Motif enrichment

Vi Dang

2022-09-11

```
#Load libraries
library(tidyverse)
library(ggplot2)
library(gridExtra)
library(grid)
library(ggpubr)
```

```
#read data
# TC motif
TC_ESharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC E30S.tsv",sep = "\t",hea
TC_ESharp<-TC_ESharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])   #252

TC_EBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC E30B.tsv",sep = "\t",hea
TC_EBroad<-TC_EBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])   #421

TC_SSharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC S30S.tsv",sep = "\t",hea
TC_SSharp<-TC_SSharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #282

TC_SBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC S30B.tsv",sep = "\t",hea
TC_SBroad<-TC_SBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #432

TC_E37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC E37S.tsv",sep = "\t",h
TC_E37Sharp<-TC_E37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #247

TC_E37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC E37B.tsv",sep = "\t",h
TC_E37Broad<-TC_E37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
```

```r
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])   #346

TC_S37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC S37S.tsv",sep = "\t",he
TC_S37Sharp<-TC_S37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #221

TC_S37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TC motif/TC S37B.tsv",sep = "\t",he
TC_S37Broad<-TC_S37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #388

#GG motif
GG_ESharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG E30S.tsv",sep = "\t",hea
GG_ESharp<-GG_ESharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #23

GG_EBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG E30B.tsv",sep = "\t",hea
GG_EBroad<- GG_EBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #27

GG_SSharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG S30S.tsv",sep = "\t",hea
GG_SSharp<-GG_SSharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #21

GG_SBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG S30B.tsv",sep = "\t",hea
GG_SBroad<-GG_SBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #64

GG_E37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG E37S.tsv",sep = "\t",he
GG_E37Sharp<- GG_E37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #23

GG_E37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG E37B.tsv",sep = "\t",he
GG_E37Broad<-GG_E37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #30

GG_S37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG S37S.tsv",sep = "\t",he
GG_S37Sharp<-GG_S37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #13

GG_S37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GG motif/GG S37B.tsv",sep = "\t",he
GG_S37Broad<-GG_S37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GG")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #49

#TATA motif
TATA_ESharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA E30S.tsv",sep = "\
TATA_ESharp<- TATA_ESharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #46

TATA_EBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA E30B.tsv",sep = "\
TATA_EBroad<- TATA_EBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #58

TATA_SSharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA S30S.tsv",sep = "\
TATA_SSharp<- TATA_SSharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #45

TATA_SBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA S30B.tsv",sep = "\
TATA_SBroad<- TATA_SBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #77

TATA_E37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA E37S.tsv",sep =
TATA_E37Sharp<- TATA_E37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #43

TATA_E37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA E37B.tsv",sep =
TATA_E37Broad<- TATA_E37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #67

TATA_S37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA S37S.tsv",sep =
TATA_S37Sharp<- TATA_S37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #36
```

```
TATA_S37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TATA motif/TATA S37B.tsv",sep =
TATA_S37Broad<- TATA_S37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TATA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #91


#AC motif
AC_ESharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC E30S.tsv",sep = "\t",hea
AC_ESharp<- AC_ESharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])   #9

AC_EBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC E30B.tsv",sep = "\t",hea
AC_EBroad<- AC_EBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #61

AC_SSharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC S30S.tsv",sep = "\t",hea
AC_SSharp<- AC_SSharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #13

AC_SBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC S30B.tsv",sep = "\t",hea
AC_SBroad<- AC_SBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #78

AC_E37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC E37S.tsv",sep = "\t",h
AC_E37Sharp<- AC_E37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #19

AC_E37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC E37B.tsv",sep = "\t",h
AC_E37Broad<- AC_E37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #57

AC_S37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC S37S.tsv",sep = "\t",h
AC_S37Sharp<- AC_S37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #16

AC_S37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/AC motif/AC S37B.tsv",sep = "\t",h
AC_S37Broad<- AC_S37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="AC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #52


#TTAC motif
TTAC_ESharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC E30S.tsv",sep = "\
TTAC_ESharp<- TTAC_ESharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])   #11


TTAC_EBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC E30B.tsv",sep = "\
TTAC_EBroad<- TTAC_EBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #50


TTAC_SSharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC S30S.tsv",sep = "\
TTAC_SSharp<- TTAC_SSharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #15


TTAC_SBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC S30B.tsv",sep = "\
TTAC_SBroad<- TTAC_SBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #48


TTAC_E37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC E37S.tsv",sep =
TTAC_E37Sharp<- TTAC_E37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #15


TTAC_E37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC E37B.tsv",sep =
TTAC_E37Broad<- TTAC_E37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #37


TTAC_S37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC S37S.tsv",sep =
TTAC_S37Sharp<- TTAC_S37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #11
```

```r
TTAC_S37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/TTAC motif/TTAC S37B.tsv",sep =
TTAC_S37Broad<- TTAC_S37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="TTAC")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #47


#GA motif
GA_ESharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA E30S.tsv",sep = "\t",hea
GA_ESharp<- GA_ESharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])   #12


GA_EBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA E30B.tsv",sep = "\t",hea
GA_EBroad<- GA_EBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #39


GA_SSharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA S30S.tsv",sep = "\t",hea
GA_SSharp<- GA_SSharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2]) #10


GA_SBroad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA S30B.tsv",sep = "\t",hea
GA_SBroad<- GA_SBroad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #33


GA_E37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA E37S.tsv",sep = "\t",h
GA_E37Sharp<- GA_E37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #11


GA_E37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA E37B.tsv",sep = "\t",h
GA_E37Broad<- GA_E37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #27


GA_S37Sharp <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA S37S.tsv",sep = "\t",h
GA_S37Sharp<- GA_S37Sharp%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #6


GA_S37Broad <- read.table("D:/PhD/TSS cluster/H99/TRASS_d17/MEME all/GA motif/GA S37B.tsv",sep = "\t",he
GA_S37Broad<- GA_S37Broad%>%filter(p.value<0.0005)%>%
  mutate(position_to_TSS=start-100)%>%
  mutate(motif="GA")%>%
  mutate(Id=str_match(sequence_name,"(\\b.+)::")[,2])  #36
```

```
#Position relative to TSS
#position of motif in Expo Sharp cluster
ESharp_motif<-bind_rows(TC_ESharp,GG_ESharp,TATA_ESharp,AC_ESharp,TTAC_ESharp,GA_ESharp)

#remove gene as identify in Percentage small set All cluster
gene_reject_E30<-read.delim("D:/PhD/TSS cluster/H99/TRASS_d17/gene_reject_E30.txt")
ESharp_motif<-ESharp_motif%>%filter(!(Id%in%gene_reject_E30$gene_reject_E30))


neworder <- c("AC","TTAC","GG","GA","TATA","TC")
table(ESharp_motif$motif)
```

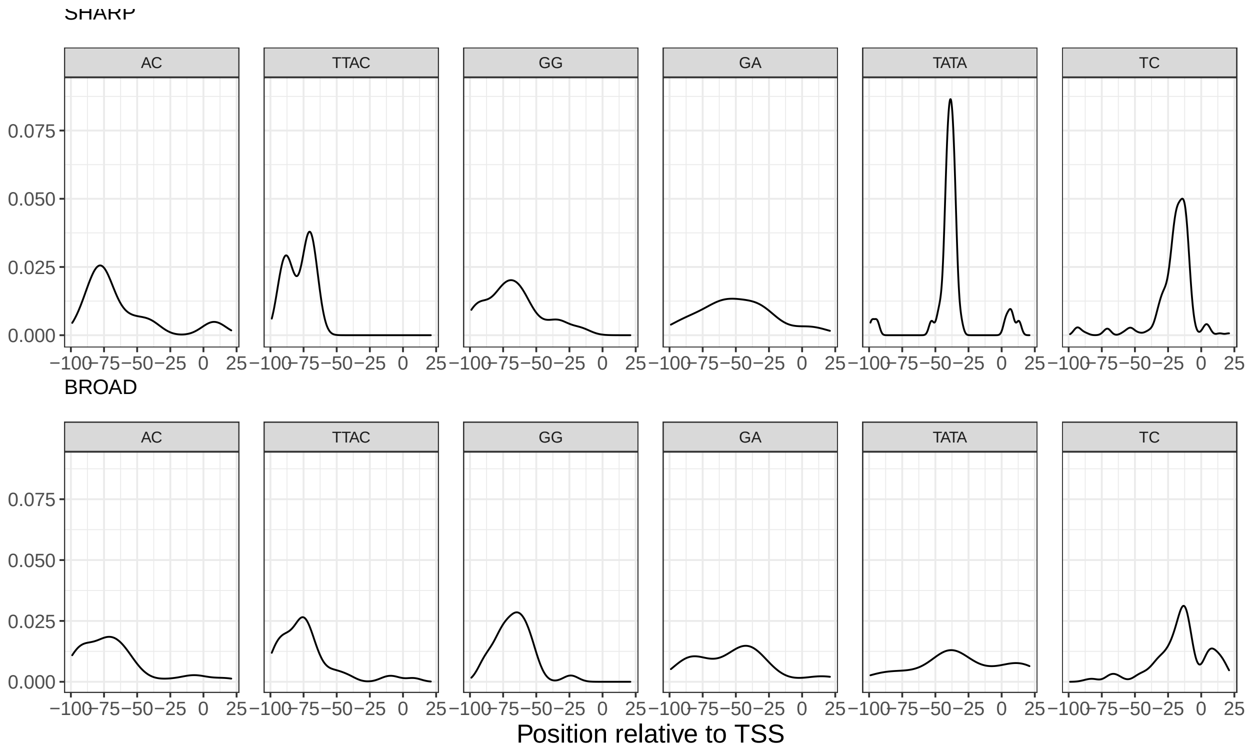
```
##
##   AC   GA   GG TATA   TC TTAC
##    9   10   21   41  241   11
```

```
ESharp_motif<-ESharp_motif%>%
  mutate(motif=factor(motif,levels=neworder))%>%
  arrange(motif)

Sharp_plot<-ESharp_motif%>%ggplot(aes(x=position_to_TSS))+
  geom_density()+
  facet_wrap(vars(motif),nrow=1)+
  scale_y_continuous(limits = c(0,0.09))+
  theme_bw()+
  theme(panel.spacing = unit(0.5, "cm"))+
  theme(plot.title = element_text(size=12,vjust=4),
        axis.title.x.bottom = element_blank(),
        axis.title.y.left = element_blank(),
        axis.text.x.bottom = element_text(size=11),
        axis.text.y.left = element_text(size=11))+
  labs(title="SHARP")



#Position of motif in Expo Broad cluster

EBroad_motif<-bind_rows(TC_EBroad,GG_EBroad,TATA_EBroad,AC_EBroad,TTAC_EBroad,GA_EBroad)

#remove gene as identify in Percentage small set All cluster
EBroad_motif<-EBroad_motif%>%filter(!(Id%in%gene_reject_E30$gene_reject_E30))

table(EBroad_motif$motif)
```

```
##
##    AC    GA    GG TATA   TC TTAC
##    60    39    27   55  392   47
```

```
neworder <- c("AC","TTAC","GG","GA","TATA","TC")

EBroad_motif<-EBroad_motif%>%
  mutate(motif=factor(motif,levels=neworder))%>%
  arrange(motif)


Broad_plot<-EBroad_motif%>%ggplot(aes(x=position_to_TSS))+
  geom_density()+
  facet_wrap(vars(motif),nrow=1)+
  scale_y_continuous(limits = c(0,0.09))+
  xlab("Position relative to TSS")+
  theme_bw()+
  theme(panel.spacing = unit(0.5, "cm"))+
  theme(plot.title = element_text(size=12,vjust=4),
        axis.title.x.bottom = element_text(size=15),
        axis.title.y.left = element_blank(),
        axis.text.x.bottom = element_text(size=11),
        axis.text.y.left = element_text(size=11))+
  labs(title="BROAD")
```

```
grid.arrange(Sharp_plot,Broad_plot, nrow=2)
```



```r
All_condition_clusters<-read.table("D:/PhD/TSS cluster/H99/TRASS_d17/All_condition_clusters.txt",sep="\
All_condition_clusters<-All_condition_clusters%>%arrange(condition,Cluster_Shape)

#Percentage of motif: need to remove left, right
Expo_30<- All_condition_clusters%>%
  filter(condition=="EXPO 30",Cluster_Shape!="NA")%>%
  mutate(TC_box = Id %in% c(TC_ESharp$Id,TC_EBroad$Id),
         GG_box = Id %in% c(GG_ESharp$Id,GG_EBroad$Id),
         TATA_box = Id %in% c(TATA_ESharp$Id,TATA_EBroad$Id),
         AC_box = Id %in% c(AC_ESharp$Id,AC_EBroad$Id),
         TTAC_box = Id %in% c(TTAC_EBroad$Id,TTAC_ESharp$Id),
         GA_box = Id %in% c(GA_ESharp$Id,GA_EBroad$Id))

Summary<-Expo_30%>%group_by(Cluster_Shape)%>%
  dplyr::summarise(nTC=sum(TC_box),perTC=100*nTC/n(),
                   nGG=sum(GG_box),perGG=100*nGG/n(),
                   nTATA=sum(TATA_box),perTATA=100*nTATA/n(),
                   nAC=sum(AC_box),perAC=100*nAC/n(),
                   nTTAC=sum(TTAC_box),perTTAC=100*nTTAC/n(),
                   nGA=sum(GA_box),perGA=100*nGA/n())
Summary
```

```
## # A tibble: 2 x 13
##   Cluste~1   nTC perTC   nGG perGG nTATA perTATA   nAC perAC nTTAC perTTAC   nGA
##   <chr>    <int> <dbl> <int> <dbl> <int>   <dbl> <int> <dbl> <int>   <dbl> <int>
## 1 Broad      154  40.7    23  6.08    44    11.6    53 14.0     48   12.7     32
## 2 Sharp       64  50.4    15 11.8     31    24.4     9  7.09    11    8.66    11
## # ... with 1 more variable: perGA <dbl>, and abbreviated variable name
## #   1: Cluster_Shape
```

```r
Expo_30%>%dplyr::summarise(nTATA_all=sum(TATA_box),
                           perTATA_all=100*nTATA_all/n())
```

```
##   nTATA_all perTATA_all
## 1        75    14.85149
```

```r
#Change the name of levels
Expo_30<-Expo_30%>%
  mutate(TC_box=if_else(TC_box==TRUE,"TC box","no TC box"),
         GG_box=if_else(GG_box==TRUE,"GG box",'no GG box'),
         TATA_box=if_else(TATA_box==TRUE,"TATA box","no TATA box"),
         AC_box=if_else(AC_box==TRUE,"AC box",'no AC box'),
         TTAC_box=if_else(TTAC_box==TRUE,"TTAC box","no TTAC box"),
         GA_box=if_else(GA_box==TRUE,"GA box",'no GA box')
  )
```

```r
#chi-squared test
#TATA box
TATA_table<-table(Expo_30$Cluster_Shape,Expo_30$TATA_box)
TATA_table
```

```
##
```

```
##         no TATA box TATA box
##   Broad         334       44
##   Sharp          96       31
```

```
chisq.test(TATA_table)
```

```
##
##  Pearson’s Chi-squared test with Yates’ continuity correction
##
## data:  TATA_table
## X-squared = 11.268, df = 1, p-value = 0.0007885
```

```
#GG box
GG_table<-table(Expo_30$Cluster_Shape,Expo_30$GG_box)
GG_table
```

```
##
##         GG box no GG box
##   Broad     23       355
##   Sharp     15       112
```

```
chisq.test(GG_table)
```

```
##
##  Pearson’s Chi-squared test with Yates’ continuity correction
##
## data:  GG_table
## X-squared = 3.6945, df = 1, p-value = 0.05459
```

```
#TC box
TC_table<-table(Expo_30$Cluster_Shape,Expo_30$TC_box)
TC_table
```

```
##
##         no TC box TC box
##   Broad       224    154
##   Sharp        63     64
```

```
chisq.test(TC_table)
```

```
##
##  Pearson’s Chi-squared test with Yates’ continuity correction
##
## data:  TC_table
## X-squared = 3.2278, df = 1, p-value = 0.0724
```

```
#TTACbox
TTAC_table<-table(Expo_30$Cluster_Shape,Expo_30$TTAC_box)
TTAC_table
```

```
##
##         no TTAC box TTAC box
##   Broad         330       48
##   Sharp         116       11
```

```
chisq.test(TTAC_table)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  TTAC_table
## X-squared = 1.1357, df = 1, p-value = 0.2866
```

```
#AC box
AC_table<-table(Expo_30$Cluster_Shape,Expo_30$AC_box)
AC_table
```

```
##
##         AC box no AC box
##   Broad     53       325
##   Sharp      9       118
```

```
chisq.test(AC_table)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  AC_table
## X-squared = 3.6251, df = 1, p-value = 0.05692
```

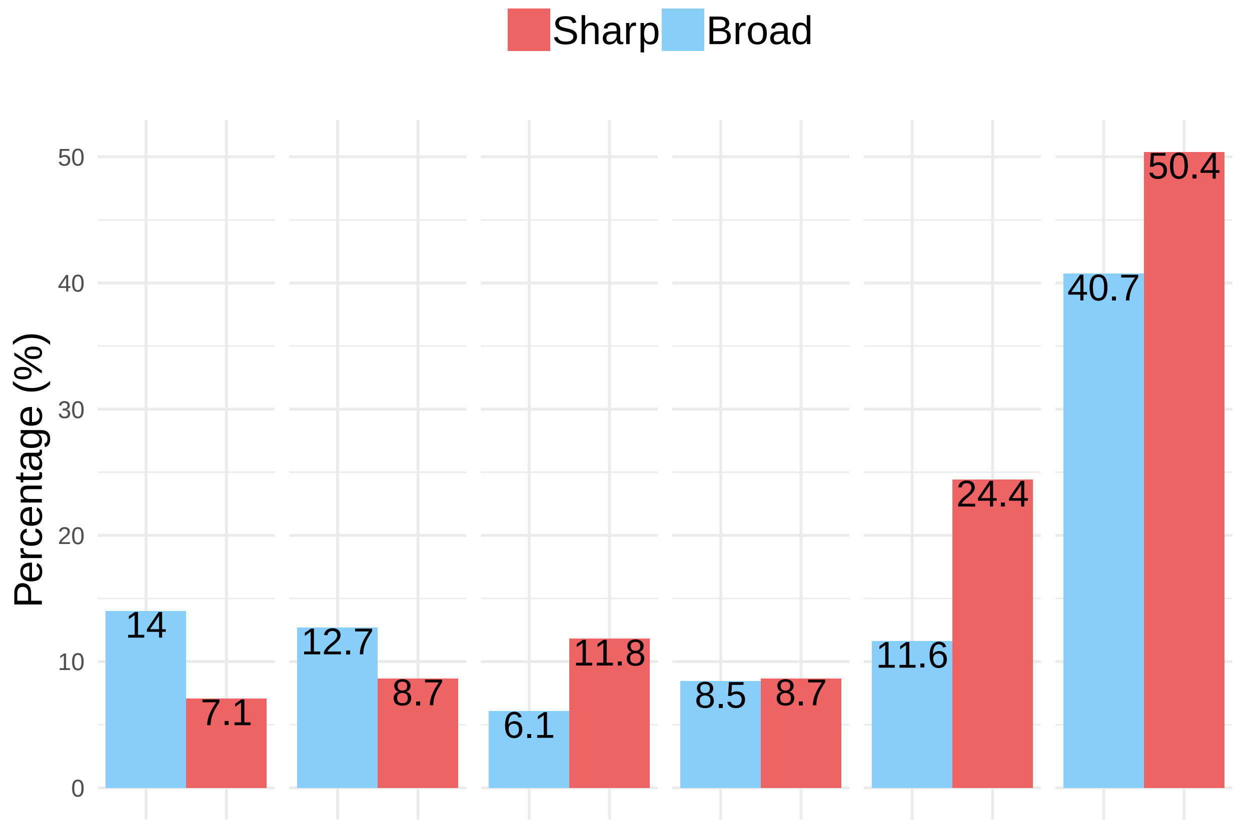
```
#GA box
GA_table<-table(Expo_30$Cluster_Shape,Expo_30$GA_box)
GA_table
```

```
##
##         GA box no GA box
##   Broad     32       346
##   Sharp     11       116
```

```
chisq.test(GA_table)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  GA_table
## X-squared = 2.2311e-29, df = 1, p-value = 1
```

```
Summary_Expo30_long<-Summary%>%
  pivot_longer(cols=c("perTC","perGG","perTATA","perAC",'perTTAC',"perGA"),
               names_to="Motif",
               values_to="Percentage")

#control order of facet_wrap
neworder<-c("perAC","perTTAC","perGG","perGA","perTATA","perTC")
Summary_Expo30_long<- Summary_Expo30_long%>%
  mutate(Motif=factor(Motif,levels=neworder))%>%
  arrange(Motif)

Summary_Expo30_long%>%ggplot(aes(x=Cluster_Shape,y=Percentage,fill=Cluster_Shape))+
  geom_col(width = 1)+
  scale_fill_manual(
    values =c("indianred2","lightskyblue"),
    breaks = c("Sharp","Broad"),
    labels = c("Sharp", "Broad ")
  )+
  geom_text(aes(label = round(Percentage,1)), size = 5, hjust = 0.5, vjust = 1, position =     "stack")
  theme_minimal()+
  ylab("Percentage (%)")+
  theme(axis.title.x = element_blank(),
        axis.text.x.bottom = element_blank(),
        axis.title.y.left = element_text(size=15),
        legend.position = "top",
        legend.title = element_text(size=0),
        legend.text = element_text(size=15))+
  facet_wrap(vars(Motif),#labeller = labeller(Motif = c("perTC"="TC box",
                             #                    "perGG" = "GG box",
                                #                 "perTATA" = "TATA box",
                                   #              "perAC" = "AC box",
                                      #           "perTTAC"="TTAC box",
                                         #        "perGA"="GA box")),
  nrow=1) +
  theme(strip.text.x = element_text(size=0))
```



```r
#Relationship between motif and gene expression
#Load gene expression data
txt_files = list.files("D:/PhD/TSS cluster/H99/regulated genes/",pattern = "\\.txt");
txt_files
```

```
##  [1] "FlucovsWT.complete.txt"  "FlucovsWT.down.txt"
##  [3] "FlucovsWT.up.txt"        "SDSvsWT.complete.txt"
##  [5] "SDSvsWT.down.txt"        "SDSvsWT.up.txt"
##  [7] "WT37vsWT.complete.txt"   "WT37vsWT.down.txt"
##  [9] "WT37vsWT.up.txt"         "WTST30vsWT.complete.txt"
## [11] "WTST30vsWT.down.txt"     "WTST30vsWT.up.txt"
```

```r
setwd("D:/PhD/TSS cluster/H99/regulated genes/")
# read multiple txt files at the same time
List <- lapply(txt_files,read.table,sep="\t",header=T,fill=TRUE)

#change name of files
newnames <- gsub('\\.', '_', txt_files)
newnames1 <- gsub('\\_txt', '', newnames)
newnames1
```

```
##  [1] "FlucovsWT_complete"  "FlucovsWT_down"      "FlucovsWT_up"
##  [4] "SDSvsWT_complete"    "SDSvsWT_down"        "SDSvsWT_up"
##  [7] "WT37vsWT_complete"   "WT37vsWT_down"       "WT37vsWT_up"
## [10] "WTST30vsWT_complete" "WTST30vsWT_down"     "WTST30vsWT_up"
```

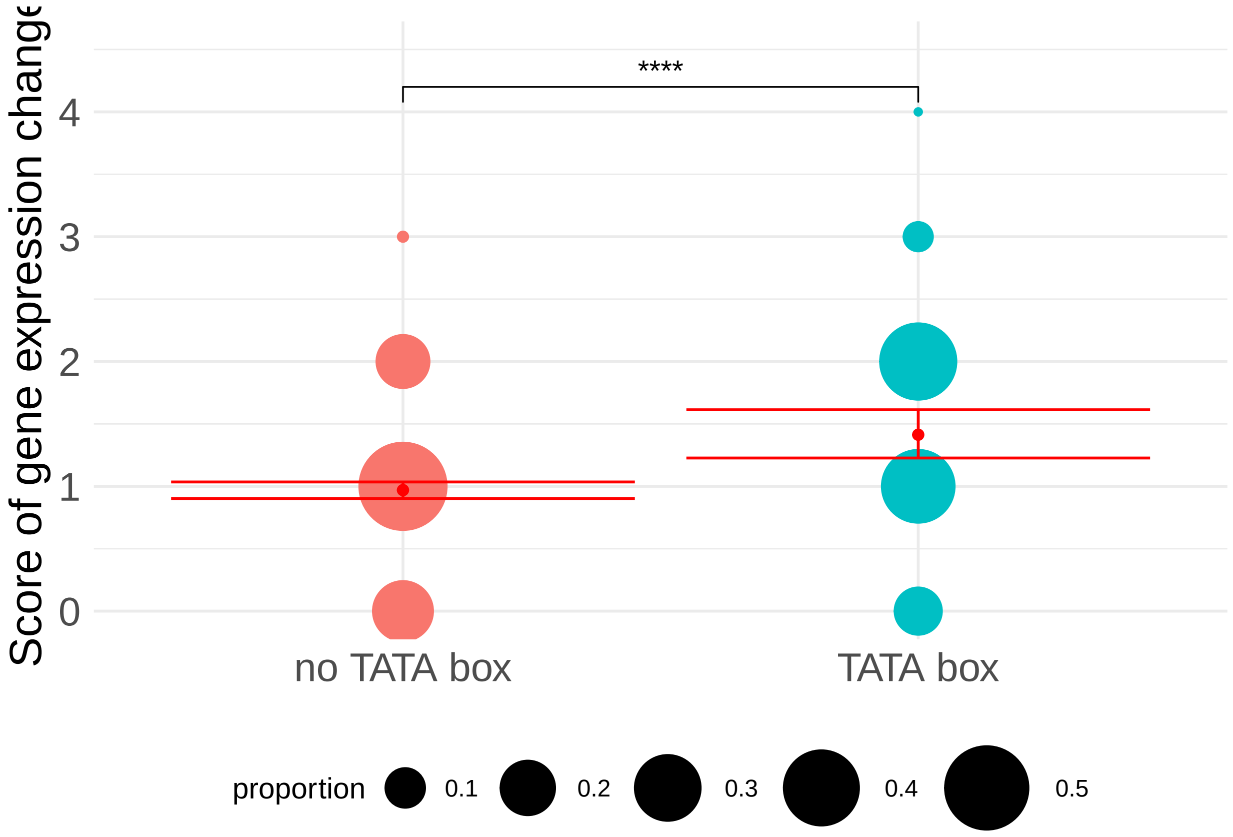
```
#Assign the names to List
names(List) <- newnames1


fluconazol<-rbind(List$FlucovsWT_down,List$FlucovsWT_up)
sds <-rbind(List$SDSvsWT_down,List$SDSvsWT_up)
stat<- rbind(List$WTST30vsWT_down,List$WTST30vsWT_up)
T37 <- rbind(List$WT37vsWT_down,List$WT37vsWT_up)

#GE change or not
Expo_30<-Expo_30%>%
  mutate(Fluconazol= Id %in% fluconazol$Id,
         SDS= Id %in% sds$Id,
         STAT=Id %in% stat$Id,
         t37= Id %in% T37$Id)

# Assign Score of gene expression change = Sum of change
Expo_30<-Expo_30%>%
  mutate(Sum_Change = Fluconazol+SDS+STAT+t37)
Expo_30%>%group_by(TATA_box)%>%
  dplyr::summarise(mean_GE_change = mean(Sum_Change))
```

```
## # A tibble: 2 x 2
##   TATA_box    mean_GE_change
##   <chr>                <dbl>
## 1 no TATA box          0.970
## 2 TATA box             1.41
```

```
head(Expo_30)
```

```
##     Chr start_max end_max         Id score strand condition Cluster_Shape
## 1     1     53175   53305 CNAG_00016     0      +   EXPO 30         Broad
## 2     1     73431   73561 CNAG_00024     0      +   EXPO 30         Broad
## 5     1    177227  177357 CNAG_00065     0      +   EXPO 30         Broad
## 7     1    255924  256054 CNAG_00092     0      +   EXPO 30         Broad
## 353   1    259681  259811 CNAG_00093     0      -   EXPO 30         Broad
## 8     1    301034  301164 CNAG_00109     0      +   EXPO 30         Broad
##        TC_box    GG_box    TATA_box    AC_box    TTAC_box    GA_box Fluconazol
## 1   no TC box no GG box no TATA box no AC box    TTAC box no GA box      FALSE
## 2      TC box no GG box no TATA box no AC box no TTAC box no GA box       TRUE
## 5      TC box no GG box no TATA box no AC box no TTAC box no GA box      FALSE
## 7      TC box no GG box no TATA box    AC box no TTAC box no GA box      FALSE
## 353 no TC box no GG box no TATA box no AC box no TTAC box no GA box      FALSE
## 8   no TC box no GG box no TATA box no AC box no TTAC box no GA box      FALSE
##       SDS  STAT   t37 Sum_Change
## 1    TRUE  TRUE FALSE          2
## 2   FALSE  TRUE FALSE          2
## 5   FALSE  TRUE FALSE          1
## 7   FALSE FALSE FALSE          0
## 353  TRUE FALSE FALSE          1
## 8   FALSE FALSE FALSE          0
```

```
#proportion plot (geom_count)
Expo_30%>%
  ggplot(aes(x=TATA_box,y=Sum_Change,colour=TATA_box))+
  geom_count(aes(size=..prop..))+
  stat_summary(fun.data = mean_cl_boot, geom = "errorbar", colour = "red") +
  stat_summary(fun = mean, geom = "point", colour = "red")+
  stat_compare_means(method = "wilcox.test",
                     method.args = list(var.equal = TRUE),
                     comparisons=list(c("no TATA box","TATA box")),label="p.signif")+
  ylab("Score of gene expression change")+
  theme_minimal()+
  scale_colour_discrete(guide = "none")+    #mute the color=TATA box legend
  scale_size_continuous(name="proportion",range = c(1,15), breaks=seq(0,0.5,by=0.1))+
  ylim(0,4.5)+
  theme(axis.title.x = element_blank(),
        axis.text.x.bottom = element_text(size=15),
        axis.text.y.left = element_text(size=15),
        axis.title.y.left = element_text(size=17,hjust=0.5),
        plot.title = element_blank(),
        legend.position = "bottom")
```



```
#Load data from DeSEQ2
complete_txt_files = list.files("D:/PhD/TSS cluster/H99/regulated genes/",pattern = "\\complete.txt")
setwd("D:/PhD/TSS cluster/H99/regulated genes/")
List <- lapply(complete_txt_files,read.table,sep="\t",header=T,fill=TRUE)
```

```r
complete_txt_files
```

```
## [1] "FlucovsWT.complete.txt"  "SDSvsWT.complete.txt"
## [3] "WT37vsWT.complete.txt"   "WTST30vsWT.complete.txt"
```

```r
newnames <- gsub('\\.', '_', complete_txt_files)
newnames1 <- gsub('\\_txt', '', newnames)
names(List)<-newnames1
FlucovsWT_complete<-List$FlucovsWT_complete
SDSvsWT_complete<-List$SDSvsWT_complete
STATvsWT_complete<-List$WTST30vsWT_complete
t37vsWT_complete<-List$WT37vsWT_complete

WT_sum_flu=sum(FlucovsWT_complete$WT)
FlucovsWT_complete<-FlucovsWT_complete%>%
  mutate(abs_scale_LFC=abs(scale(log2FoldChange)))%>%
  mutate(normalized_WT=1000000*WT/WT_sum_flu)


WT_sum_SDS=sum(SDSvsWT_complete$WT)
SDSvsWT_complete<-SDSvsWT_complete%>%
  mutate(abs_scale_LFC=abs(scale(log2FoldChange)))%>%
  mutate(normalized_WT=1000000*WT/WT_sum_SDS)

WT_sum_STAT=sum(STATvsWT_complete$WT)
STATvsWT_complete<-STATvsWT_complete%>%
  mutate(abs_scale_LFC=abs(scale(log2FoldChange)))%>%
  mutate(normalized_WT=1000000*WT/WT_sum_STAT)

WT_sum_t37=sum(t37vsWT_complete$WT)
t37vsWT_complete<-t37vsWT_complete%>%
  mutate(abs_scale_LFC=abs(scale(log2FoldChange)))%>%
  mutate(normalized_WT=1000000*WT/WT_sum_t37)
```

```r
#Expression mean and change level
Fluco<-FlucovsWT_complete%>%select(c(Id,normalized_WT,Fluco,abs_scale_LFC))%>%
  rename(WTa=normalized_WT)%>%
  rename(Fluconazol_abs_scale_LFC=abs_scale_LFC)
SDS<-SDSvsWT_complete%>%select(c(Id,normalized_WT,SDS,abs_scale_LFC))%>%
  rename(WTb=normalized_WT)%>%
  rename(SDS_abs_scale_LFC=abs_scale_LFC)
STAT<-STATvsWT_complete%>%select(c(Id,normalized_WT,WTST30,abs_scale_LFC))%>%
  rename(WTc=normalized_WT)%>%
  rename(STAT_abs_scale_LFC=abs_scale_LFC)
temp<-t37vsWT_complete%>%select(c(Id,normalized_WT,WT37,abs_scale_LFC))%>%
  rename(WTd=normalized_WT)%>%
  rename(temp_abs_scale_LFC=abs_scale_LFC)

Mean_vs_change<-merge(Fluco,SDS,by="Id")%>%merge(STAT,by="Id")%>%merge(temp,by="Id")%>%filter(Id!="__no

Mean_vs_change<-Mean_vs_change%>%mutate(log_expression_mean=log((WTa+WTb+WTc+WTd+Fluco+SDS+WTST30+WT37)
  mutate(avg_abs_LFC=(Fluconazol_abs_scale_LFC+SDS_abs_scale_LFC+STAT_abs_scale_LFC+temp_abs_scale_LFC)
  mutate(mean_log_expression_WT= (log2(WTa)+log2(WTb)+log2(WTc)+log2(WTc))/4)
```

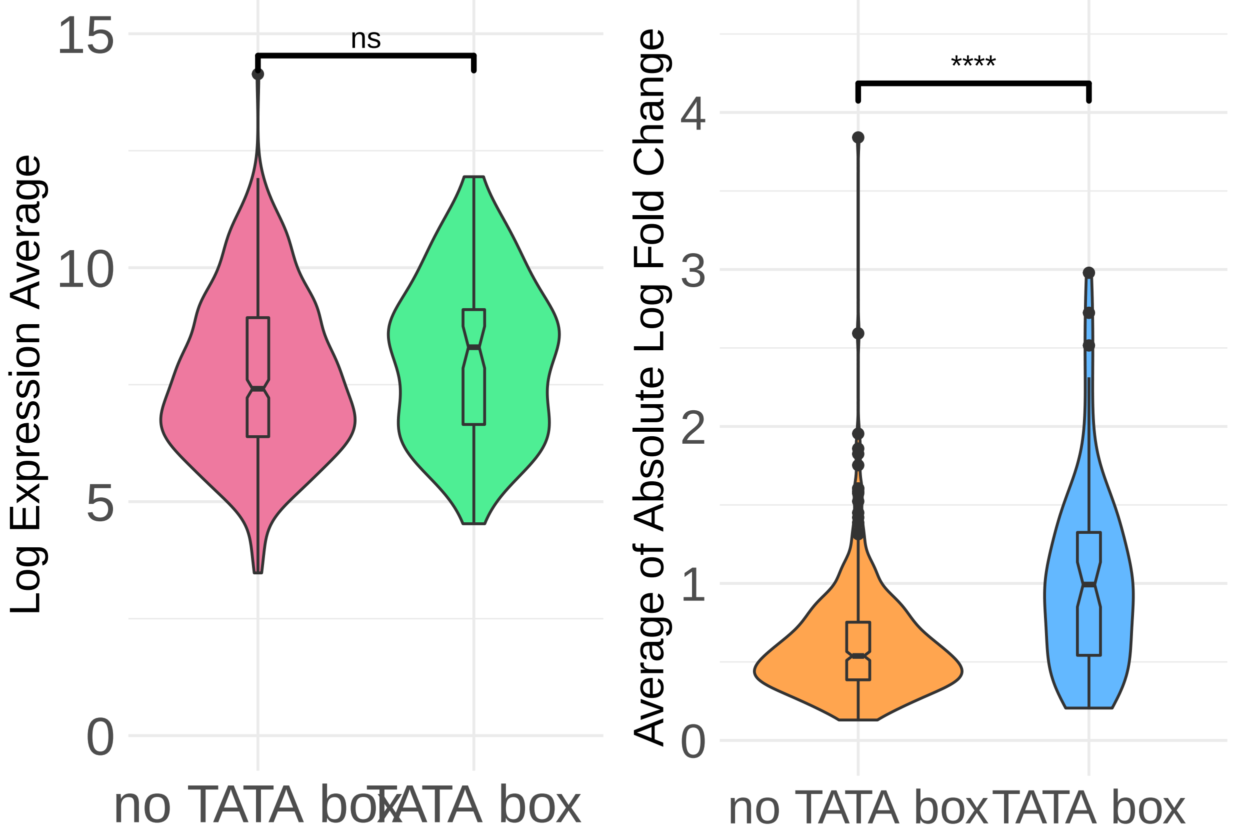
```r
#Take gene from Expo_30
Mean_vs_change <- Mean_vs_change %>%
  right_join(Expo_30[,c("Id","Cluster_Shape","TATA_box","TC_box","GG_box","TTAC_box","AC_box","GA_box")]
             by="Id")

Mean_vs_change<-Mean_vs_change%>%
  mutate(mean_log_expression_WT=round(mean_log_expression_WT,3))%>%
  mutate(avg_abs_LFC=round(avg_abs_LFC,3))
head(Mean_vs_change)
```

```
##           Id       WTa Fluco Fluconazol_abs_scale_LFC        WTb   SDS
## 1 CNAG_00016  111.1171  2927                0.4281820  111.10492  2609
## 2 CNAG_00024 1019.7123 15778                1.5816172 1019.62845 25317
## 3 CNAG_00057  684.5586 24066                1.5222895  684.85818 11138
## 4 CNAG_00060   38.3943  1037                0.5252611   38.39808  1454
## 5 CNAG_00065  412.0681  7659                0.8889449  412.05160 10535
## 6 CNAG_00092  169.4630  4373                0.3520930  169.42980  6036
##   SDS_abs_scale_LFC        WTc WTST30 STAT_abs_scale_LFC       WTd  WT37
## 1         0.6466970  111.14964   1025          0.9226299 110.30411  2565
## 2         0.4698233 1019.93146   8055          1.1637873 996.45989 31286
## 3         1.8184855  684.44777  13796          0.2982973 715.13170 24029
## 4         0.8744170   38.34281    698          0.1371652  40.55686  1253
## 5         0.3769646  411.97855   4356          0.7085080 415.78191 10623
## 6         0.6798559  169.39510   2205          0.3851624 171.97952  4686
##   temp_abs_scale_LFC log_expression_mean avg_abs_LFC mean_log_expression_WT
## 1          0.5583391            7.086913       0.639                  6.796
## 2          0.2029380            9.264967       0.855                  9.994
## 3          0.3752363            9.156386       1.004                  9.419
## 4          0.1625007            6.353868       0.425                  5.262
## 5          0.3192313            8.378646       0.573                  8.687
## 6          0.1574818            7.717589       0.394                  7.404
##   Cluster_Shape    TATA_box    TC_box    GG_box    TTAC_box    AC_box    GA_box
## 1         Broad no TATA box no TC box no GG box    TTAC box no AC box no GA box
## 2         Broad no TATA box    TC box no GG box no TTAC box no AC box no GA box
## 3         Sharp no TATA box no TC box no GG box no TTAC box no AC box no GA box
## 4         Sharp no TATA box no TC box no GG box no TTAC box no AC box no GA box
## 5         Broad no TATA box    TC box no GG box no TTAC box no AC box no GA box
## 6         Broad no TATA box    TC box no GG box no TTAC box    AC box no GA box
```

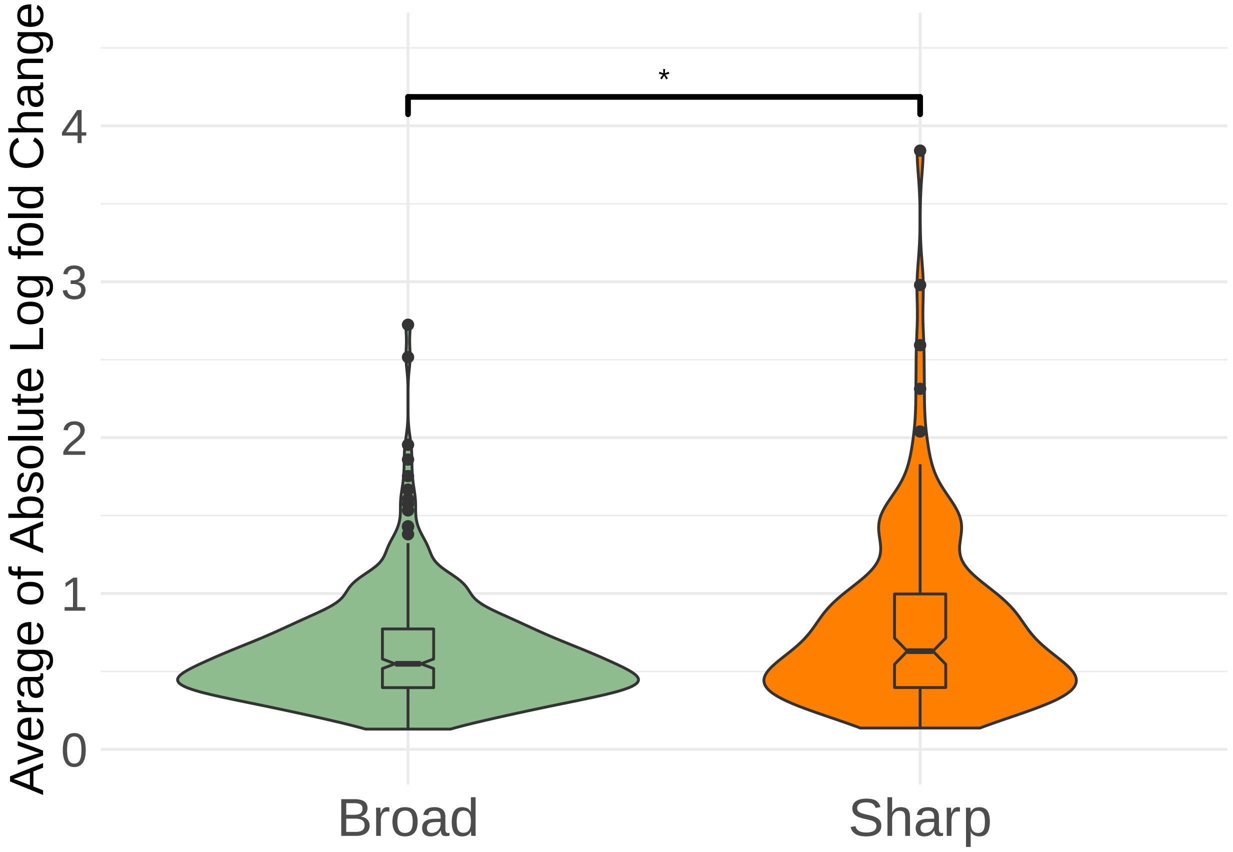
```r
#Relationship between TATA box and gene expression change
#Box plot for mean expression
p1<-Mean_vs_change%>%
  ggplot(aes(x=TATA_box,y=mean_log_expression_WT,fill=TATA_box))+
  scale_fill_manual(values=c("no TATA box"="palevioletred2","TATA box"="seagreen2"),)+
  geom_violin()+
  geom_boxplot(width=0.1,notch=T)+
  stat_compare_means(comparisons=list(c("TATA box","no TATA box")),
                     label="p.signif",label.y = 14, bracket.size = 1)+  #wilcoxon test
  ylab("Log Expression Average")+
  theme_minimal()+
  ylim(0,15)+
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(size=20),
        axis.title.y = element_text(size=16),
        axis.text.y = element_text(size=20),
        legend.position = "none")
```

```
#Box_plot for mean of absolute LFC
p2<-Mean_vs_change%>%
  ggplot(aes(x=TATA_box,y=avg_abs_LFC,fill=TATA_box))+
  geom_violin()+
  geom_boxplot(width=0.1,notch=T)+
  scale_fill_manual(values=c("no TATA box"="tan1","TATA box"="steelblue1"),)+
  stat_compare_means(comparisons=list(c("TATA box","no TATA box")),   #wilcoxon test
                     label="p.signif",label.y = 4, bracket.size = 1
  )+
  ylab("Average of Absolute Log Fold Change")+
  ylim(c(0,4.5))+
  theme_minimal()+
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(size=18),
        axis.title.y = element_text(size=16),
        axis.text.y = element_text(size=18),
        legend.position = "none")
```

```
ggarrange(p1, p2,nrow=1,ncol=2)
```

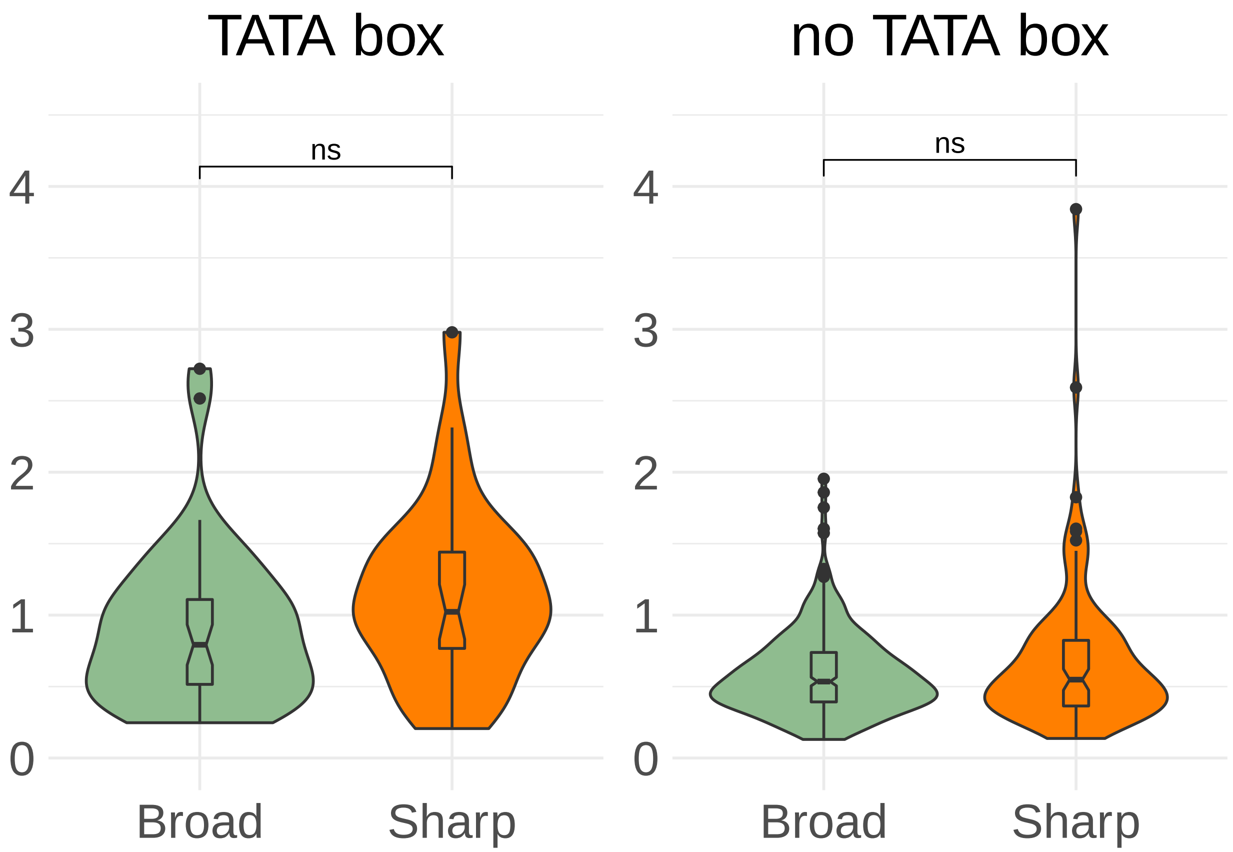


```
#CLUSTER SHAPE
#Both TATA box and no TATA box
pBoth_LFC<-Mean_vs_change%>%
  ggplot(aes(x=Cluster_Shape,y=avg_abs_LFC,fill=Cluster_Shape))+
  geom_violin()+
  geom_boxplot(width=0.1,notch=T)+
  scale_fill_manual(
    values =c("darkorange1","darkseagreen"),
    breaks = c("Sharp","Broad"),
    labels = c("Sharp Clusters", "Broad Cluster")
  )+
  stat_compare_means(comparisons=list(c("Sharp","Broad")),label="p.signif",label.y = 4, bracket.size =
  theme_minimal()+
  theme(legend.position = "none")+
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(size=20),
        axis.title.y=element_text(size=18),
        axis.text.y.left = element_text(size=18))+
  ylim(c(0,4.5))+
  ylab("Average of Absolute Log fold Change")+
  theme(plot.title = element_text(hjust = 0.5,size=22))
pBoth_LFC
```

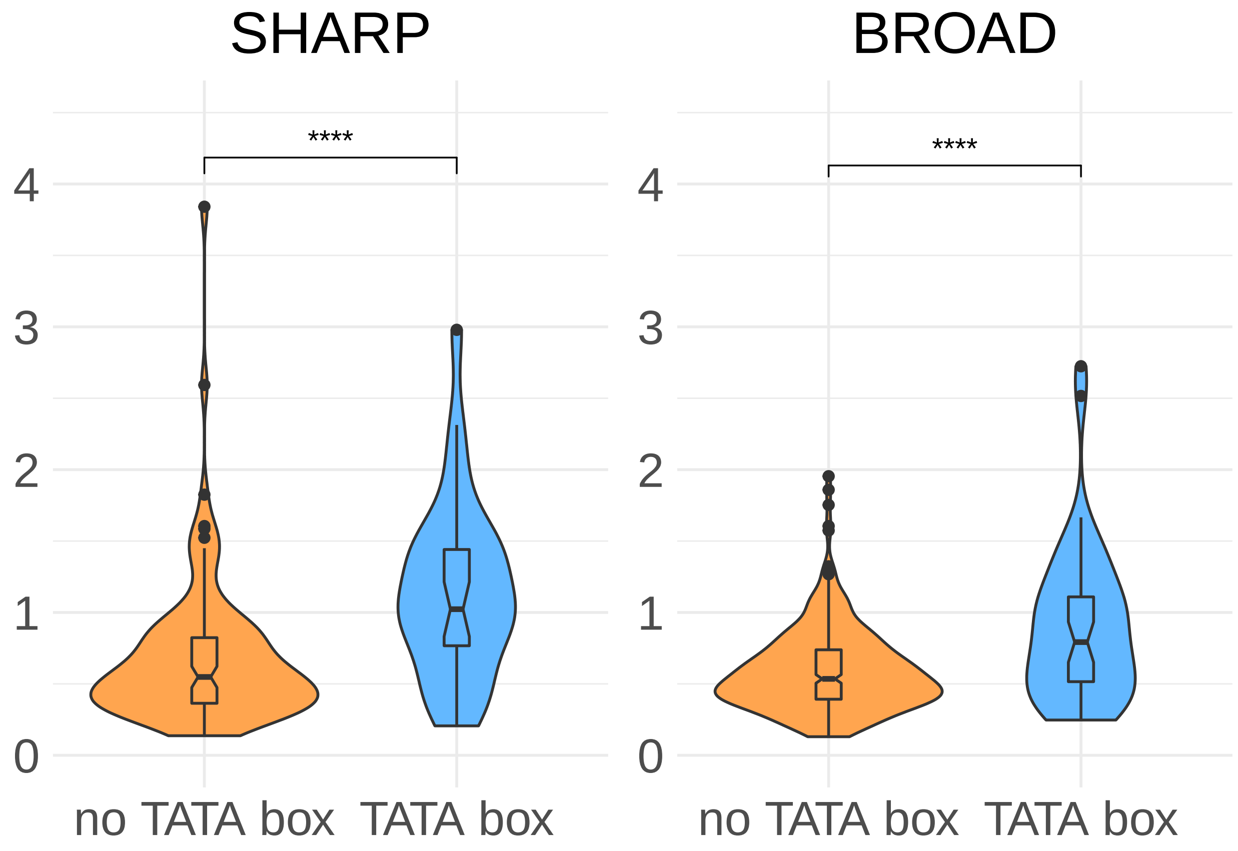


```
#Stratification for TATA box-containing genes and non TATA box genes
#TATA box
```

```
p_TATA_LFC<-Mean_vs_change%>%
  filter(TATA_box=="TATA box")%>%
  ggplot(aes(x=Cluster_Shape,y=avg_abs_LFC,fill=Cluster_Shape))+
  geom_violin()+
  geom_boxplot(width=0.1,notch=T)+
  scale_fill_manual(
    values =c("darkorange1","darkseagreen"),
    breaks = c("Sharp","Broad"),
    labels = c("Sharp Clusters", "Broad Cluster")
  )+
  stat_compare_means(comparisons=list(c("Sharp","Broad")),label="p.signif",label.y = 4,method = "t.test
  theme_minimal()+
  theme(legend.position = "none")+
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(size=18),
        axis.title.y=element_blank(),
        axis.text.y.left = element_text(size=18))+
  ylim(c(0,4.5))+
  labs(title = "TATA box")+
  theme(plot.title = element_text(hjust = 0.5,size=22))



#non TATA
p_nonTATA_LFC<-Mean_vs_change%>%
  filter(TATA_box=="no TATA box")%>%
  ggplot(aes(x=Cluster_Shape,y=avg_abs_LFC,fill=Cluster_Shape))+
  geom_violin()+
  geom_boxplot(width=0.1,notch=T)+
  scale_fill_manual(
    values =c("darkorange1","darkseagreen"),
    breaks = c("Sharp","Broad"),
    labels = c("Sharp Clusters", "Broad Cluster")
  )+
  stat_compare_means(comparisons=list(c("Sharp","Broad")),label="p.signif",label.y = 4,method = "t.test
  theme_minimal()+
  theme(legend.position = "none")+
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(size=18),
        axis.title.y=element_blank(),
        axis.text.y.left = element_text(size=18))+
  ylim(c(0,4.5))+
  labs(title = "no TATA box")+
  theme(plot.title = element_text(hjust = 0.5,size=22))


ggarrange(p_TATA_LFC,p_nonTATA_LFC,nrow = 1,ncol = 2)
```



```
#TATA box: do stratification with Broad and Sharp cluster
#Only Sharp
p_sharp_LFC<-Mean_vs_change%>%
  filter(Cluster_Shape=="Sharp")%>%
  ggplot(aes(x=TATA_box,y=avg_abs_LFC,fill=TATA_box))+
  geom_violin()+
  geom_boxplot(width=0.1,notch = T)+
  scale_fill_manual(values=c("no TATA box"="tan1","TATA box"="steelblue1"))+
  stat_compare_means(comparisons=list(c("no TATA box","TATA box")),label="p.signif",label.y = 4)+
  theme_minimal()+
  theme(legend.position = "none")+
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(size=18),
        axis.title.y.left = element_blank(),
        axis.text.y.left = element_text(size=18))+
  ylim(0,4.5)+
  labs(title = "SHARP")+
  theme(plot.title = element_text(hjust = 0.5,size=22))


#BRoad
p_Broad_LFC<-Mean_vs_change%>%
  filter(Cluster_Shape=="Broad")%>%
  ggplot(aes(x=TATA_box,y=avg_abs_LFC,fill=TATA_box))+
  geom_violin()+
  geom_boxplot(width=0.1,notch=T)+
  scale_fill_manual(values = c("no TATA box"="tan1","TATA box"="steelblue1"))+
  stat_compare_means(comparisons=list(c("no TATA box","TATA box")),label="p.signif",label.y = 4)+
  theme_minimal()+
  ylim(0,4.5)+
  theme(legend.position = "none")+
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(size=18),
        axis.title.y.left = element_blank(),
        axis.text.y.left = element_text(size=18))+
  labs(title = "BROAD")+
  theme(plot.title = element_text(hjust = 0.5,size=22))


ggarrange(p_sharp_LFC,p_Broad_LFC,nrow = 1,ncol = 2)
```


SHARP
BROAD
****
****
no TATA box
TATA box
no TATA box
TATA box