

adapter_trim.T7.PE.multiprocessor.py

- For paired-end data trimming
- Packages required: Bio.Seq, fuzzywuzzy, Levenshtein, multiprocessing, pandas

```
Usage: python adapter_trim.T7.PE.multiprocessor.py -r1 R1.fastq.gz -r2
R2.fastq.gz -p 5
```

adapter_trim.T7.SE.multiprocessor.py

- For single-end data trimming
- Packages required: Bio.Seq, fuzzywuzzy, Levenshtein, multiprocessing, pandas

```
Usage: python adapter_trim.T7.SE.multiprocessor.py -i R1.fastq.gz -p 5
```

atacseqqc_for_nucleosome_dist.r

- To get nucleosome-free region and mono-nucleosome region and plot the distribution around tss.

```
mkdir NMK_1_split
Rscript atacseqqc.r NAK1_merge.pe.sort.rmdup.bam
NMK_1_L001.pe.sort.rmdup.pdf NMK_1_split &
```

correlation_basedon_peaks.r

- calculate the correlation between two bam files

```
cat Omni_ATAC_Kid.bed THS_Kid_50nM.merge.bed |sort -k1,1 -k2,2n |bedtools
merge -i stdin > Omni_new.vs.THS50nM.bed
bedtools multicov -bams Omni_ATAC_Kid_1.pe.q10.sort.rmdup.bam
Omni_ATAC_Kid_2.pe.q10.sort.rmdup.bam THS_Kid_50nM_1.q0.sort.rmdup.bam
THS_Kid_50nM_2.q0.sort.rmdup.bam -bed Omni_new.vs.THS50nM.bed >
Omni_new.vs.THS50nM.bed.cov
Rscript corr.r Omni_new.vs.THS50nM.bed.cov Omni_new FFPE
```

corr_for_duplication.r

- calculate the correlation between two replicates

```
bedtools multicov -bams trim_kidney_1.bam.sort.rmdup.bam
trim_kidney_2.bam.sort.rmdup.bam -bed FFPE_ATAC.bed > FFPE_ATAC.bed.cov
```

```
Rscript corr_dup.r FFPE_ATAC.bed.cov
```

footprint.r

- CTCF footprint calculation

```
Rscript footprint.r Kidney_SD_ATAC_1.rmdup.bam SD_1 SD_1
```

fragment_length_dist.pl

- get the fragment length distribution

```
bam=$1
out=`echo $bam|sed s/.bam//g`
perl /home/xingqichen/SOFTWARE/Code/ATAC-seq/ATAC-
seq/Code/fragment_length_dist.pl $bam $out.fragL.txt
sort -n $out.fragL.txt | uniq -c > $out.fragL.sort.txt
```

FRiP.sh

- get the FRiP value from each bam file

```
sh FRiP.sh
```

FRiP.r

- plot the bargraph for FRiP value of each sample

```
Rscript FRiP.r
```

genomic_annotation_peaks.r

- genomic annotation for all peak files

```
Rscript FRiP.r
```

length_dist.plot.line.py

- plot linegraph for fragment length distribution

multicoverage.sh

- calculate the readcounts within peaks for each sample

plot_tss.r

- plot tss enrichment

```
out=`echo $1|sed s/_bowtie2.sort.rmdup.bam//g`
bamCoverage -b $1 -o $out.bw --numberOfProcessors 20 --binSize 20
computeMatrix reference-point -S $out.bw -R
/disk1/pengweixing/database/mm9.refGene.gtf --beforeRegionStartLength
2000 --afterRegionStartLength 2000 --binSize 20 --missingDataAsZero --
sortRegions descend --skipZeros -o $out.reference-point.mat.gz -p 20 --
outFileNameMatrix $out.Matrix
plotProfile --matrixFile $out.reference-point.mat.gz --outFileName
$out.reference-point.mat.pdf
cat $out.Matrix |sed -n '4,100000p' > $out.Matrix2
Rscript /disk1/pengweixing/pipeline/tss_enrich/plot_tss.r $out.Matrix2
```

preseq.sh

- library complexity calculation based on preseq software

preseq_plot4.py

- plot library complexity distribution

relativetotss_barplot_percent.r

- bar plot for peaks relative to tss enrichment

relativetotss.r

- default plot for peaks relative to tss enrichment

venn.pair.r

- plot venn graph for pair samples

venn.r

- plot venn graph for three samples

volcano.r and differential_peaks.r

- differential analysis and volcano plot

```
/disk1/pengweixing/software/R-3.6.3/bin/Rscript
/disk1/pengweixing/pipeline/differential_peak/diff.r
THS_Kid_50nM.0mni.bed.cov name.txt
```

```
/usr/bin/Rscript /disk1/pengweixing/pipeline/differential_peak/volcano2.r  
all_peak_diff.csv THS SD
```

fastqtopeak.sh

- for mapping and peak calling