	typical bin size for profile-bins
variable_of_interest	string	tissue_type	interested variable for differential analysis
project_name	string	LUAD_H3K27ac_ChIP_seq	name of project

> Advanced parameters

close submit

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

PrivateCloud

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

 luad-h3k27ac-chip-seq-20220821

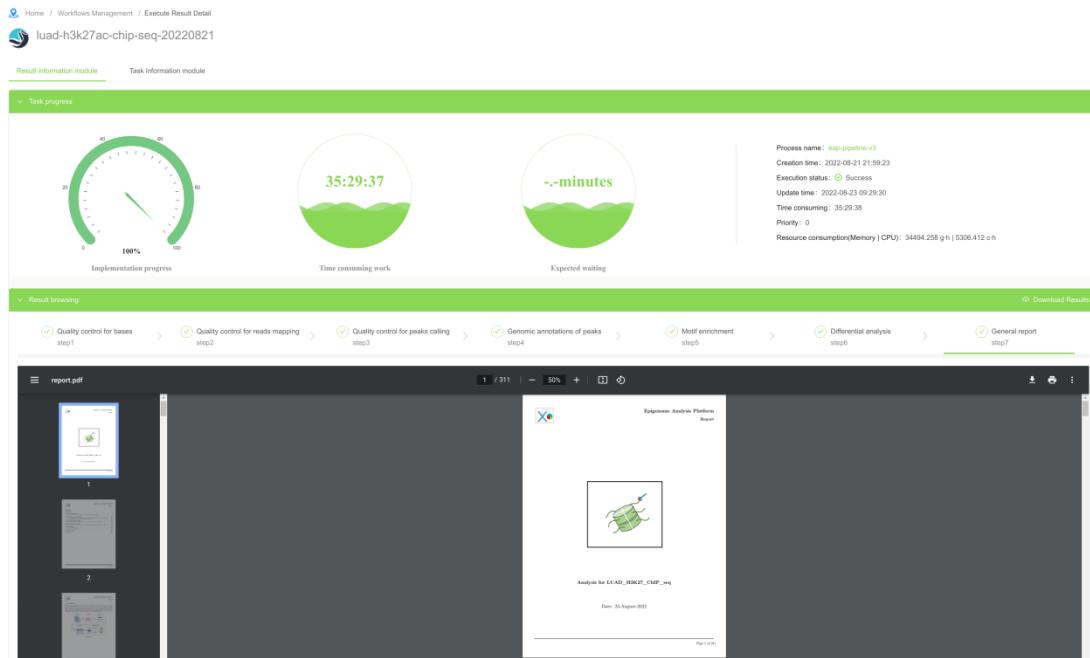


After data preprocessing, a PDF report will be generated, including quality control plots and summary statistics (**Menu bar: Workflows Management > Execute Result**). User can filter poor quality samples based on this summary report.

PrivateCloud

Process name	Task name	Execution status	Implementation	Time consuming	Creator	Creation time	Operation
ATAC-seq/ChIP-seq analysis pipeline-v3	luad-h3k27ac-chip-seq-20220821	Succeeded	<div style="width: 100%;">100%</div>	35:29:37	epgu-01	2022-08-21 21:59:23	<a href="#">delete</a> <a href="#">clear cache</a>
ATAC-seq/ChIP-seq analysis pipeline-v3	eap-pipeline-v3-08-18-152545	Succeeded	<div style="width: 100%;">100%</div>	03:23:35	epgu-01	2022-08-18 15:27:00	<a href="#">delete</a> <a href="#">clear cache</a>
ChIP-seq analysis pipeline-v2	chip-analysis-v2-08-02-124031	Succeeded	<div style="width: 100%;">100%</div>	01:28:33	epgu-01	2022-08-02 12:41:50	<a href="#">delete</a> <a href="#">clear cache</a>
ATAC-seq analysis pipeline-v2	atac-analysis-v2-08-02-110556	Succeeded	<div style="width: 100%;">100%</div>	00:57:33	epgu-01	2022-08-02 11:06:37	<a href="#">delete</a> <a href="#">clear cache</a>

Total 4 items < 1 > 10 / page



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**.

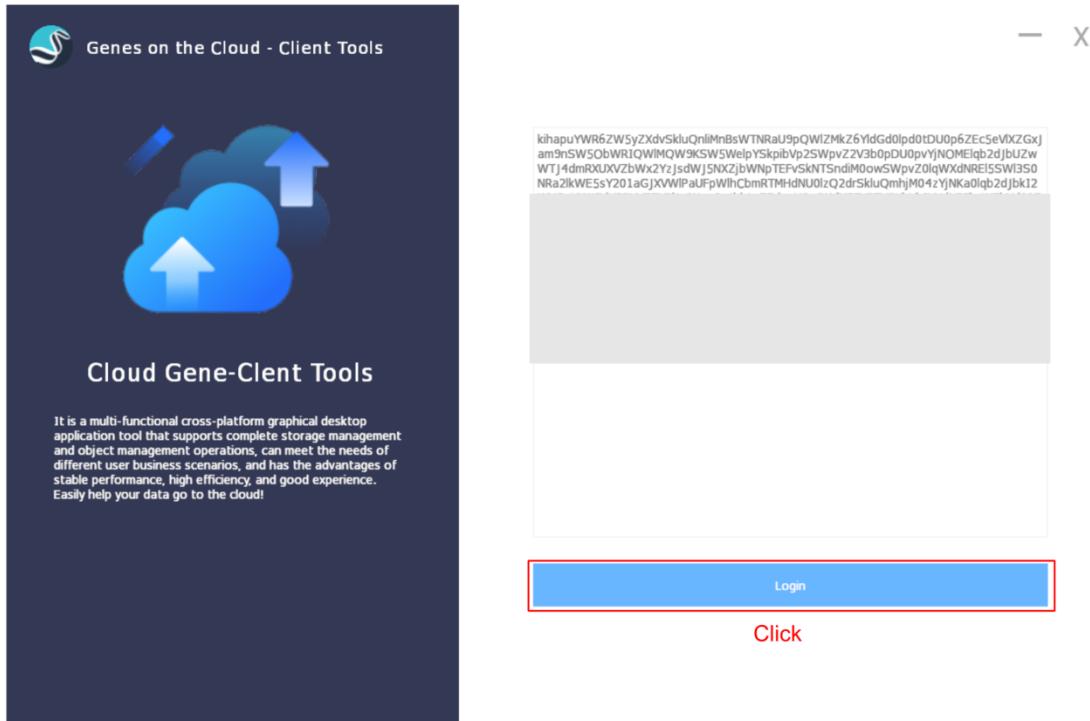
The screenshot shows the 'Storage Management / My Data' page under the 'PrivateCloud' section. It lists files in a bucket, with a red box highlighting the 'Copy authorization code' button next to the file 'LUAD\_H3K27ac\_ChIP\_seq\_output'.

#	Object name	Storage category	Size	Last modification time	Operation
1	input/	Standard storage	--	2022-07-05 02:15:11	<input type="checkbox"/> Delete   More
2	.config/	Standard storage	--	2022-05-27 03:24:53	<input type="checkbox"/> Delete   More
3	xiedong_outdir/	Standard storage	--	2022-08-11 11:45:44	<input type="checkbox"/> Delete   More
4	xiedong_dir/	Standard storage	--	2022-08-05 08:22:11	<input type="checkbox"/> Delete   More
5	.cache/	Standard storage	--	2022-05-27 03:24:53	<input type="checkbox"/> Delete   More
6	PDAC_organoid_ATAC_seq/	Standard storage	--	2022-08-22 14:21:48	<input type="checkbox"/> Delete   More
7	output/	Standard storage	--	2022-08-09 06:15:12	<input type="checkbox"/> Delete   More
8	.bash_history	Standard storage	0.28 KB	2022-05-27 03:25:12	<input type="checkbox"/> Download   More
9	LUAD_H3K27ac_ChIP_seq/	Standard storage	--	2022-08-21 14:00:12	<input type="checkbox"/> Delete   More
10	opt/	Standard storage	--	2022-08-20 08:18:41	<input type="checkbox"/> Delete   More
11	LUAD_H3K27ac_ChIP_seq_output/	Standard storage	--	2022-08-23 01:29:30	<input type="checkbox"/> Delete   More

Total 11 items

**Copy authorization code**

**Click**

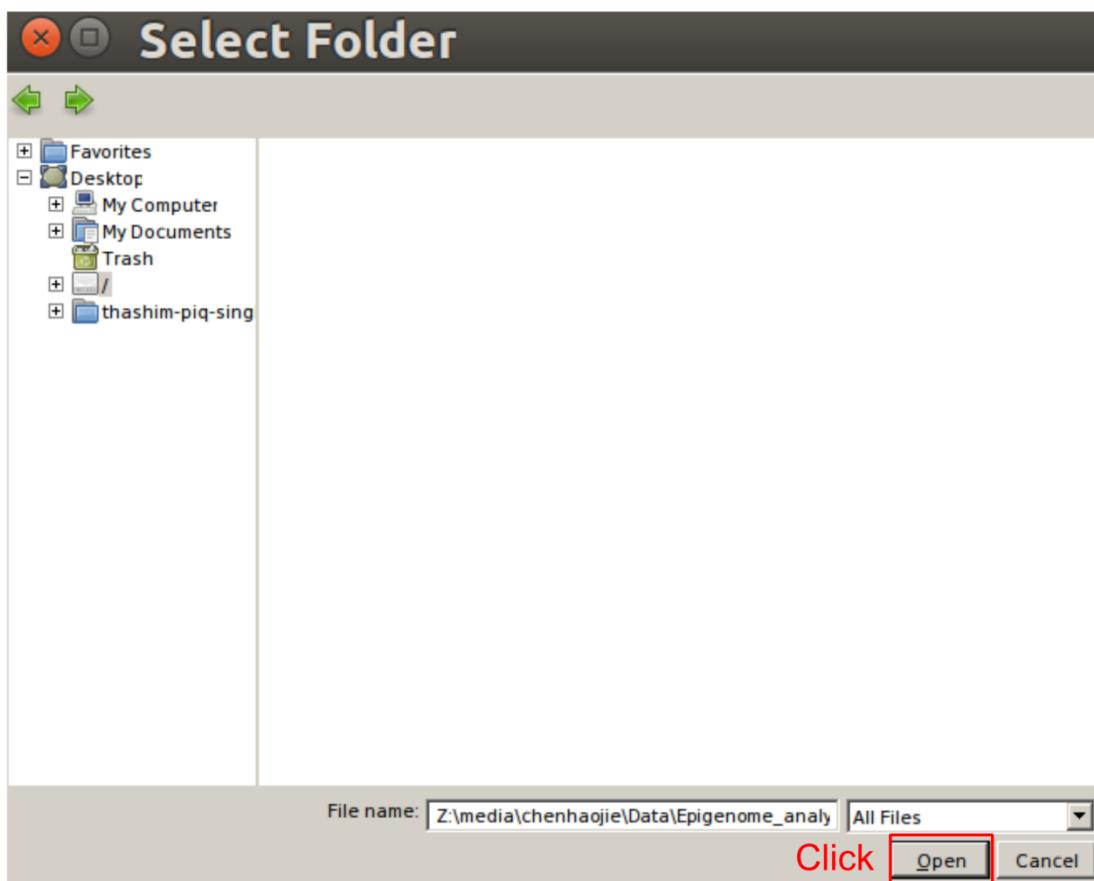


Select files (or folder) and click on **download**.

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

	object name	storage type	size	Last Modified	operate
<input type="checkbox"/>	Motif_enrichment_de_novo_motif.txt	Private Cloud	7 KB	2022/08/23 09:02:43	
<input checked="" type="checkbox"/>	report.pdf	Private Cloud	34571 KB	2022/08/23 09:07:37	
<input type="checkbox"/>	step1_fastqc_and_trim_galore	Private Cloud	--	2022/08/21 21:59:45	
<input type="checkbox"/>	step7_differential_analysis	Private Cloud	--	2022/08/21 21:59:46	
<input type="checkbox"/>	step8_functional_enrichment	Private Cloud	--	2022/08/21 21:59:46	
<input type="checkbox"/>	Quality_control_for_bases.txt	Private Cloud	132 KB	2022/08/23 09:02:14	
<input type="checkbox"/>	step4_motif_enrichment	Private Cloud	--	2022/08/21 21:59:46	
<input type="checkbox"/>	Quality_control_for_peaks_calling1.txt	Private Cloud	4 KB	2022/08/23 09:02:42	
<input type="checkbox"/>	Quality_control_for_mapping2.txt	Private Cloud	7 KB	2022/08/23 09:02:34	
<input type="checkbox"/>	step6_reads_counting	Private Cloud	--	2022/08/21 21:59:46	

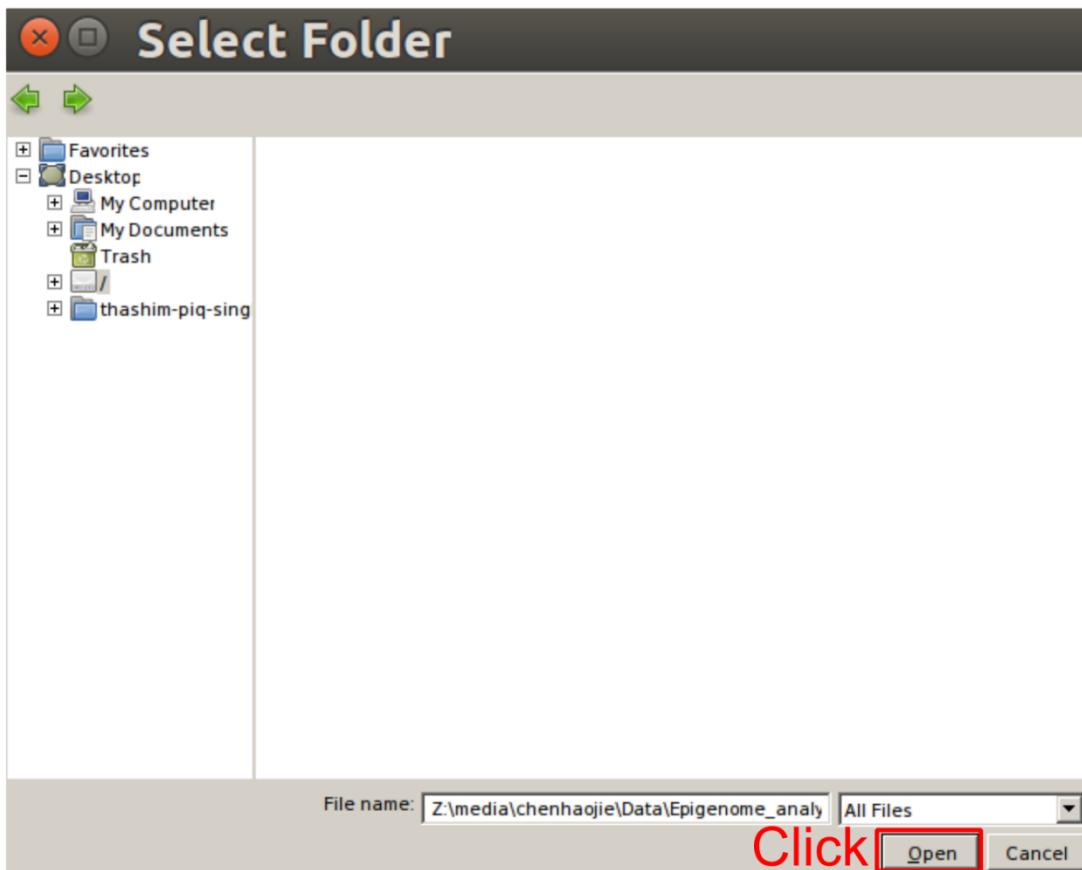
Select a local folder and click on Open to download the files.



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

	object name	storage type	size	Last Modified	operate
<input type="checkbox"/>	step4_motif_enrichment	Private Cloud	--	2022/08/21 21:59:46	
<input type="checkbox"/>	Quality_control_for_peaks_calling1.txt	Private Cloud	4 KB	2022/08/23 09:02:42	
<input type="checkbox"/>	Quality_control_for_mapping2.txt	Private Cloud	7 KB	2022/08/23 09:02:34	
<input checked="" type="checkbox"/>	step6_reads_counting	Private Cloud	--	2022/08/21 21:59:46	
<input type="checkbox"/>	result.css	Private Cloud	1017 B	2022/08/23 09:29:30	
<input type="checkbox"/>	Motif_enrichment.html	Private Cloud	3 KB	2022/08/23 09:29:30	
<input type="checkbox"/>	Quality_control_for_peaks_calling.html	Private Cloud	4 KB	2022/08/23 09:29:30	
<input type="checkbox"/>	Differential_analysis.txt	Private Cloud	528 B	2022/08/23 09:07:36	
<input type="checkbox"/>	step5_peaks_annotations	Private Cloud	--	2022/08/21 21:59:46	
<input type="checkbox"/>	step2_mapping	Private Cloud	--	2022/08/21 21:59:46	

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**.

Numbering	object name
1740	Download server files:/epgu-01/LUAD_H3K27ac_ChIP_seq_output/step6_reads_counting/distal_peak_regions_2000bp.txt
1741	Download server files:/epgu-01/LUAD_H3K27ac_ChIP_seq_output/step6_reads_counting/proximal_peak_regions_2000bp.txt
1742	Download server files:/epgu-01/LUAD_H3K27ac_ChIP_seq_output/step6_reads_counting/NA_profile_bins.xls
1743	Download server files:/epgu-01/LUAD_H3K27ac_ChIP_seq_output/report.pdf

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
job10-0	CCE.Job	创建成功 ##SuccessfulCreate
job10-0	CCE.Job	等待执行完毕 ##WaitJobFinished
job11-0	CCE.Job	创建成功 ##SuccessfulCreate
job11-0	CCE.Job	等待执行完毕 ##WaitJobFinished
luad-h3k27ac-seq-20220821166109036216745	GCS.Action	流程执行成功 ##ExecutionSuccess

Completion time Occurrences Status reason

2022-08-23 08:55:58 1 创建k8s Job对象 job10-0-c32e0104ac1043c2 成功. ## Create Job Instance(job10-0-c32e0104ac1043c2) Successfully.

2022-08-23 08:55:59 1 等待任务 job10-0-c32e0104ac1043c2 执行完成 ## Waiting job job10-0-c32e0104ac1043c2 finished

2022-08-23 08:57:59 1 创建k8s Job对象 job11-0-c32e0104ac1043c2 成功. ## Create Job Instance(job11-0-c32e0104ac1043c2) Successfully.

2022-08-23 08:57:59 1 等待任务 job11-0-c32e0104ac1043c2 执行完成 ## Waiting job job11-0-c32e0104ac1043c2 finished

2022-08-23 09:29:00 1 执行create成功 ## Action(create) Success.

Click 

### Chat dialog



No Data

Total 0 items



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 TF activity 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. [\[back\]](#)

The screenshot shows a grid of five tool cards:

- differential-analysis**: Identifies differential epigenetic modification sites between two groups of samples. Last run: shaoab, 2022-12-04 20:37:21. Buttons: Delete, Edit, Run.
- GFF & enrichment analysis**: Identifies TF motifs that are enriched in one set of peaks regions relative to another set. Last run: shaoab, 2022-12-04 21:35:01. Buttons: Delete, Edit, Run.
- Hypervariable-analysis**: Identifies hypervariable signals across samples. This tool performs hypervariable analysis in proximal and distal peak regions respectively and principal component analysis based on these identified hypervariable signals. Last run: shaoab, 2022-12-04 22:27:51. Buttons: Delete, Edit, Run.
- differential TF activity-analysis**: Identifies TF motif associated epigenetic alterations. This tool performs motif enrichment analysis within hypervariable peak regions and estimates the TF activity by aggregating the signals of peaks associated with the corresponding TF motif. Then identifies the condition specific TF motifs based on the TF activity. Last run: shaoab, 2022-12-04 22:18:20. Buttons: Delete, Edit, Run.
- clustering-analysis**: Grouping a set of samples into clusters based on hypervariable signals and creates a new meta-data file with cluster labels. User could perform differential analysis, differential motif enrichment analysis and differential TF activity analysis based on this clustering result. Last run: shaoab, 2022-12-04 22:24:33. Buttons: Delete, Edit, Run.

**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.

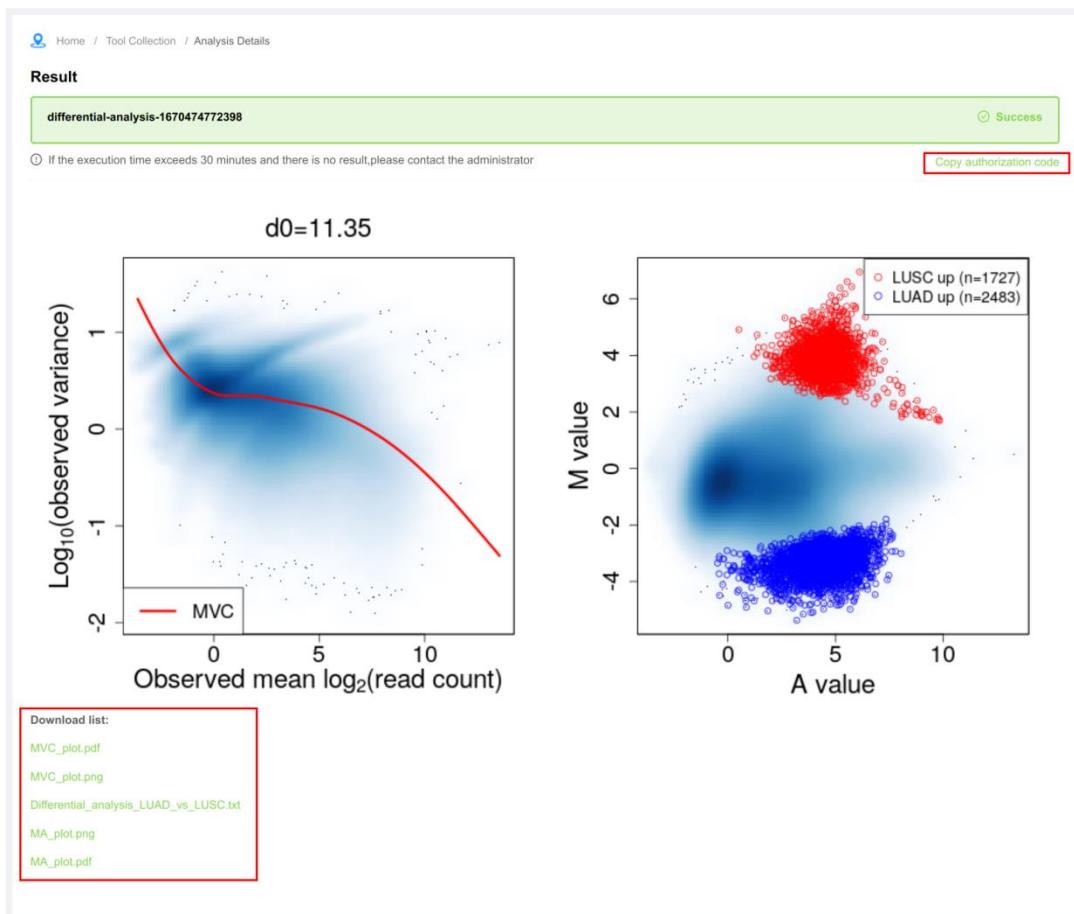
The screenshot shows the "differential-analysis" parameter setting form:

- Task Information** (highlighted in green): Inputdir: input/differential-analysis/1670474772398; Outputdir: output/differential-analysis/1670474772398. Buttons: Download Template, Run, Reset.
- Parameter setting** tab: Input file: peaks.txt (or upload from pc); Metadata file: infos.csv (or upload from pc); Variable of interest: cancer\_type; Condition1: LUAD; Condition2: LUSC; Log2FC: 1; Adjusted p-value: 0.001.
- IO demo** tab: Not visible in the screenshot.

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-analysis-1670474772398	Success	00:03:19	admin	2022-12-08 12:50:41	<span style="background-color: red;">More &gt;</span>

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 screenshot shows the 'Analysis Tool' interface with the following configuration:

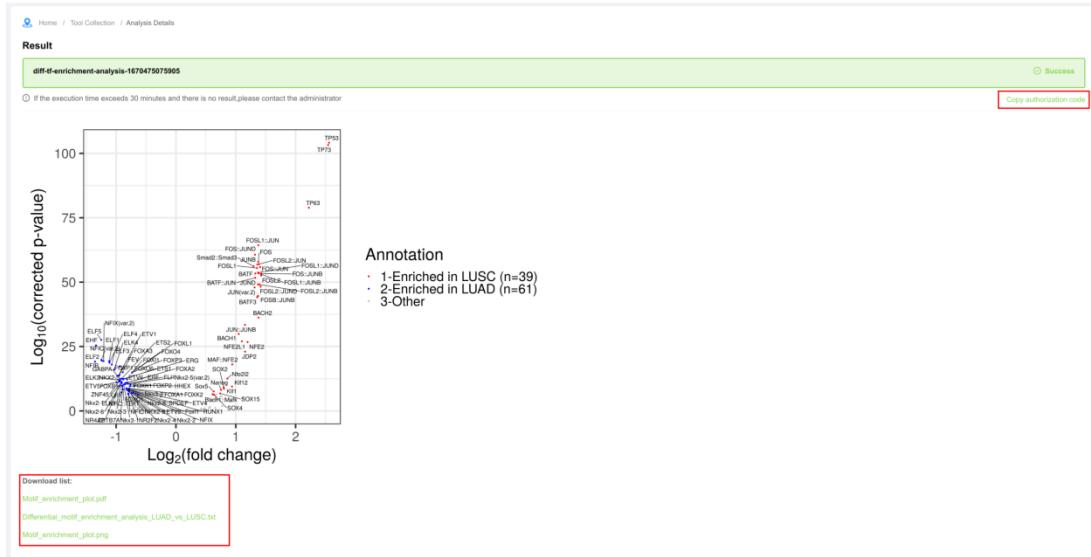
- Parameter setting:** IO demo
- Task Information:**
  - Inputdir: input/diff-tf-enrichment-analysis/1670475075905
  - Outputdir: output/diff-tf-enrichment-analysis/1670475075905
  - Input file: Differential\_analysis\_LUAD\_vs\_LUSC.txt
  - Reference genome: hg19
  - Adjusted p-value: 0.001
  - Log2FC: 1
  - The number of top ranked TF motifs to show: 100
- Buttons:** Run (highlighted with a red rectangle), Reset

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

The screenshot shows the 'Result List' interface with the following table:

task name	Status	Time Consuming	Creator	Create Date	Operation
diff-tf-enrichment-analysis-1670475075905	Success	00:15:13	admin	2022-12-08 12:59:47	<a href="#">Detail</a> <a href="#">More</a>
differential-analysis-167047472398	Success	00:03:19	admin	2022-12-08 12:50:41	<a href="#">Detail</a> <a href="#">More</a>

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 (number of samples-1)/3, default value is (number of samples-1)/3), then making some assessments with t-SNE visualization plots. Running the analysis tool by clicking on button **Run**. [\[back\]](#)

The screenshot shows the 'hypervariable-analysis' tool interface. It includes fields for Inputdir, Outputdir, Proximal peak regions file, Distal peak regions file, Metadata file, Categorical variable, The number of PCs, Perplexity, and Adjusted p-value. A red box highlights the 'Run' button at the bottom left.

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
hypervariable-analysis-1670484980625	Success	00:02:28	admin	2022-12-08 15:52:34	<a href="#">Detail</a> <a href="#">More</a>
differential-enrichment-analysis-1670476075905	Success	00:15:13	admin	2022-12-08 12:59:47	<a href="#">Detail</a> <a href="#">More</a>
differential-analysis-16704772398	Success	00:03:19	admin	2022-12-08 12:50:41	<a href="#">Detail</a> <a href="#">More</a>

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. [\[back\]](#)

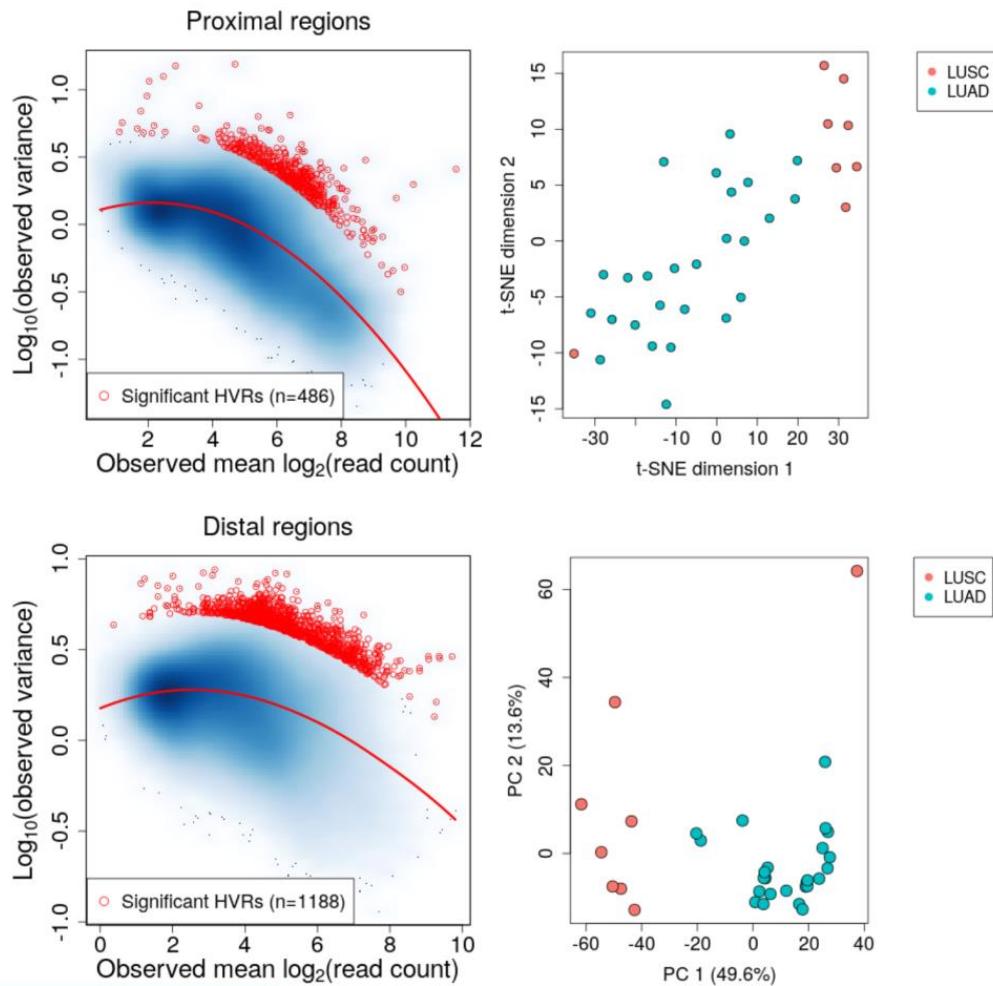
## Result

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 TF activity analysis:** Differential TF activity 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**).

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

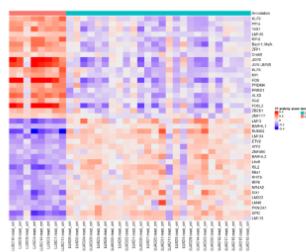
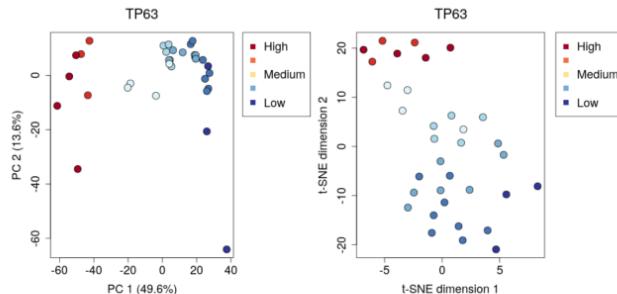
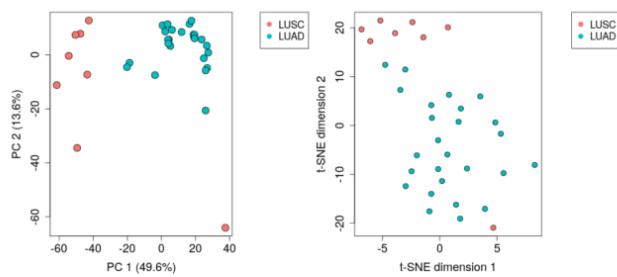
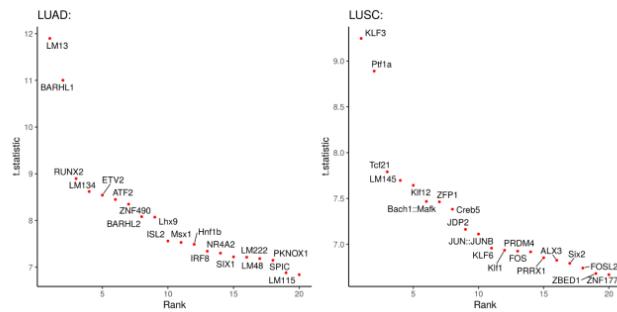
appear. Results with file size below 40MB could be downloaded directly in the web page by clicking the links from the [Download list](#). Other results should be downloaded through the client by copying the [authorization code](#) and pasting it in client. [\[back\]](#)

**Result**

differential-tf-activity-analysis-1672730260794

Success

① 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 screenshot shows the 'clustering-analysis' tool interface. At the top, there are tabs for 'Parameter setting' and 'IO demo'. Below the tabs, the title 'clustering-analysis' is displayed. The main area contains several input fields and dropdown menus:

- Inputdir:** input/clustering-analysis/1672751949953
- Outputdir:** output/clustering-analysis/1672751949953
- Proximal peak regions file:** Proximal\_hypervariable\_ or upload from pc or Proximal\_hypervariable\_analysis.RData
- Distant peak regions file:** Distant\_hypervariable\_ or upload from pc or Distant\_hypervariable\_analysis.RData
- Metadata file:** infos.csv or upload from pc or infos.csv
- Categorical variable:** cancer\_type
- Name of output clustering result file:** Test
- The number of clusters:** 2
- Adjusted p-value:** 0.001
- The number of PC:** 0

At the bottom left is a red-bordered 'Run' button, and at the bottom right is a 'Reset' button.

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

The screenshot shows the 'Result List' page. At the top, there is a search bar with placeholder 'task name: please enter task name' and a search button. Below the search bar is a table with the following columns: Task Name, Status, Time Consuming, Creator, Create Date, and Operation.

Task Name	Status	Time Consuming	Creator	Create Date	Operation
clustering-analysis-167050999976	Success	00:02:03	admin	2022-12-08 22:49:57	<a href="#">Detail</a> <a href="#">More</a>
differential-R-activity-analysis-167050373528	Success	00:08:53	admin	2022-12-08 21:14:08	<a href="#">Detail</a> <a href="#">More</a>
hypervariable-analysis-1670484980625	Success	00:02:28	admin	2023-12-08 19:52:34	<a href="#">Detail</a> <a href="#">More</a>
diff-it-enrichment-analysis-1670475075905	Success	00:15:13	admin	2022-12-08 12:59:47	<a href="#">Detail</a> <a href="#">More</a>
differential-analysis-167047477238	Success	00:03:19	admin	2022-12-08 12:50:41	<a href="#">Detail</a> <a href="#">More</a>

At the bottom right of the table, it says 'Total 5 items < 1 > 10 / page ▾'. One of the 'Detail' links in the last row is highlighted with a red rectangle.

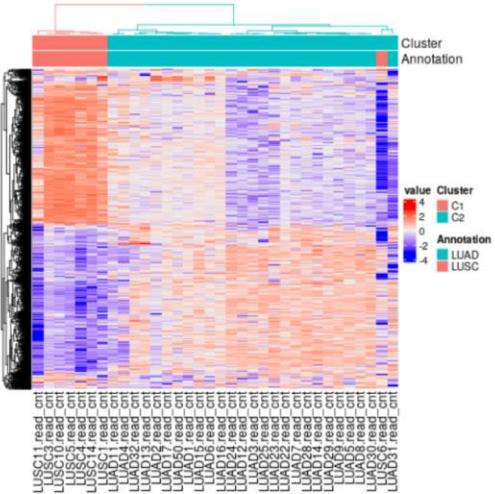
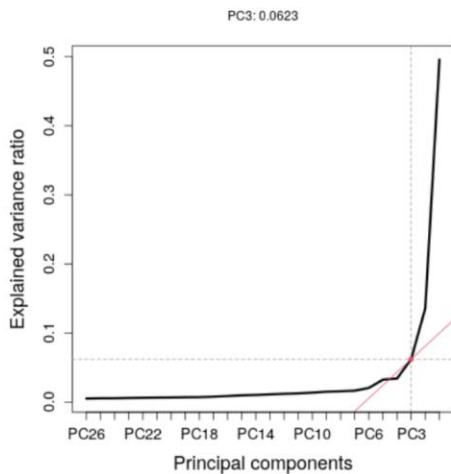
After successful completion of the submitted task. A rank plot, a dendrogram plot and a result table will be generated and can be accessed by clicking on Detail (Rectangle in Red in the image above). The rank plot indicates the recommended optimal number of principal components used for hierarchical 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

① 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\\_result.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 IO demo

### signature-genes-score-estimation

Task Information

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

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

\* Input file : Proximal\_hypenatatic\_ or upload from pc ⌘ Proximal\_hypenavable\_analysis.RData

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

\* Metadata file : Test\_clustering\_result.txt or upload from pc ⌘ Test\_clustering\_result.txt

\* Allsignature 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

PrivateCloud

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	Detail More
clustering-analysis-167272535928	Success	00:01:57	admin	2023-01-03 18:44:04	Detail More
differential-d-activity-analysis-1672720260794	Success	00:07:40	admin	2023-01-03 15:18:21	Detail More
hypervariable-analysis-1672728067237	Success	00:02:40	admin	2023-01-03 14:44:20	Detail More
diff-d-enrichment-analysis-1672727176269	Success	00:15:12	admin	2023-01-03 14:28:49	Detail More
differential-analysis-167272870354	Success	00:02:22	admin	2023-01-03 14:22:40	Detail More

Total 6 items < 1 > 10 / 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. [\[back\]](#)

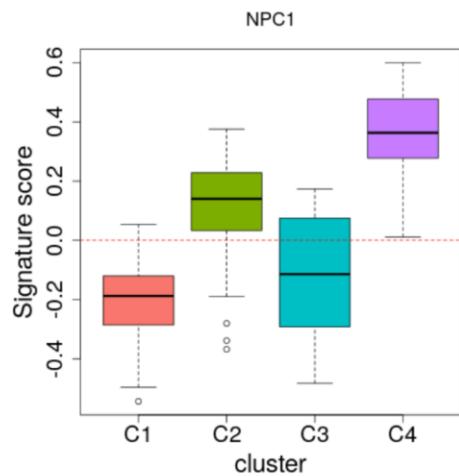
## Result

**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

These files are available on **Home page** by clicking on [Tutorial](#). [\[back\]](#)

