ssh -X [$USER@eddie.ecdf.ed.ac.uk](mailto:$USER@eddie.ecdf.ed.ac.uk)

qlogin -l h\_vmem=4G

conda env list

conda activate myrp-env

conda create --name myrp2 python

export LC\_COLLATE=C

scp [s2304881@eddie.ecdf.ed.ac.uk:/home/s2304881/rp/macs\_ATAC/\*](mailto:s2304881@eddie.ecdf.ed.ac.uk:/home/s2304881/rp/macs_ATAC/*)

[s2304881@129.215.237.85:/localdisk/home/s2304881/RP/ATAC\_FE.bdg](mailto:s2304881@129.215.237.85:/localdisk/home/s2304881/RP/ATAC_FE.bdg%20)

source /localdisk/home/ubuntu-software/Anaconda3/bin/activate myrp

####################

gu -g hs --keep-dup all -B -q 0.01 -n ATAC -t GM12878\_\*.bam --outdir macs\_GM12878

Format of tag file, "AUTO", "BED" or "ELAND" or "ELANDMULTI" or "ELANDEXPORT" or "SAM" or "BAM" or "BOWTIE" or "BAMPE" or "BEDPE". The default AUTO option will let MACS decide which format (except for BAMPE and BEDPE which should be implicitly set) the file is. Please check the definition in README. Please note that if the format is set as BAMPE or BEDPE, MACS2 will call its special Paired-end mode to call peaks by piling up the actual ChIPed fragments defined by both aligned ends, instead of predicting the fragment size first and extending reads. Also please note that the BEDPE only contains three columns and is NOT the same BEDPE format used by BEDTOOLS. DEFAULT: "AUTO"

macs2 bdgcmp -t macs\_GM12878/ATAC\_treat\_pileup.bdg -c macs\_GM12878/ATAC\_control\_lambda.bdg -m FE -o ATAC\_FE.bdg

Method to use while calculating a score in any bin by comparing treatment value and control value. Available choices are: ppois, qpois, subtract, logFE, logLR, and slogLR. They represent Poisson Pvalue (-log10(pvalue) form) using control as lambda and treatment as observation, q-value through a BH process for poisson pvalues, subtraction from treatment, linear scale fold enrichment, log10 fold enrichment (need to set pseudocount), log10 likelihood between ChIP-enriched model and open chromatin model(need to set pseudocount), symmetric log10 likelihood between two ChIP-enrichment models, or maximum value between the two tracks. Default option is ppois.

cp macs\_GM12878/ATAC\_peaks.narrowPeak GM12878\_macspeaks.bed

sort -k1,1 -k2,2n ATAC\_FE.bdg > ATAC\_FE.bdg.tmp

mv ATAC\_FE.bdg.tmp GM12878\_FE.bdg

rm ATAC\_FE.bdg

bedGraphToBigWig GM12878\_FE.bdg ../chrom\_sizes38 GM12878\_FE.bigWig

####################

nthreads=8

effect\_genome\_size=3099734149

samtools merge GM12878\_merged.bam GM12878\_1.bam GM12878\_2.bam

samtools index GM12878\_merged.bam

bamCoverage --numberOfProcessors $nthreads --binSize 10 --normalizeUsing RPGC --effectiveGenomeSize $effect\_genome\_size --bam GM12878\_merged.bam -o GM12878\_merged.bigWig

bigWigToBedGraph GM12878\_merged.bigWig GM12878\_merged.bdg

####################

signal\_file="GM12878\_merged.bdg"

peaks\_file="GM12878\_macspeaks.bed"

out\_file="GM12878\_finalpeaks.bed"

hist\_gp="GM12878\_hist.gp"

minHeight=10

minHeightBuffer=5

minSep=1000 # bp

minWidth=150 # bp

step=50 # bp

smooth=100 # bp

####################

Make sure the data is sorted.

awk '{OFS="\t";print $1,0,$2}' $chromsizes | sort-bed - > hg38.bounds.bed

####################

# Make histograms to estimate the lower threshold

awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' $signal\_file | sort -k1,1n > GM12878\_merged\_hist.dat

bedtools intersect -wa -a $signal\_file -b GM12878\_macspeaks.bed | awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' | sort -k1,1n > GM12878\_macspeaks\_hist.dat

echo "

set logscale xy 10

p 'GM12878\_merged\_hist.dat' u 1:2 w lp,\

'GM12878\_macspeaks\_hist.dat' u 1:2 w lp

" > $hist\_gp

gnuplot --persist $hist\_gp

rm ATAC\_hist.gp

# The two distributions look similar when the score is > 1

####################

####################

The BDG files.

Chrom ChromStart ChromEnd DataValue

####################

macs2 callpeak -f BAMPE -g hs --keep-dup all -B -q 0.01 -n ATAC -t K562\_\*.bam --outdir macs\_K562

macs2 bdgcmp -t macs\_K562/ATAC\_treat\_pileup.bdg -c macs\_K562/ATAC\_control\_lambda.bdg -m FE -o K562\_FE.bdg

cp macs\_K562/ATAC\_peaks.narrowPeak K562\_macspeaks.bed

sort -k1,1 -k2,2n K562\_FE.bdg > K562\_FE.bdg.tmp

mv K562\_FE.bdg.tmp K562\_FE.bdg

bedGraphToBigWig K562\_FE.bdg ../chrom\_sizes38 K562\_FE.bigWig

####################

nthreads=8

effect\_genome\_size=3096649726 # For GRCh38

samtools merge K562\_merged.bam K562\_1.bam K562\_2.bam

samtools index K562\_merged.bam

bamCoverage --numberOfProcessors $nthreads --binSize 10 --normalizeUsing RPGC --effectiveGenomeSize $effect\_genome\_size --bam K562\_merged.bam -o K562\_merged.bigWig

bigWigToBedGraph K562\_merged.bigWig K562\_merged.bdg

####################

signal\_file="K562\_merged.bdg"

peaks\_file="K562\_macspeaks.bed"

out\_file="K562\_finalpeaks.bed"

minHeight=10

minHeightBuffer=5

minSep=1000 # bp

minWidth=150 # bp

step=50 # bp

smooth=100 # bp

####################

Make sure the data is sorted.

awk '{OFS="\t";print $1,0,$2}' $chromsizes | sort-bed - > hg38.bounds.bed

####################

# Make histograms to estimate the lower threshold

awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' $signal\_file | sort -k1,1n > K562\_merged\_hist.dat

bedtools intersect -wa -a $signal\_file -b K562\_macspeaks.bed | awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' | sort -k1,1n > K562\_macspeaks\_hist.dat

echo "

set logscale xy 10

p 'K562\_merged\_hist.dat' u 1:2 w lp,\

'K562\_macspeaks\_hist.dat' u 1:2 w lp

" > $hist\_gp

hist\_gp="K562\_hist.gp"

gnuplot --persist $hist\_gp

rm ATAC\_hist.gp

# The two distributions look similar when the score is > 1

####################

####################

The BDG files.

Chrom ChromStart ChromEnd DataValue

####################

macs2 callpeak -f BAMPE -g hs --keep-dup all -B -q 0.01 -n ATAC -t IMR90\_\*.bam --outdir macs\_IMR90

macs2 bdgcmp -t macs\_IMR90/ATAC\_treat\_pileup.bdg -c macs\_IMR90/ATAC\_control\_lambda.bdg -m FE -o IMR90\_FE.bdg

cp macs\_IMR90/ATAC\_peaks.narrowPeak IMR90\_macspeaks.bed

sort -k1,1 -k2,2n IMR90\_FE.bdg > IMR90\_FE.bdg.tmp

mv IMR90\_FE.bdg.tmp IMR90\_FE.bdg

bedGraphToBigWig IMR90\_FE.bdg ../chrom\_sizes38 IMR90\_FE.bigWig

####################

nthreads=8

effect\_genome\_size=3096649726 # For GRCh38

samtools merge IMR90\_merged.bam IMR90\_1.bam IMR90\_2.bam

samtools index IMR90\_merged.bam

bamCoverage --numberOfProcessors $nthreads --binSize 10 --normalizeUsing RPGC --effectiveGenomeSize $effect\_genome\_size --bam IMR90\_merged.bam -o IMR90\_merged.bigWig

bigWigToBedGraph IMR90\_merged.bigWig IMR90\_merged.bdg

####################

signal\_file="IMR90\_merged.bdg"

peaks\_file="IMR90\_macspeaks.bed"

out\_file="IMR90\_finalpeaks.bed"

minHeight=10

minHeightBuffer=5

minSep=1000 # bp

minWidth=150 # bp

step=50 # bp

smooth=100 # bp

####################

Make sure the data is sorted.

awk '{OFS="\t";print $1,0,$2}' $chromsizes | sort-bed - > hg38.bounds.bed

####################

# Make histograms to estimate the lower threshold

awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' $signal\_file | sort -k1,1n > IMR90\_merged\_hist.dat

bedtools intersect -wa -a $signal\_file -b IMR90\_macspeaks.bed | awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' | sort -k1,1n > IMR90\_macspeaks\_hist.dat

echo "

set logscale xy 10

p 'IMR90\_merged\_hist.dat' u 1:2 w lp,\

'IMR90\_macspeaks\_hist.dat' u 1:2 w lp

" > $hist\_gp

hist\_gp="IMR902\_hist.gp"

gnuplot --persist $hist\_gp

rm ATAC\_hist.gp

# The two distributions look similar when the score is > 1

####################

####################

The BDG files.

Chrom ChromStart ChromEnd DataValue

####################

macs2 callpeak -f BAMPE -g mm --keep-dup all -B -q 0.01 -n ATAC -t mouse\*.bam --outdir macs\_mouse

macs2 bdgcmp -t macs\_mouse/ATAC\_treat\_pileup.bdg -c macs\_mouse/ATAC\_control\_lambda.bdg -m FE -o mouse\_FE.bdg

cp macs\_mouse/ATAC\_peaks.narrowPeak mouse\_macspeaks.bed

sort -k1,1 -k2,2n mouse\_FE.bdg > mouse\_FE.bdg.tmp

mv mouse\_FE.bdg.tmp mouse\_FE.bdg

bedGraphToBigWig mouse\_FE.bdg ./mm10.chrom.sizes mouse\_FE.bigWig

####################

nthreads=8

effect\_genome\_size= 2730855475 #For GRCm38

samtools merge mouse\_merged.bam mouse1.bam mouse2.bam

samtools index mouse\_merged.bam

bamCoverage --numberOfProcessors $nthreads --binSize 10 --normalizeUsing RPGC --effectiveGenomeSize $effect\_genome\_size –bam mouse\_merged.bam -o mouse\_merged.bigWig

bigWigToBedGraph mouse\_merged.bigWig mouse\_merged.bdg

####################

signal\_file="mouse\_merged.bdg"

peaks\_file="mouse\_macspeaks.bed"

out\_file="mouse\_finalpeaks.bed"

minHeight=10

minHeightBuffer=5

minSep=1000 # bp

minWidth=150 # bp

step=50 # bp

smooth=100 # bp

####################

Make sure the data is sorted.

awk '{OFS="\t";print $1,0,$2}' $chromsizes | sort-bed - > hg38.bounds.bed

####################

# Make histograms to estimate the lower threshold

awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' $signal\_file | sort -k1,1n > mouse\_merged\_hist.dat

bedtools intersect -wa -a $signal\_file -b mouse\_macspeaks.bed | awk '{a[$4]++}END{for (i in a) {print i,a[i]}}' | sort -k1,1n > mouse\_macspeaks\_hist.dat

echo "

set logscale xy 10

p 'mouse\_merged\_hist.dat' u 1:2 w lp,\

'mouse\_macspeaks\_hist.dat' u 1:2 w lp

" > $hist\_gp

hist\_gp="mouse\_hist.gp"

gnuplot --persist $hist\_gp

rm ATAC\_hist.gp

# The two distributions look similar when the score is > 1

####################