CherylUpdateEmail

Awais Choudhry

19/04/2017

## 19/04/2017

Today i was creating some plots and graphs to discuss with my co-supervisior Cheryl Shouhbridge and Jimmy breen

Below is the code i wrote to interrogate different spacings and orientations of 6mer motifs in different arrangements.

Edit:Upon second examination it is quite long and very repetitive

This can be shorter by creating a list object and lapplying every fnction and converting to dataframe.

## Generation for PWM for Planidromic sequences  
library(seqLogo)  
library(magrittr)  
library(GenomicRanges)  
library(ggplot2)  
library(magrittr)  
library(tibble)  
library(pander)  
library(reshape2)  
library(plyr)  
library(MotifDb)  
library(BSgenome.Mmusculus.UCSC.mm9)  
library(JASPAR2016)  
  
enhancerGrange <- import( con = "~/DataFiles/Enhancer Tracks/Mouse/Enhanceresmm9.bed")  
UCSCgenes<- import("~/Scripts/March/mm9.bed")  
promoters<- promoters(UCSCgenes)  
  
arx6MerPWMnospace<- MotifDb::query(MotifDb, "arx")[[6]]  
arx6MerPWM1space<-rbind( A=c(0,1,1,0,0,1,0.25,0,1,1,0,0,1),   
 C=c(0,0,0,0,0,0,0.25,0),  
 G=c(0,0,0,0,0,0,0.25,0) ,  
 T=c(1,0,0,1,1,0,0.25,1,0,0,1,1,0))  
  
arx6MerPWM2space<-rbind( A=c(0,1,1,0,0,1,0.25,0.25,0,1,1,0,0,1),   
 C=c(0,0,0,0,0,0,0.25,0.25),  
 G=c(0,0,0,0,0,0,0.25,0.25,0) ,  
 T=c(1,0,0,1,1,0,0.25,0.25,1,0,0,1,1,0))  
  
arx6MerPWM3space<-rbind( A=c(0,1,1,0,0,1,0.25,0.25,0.25,0,1,1,0,0,1),  
 C=c(0,0,0,0,0,0,0.25,0.25,0.25),  
 G=c(0,0,0,0,0,0,0.25,0.25,0.25,0) ,  
 T=c(1,0,0,1,1,0,0.25,0.25,0.25,1,0,0,1,1,0))  
  
arx6MerPWM4space<-rbind( A=c(0,1,1,0,0,1,0.25,0.25,0.25,0.25,0,1,1,0,0,1),  
 C=c(0,0,0,0,0,0,0.25,0.25,0.25,0.25),  
 G=c(0,0,0,0,0,0,0.25,0.25,0.25,0.25) ,  
 T=c(1,0,0,1,1,0,0.25,0.25,0.25,0.25,1,0,0,1,1,0))  
  
grangeplaindromic1space<-matchPWM(arx6MerPWM1space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
grangeplaindromic2space<-matchPWM(arx6MerPWM2space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
grangeplaindromic3space<-matchPWM(arx6MerPWM3space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
grangeplaindromic4space<-matchPWM(arx6MerPWM4space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
##Databale results  
  
planindromicDataTable<-rbind(cbind(  
length(grangeplaindromic1space),  
Arx6mer<-sum(countOverlaps(grangeplaindromic1space, UCSCgenes)),  
sum(countOverlaps(grangeplaindromic1space, promoters)),  
sum(countOverlaps(grangeplaindromic1space, enhancerGrange))),  
   
cbind(length(grangeplaindromic2space),  
sum(countOverlaps(grangeplaindromic2space, UCSCgenes)),  
sum(countOverlaps(grangeplaindromic2space, promoters)),  
sum(countOverlaps(grangeplaindromic2space, enhancerGrange)))  
,  
cbind(  
numberOfArxSitesPlaindromic3Space<-length(grangeplaindromic3space),  
sum(countOverlaps(grangeplaindromic3space, UCSCgenes)),  
sum(countOverlaps(grangeplaindromic3space, promoters)),  
sum(countOverlaps(grangeplaindromic4space, enhancerGrange))),  
  
cbind(  
numberOfArxSitesPlaindromic4Space<-length(grangeplaindromic4space),  
sum(countOverlaps(grangeplaindromic4space, UCSCgenes)),  
sum(countOverlaps(grangeplaindromic4space, promoters)),  
sum(countOverlaps(grangeplaindromic4space, enhancerGrange)))  
)%>%as.data.frame()  
  
colnames(planindromicDataTable)<- c("Total",  
 "Motifs in genes",  
 "Motifs in Promoters",  
 "Motifs in Enhancers")  
rownames(planindromicDataTable)<- c("1 Space",  
 "2 Space",  
 "3 Space",  
 "4 Space")  
planindromicDataTable%>%pander()  
planindromicDataTable<- rownames\_to\_column(planindromicDataTable)  
  
ggplot(planindromicDataTable, aes(x=rowname, y=Total))+  
 geom\_bar(stat = "identity")+  
 theme\_bw()  
  
  
### Tandeom Sites  
  
arxTandem1Space<-rbind( A=c(0,1,1,0,0,1,0.25,0,1,1,0,0,1),  
 C=c(0,0,0,0,0,0,0.25,0),  
 G=c(0,0,0,0,0,0,0.25,0) ,  
 T=c(1,0,0,1,1,0,0.25,1,0,0,1,1,0))  
arxTandem2Space<-rbind( A=c(0,1,1,0,0,1,0.25,0.25,0,1,1,0,0,1),  
 C=c(0,0,0,0,0,0,0.25,0.25,0),  
 G=c(0,0,0,0,0,0,0.25,0.25,0) ,T=c(1,0,0,1,1,0,0.25,0.25,1,0,0,1,1,0))  
  
arxTandem3Space<-rbind( A=c(0,1,1,0,0,1,0.25,0.25,0.25,0,1,1,0,0,1),  
 C=c(0,0,0,0,0,0,0.25,0.25,0.25,0),  
 G=c(0,0,0,0,0,0,0.25,0.25,0.25,0) ,  
 T=c(1,0,0,1,1,0,0.25,0.25,0.25,1,0,0,1,1,0))  
  
arxTandem4Space<-rbind( A=c(0,1,1,0,0,1,0.25,0.25,0.25,0.25,0,1,1,0,0,1),   
 C=c(0,0,0,0,0,0,0.25,0.25,0.25,0.25,0),  
 G=c(0,0,0,0,0,0,0.25,0.25,0.25,0.25,0) ,  
 T=c(1,0,0,1,1,0,0.25,0.25,0.25,0.25,1,0,0,1,1,0))  
  
grangeTandem1space<-matchPWM(arxTandem1Space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
grangeTandem2space<-matchPWM(arxTandem2Space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
grangeTandem3space<-matchPWM(arxTandem3Space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
grangeTandem4space<-matchPWM(arxTandem4Space, BSgenome.Mmusculus.UCSC.mm9, "90%")  
  
##Tandem DataTable  
tandemDataTable<-rbind(cbind(  
numberofTandem1spaceSites<-length(grangeTandem1space),  
dataTable1SpaceGenes<-sum(countOverlaps(grangeTandem1space, UCSCgenes)),  
dataTable1SpacePromoters<-sum(countOverlaps(grangeTandem1space, promoters)),  
dataTable1SpaceEnhancer<-sum(countOverlaps(grangeTandem1space, enhancerGrange))  
), cbind(  
numberofTandem2spaceSites<-length(grangeTandem2space),  
dataTable2SpaceGenes<-sum(countOverlaps(grangeTandem2space, UCSCgenes)),  
dataTable2SpacePromoters<-sum(countOverlaps(grangeTandem2space, promoters)),  
dataTable2SpaceEnhancer<-sum(countOverlaps(grangeTandem2space, enhancerGrange))  
), cbind(  
numberofTandem3spaceSites<-length(grangeTandem3space),  
dataTable3SpaceGenes<-sum(countOverlaps(grangeTandem3space, UCSCgenes)),  
dataTable3SpacePromoters<-sum(countOverlaps(grangeTandem3space, promoters)),  
dataTable3SpaceEnhancer<-sum(countOverlaps(grangeTandem3space, enhancerGrange))  
), cbind(  
numberofTandem4spaceSites<-length(grangeTandem4space),  
dataTable4SpaceGenes<-sum(countOverlaps(grangeTandem4space, UCSCgenes)),  
dataTable4SpacePromoters<-sum(countOverlaps(grangeTandem4space, promoters)),  
dataTable4SpaceEnhancer<-sum(countOverlaps(grangeTandem4space, enhancerGrange))  
)  
)%>%as.data.frame  
  
colnames(tandemDataTable)<- c("Total",   
 "Motifs in genes",  
 "Motifs in promoters",  
 "Motifs in enhancers")  
rownames(tandemDataTable)<-c("1 Space",  
 "2 Space",  
 "3 Space",  
 "4 Space")  
tandemDataTable%>%pander()  
  
tandemDataTable<-rownames\_to\_column(tandemDataTable)  
  
ggplot(tandemDataTable, aes(x= rowname, y= Total))+  
 geom\_bar(stat="identity")+  
 theme\_bw()

```