The Analysis Pipeline of Peak Calling in Different Plant Species

**1 Building work environment**

**1.1 The software needs to install**

* sratoolkit.2.
* STAR
* fastp
* samtools
* MACS2

**1.2 R and R package**

* R 4.2
* exomePeak2

#### 1.3 Database

### take *Arabidopsis thaliana* as an exampl*e*

## reference sequence download

wget <https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/fasta/arabidopsis_thaliana/cds/Arabidopsis_thaliana.TAIR10.cds.all.fa.gz>

## or deliver tasks to the background server

nohup wget <https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/fasta/arabidopsis_thaliana/cds/Arabidopsis_thaliana.TAIR10.cds.all.fa.gz> &

## GTF format annotation file download

wget <https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/gtf/arabidopsis_thaliana/Arabidopsis_thaliana.TAIR10.57.gtf.gz>

or

nohup wget <https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/gtf/arabidopsis_thaliana/Arabidopsis_thaliana.TAIR10.57.gtf.gz> &

## MeRIP-seq datasets download

#srrid need to be prepared

# cat download.sh

#!/usr/bin/bash

cat srrid|while read id

do

$software\_path/sratoolkit.2.11.0-centos\_linux64/bin/prefetch $id

Done

# run download.sh

nohup sh download.sh software\_path &

### 2 Get BAM formatted files

### run get\_bam.sh to get BAM formatted files

sh get\_bam.sh species dir\_path soft\_path

### rename the bam files

##Convert srrid\_name bam file to srxid\_name bam file, the file sra2srx\_id is needed, here take 12 MeRIP-seq samples of Arabidopsis thaliana as an example

## cat sra2srx\_id

SRR13522944 SRX9933631

SRR13522945 SRX9933630

SRR13522946 SRX9933629

SRR13522947 SRX9933628

SRR13522948 SRX9933627

SRR13522949 SRX9933626

SRR13522950 SRX9933625

SRR13522951 SRX9933624

SRR13522952 SRX9933623

SRR13522953 SRX9933622

SRR13522954 SRX9933621

SRR13522955 SRX9933620

## get rename\_bam\_file.sh

awk '{print "mv""\t"$1"Aligned.sortedByCoord.out.bam""\t"$2".bam"}'

sra2srx\_id >>rename\_bam\_file.sh

## run rename\_bam\_file.sh

sh rename\_bam\_file.sh

### 3 Peak calling use MACS2

### take 12 MeRIP-seq samples of Arabidopsis thaliana as an example

## cat id.txt

IP Input

SRX9933630 SRX9933620

SRX9933631 SRX9933626

SRX9933621 SRX9933629

SRX9933625 SRX9933622

SRX9933627 SRX9933623

SRX9933628 SRX9933624

## cat macs2.sh

#!/usr/bin/bash

software\_path=$1

inputs=(SRX9933620 SRX9933626 SRX9933629 SRX9933622 SRX9933623 SRX9933624)

ips=(SRX9933630 SRX9933631 SRX9933621 SRX9933625 SRX9933627 SRX9933628)

for(( i=0;i<${#inputs[@]};i++)) do

inputSample=${inputs[$i]}.bam

ipSample=${ips[$i]}.bam

$software\_path/macs2 callpeak -t $ipSample -c $inputSample -f BAM --nomodel --extsize 150 --outdir macs2\_out -B -n ${ips[$i]}\_${inputs[$i]} -q 0.05

done

## run macs2.sh

nohup sh macs2.sh oftware\_path &

### 4 Peak calling use exomePeak2

### take 12 MeRIP-seq samples of *Arabidopsis thaliana* as an example

## cat id.txt

IP Input

SRX9933630 SRX9933620

SRX9933631 SRX9933626

SRX9933621 SRX9933629

SRX9933625 SRX9933622

SRX9933627 SRX9933623

SRX9933628 SRX9933624

## cat exomePeak2.r

library("exomePeak2")

sample\_list <- read.csv("id.txt",header = TRUE,sep = "\t")

sample\_list <- split(sample\_list,sample\_list[,1])

for (id in names(sample\_list)){

exp\_name <- paste0("IP",sample\_list[[id]]$IP,"-input",sample\_list[[id]]$Input)

ip\_sample <- paste0(sample\_list[[id]]$IP,".bam")

input\_sample <- paste0(sample\_list[[id]]$Input,".bam")

exomePeak2(bam\_ip =ip\_sample, bam\_input =input\_sample, gff ="ath.gtf", save\_dir = getwd(),experiment\_name = paste0("exomePeak2\_output",exp\_name))

}

#notice :ath is the spell in a simplified form of *Arabidopsis thaliana*

## run exomePeak2.r

nohup Rscript exomePeak2.r &

### 5 Annotation Analysis

* This process was already in the RNAmod database, please refer to [RNAmod | Home (bioinformatics.sc.cn)](https://bioinformatics.sc.cn/RNAmod/index.php)
* If you use this analysis pipeline, please cite:

Liu, Q. and R.I. Gregory, *RNAmod: an integrated system for the annotation of mRNA modifications.* Nucleic Acids Res, 2019. **47**(W1): p. W548-W555.

### 6 Citation

If you make use of the data and web-server presented here, please **cite our PRMD paper**[**(2023)**](http://61.147.117.195/PRMD/download.php) in addition to the primary data sources.

The PRMD data files can be freely downloaded and used in accordance with the GNU Public License and the license of primary data sources.