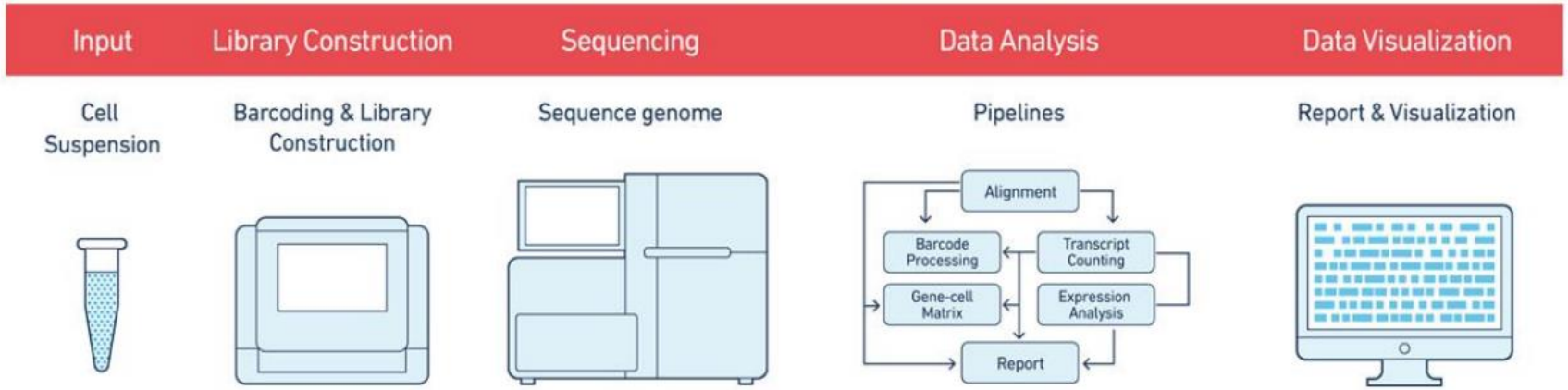
ScRNAPip: a systematic and dynamic pipeline for single-cell RNA sequencing analysis

Limin Xu[1,2,#], Jing Zhang[3,#], Yiqian He[1], Qianqian Yang[1], Tianhao Mu[3], Qiushi Guo[1,2], Yingqiang Li[3], Tian Tong[3], Shifu Chen[3,*], Richard D Ye[1,4,*]

[1]The Chinese University of Hong Kong, Shenzhen Futian Biomedical Innovation R&D Center, Shenzhen, China
[2]School of Medicine, The Chinese University of Hong Kong, Shenzhen, China
[3]HaploX Biotechnology, Shenzhen, China
[4]Kobilka Institute of Innovative Drug Discovery, School of Medicine, The Chinese University of Hong Kong, Shenzhen, China

Xu, Limin, Jing Zhang, Yiqian He, Qianqian Yang, Tianhao Mu, Qiushi Guo, Yingqiang Li, Tian Tong, Shifu Chen, and Richard D.Ye. 2023. "ScRNAPip: A Systematic and Dynamic Pipeline for Single-Cell RNA Sequencing Analysis." iMeta e132. https://doi.org/10.1002/imt2.132

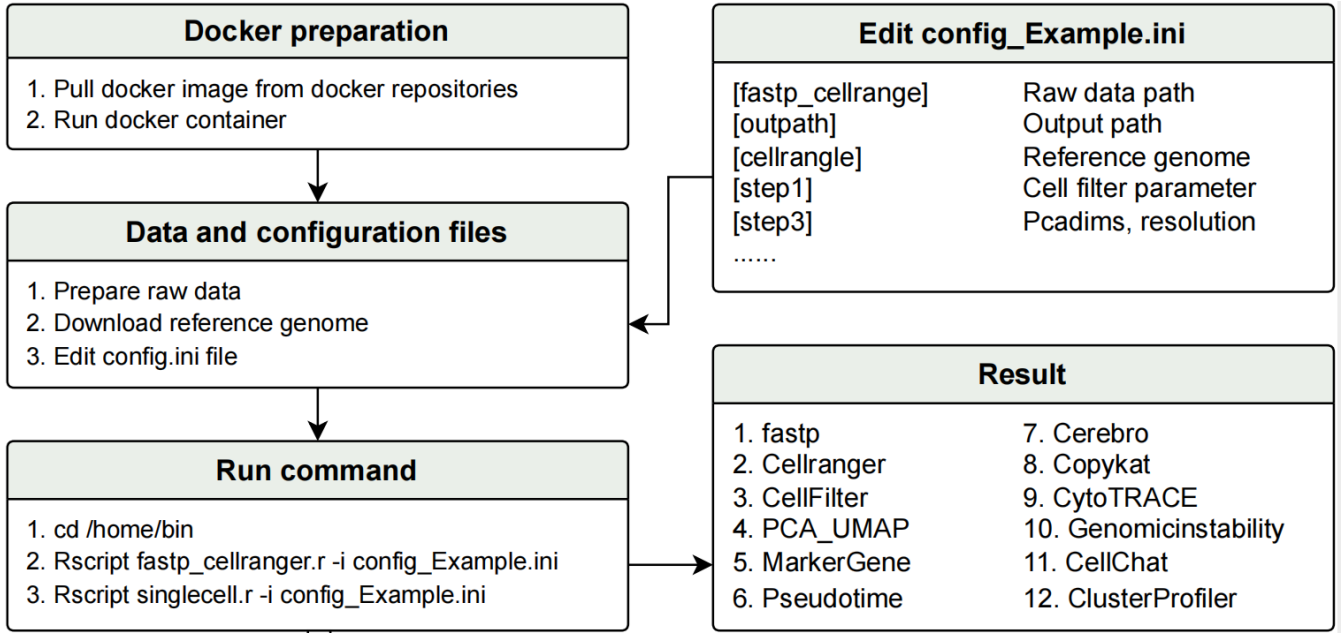
Introduction



- Currently, more than 1,500 tools have been developed for scRNA-seq data analysis in over 30 categories. This might be a great challenge for researchers in choosing appropriate tools for data analysis.

- In this study, we aim to establish a systematic, dynamic, and reproducible workflow and guide users through the key steps in the scRNA-seq analysis.

Workflow



Step 1: Pull Docker image

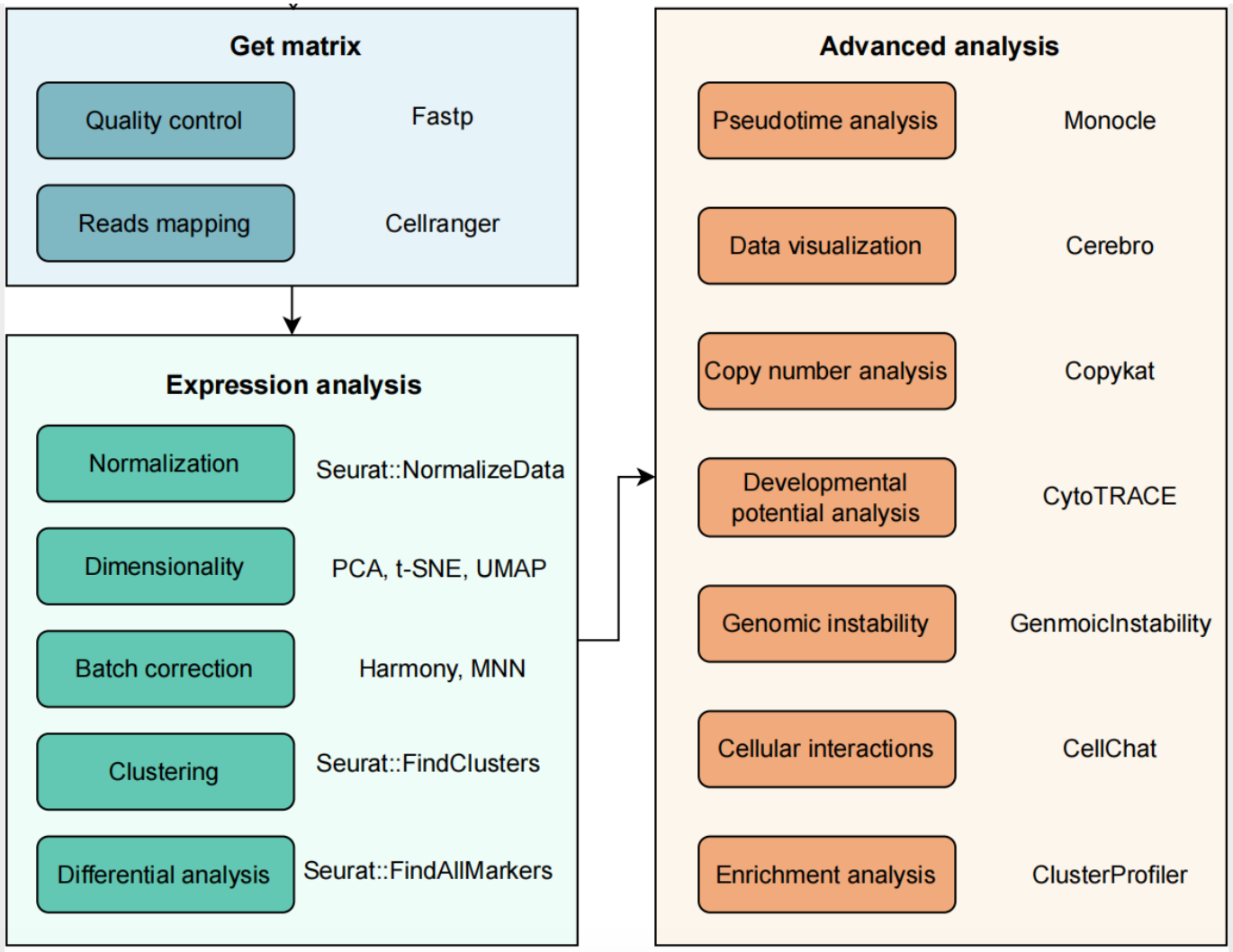
Step 2: Data and configuration file preparation

Step 3: Edit configuration file

Step 4: Execute process

Step 5: View results

Result

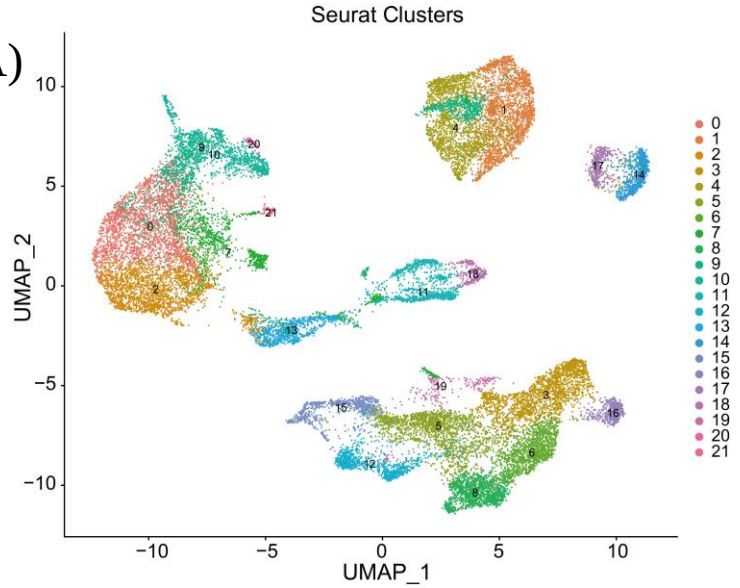


ScRNA-seq analysis consists of several steps:

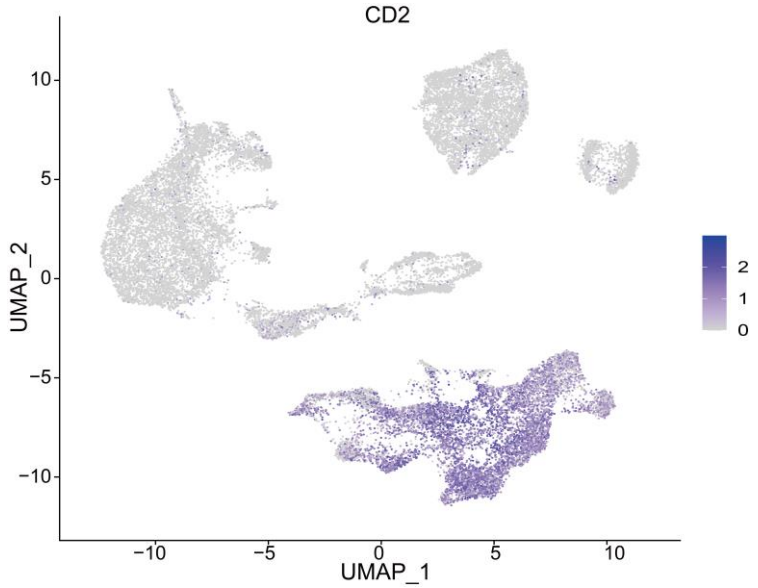
1) Raw data filtering
2) Mapping to reference genome to generate expression matrix
3) Filtering multiple, apoptotic, and poor-quality cells
4) Normalization
5) Dimensionality reduction and clustering
6) Differential analysis
7) Data visualization
8) Circos plot analysis
9) Copy number analysis
10) Developmental potential analysis
11) Genomic instability analysis
12) Cellular interactions
13) Enrichment analysis.

Result

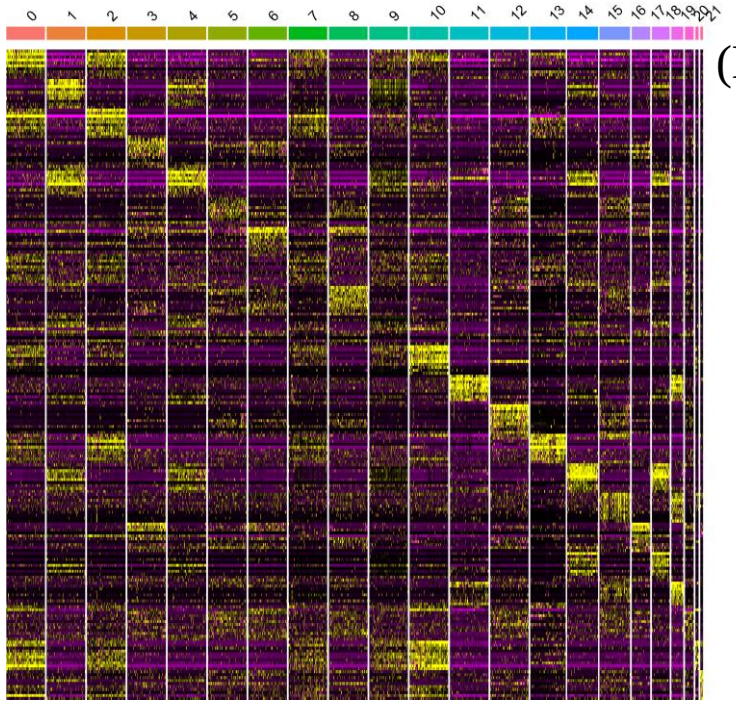
(A)



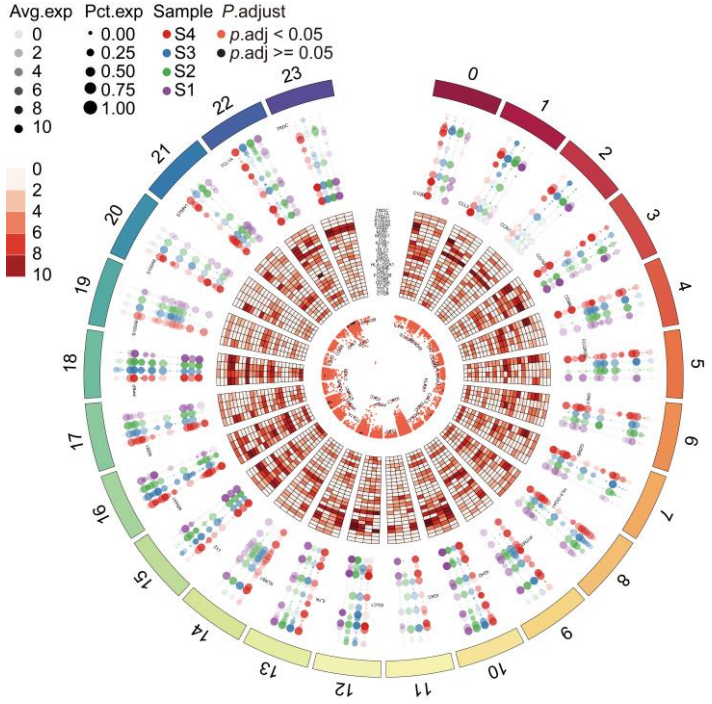
(B)



(C)



(D)



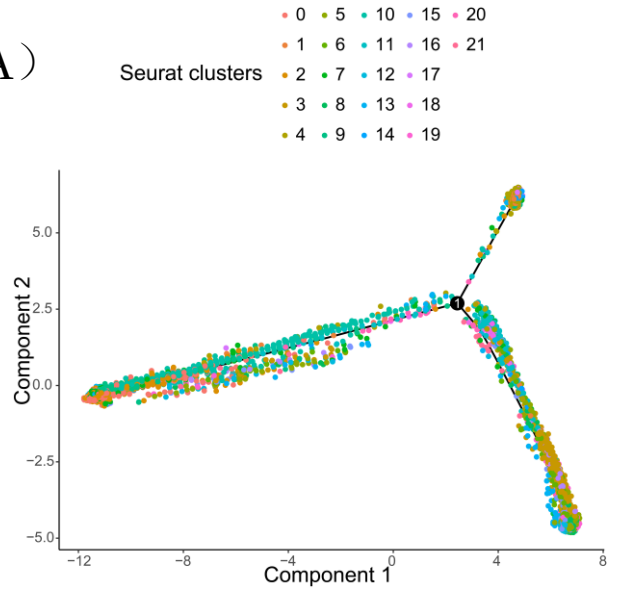
(A) UMAP plot displaying cell clustering in single-cell data.

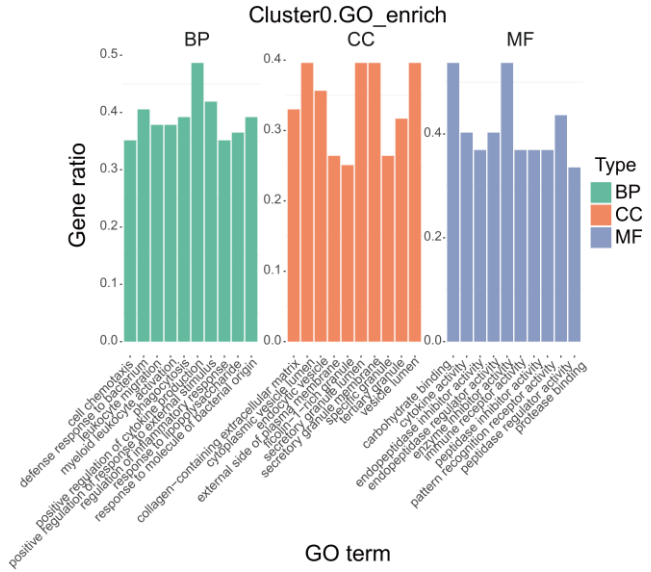
(B) Expression levels of the CD2 marker gene in individual cells.

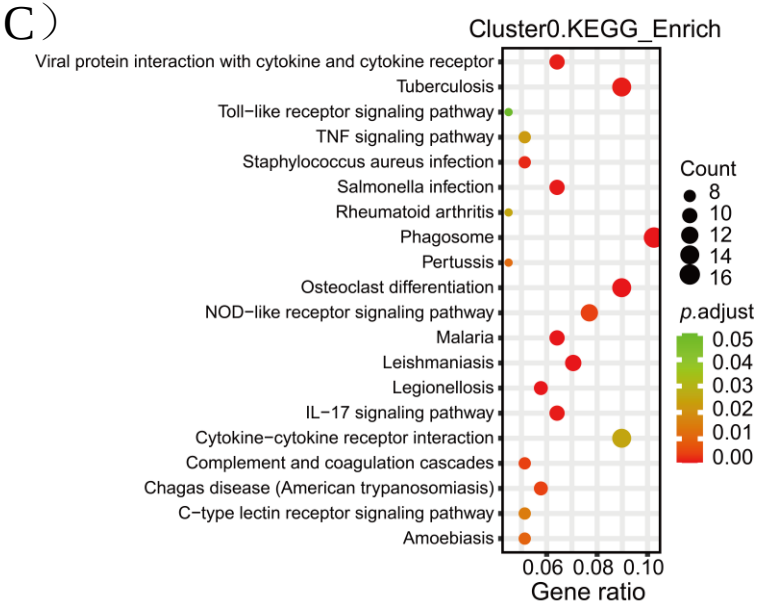
(C) Marker gene heat map, each cluster selects the top 10 marker genes according to *p*-value to draw heat map.

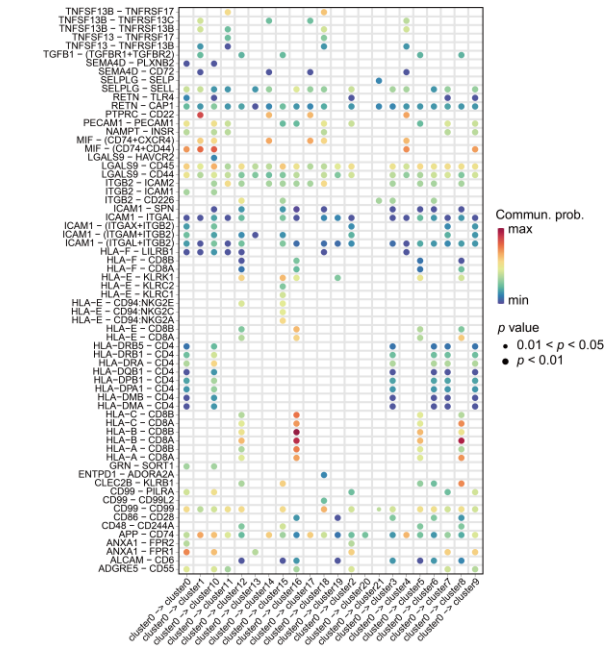
(D) Circos plot displaying marker genes for all clusters or cell types, with bubble plots, heatmaps, and scatter plots arranged on different tracks.

Result









(A) Cell trajectory built using Monocle.

(B) The bar plot shows significant enrichment GO term.

(C) The bubble plot shows the significantly enriched KEGG pathway. The size of the dot represents significantly enriched gene numbers of the pathway, and the color represents the *p*-value adjust.

(D) All the significant ligand-receptor pairs with cluster0 as the ligand. The dot color and size represent the calculated communication probability and *p*-values.

Demonstration

Pull the image of ScRNAPip

```
@gm55b-1:~>docker pull zhangjing12/scrnapip
```

Confirm your own ID information

```
@gm55b-1:~>id
uid=20022(usr) gid=20022(usr) groups=20022(usr),979(docker)
@gm55b-1:~>
```

Run the image

```
@gm55b-1:~>docker run -d -e USERID=20022 -e GROUPID=20022 -v /thinker:/thinker zhangjing12/scrnapip
```

Enter the Docker container

```
@gm55b-1:~>docker exec -ti --user 20022:20022 7b4e02613359 bash
```

Demonstration

Modify the configuration file:

Note:
An example configuration file is located in the path within the image
/home/bin/config_Example.ini

```
[fastp_cellrange]
SR39.R1=["/scRNAtest/data/rawdata/SRR8448139_S1_L001_R1_001.fastq.gz"]
SR39.R2=["/scRNAtest/data/rawdata/SRR8448139_S1_L001_R2_001.fastq.gz"]
SR40.R1=["/scRNAtest/data/rawdata/SRR8448140_S1_L001_R1_001.fastq.gz"]
SR40.R2=["/scRNAtest/data/rawdata/SRR8448140_S1_L001_R2_001.fastq.gz"]
SR41.R1=["/scRNAtest/data/rawdata/SRR8448141_S1_L001_R1_001.fastq.gz"]
SR41.R2=["/scRNAtest/data/rawdata/SRR8448141_S1_L001_R2_001.fastq.gz"]
SR42.R1=["/scRNAtest/data/rawdata/SRR8448142_S1_L001_R1_001.fastq.gz"]
SR42.R2=["/scRNAtest/data/rawdata/SRR8448142_S1_L001_R2_001.fastq.gz"]


[indata]
SR39="/scRNAtest/outfile/00.cellranger/SR39/SR39/outs/filtered_feature_bc_matrix"
SR40="/scRNAtest/outfile/00.cellranger/SR40/SR40/outs/filtered_feature_bc_matrix"
SR41="/scRNAtest/outfile/00.cellranger/SR41/SR41/outs/filtered_feature_bc_matrix"
SR42="/scRNAtest/outfile/00.cellranger/SR42/SR42/outs/filtered_feature_bc_matrix"
#       NC4="/celltest/00.cellranger/NC4/outs/filtered_feature_bc_matrix"
[outpath]
outpath="/scRNAtest/outfile"
[tempdata]
#传入的rds的路径，优先考虑默认路径下，其次再在该路径下寻找
tempdata="workout"

[run]
#运行那些步骤，需要运行的为true，不需要的为false或删除
fastp=true
cellrangle=true
step1=true
step2=true
step3=true
step4=true
step5=true
step6=true
step7=true
step8=true
step9=true
step10=true
step11=true
```

Demonstration

Run the cellranger command

```
rstudio@245c9361c1d0:/$ nohup Rscript /home/bin/fastp_cellranger.r  -i config_Example.ini 2>&1 > nohup.log &
```

Run the analysis step command

```
rstudio@245c9361c1d0:/$ nohup Rscript /home/bin/singlecell.r  -i config_Example.ini  2>&1 > nohup.log &
```

Results

```
@gm55b-1:~>cd workout
@gm55b-1:~>ls
01.Fastp        03.CellFilter  05.MarkerGene  08.copykat    10.genomicinstably  12.ClusterProfiler  temp
02.Cellranger   04.PCA_UMAP    07.Cerebro     09.cytoTRACE  11.cellchat         shell
```

Summary

- ScRNAPip provides a one-stop, user-friendly workflow.
- Comprehensively covering the entire process from raw sequencing data to personalized analysis, generating results that are easy to read and interpret.
- The selection of parameters for analysis is flexible and applicable to multiple types of cancer.

◆ ScRNAPip is publicly available via GitHub (https://github.com/OpenGene/scrnapip). A detailed description of software usage and result interpretation can be found on the website.

iMeta: Integrated meta-omics to change the understanding of the biology and environment

WILEY

"***iMeta***" is an open-access Wiley partner journal launched by scientists of the Chinese Academy of Sciences. iMeta aims to promote metagenomics, microbiome, and bioinformatics research by publishing original research, methods, or protocols, and reviews. The goal is to publish high-quality papers (Top 10%, IF > 15) targeting a broad audience. Unique features include video submission, reproducible analysis, figure polishing, APC waiver, and promotion by social media with 500,000 followers. Three issues were released in March, June , and September 2022.

Society: http://www.imeta.science
Publisher: https://wileyonlinelibrary.com/journal/imeta
Submission: https://mc.manuscriptcentral.com/imeta

office@imeta.science
iMeta

iMetaScience
iMetaScience