SI-Size-Figure 1A

Vi Dang

2022-09-10

```
#Set up directory
my_directory <- file.path("D:", "PhD", "git_PhD","TSS-cluster-Classification")
setwd(my_directory)
```

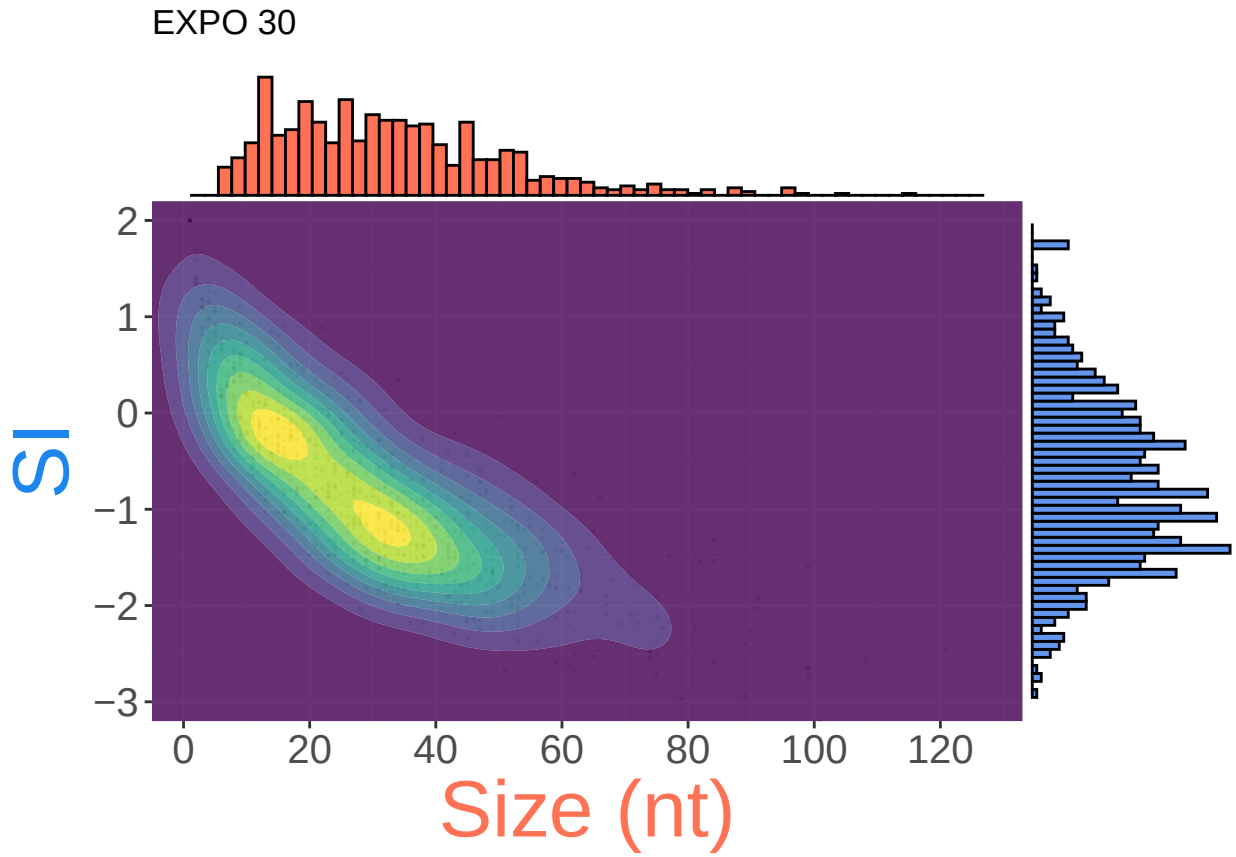
```
#Libraries Loading
library(tidyverse)
library(ggplot2)
library(ggExtra)
library(ggpubr)
library(FactoMineR)
library(MixtureInf)
```

```
#Import data
Expo_For<-read.table("241EXPO.d17.fwd.norm.txt",header=T)%>%
  mutate(strand="+",GrowthPhase="EXPO",temperature="30")
Expo_Rev<-read.table("241EXPO.d17.rev.norm.txt",header=T)%>%
  mutate(strand="-",GrowthPhase="EXPO",temperature="30")
Stat_For<-read.table("241STAT.d17.fwd.norm.txt",header=T)%>%
  mutate(strand="+",GrowthPhase="STAT",temperature="30")
Stat_Rev<-read.table("241STAT.d17.rev.norm.txt",header=T)%>%
  mutate(strand="-",GrowthPhase="STAT",temperature="30")
Expo_Combined<-rbind(Expo_For,Expo_Rev)
Stat_Combined<-rbind(Stat_For,Stat_Rev)
ExpoStat_Combined<-rbind(Expo_Combined,Stat_Combined)
head(ExpoStat_Combined)
```

```
##           Id Chr  start   stop pos_max max_weight         SI Size strand
## 1 CNAG_00016   1  53265  53303   53275     881.44 -0.8653976   39      +
## 2 CNAG_00024   1  73516  73540   73531    1234.28 -0.5014383   25      +
## 3 CNAG_00034   1 101064 101068  101064    2275.42  1.0506036    5      +
## 4 CNAG_00061   1 169285 169293  169286   18863.74  0.5127141    9      +
## 5 CNAG_00065   1 177327 177369  177327    2502.00 -1.9717574   43      +
## 6 CNAG_00067   1 184020 184086  184058    1108.42 -2.0819332   67      +
##   GrowthPhase temperature
## 1        EXPO          30
## 2        EXPO          30
## 3        EXPO          30
## 4        EXPO          30
## 5        EXPO          30
## 6        EXPO          30
```

```
#Plot EXPO data
E<-ggplot(Expo_Combined,aes(x=Size,y=SI))

pE<-E+
  geom_point(alpha=0.15,size=0.1)+
  geom_density_2d_filled(alpha=0.8,bins=10)+
  scale_x_continuous(expand = c(0, 0),lim=c(-5,133),breaks = sort(c(seq(0, 130, by = 20))) )+
  scale_y_continuous(expand = c(0, 0),lim=c(-3.2,2.2),breaks=sort(c(seq(-3,2,by=1))))+
  theme(legend.position = "none")+
  theme(axis.text.x = element_text(size=15),
        axis.text.y = element_text(size=15),
        axis.title.x = element_text(size=30,color = "coral1"),
        axis.title.y = element_text(size=30,color="dodgerblue2"))+
  xlab("Size (nt)")+
  labs(title="EXPO 30")

S<-ggplot(Stat_Combined,aes(x=Size,y=SI))
pS<-S+
  geom_point(alpha=0.15,size=0.1)+
  geom_density_2d_filled(alpha=0.8,bins=10)+
  scale_x_continuous(expand = c(0, 0),lim=c(-5,133),breaks = sort(c(seq(0, 130, by = 20))) )+
  scale_y_continuous(expand = c(0, 0),lim=c(-3.2,2.2),breaks=sort(c(seq(-3,2,by=1))))+
  theme(legend.position = "none")+
  theme(axis.text.x = element_text(size=15),
        axis.text.y = element_text(size=15),
        axis.title.x = element_text(size=30,color = "coral1"),
        axis.title.y = element_text(size=30,color="dodgerblue2"))+
  labs(title="STAT 30")



#add margin distribution
E_plot<-ggMarginal(pE,size=4,type="histogram", xparams = list(bins=60,fill="coral1"),yparams = list(bins
S_plot<-ggMarginal(pS,size=3,type="histogram", xparams = list(bins=60,fill="coral1"),yparams = list(bins

#Both Expo and Stat plot
#ggarrange(E_plot, S_plot, ncol = 2, nrow = 1)
```

```
#Figure 1A
E_plot
```


EXPO 30
SI
2
1
0
−1
−2
−3
0
20
40
60
80
100
120
Size (nt)