

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.

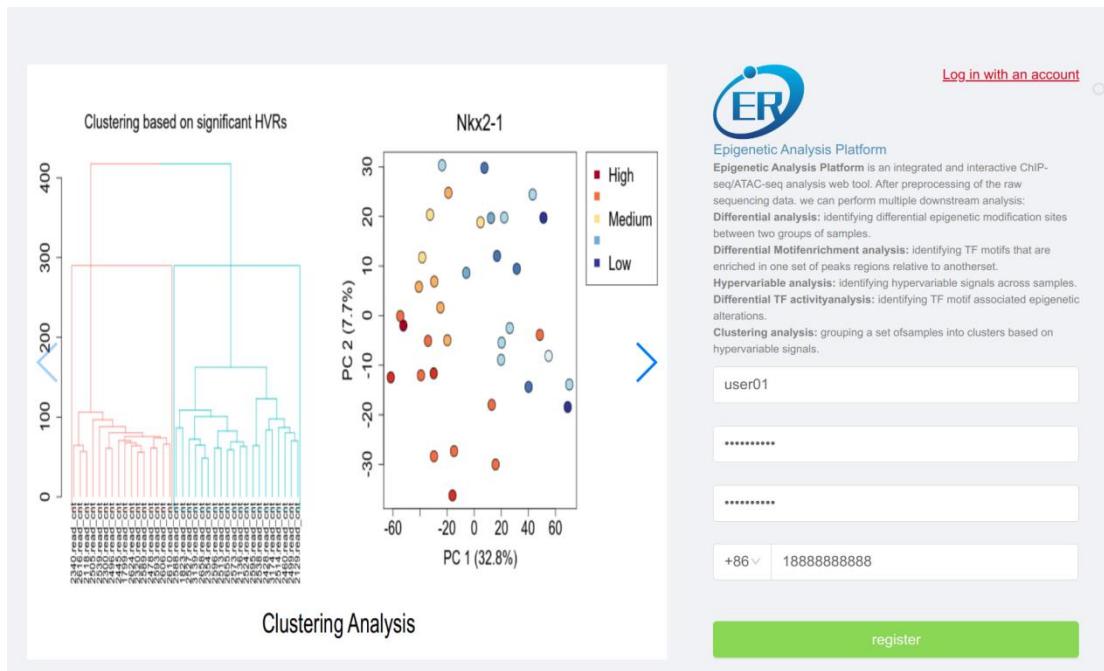
[Section 3. ChIP/ATAC-seq data preprocessing:](#) Transform raw sequencing data into analysis ready data.

[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 bellow:



Section 2. Upload raw sequencing data and metadata

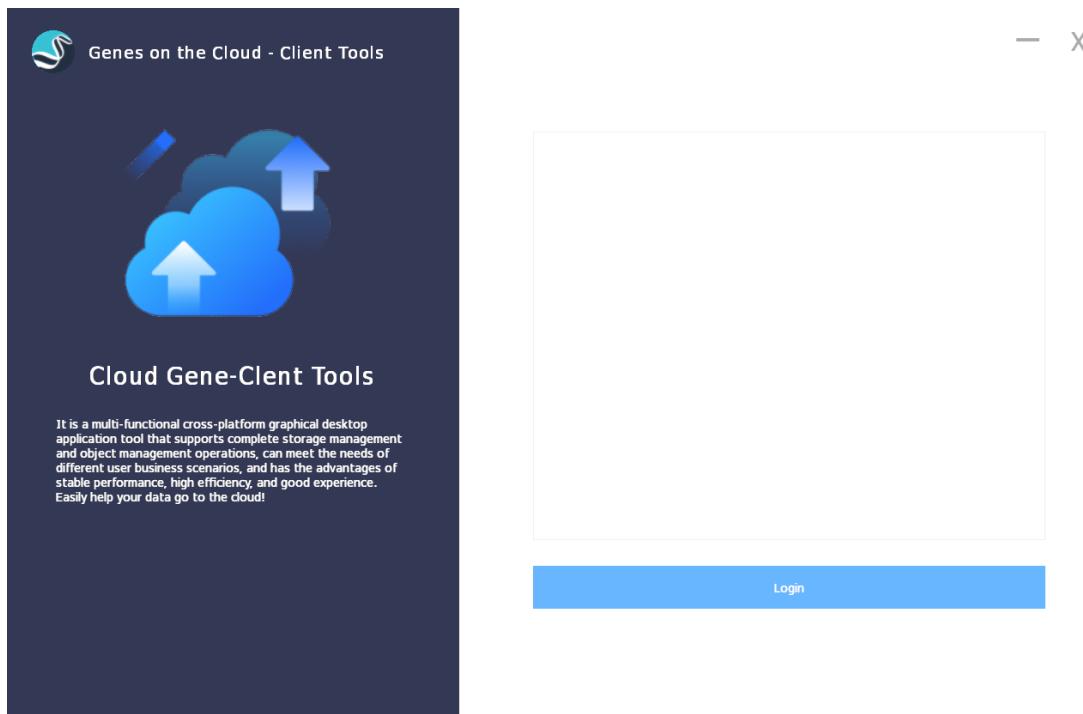
User can perform the analysis tools implemented in EAP from either raw sequencing data (i.e. FASTQ format files) or analysis ready data (following the ChIP/ATAC-seq data processing guideline in MAnorm2-utils: <https://pypi.org/project/MAnorm2-utils/>). Before running analysis tools, user need to prepare 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 More
2	.config/	Standard storage	--	2022-05-27 03:24:53	<input type="button"/> Delete More
3	xiedong_outdir/	Standard storage	--	2022-08-11 11:45:44	<input type="button"/> Delete More
4	xiedong_dir/	Standard storage	--	2022-08-05 08:22:11	<input type="button"/> Delete More
5	.cache/	Standard storage	--	2022-05-27 03:24:53	<input type="button"/> Delete More
6	HeLa_Sy_S_ChIP/	Standard storage	--	2022-08-19 11:35:34	<input type="button"/> Delete More
7	output/	Standard storage	--	2022-08-09 06:15:12	<input type="button"/> Delete More
8	H1/	Standard storage	--	2022-07-19 05:21:44	<input type="button"/> Delete More
9	localdata/	Standard storage	--	2022-08-11 03:57:00	<input type="button"/> Delete More
10	.bash_history	Standard storage	0.28 KB	2022-05-27 03:25:12	<input type="button"/> Download 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 (/).

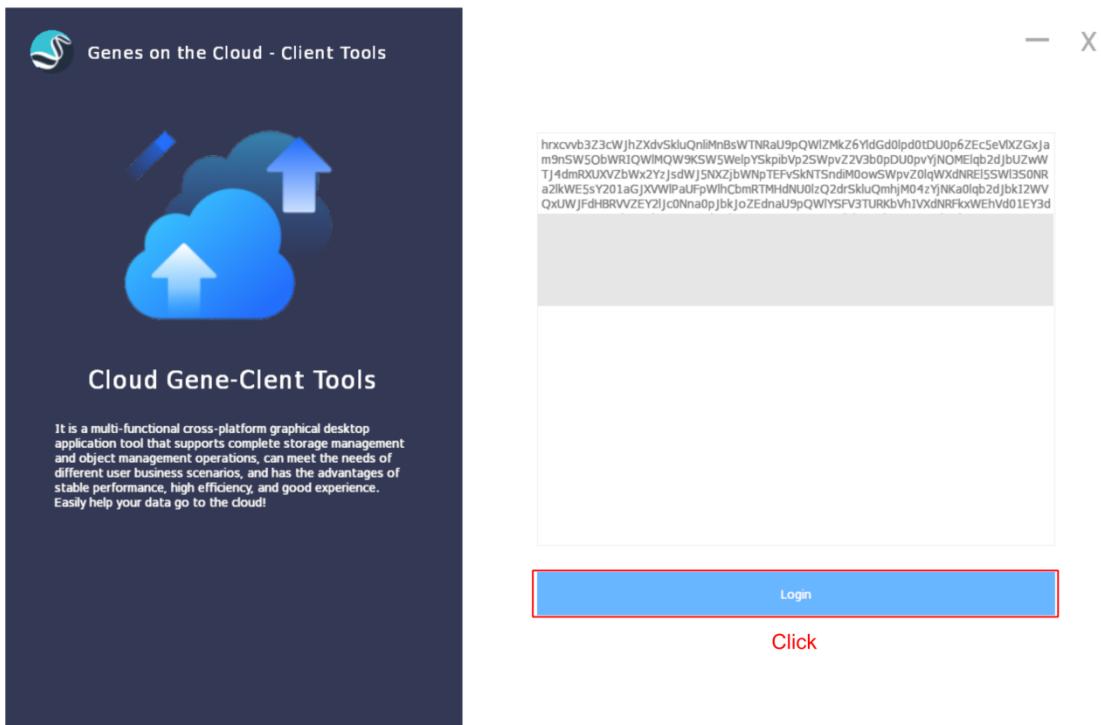
[关闭](#) [OK](#)

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**.

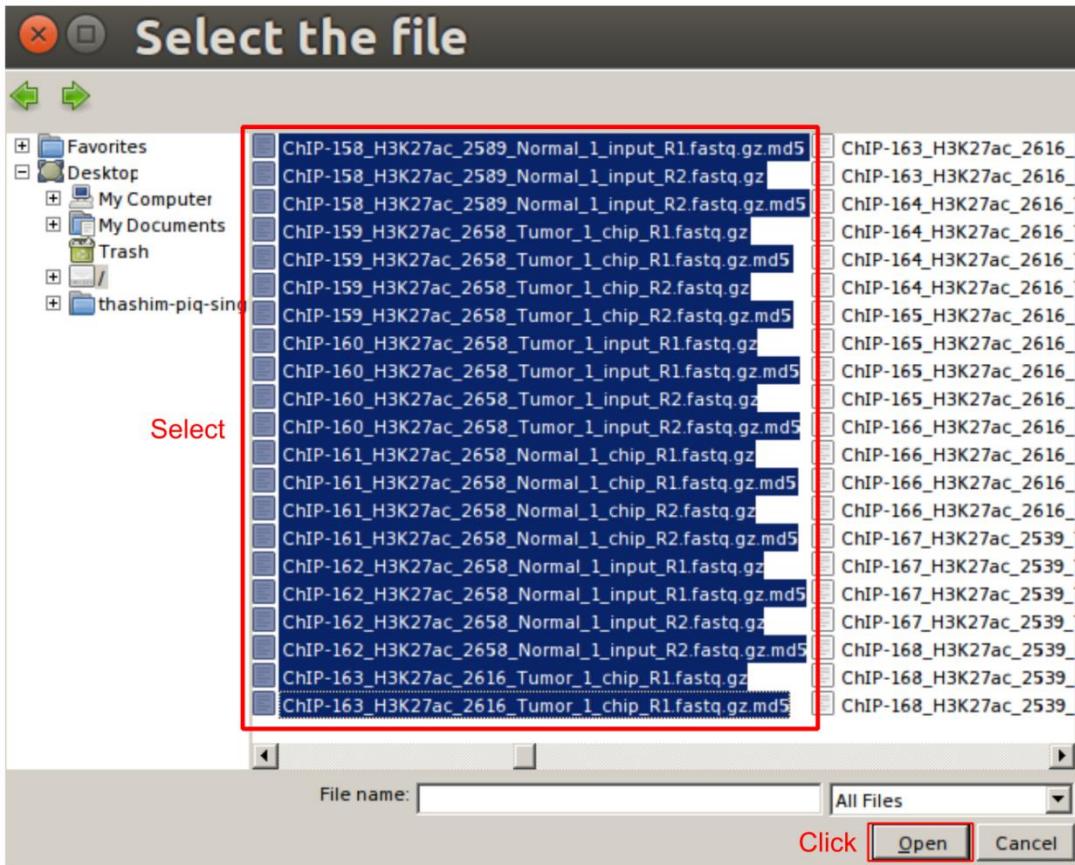
The screenshot shows a cloud storage interface with the following details:

- 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	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
- 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 /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-158_H3K27ac_2589_Normal_1_input_R1.fastq.gz.md5	upload	completed	82 B	100%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-158_H3K27ac_2589_Normal_1_input_R2.fastq.gz.md5	upload	completed	82 B	100%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-159_H3K27ac_2658_Tumor_1_chip_R1.fastq.gz	upload	uploading...	4003480 KB	16%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-159_H3K27ac_2658_Tumor_1_chip_R2.fastq.gz	upload	uploading...	4036326 KB	15%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-160_H3K27ac_2658_Tumor_1_input_R1.fastq.gz	upload	uploading...	4113169 KB	14%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-160_H3K27ac_2658_Tumor_1_input_R2.fastq.gz.md5	upload	completed	82 B	100%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-161_H3K27ac_2658_Normal_1_chip_R1.fastq.gz.md5	upload	uploading...	5000406 KB	13%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-161_H3K27ac_2658_Normal_1_input_R1.fastq.gz	upload	pause	81 B	0%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-162_H3K27ac_2658_Normal_1_input_R2.fastq.gz	upload	uploading...	4036326 KB	0%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_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 /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-163_H3K27ac_2616_Tumor_1_chip_R1.fastq.gz	upload	pause	3407234 KB	0%	2022/08/20 21:17:56	
Upload local files /rpgu-01/UPLOAD_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 /rpgu-01/UPLOAD_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 /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-163_H3K27ac_2658_Tumor_1_input_R1.fastq.gz	upload	uploading...	3717122 KB	23%	2022/08/20 21:17:45	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-159_H3K27ac_2658_Tumor_1_input_R1.fastq.gz.md5	upload	completed	81 B	100%	2022/08/20 21:17:36	
Upload local files /rpgu-01/UPLOAD_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 /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-159_H3K27ac_2658_Tumor_1_chip_R2.fastq.gz	upload	completed	81 B	100%	2022/08/20 21:17:30	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-159_H3K27ac_2658_Tumor_1_input_R2.fastq.gz.md5	upload	uploading...	3767163 KB	35%	2022/08/20 21:17:27	
Upload local files /rpgu-01/UPLOAD_H3K27ac_CHIP_seq/ChIP-153_H3K27ac_2658_Tumor_1_chip_R1.fastq.gz	upload	completed	83 B	100%	2022/08/20 21:17:27	

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).

Welcome to epigenetics zone platform

Workflows Management > Workflows Task

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

EAP md5sum pipeline-v2

Version number: V2 Creation time: 2024-03-20 14:29:58 Pipeline ID: eap-md5sum-pipeline-v2

This procedure involves using md5sum to verify the integrity of uploaded files... <More>

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).

Welcome to epigenetics zone platform

Workflows Management > One key analysis

Task information

Execution parameters

Parameter name	type	Parameter value	parameter
Inputdir	string	input/eap	dir of files to be checked
Outputdir	string	output/eap	outdir of md5sum_summary.txt
md5testfile	string	md5sum.t or upload from pc	md5sum file generate by user follow scripts: cd {dir_of_files} && md5sum * > md5sum.txt, should be put into the Inputdir which containing files to be checked

Advanced parameters

close submit Click

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

Welcome to epigenetics zone platform

Workflows Management > Execute Result

Process name	Task name	Execution status	Implementation	Time consuming	Creator	Creation time	Operation
ATAC-seq analysis pipeline-v2	atac-analysis-v2-02-23-205847	Succeeded	<div style="width: 100%;"><div style="width: 100%;"> </div></div>	11:22:04	admin	2024-02-23 2	delete clear cache
ATAC-seq/ChIP-seq analysis pipeline-v3	eap-pipeline-v3-02-23-205627	Succeeded	<div style="width: 100%;"><div style="width: 100%;"> </div></div>	19:46:35	admin	2024-02-23 2	delete clear cache
EAP md5sum pipeline-v2	eap-md5sum-pipeline-v2-02-23-202447	Succeeded	<div style="width: 100%;"><div style="width: 100%;"> </div></div>	00:14:31	admin	2024-02-23 2	delete clear cache

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.

The screenshot shows the Epigenetics zone platform interface. The top navigation bar includes links for Home, Tool List, Execute Result, Workflows Task, Workflows Config, Workflows Detail, One key analysis, and My Data. The My Data section is active. Below this, a sub-menu for Storage Management is open, showing a list of files in a PrivateCloud bucket. The file 'md5sum_summary.txt' is listed, showing its details: Standard storage, 1.08 KB, last modified 2024-02-23 20:41:22. A red box highlights the 'Download' button. The terminal window below shows the contents of 'md5sum_summary.txt'.

```

1 02cd08353be1041fd89b2dc8a4ed0383 Unknown_BV820-001A0001_1.fq.gz OK
2 fd99225f620668954b7268db6c7d27d2 Unknown_BV820-001A0001_2.fq.gz OK
3 28825f86c2a3d4796a9c5994d7c048b1 Unknown_BV820-001A0002_1.fq.gz OK
4 79ef31abdd8e51ba4ff1ce7505bddb68 Unknown_BV820-001A0002_2.fq.gz OK
5 b74a6ef90cae8e43d211efa47f80e3ed Unknown_BV820-001A0003_1.fq.gz OK
6 0b6a25841eb61eb12ebadb9a3f0088b1 Unknown_BV820-001A0003_2.fq.gz OK
7 7324a5ac91b39e32a9b4beeff72bb0d7 Unknown_BV820-001A0004_1.fq.gz OK
8 790ae69bd01f6137d7e0af6a5e6df25b Unknown_BV820-001A0004_2.fq.gz OK
9 715e60176d3bab98f210ade0a38a347c Unknown_BV820-001A0005_1.fq.gz OK
10 c65192cc2c0bf3be0d230ce69dc4040e Unknown_BV820-001A0005_2.fq.gz OK
11 5d347c169cc2c1916e0c6624472025a8 Unknown_BV820-001A0006_1.fq.gz OK
12 a6d0c74f1f4bff632727078fa675ea66 Unknown_BV820-001A0006_2.fq.gz OK
13 f20d234938dc9d306b5521e371ad12e7 Unknown_BV820-001A0007_1.fq.gz OK
14 98832a0f6c7febfb658202e9a1675b0ad Unknown_BV820-001A0007_2.fq.gz OK
15 4fd0031200afdc4348f9d6332cdf2e8b Unknown_BV820-001A0008_1.fq.gz OK
16 L07f917033bcdd552c085534dafcb00c Unknown_BV820-001A0008_2.fq.gz OK
17 Total:16; OK:16; NO:0; Not exist: 0

```

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	LUAD_H3K27ac_ChIP_seq_output/	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).



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:	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: <input checked="" type="radio"/> Use last pulled image(recommend) <input type="radio"/> 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
Category title (click Change)			
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_H3K27_ 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_H3K27_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**)

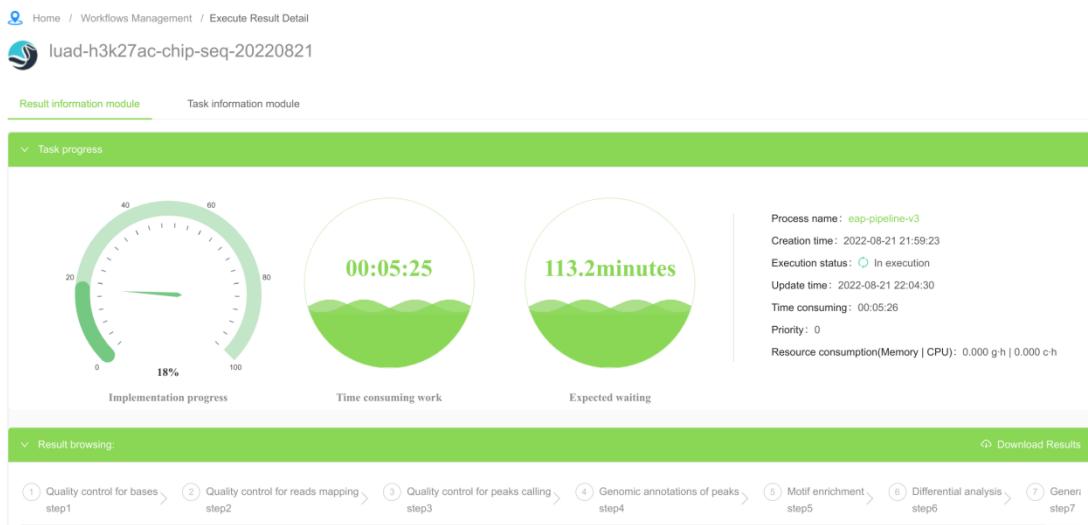
Home / Workflows Management / Excute Result

PrivateCloud

<input type="checkbox"/> 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	Running	<div style="width: 18%;">18%</div>	00:03:06	epgu-01	2022-08-21 21:59:23 delete clear cache
<input type="checkbox"/>	ATAC-seq/ChIP-seq analysis pipeline-v3	eap-pipeline-v3-08-19-133854	Succeeded	<div style="width: 100%; background-color: green;">100%</div>	01:14:05	epgu-01	2022-08-19 13:40:48 delete clear cache
<input type="checkbox"/>	ATAC-seq/ChIP-seq analysis pipeline-v3	eap-pipeline-v3-08-18-152545	Succeeded	<div style="width: 100%; background-color: green;">100%</div>	03:23:35	epgu-01	2022-08-18 15:27:00 delete clear cache
<input type="checkbox"/>	ChIP-seq analysis pipeline-v2	chip-analysis-v2-08-02-124031	Succeeded	<div style="width: 100%; background-color: green;">100%</div>	01:28:33	epgu-01	2022-08-02 12:41:50 delete clear cache
<input type="checkbox"/>	ATAC-seq analysis pipeline-v2	atac-analysis-v2-08-02-110566	Succeeded	<div style="width: 100%; background-color: green;">100%</div>	00:57:33	epgu-01	2022-08-02 11:06:37 delete clear cache

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 > Execute 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
ATAC-seq/ChIP-seq analysis pipeline-v3	luad-h3k27ac-chip-seq-20220821	Succeeded	<div style="width: 100%;"><div style="width: 100%;">Click</div></div>	35:29:37	epgu-01	2022-08-21 21:59:23	delete clear cache
ATAC-seq/ChIP-seq analysis pipeline-v3	eap-pipeline-v3-08-18-152545	Succeeded	<div style="width: 100%;"><div style="width: 100%;">Click</div></div>	03:23:35	epgu-01	2022-08-18 15:27:00	delete clear cache
ChIP-seq analysis pipeline-v2	chip-analysis-v2-08-02-124031	Succeeded	<div style="width: 100%;"><div style="width: 100%;">Click</div></div>	01:28:33	epgu-01	2022-08-02 12:41:50	delete clear cache
ATAC-seq analysis pipeline-v2	atac-analysis-v2-08-02-110556	Succeeded	<div style="width: 100%;"><div style="width: 100%;">Click</div></div>	00:57:33	epgu-01	2022-08-02 11:06:37	delete clear cache

Total 4 items < 1 > 10 / page

Home / Workflows Management / Execute Result Detail

luad-h3k27ac-chip-seq-20220821

Result information module **Task information module**

Task progress

Implementation progress: 100%
Time consuming work: 35:29:37
Expected waiting: ~-- minutes

Process name: eap-pipeline-v3
Creation time: 2022-08-21 21:59:23
Execution status: Success
Update time: 2022-08-23 09:29:30
Time consuming: 35:29:38
Priority: 0
Resource consumption(Memory | CPU): 34494.258 g h | 5306.412 c h

Result browsing:

- ✓ Quality control for bases step1
- ✓ Quality control for reads mapping step2
- ✓ Quality control for peaks calling step3
- ✓ Genomic annotations of peaks step4
- ✓ Motif enrichment step5
- ✓ Differential analysis step6
- ✓ General report step7

[Download Results](#)

report.pdf

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**.

Epigenetics Welcome to epigenetics zone platform

Home Workflows Task Execute Result My Data

Storage Management / My Data

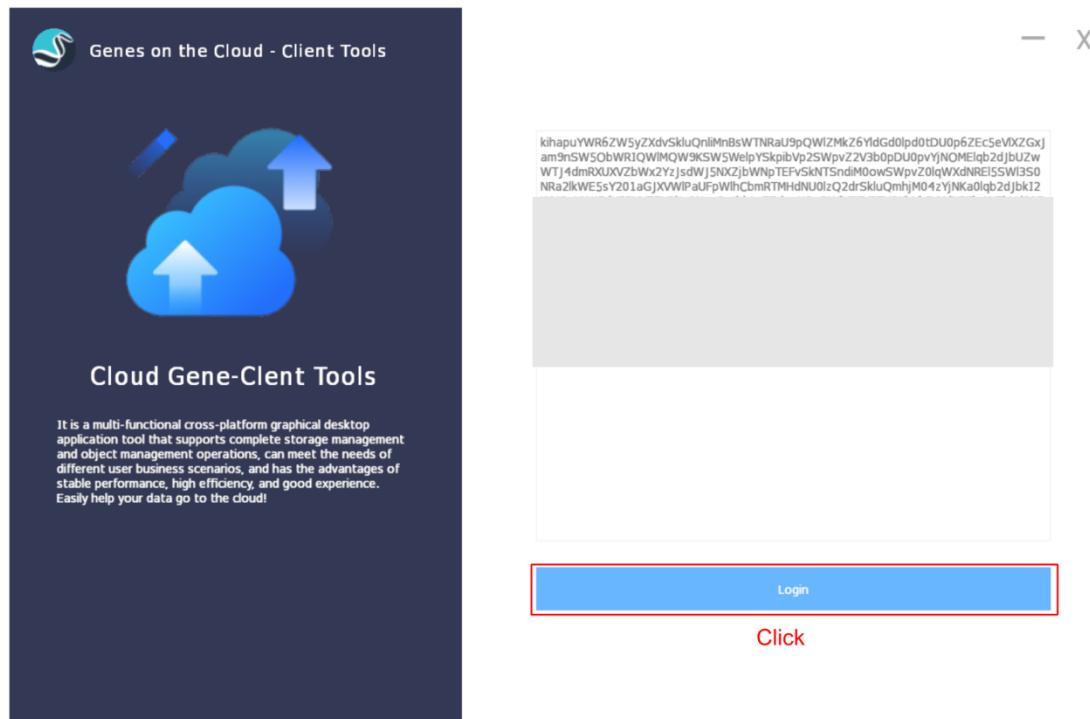
Data list-bucket: PrivateCloud

#	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	--	2022-05-21 14:00:12	<input type="checkbox"/> Delete More
10	oip/	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 path

Click



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

Example for download files (report.pdf):

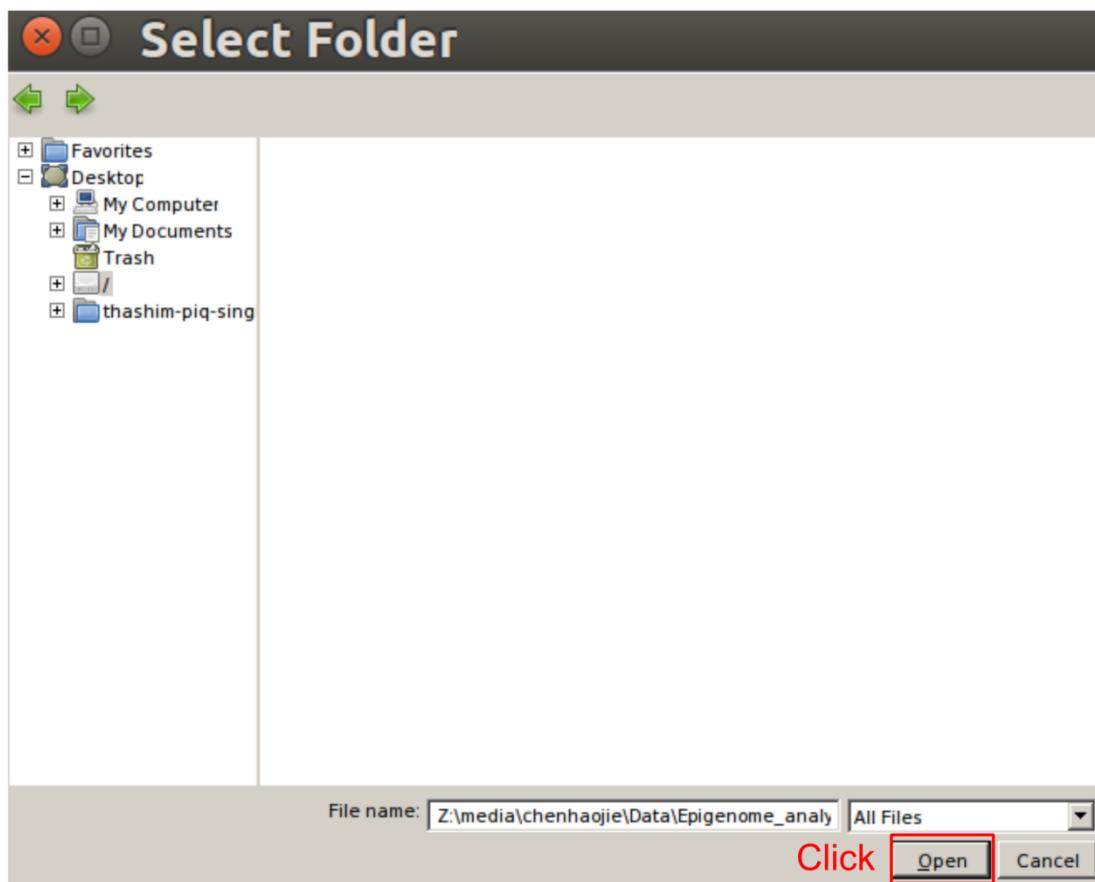
Storage management epGU-01 / LUAD_H3K27ac_ChIP_seq_output /

task management Private cloud storage Click

Download details

	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):

Storage management

task management

Download details

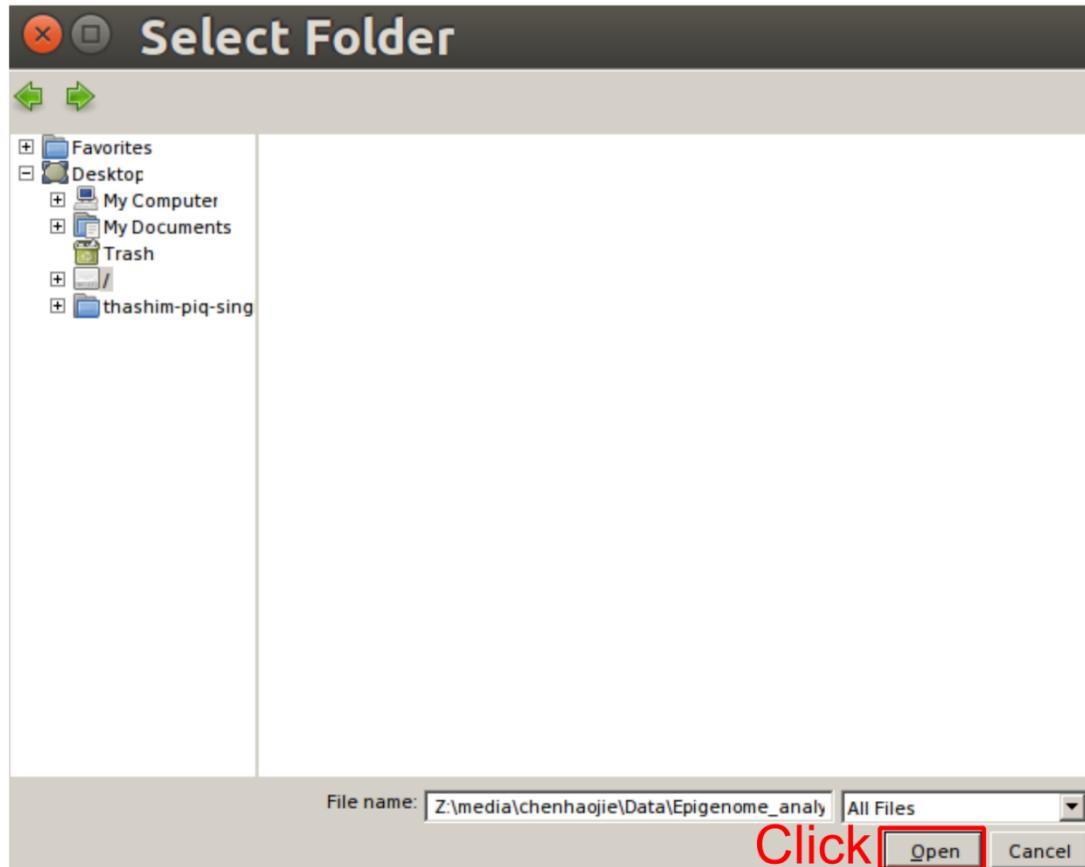
Private cloud storage Click

object name storage type size Last Modified operate

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

Select

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**.

The screenshot shows a software interface titled 'Genes on the Cloud - Client Tools' with a sub-section 'Cloud Gene-Client Tools'. A message at the top right indicates 9999 days left in the validity period. The main area has a sidebar with 'Storage management', 'task management', and 'Download details'. The 'task management' section is active, displaying a table titled 'Numbering' with four rows of data. The table columns are 'Numbering' and 'object name'. The data rows are:

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. [\[back\]](#)

The screenshot shows a 'Execute Result Detail' page for a workflow named 'luad-h3k27ac-chip-seq-20220821'. The title 'Click' is highlighted with a red box. Below it, there are two tabs: 'Result information module' and 'Task information module', with 'Task information module' being the active tab.

The following is a reference to Huawei cloud data!

Below this, there is a table with three tabs: 'Process events', 'Task event', and 'Input'. The 'Task event' tab is active. The table has columns: Task name, Task type, Process events, Completion time, Occurrences, and Status reason.

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k8s 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
job11-0	CCE.Job	创建成功 ##SuccessfulCreate	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
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.

At the bottom right of the table, there is a green button labeled 'Click' with a red box around it. Navigation buttons < 1 > are also visible.

Chat dialog

X



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 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 six analysis tools in the Tool Collection Result List:

- differential-analysis**: Identifies differential epigenetic modification sites between two groups of samples. Last run: 2022-12-04 20:37:21.
- gPRM-enrichment-analysis**: Identifies TF motifs that are enriched in one set of peaks regions relative to another set. Last run: 2022-12-04 21:05:01.
- hypervariable-analysis**: Identifies hypervariable signals across samples. Last run: 2022-12-04 22:27:51.
- differential-PRM-activity-analysis**: Identifies TF motif associated epigenetic alterations. Last run: 2022-12-04 22:18:20.
- clustering-analysis**: Groups a set of samples into clusters based on hypervariable signals and creates a new meta data file with cluster labels. Last run: 2022-12-04 22:34:33.
- Another unnamed clustering-analysis tool, last run: 2022-12-04 22:34:33.

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 parameter setting for the differential-analysis tool:

- Task Information** (highlighted in green):
 - Inputdir: input/differential-analysis/1670474772398
 - Outputdir: output/differential-analysis/1670474772398
 - Input file: peaks.txt (highlighted with a red rectangle)
 - Metadata file: infos.csv
 - Variable of interest: cancer_type
 - Condition1: LUAD
 - Condition2: LUSC
 - Log2FC: 1
 - Adjusted p-value: 0.001
- Buttons:** Run (highlighted with a red rectangle) and Reset.

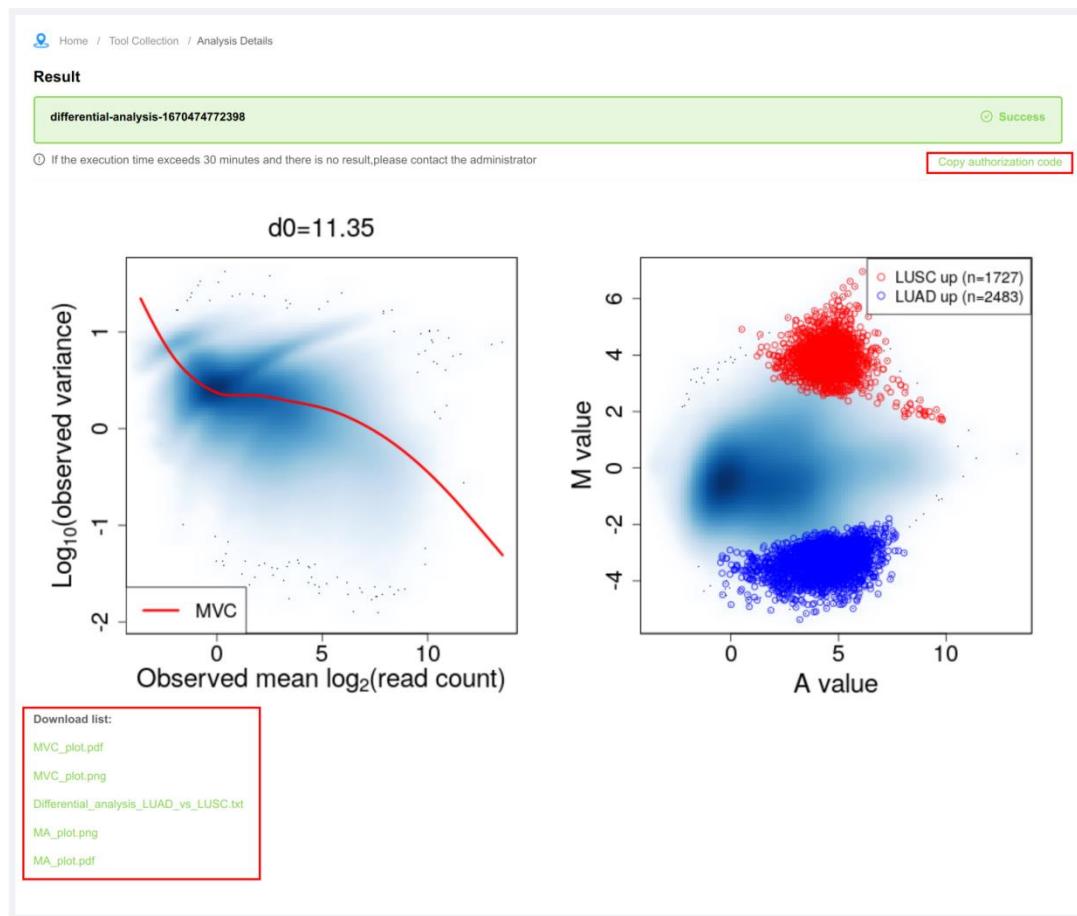
The progress of running task could be monitored in the Result List from Tool Collection (**Menu bar: Tool Collection > Result List**).

The screenshot shows the Result List table:

Task Name	Status	Time Consuming	Creator	Create Date	Operation
differential-analysis-1670474772398	Success	00:03:19	admin	2022-12-08 12:50:41	Detail More

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.

Home / Tool Collection / Analysis Tool

Parameter setting IO demo

diff-f-enrichment-analysis

Task Information

Inputdir : input/diff-f-enrichment-analysis/1670475075905
Outputdir : output/diff-f-enrichment-analysis/1670475075905

Input file : Differential_analysis_LU or upload from pc Differential_analysis_LUAD_vs_LUAD.txt

Reference genome : hg19
Adjusted p-value : 0.001
Log2FC : 1
The number of top ranked TF motifs to show : 100

Submit **Reset**

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

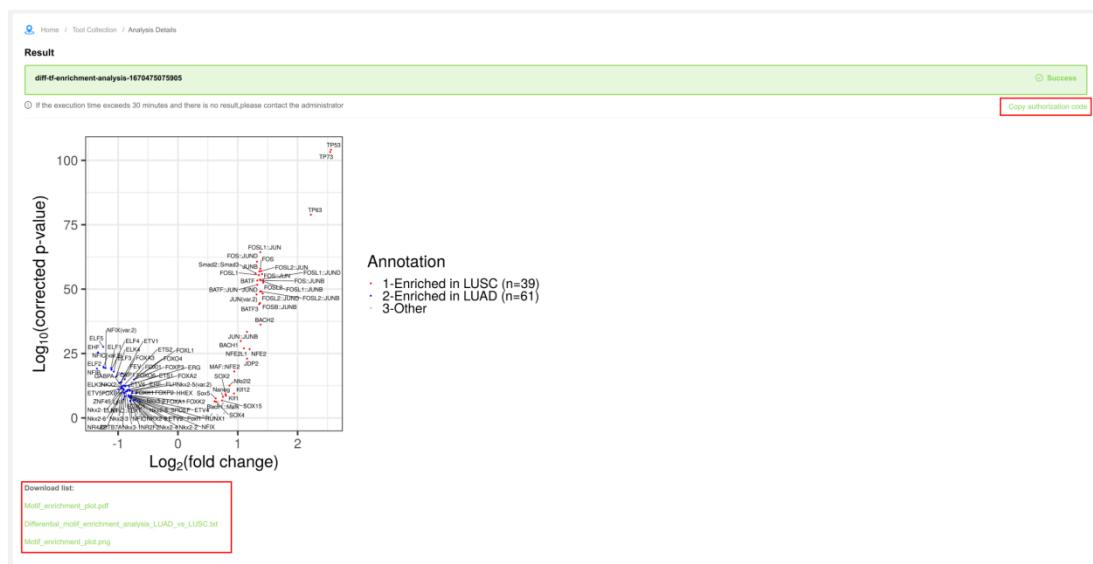
Home / Tool Collection / Result List

PrivateCloud

task name:	search	reset
<input type="checkbox"/> Task Name		
<input type="checkbox"/> diff-f-enrichment-analysis-1670475075905	Success	00:15:13
<input type="checkbox"/> differential-analysis-1670474772398	Success	00:03:19

Total 2 items < 10 / page >

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 'Analysis Tool' interface with the following configuration:

- Task Information:**
 - Inputdir: input/hypervariable-analysis/1670484980625
 - Outputdir: output/hypervariable-analysis/1670484980625
 - Proximal peak regions file: proximal_peaks_5kb.txt (selected)
 - Distal peak regions file: distal_peaks_5kb.txt (selected)
 - Metadata file: infos.csv (selected)
 - Categorical variable: cancer_type (selected)
 - The number of PCs: 3
 - Perplexity: 0
 - Adjusted p-value: 0.001
- 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 of completed tasks:

Task Name	Status	Time Consuming	Creator	Create Date	Operation
hypervariable-analysis-1670484980625	Success	00:02:28	admin	2022-12-08 15:52:34	Detail More
diffl enrichment-analysis-1670474757905	Success	00:15:13	admin	2022-12-08 12:58:47	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. 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. [\[back\]](#)

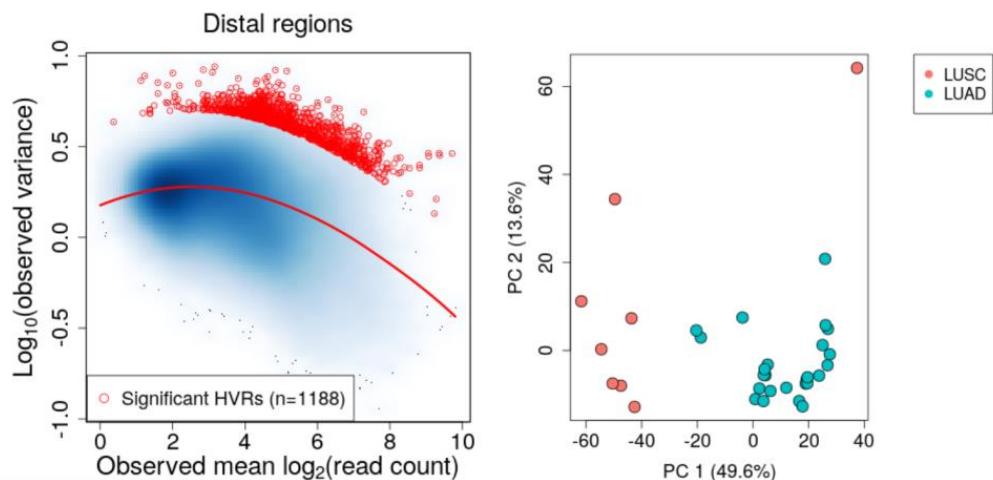
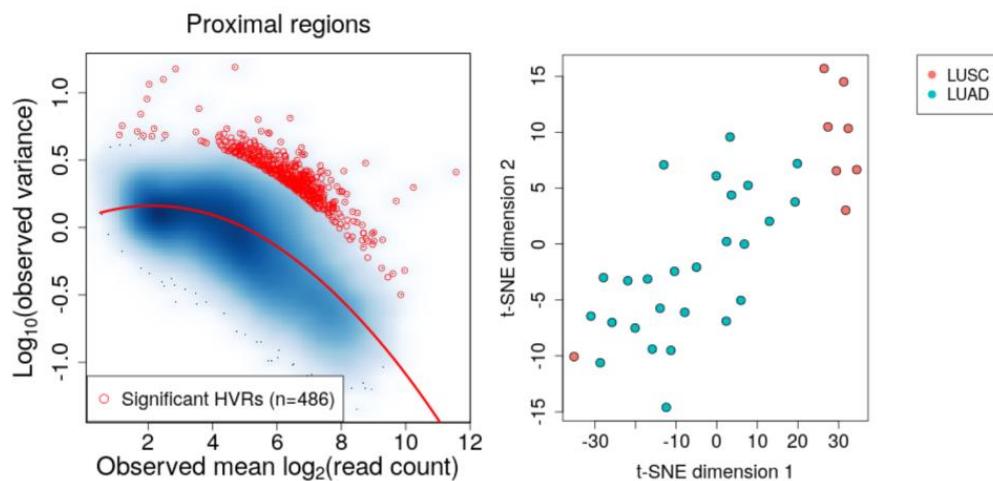
Result

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**).

Task Name	Status	Time Consuming	Creator	Create Date	Operation
differential-tf-activity-analysis-1670503735285	Success	00:08:53	admin	2022-12-08 21:14:08	
hypervariable-analysis-1670484880625	Success	00:02:28	admin	2022-12-08 15:52:34	
diff-t-enrichment-analysis-1670475075905	Success	00:15:13	admin	2022-12-08 12:59:47	
differential-analysis-1670474772398	Success	00:03:19	admin	2022-12-08 12:50:41	

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. [\[back\]](#)

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. It includes fields for Inputdir (input/clustering-analysis/1672751949953) and Outputdir (output/clustering-analysis/1672751949953). There are dropdown menus for Proximal peak regions file (Proximal_hypervariable_), Distal peak regions file (Distal_hypervariable_.RData), Metadata file (infos.csv), and Categorical variable (cancer_type). Below these are fields for Name of output clustering result file (Test), The number of clusters (2), Adjusted p-value (0.001), and The number of PC (0). At the bottom is a red-bordered 'Run' 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 with a table of completed tasks. The columns include Task Name, Status, Time Consuming, Creator, Create Date, and Operation. One row for 'clustering-analysis-167050999976' has its 'Detail' link highlighted with a red rectangle. The table shows 5 items in total.

Task Name	Status	Time Consuming	Creator	Create Date	Operation
clustering-analysis-167050999976	Success	00:02:03	admin	2022-12-08 22:49:57	Detail More
differential-f-activity-analysis-167050373528	Success	00:08:53	admin	2022-12-08 21:14:08	Detail More
hypervariable-analysis-167048490625	Success	00:02:28	admin	2023-12-08 19:52:34	Detail More
diff-f-enrichment-analysis-167047075905	Success	00:15:13	admin	2022-12-08 12:59:47	Detail More
differential-analysis-167047472388	Success	00:03:19	admin	2022-12-08 12:50:41	Detail More

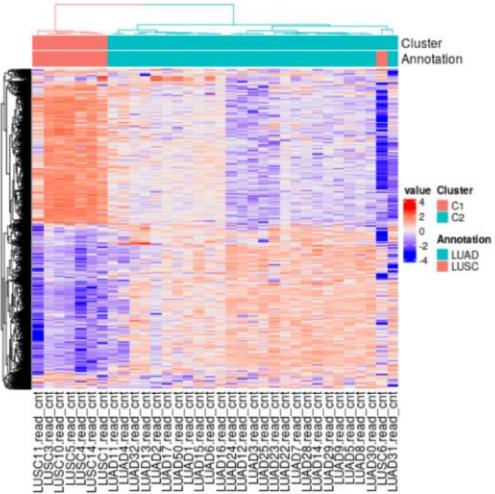
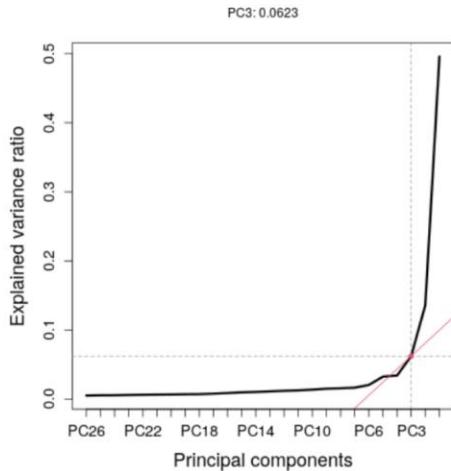
After successful completion of the submitted task. A rank plot, a dendrogram plot and a result table will be generated and can be accessed by clicking on Detail (Rectangle in Red in the image above). The rank plot indicates the recommended optimal number of principal components used for hierarchical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

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_hypenatice_ 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

* Signature genes : NPC1.gmt or upload from pc ⚡ NPC1.gmt

* Variable_of_interest : cluster

Submit 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. [\[back\]](#)

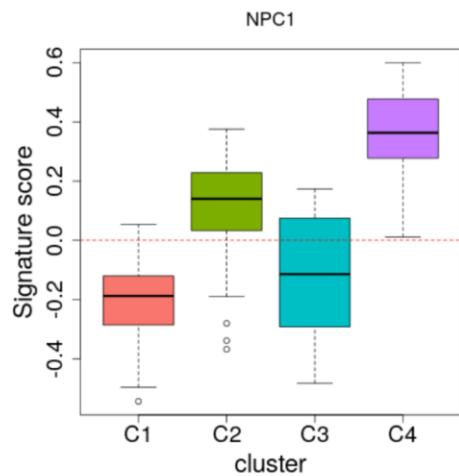
Result

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](#)

Administrators' email address:

Zhen Shao shaozhen@picb.ac.cn

Haojie Chen chenhaojie@picb.ac.cn

Zhijie Guo guozhijie@picb.ac.cn

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

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

