################## ChIP-seq analysis ##################

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

#### Part1. Trimming, filtering and mapping reads. ####

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

# Step1. Trimming the raw read files using Trim Galore! v0.6.6

trim\_galore --illumina -o ./cleandata --paired ./Bd\_Input\_R1.fastq.gz ./Bd\_Input\_R2.fastq.gz

trim\_galore --illumina -o ./cleandata --paired ./Bd\_H3K4me3\_1\_R1.fastq.gz ./Bd\_H3K4me3\_1\_R2.fastq.gz

trim\_galore --illumina -o ./cleandata --paired ./Bd\_H3K4me3\_2\_R1.fastq.gz ./Bd\_H3K4me3\_2\_R2.fastq.gz

trim\_galore --illumina -o ./cleandata --paired ./Bd\_H3K27me3\_1\_R1.fastq.gz ./Bd\_H3K27me3\_1\_R2.fastq.gz

trim\_galore --illumina -o ./cleandata --paired ./Bd\_H3K27me3\_2\_R1.fastq.gz ./Bd\_H3K27me3\_2\_R2.fastq.gz

# Step2. Mapping the reads files using Bowtie2 v2.4.2

bowtie2 -x ./ref/genome.fasta -1 ./cleandata/Bd\_Input\_R1.fq.gz -2 ./cleandata/Bd\_Input\_R2.fq.gz 2>Input.mapping.metrics.txt | samtools view -o Input.bam

bowtie2 -x ./ref/genome.fasta -1 ./cleandata/Bd\_H3K4me3\_1\_R1.fq.gz -2 ./cleandata/Bd\_H3K4me3\_1\_R2.fq.gz 2>IP\_1.mapping.metrics.txt | samtools view -o IP\_1.bam

bowtie2 -x ./ref/genome.fasta -1 ./cleandata/Bd\_H3K4me3\_2\_R1.fq.gz -2 ./cleandata/Bd\_H3K4me3\_2\_R2.fq.gz 2>IP\_2.mapping.metrics.txt | samtools view -o IP\_2.bam

bowtie2 -x ./ref/genome.fasta -1 ./cleandata/Bd\_H3K27me3\_1\_R1.fq.gz -2 ./cleandata/Bd\_H3K27me3\_1\_R2.fq.gz 2>IP\_3.mapping.metrics.txt | samtools view -o IP\_3.bam

bowtie2 -x ./ref/genome.fasta -1 ./cleandata/Bd\_H3K27me3\_2\_R1.fq.gz -2 ./cleandata/Bd\_H3K27me3\_2\_R2.fq.gz 2>IP\_4.mapping.metrics.txt | samtools view -o IP\_4.bam

# Step3. Filtering the reads files using SAMtools v1.3.1, Picard v2.25.1, and deepTools v3.5.1

## 3.1 Filtering unmapped and non-uniquely mapped reads\_Sort Aligned File

samtools sort -o Input.s.bam Input.bam

samtools sort -o IP\_1.s.bam IP\_1.bam

samtools sort -o IP\_2.s.bam IP\_2.bam

samtools sort -o IP\_3.s.bam IP\_3.bam

samtools sort -o IP\_4.s.bam IP\_4.bam

## 3.2 Filtering unmapped and non-uniquely mapped reads\_Index Aligned File

samtools index Input.s.bam

samtools index IP\_1.s.bam

samtools index IP\_2.s.bam

samtools index IP\_3.s.bam

samtools index IP\_4.s.bam

## 3.3 Filtering unmapped and non-uniquely mapped reads\_Filter

samtools view -F 3852 -o Input.af.bam Input.s.bam

samtools view -F 3852 -o IP\_1.af.bam IP\_1.s.bam

samtools view -F 3852 -o IP\_2.af.bam IP\_2.s.bam

samtools view -F 3852 -o IP\_3.af.bam IP\_3.s.bam

samtools view -F 3852 -o IP\_4.af.bam IP\_4.s.bam

# 3.4 Filtering improperly paired reads\_Index Aligned File

samtools index Input.af.bam

samtools index IP\_1.af.bam

samtools index IP\_2.af.bam

samtools index IP\_3.af.bam

samtools index IP\_4.af.bam

# 3.5 Filtering improperly paired reads\_Filter

samtools view -b -f 2 Input.af.bam -o Input.pf.bam

samtools view -b -f 2 IP\_1.af.bam -o IP\_1.pf.bam

samtools view -b -f 2 IP\_2.af.bam -o IP\_2.pf.bam

samtools view -b -f 2 IP\_3.af.bam -o IP\_3.pf.bam

samtools view -b -f 2 IP\_4.af.bam -o IP\_4.pf.bam

## 3.6 Filtering low quality reads\_Index Aligned File

samtools index Input.pf.bam

samtools index IP\_1.pf.bam

samtools index IP\_2.pf.bam

samtools index IP\_3.pf.bam

samtools index IP\_4.pf.bam

## 3.7 Filtering low quality reads\_Filter

samtools view -b -q 30 Input.pf.bam -o Input.qf.bam

samtools view -b -q 30 IP\_1.pf.bam -o IP\_1.qf.bam

samtools view -b -q 30 IP\_2.pf.bam -o IP\_2.qf.bam

samtools view -b -q 30 IP\_3.pf.bam -o IP\_3.qf.bam

samtools view -b -q 30 IP\_4.pf.bam -o IP\_4.qf.bam

## 3.8 Filtering PCR duplicates\_Index Aligned File

samtools index Input.qf.bam

samtools index IP\_1.qf.bam

samtools index IP\_2.qf.bam

samtools index IP\_3.qf.bam

samtools index IP\_4.qf.bam

## 3.9 Filtering PCR duplicates\_Filter

picard MarkDuplicates INPUT=Input.qf.bam OUTPUT=Input.df.bam METRICS\_FILE=Input.df.metrics.txt VALIDATION\_STRINGENCY=LENIENT REMOVE\_DUPLICATES=true

picard MarkDuplicates INPUT=IP\_1.qf.bam OUTPUT=IP\_1.df.bam METRICS\_FILE=IP\_1.df.metrics.txt VALIDATION\_STRINGENCY=LENIENT REMOVE\_DUPLICATES=true

picard MarkDuplicates INPUT=IP\_2.qf.bam OUTPUT=IP\_2.df.bam METRICS\_FILE=IP\_2.df.metrics.txt VALIDATION\_STRINGENCY=LENIENT REMOVE\_DUPLICATES=true

picard MarkDuplicates INPUT=IP\_3.qf.bam OUTPUT=IP\_3.df.bam METRICS\_FILE=IP\_3.df.metrics.txt VALIDATION\_STRINGENCY=LENIENT REMOVE\_DUPLICATES=true

picard MarkDuplicates INPUT=IP\_4.qf.bam OUTPUT=IP\_4.df.bam METRICS\_FILE=IP\_4.df.metrics.txt VALIDATION\_STRINGENCY=LENIENT REMOVE\_DUPLICATES=true

## 3.10 Read counts were counts per million (CPM) normalized and converted to bigWig format\_Index Aligned File

samtools index Input.df.bam

samtools index IP\_1.df.bam

samtools index IP\_2.df.bam

samtools index IP\_3.df.bam

samtools index IP\_4.df.bam

## 3.11 Read counts were CPM normalized and converted to bigWig format

bamCoverage -b Input.df.bam -o Input.df.bw -p 10 --normalizeUsing CPM --binSize 10 --extendReads 200

bamCoverage -b IP\_1.df.bam -o IP\_1.df.bw -p 10 --normalizeUsing CPM --binSize 10 --extendReads 200

bamCoverage -b IP\_2.df.bam -o IP\_2.df.bw -p 10 --normalizeUsing CPM --binSize 10 --extendReads 200

bamCoverage -b IP\_3.df.bam -o IP\_3.df.bw -p 10 --normalizeUsing CPM --binSize 10 --extendReads 200

bamCoverage -b IP\_4.df.bam -o IP\_4.df.bw -p 10 --normalizeUsing CPM --binSize 10 --extendReads 200

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

#### Part2. Peak calling, annotation, and visualization ####

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

# Step1. Peak calling using MACS2 v2.1.1

## 1.1 Peak calling for H3K4me3 and H3K27me3

macs2 callpeak -t ./IP\_1.df.bam -c ./Input.df.bam -f BAM -n IP1 -g 4.53e8 --keep-dup all

macs2 callpeak -t ./IP\_2.df.bam -c ./Input.df.bam -f BAM -n IP1 -g 4.53e8 --keep-dup all

macs2 callpeak -t ./IP\_3.df.bam -c ./Input.df.bam -f BAM -n IP3 --broad -g 4.53e8 --broad-cutoff 0.1 --bdg --keep-dup all

macs2 callpeak -t ./IP\_4.df.bam -c ./Input.df.bam -f BAM -n IP4 --broad -g 4.53e8 --broad-cutoff 0.1 --bdg --keep-dup all

## 1.2 Determining reproducible peaks between replicates using irreproducible discovery rate (IDR) for H3K4me3

macs2 callpeak -t ./IP\_1.zf.bam ./IP\_2.zf.bam -c ./Input.zf.bam -f BAM -n H3K4me3 -g 4.2e8 --keep-dup all -p 0.01

idr --samples IP1\_peaks.narrowPeak IP2\_peaks.narrowPeak --input-file-type narrowPeak --rank p.value --output-file H3K4me3\_1-idr --plot --log-output-file H3K4me3\_1.idr.log

cut -f 1-10 H3K4me3\_1-idr | sort -k1,1 -k2,2n -k3,3n >H3K4me3.idr.narrowPeak

## 1.3 Determining reproducible peaks between replicates using the ‘findOverlapsOfPeaks’ function in the R package ChIPpeakAnno for H3K27me3

library(ChIPpeakAnno)

H3K27me3\_1\_peaks <- read.delim("H3K27me3\_1\_peaks.txt")

H3K27me3\_2\_peaks <- read.delim("H3K27me3\_2\_peaks.txt")

H3K27me3\_1\_peaks <- toGRanges(H3K27me3\_1\_peaks, format="broadPeak")

H3K27me3\_2\_peaks <- toGRanges(H3K27me3\_2\_peaks, format="broadPeak")

ol <- findOverlapsOfPeaks(H3K27me3\_1\_peaks, H3K27me3\_2\_peaks, connectedPeaks= "merge")

ol <- addMetadata(ol, colNames="score", FUN=mean)

makeVennDiagram(ol, fill=c("#009E73", "#F0E442"), # circle fill color

 col=c("#D55E00", "#0072B2"), #circle border color

 cat.col=c("#D55E00", "#0072B2"),connectedPeaks= "merge")

write.csv(ol$peaklist[["H3K27me3\_1\_peaks///H3K27me3\_2\_peaks"]],"overlaps\_H3K27me3.broadPeak")

## 1.4 Determining bivalent domain

bedtools intersect -a ./H3K4me3.idr.narrowPeak -b ./overlaps\_H3K27me3.broadPeak|awk '{if(($3-$2+1)>=100) print}'|sort -k1,1 -k2,2n>./bivalent\_domain.txt

# Step2. Peak annotation using the R package ChIPseeker v1.26.2

library(ChIPseeker)

library(GenomicFeatures)

# build Txdb

gtfFile = "GCF\_000789215.1\_ASM78921v2\_genomic.gtf"

db = makeTxDbFromGFF(gtfFile, format = "gtf")

saveDb(db, file="bd.txdb.sqlite")

# read ChIP-seq peak bed file

peak\_H3K4me3 = readPeakFile("H3K4me3.idr.narrowPeak")

peak\_H3K27me3 = readPeakFile("overlaps\_H3K27me3.broadPeak")

# annotation

H3K4me3\_peakAnno = annotatePeak(peak\_H3K4me3,

 tssRegion = c(-3000, 3000),

 TxDb = db,

 verbose = F)

H3K4me3\_peakAnno.df = as.data.frame(H3K4me3\_peakAnno)

H3K27me3\_peakAnno = annotatePeak(peak\_H3K27me3,

 tssRegion = c(-3000, 3000),

 TxDb = db,

 verbose = F)

H3K27me3\_peakAnno.df = as.data.frame(H3K27me3\_peakAnno)

# Output results

write.table(H3K4me3\_peakAnno, file = "H3K4me3\_peakAnno.txt",sep = '\t', quote = FALSE, row.names = FALSE)

write.table(H3K27me3\_peakAnno, file = "H3K27me3\_peakAnno.txt",sep = '\t', quote = FALSE, row.names = FALSE)

#Plot the results

tiff("H3K4me3\_peakAnno\_3.tiff")

p1 = plotAnnoPie(H3K4me3\_peakAnno)

dev.off()

tiff("H3K27me3\_peakAnno\_3.tiff")

p2 = plotAnnoPie(H3K27me3\_peakAnno)

dev.off()

# Step3. Peak visualization using the R package ChIPseeker v1.26.2

library(ChIPseeker)

library(GenomicFeatures)

# build Txdb

gtfFile = "GCF\_000789215.1\_ASM78921v2\_genomic.gtf"

db = makeTxDbFromGFF(gtfFile, format = "gtf")

saveDb(db, file="bd.txdb.sqlite")

# read ChIP-seq peak bed file

peak\_H3K4me3 = readPeakFile("H3K4me3.idr.narrowPeak")

peak\_H3K27me3 = readPeakFile("overlaps\_H3K27me3.broadPeak")

head(peak\_H3K4me3, 4)

head(peak\_H3K27me3, 4)

# visualization for H3K4me3

promoter <- getPromoters(TxDb=db, upstream=3000, downstream=3000)

tagMatrix <- getTagMatrix(peak\_H3K4me3, windows=promoter)

plotAvgProf(tagMatrix, xlim=c(-3000, 3000), conf = 0.95, resample = 1000)

# visualization for H3K27me3

plotPeakProf2(peak = peak\_H3K27me3, upstream = rel(0.20), downstream = rel(0.20),

 conf = 0.95, by = "gene", type = "body", nbin = 800,

 TxDb = db, weightCol = "V5",ignore\_strand = F)

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

#### Part3. GO analysis for genes associated with Cluster 1-3 #####

#### and genes with H3K4me3-only, H3K27me3-only and bivalent domains at the TSSs,#####

#### and genes with H3K27me3 at the gene body regions. #####

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

## R code for Gene Ontology term enrichment analysis using the R package clusterProfiler v4.0.5

library(clusterProfiler)

library(tidyverse)

library(cowplot)

library(dplyr)

library(ggnewscale)

library(ggplot2)

library(stringr)

## get gene IDs from peak-tss association table

genes\_Cluster1 = read.csv("ass.Cluster1.csv")

genes\_Cluster1 = as.character(genes\_Cluster1$geneId)

genes\_Cluster1 = unique(genes\_Cluster1)

genes\_Cluster2 = read.csv("ass.Cluster2.csv")

genes\_Cluster2 = as.character(genes\_Cluster2$geneId)

genes\_Cluster2 = unique(genes\_Cluster2)

genes\_Cluster3 = read.csv("ass.Cluster3.csv")

genes\_Cluster3 = as.character(genes\_Cluster3$geneId)

genes\_Cluster3 = unique(genes\_Cluster3)

genes\_H3K4me3\_onlyTSS = read.csv("ass.H3K4me3\_onlyTSS.csv")

genes\_H3K4me3\_onlyTSS = as.character(genes\_H3K4me3\_onlyTSS$geneId)

genes\_H3K4me3\_onlyTSS = unique(genes\_H3K4me3\_onlyTSS)

genes\_H3K27me3\_onlyTSS = read.csv("ass.H3K27me3\_onlyTSS.csv")

genes\_H3K27me3\_onlyTSS = as.character(genes\_H3K27me3\_onlyTSS$geneId)

genes\_H3K27me3\_onlyTSS = unique(genes\_H3K27me3\_onlyTSS)

genes\_H3K27me3\_genebody = read.csv("ass.H3K27me3\_genebody.csv")

genes\_H3K27me3\_genebody = as.character(genes\_H3K27me3\_genebody$geneId)

genes\_H3K27me3\_genebody = unique(genes\_H3K27me3\_genebody)

genes\_bivalent = read.csv("ass.bivalent.csv")

genes\_bivalent = as.character(genes\_bivalent$geneId)

genes\_bivalent = unique(genes\_bivalent)

# read emapper

emapper = read.csv("GO\_all.csv")

GO = select(emapper, GID, GO) %>%

 separate\_rows(GO, sep = ',') %>%

 filter(str\_detect(GO, 'GO'))

# read term2name

term2name = read.csv("go\_term2name.csv")

# GO enrichment

term2gene\_GO = data.frame(TERM=GO$GO, GENE=GO$GID)

## genes\_Cluster1

enGO\_genes\_Cluster1 = enricher(genes\_Cluster1,

 TERM2GENE = term2gene\_GO,

 TERM2NAME = term2name,

 qvalueCutoff = 0.05,

 pAdjustMethod = "BH")

GO\_genes\_Cluster1\_summary = as.data.frame(enGO\_genes\_Cluster1)

write.csv(GO\_genes\_Cluster1\_summary,file = "GO\_genes\_Cluster1.csv")

barplot(enGO\_genes\_Cluster1, showCategory = 20, font.size = 20, x = "GeneRatio")

## genes\_Cluster2

enGO\_genes\_Cluster2 = enricher(genes\_Cluster2,

 TERM2GENE = term2gene\_GO,

 TERM2NAME = term2name,

 qvalueCutoff = 0.05,

 pAdjustMethod = "BH")

GO\_genes\_Cluster2\_summary = as.data.frame(enGO\_genes\_Cluster2)

write.csv(GO\_genes\_Cluster2\_summary,file = "GO\_genes\_Cluster2.csv")

barplot(enGO\_genes\_Cluster2, showCategory = 20, font.size = 20, x = "GeneRatio")

## genes\_Cluster3

enGO\_genes\_Cluster3 = enricher(genes\_Cluster3,

 TERM2GENE = term2gene\_GO,

 TERM2NAME = term2name,

 qvalueCutoff = 0.05,

 pAdjustMethod = "BH")

GO\_genes\_Cluster3\_summary = as.data.frame(enGO\_genes\_Cluster3)

write.csv(GO\_genes\_Cluster3\_summary,file = "GO\_genes\_Cluster3.csv")

barplot(enGO\_genes\_Cluster3, showCategory = 20, font.size = 20, x = "GeneRatio")

## genes\_H3K4me3\_onlyTSS

enGO\_genes\_H3K4me3\_onlyTSS = enricher(genes\_H3K4me3\_onlyTSS,

 TERM2GENE = term2gene\_GO,

 TERM2NAME = term2name,

 qvalueCutoff = 0.05,

 pAdjustMethod = "BH")

GO\_genes\_H3K4me3\_onlyTSS\_summary = as.data.frame(enGO\_genes\_H3K4me3\_onlyTSS)

write.csv(GO\_genes\_H3K4me3\_onlyTSS\_summary,file = "GO\_genes\_H3K4me3\_onlyTSS.csv")

barplot(enGO\_genes\_H3K4me3\_onlyTSS, showCategory = 20, font.size = 20, x = "GeneRatio")

## genes\_H3K27me3\_onlyTSS

enGO\_genes\_H3K27me3\_onlyTSS = enricher(genes\_H3K27me3\_onlyTSS,

 TERM2GENE = term2gene\_GO,

 TERM2NAME = term2name,

 qvalueCutoff = 0.05,

 pAdjustMethod = "BH")

GO\_genes\_H3K27me3\_onlyTSS\_summary = as.data.frame(enGO\_genes\_H3K27me3\_onlyTSS)

write.csv(GO\_genes\_H3K27me3\_onlyTSS\_summary,file = "GO\_genes\_H3K27me3\_onlyTSS.csv")

barplot(enGO\_genes\_H3K27me3\_onlyTSS, showCategory = 20, font.size = 20, x = "GeneRatio")

## genes\_H3K27me3\_genebody

enGO\_genes\_H3K27me3\_genebody = enricher(genes\_H3K27me3\_genebody,

 TERM2GENE = term2gene\_GO,

 TERM2NAME = term2name,

 qvalueCutoff = 0.05,

 pAdjustMethod = "BH")

GO\_genes\_H3K27me3\_genebody\_summary = as.data.frame(enGO\_genes\_H3K27me3\_genebody)

write.csv(GO\_genes\_H3K27me3\_genebody\_summary,file = "GO\_genes\_H3K27me3\_genebody.csv")

barplot(enGO\_genes\_H3K27me3\_genebody, showCategory = 20, font.size = 20, x = "GeneRatio")

## genes\_bivalent

enGO\_genes\_bivalent = enricher(genes\_bivalent,

 TERM2GENE = term2gene\_GO,

 TERM2NAME = term2name,

 qvalueCutoff = 0.05,

 pAdjustMethod = "BH")

GO\_genes\_bivalent\_summary = as.data.frame(enGO\_genes\_bivalent)

write.csv(GO\_genes\_bivalent\_summary,file = "GO\_genes\_bivalent.csv")

barplot(enGO\_genes\_bivalent, showCategory = 20, font.size = 20, x = "GeneRatio")

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

#### Part4. Motif analysis #####

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

## H3K4me3

sort -k8,8nr ./H3K4me3.idr.narrowPeak | head -n 500 | awk -v OFS="\t" '{print $1,$2+$10,$2+$10+1}' >summit.bed

awk -v OFS="\t" '{print $1,$2-50,$2+50}' summit.bed | awk '$2>=0' >summit.l50.r50.bed

bedtools getfasta -fi ./ref/GCF\_000789215.1\_ASM78921v2\_genomic.fna -bed summit.l50.r50.bed >summit.l50.r50.fa

## download motif database

wget http://meme-suite.org/meme-software/Databases/motifs/motif\_databases.12.19.tgz

tar -xzf motif\_databases.12.19.tgz

meme-chip -o summit.l50.r50.meme-chip -db ./motif\_databases/JASPAR/JASPAR2018\_CORE\_insects\_non-redundant.meme -seed 10 -meme-minw 8 -meme-maxw 10 -ccut 0 summit.l50.r50.fa

tomtom -oc summit.l50.r50.meme-chip.motif0.tomtom -min-overlap 5 -bfile ./summit.l50.r50.meme-chip/background ./summit.l50.r50.meme-chip/combined.meme motif\_databases/FLY/fly\_factor\_survey.meme

bedtools getfasta -fi ./ref/GCF\_000789215.1\_ASM78921v2\_genomic.fna -bed ./H3K4me3.idr.narrowPeak >peak.fa

fimo --parse-genomic-coord --oc summit.l50.r50.meme-chip.motif0.fimo --bgfile summit.l50.r50.meme-chip/background ./summit.l50.r50.meme-chip/combined.meme peak.fa

## H3K27me3

sort -k8,8nr ./overlaps\_H3K27me3.broadPeak | head -n 500 | awk -v OFS="\t" '{print $1,$2+$10,$2+$10+1}' >summit.bed

awk -v OFS="\t" '{print $1,$2-50,$2+50}' summit.bed | awk '$2>=0' >summit.l50.r50.bed

bedtools getfasta -fi ./ref/GCF\_000789215.1\_ASM78921v2\_genomic.fna -bed summit.l50.r50.bed >summit.l50.r50.fa

## download motif database

wget http://meme-suite.org/meme-software/Databases/motifs/motif\_databases.12.19.tgz

tar -xzf motif\_databases.12.19.tgz

meme-chip -o summit.l50.r50.meme-chip -db ./motif\_databases/JASPAR/JASPAR2018\_CORE\_insects\_non-redundant.meme -seed 10 -ccut 0 summit.l50.r50.fa

tomtom -oc summit.l50.r50.meme-chip.motif0.tomtom -min-overlap 5 -bfile ./summit.l50.r50.meme-chip/background ./summit.l50.r50.meme-chip/combined.meme motif\_databases/FLY/fly\_factor\_survey.meme

bedtools getfasta -fi ./ref/GCF\_000789215.1\_ASM78921v2\_genomic.fna -bed ./overlaps\_H3K27me3.broadPeak >peak.fa

fimo --parse-genomic-coord --oc summit.l50.r50.meme-chip.motif0.fimo --bgfile summit.l50.r50.meme-chip/background ./summit.l50.r50.meme-chip/combined.meme peak.fa

## bivalent domains

sort -k8,8nr ./bivalent\_domain.txt | head -n 500 | awk -v OFS="\t" '{print $1,$2+$10,$2+$10+1}' >summit.bed

awk -v OFS="\t" '{print $1,$2-50,$2+50}' summit.bed | awk '$2>=0' >summit.l50.r50.bed

bedtools getfasta -fi ./ref/GCF\_000789215.1\_ASM78921v2\_genomic.fna -bed summit.l50.r50.bed >summit.l50.r50.fa

## download motif database

wget http://meme-suite.org/meme-software/Databases/motifs/motif\_databases.12.19.tgz

tar -xzf motif\_databases.12.19.tgz

meme-chip -o summit.l50.r50.meme-chip -db ./motif\_databases/JASPAR/JASPAR2018\_CORE\_insects\_non-redundant.meme -seed 10 -meme-minw 8 -meme-maxw 10 -ccut 0 summit.l50.r50.fa

tomtom -oc summit.l50.r50.meme-chip.motif0.tomtom -min-overlap 5 -bfile ./summit.l50.r50.meme-chip/background ./summit.l50.r50.meme-chip/combined.meme motif\_databases/FLY/fly\_factor\_survey.meme

bedtools getfasta -fi ./ref/GCF\_000789215.1\_ASM78921v2\_genomic.fna -bed ./overlaps\_H3K27me3.broadPeak >peak.fa

fimo --parse-genomic-coord --oc summit.l50.r50.meme-chip.motif0.fimo --bgfile summit.l50.r50.meme-chip/background ./summit.l50.r50.meme-chip/combined.meme peak.fa

################## RNA-seq analysis ##################

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

#### Part1. Trimming, filtering and mapping reads. ####

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

# Step1. Trimming the raw read files using Trim Galore! v0.6.6

trim\_galore --illumina -o ./cleandata --paired ./Bd\_1\_R1.fastq.gz ./Bd\_1\_R2.fastq.gz

trim\_galore --illumina -o ./cleandata --paired ./Bd\_2\_R1.fastq.gz ./Bd\_2\_R2.fastq.gz

trim\_galore --illumina -o ./cleandata --paired ./Bd\_3\_R1.fastq.gz ./Bd\_3\_R2.fastq.gz

# Step2. Mapping the reads files using HISAT2 v2.1.0

## 2.1 Building a HISAT2 index

hisat2-build -f ./ref/genome.fasta ./ref/genome 1>hisat2-build.log 2>&1

## 2.2 Mapping the reads files

hisat2 --new-summary -p 10 -k 1 --rna-strandness RF -x ./ref/genome -1 ./cleandata/Bd\_1\_R1\_val\_1.fq.gz -2 ./cleandata/Bd\_1\_R2\_val\_2.fq.gz -S Bd\_1.sam 1>Bd\_1.log 2>&1

hisat2 --new-summary -p 10 -k 1 --rna-strandness RF -x ./ref/genome -1 ./cleandata/Bd\_2\_R1\_val\_1.fq.gz -2 ./cleandata/Bd\_2\_R2\_val\_2.fq.gz -S Bd\_2.sam 1>Bd\_2.log 2>&1

hisat2 --new-summary -p 10 -k 1 --rna-strandness RF -x ./ref/genome -1 ./cleandata/Bd\_3\_R1\_val\_1.fq.gz -2 ./cleandata/Bd\_3\_R2\_val\_2.fq.gz -S Bd\_3.sam 1>Bd\_3.log 2>&1

# Step3. Sorting reads and building index using SAMtools v1.3.1

## 3.1 Sorting reads

samtools sort -o Bd\_1.bam Bd\_1.sam

samtools sort -o Bd\_2.bam Bd\_2.sam

samtools sort -o Bd\_3.bam Bd\_3.sam

## 3.2 Building index

samtools index Bd\_1.bam

samtools index Bd\_2.bam

samtools index Bd\_3.bam

# Step4. Filtering the reads files using Picard v2.25.1

## 4.1

samtools sort -n -o Bd\_1\_namesort.bam Bd\_1.bam

samtools sort -n -o Bd\_2\_namesort.bam Bd\_2.bam

samtools sort -n -o Bd\_3\_namesort.bam Bd\_3.bam

## 4.2

samtools fixmate Bd\_1\_namesort.bam Bd\_1\_fixmate.bam

samtools fixmate Bd\_2\_namesort.bam Bd\_2\_fixmate.bam

samtools fixmate Bd\_3\_namesort.bam Bd\_3\_fixmate.bam

## 4.3

samtools sort -o Bd\_1\_positionsort.bam Bd\_1\_fixmate.bam

samtools sort -o Bd\_2\_positionsort.bam Bd\_2\_fixmate.bam

samtools sort -o Bd\_3\_positionsort.bam Bd\_3\_fixmate.bam

## 4.4

samtools rmdup -S Bd\_1\_positionsort.bam Bd\_1\_markdup.bam

samtools rmdup -S Bd\_2\_positionsort.bam Bd\_2\_markdup.bam

samtools rmdup -S Bd\_3\_positionsort.bam Bd\_3\_markdup.bam

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

#### Part2. Quantifying the reads count. ####

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

htseq-count -f bam -r name -s reverse -a 10 -t exon -i gene\_id -m union ./Bd\_1\_markdup.bam ./ref/genome.gtf > Bd\_1\_htseq.txt

htseq-count -f bam -r name -s reverse -a 10 -t exon -i gene\_id -m union ./Bd\_2\_markdup.bam ./ref/genome.gtf > Bd\_2\_htseq.txt

htseq-count -f bam -r name -s reverse -a 10 -t exon -i gene\_id -m union ./Bd\_3\_markdup.bam ./ref/genome.gtf > Bd\_3\_htseq.txt