ProjReport20160529

hy

5/29/2016

# Goal:This project is to study the non-promoter H3K4me3 in K562.

# Data discription

There are ChIP-Seq data(H3K4me3, H3K4me1, and H3K27ac) of multiple cell lines from ENCODE,including K562 and H1hesc, There are 6 paired-end RNA-Seq data, including 2 K562 and 4 H1hesc.

# Preprocessing

run Macs 1.3.7(default) to get peaks for ChIP-Seq data use DESeq2 to normalize these sample's read count and calculate the log2-tranformed change of gene expression form K562 to H1hesc

# A overview of H3K4me3 distribution in the whole genome among different cell lines

## we check the H3K4me3 peak distribution in different cell lines

preprocess h3k4me3 data, get top 15k peaks by decreaseing p-vale and each state in different cell lines

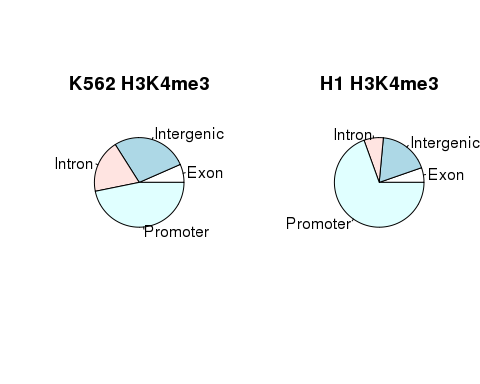
look into the H3K4me3 distribution in the whole genome to inverstigate the difference of H3K4me3 between normal and cancer cell lines

Although the H3K4me3 distribution in A549 samples is variant, the reslut show the tendency that **cancer cell lines seem to have more H3K4me3 in non-promoter region than normal cell lines.**

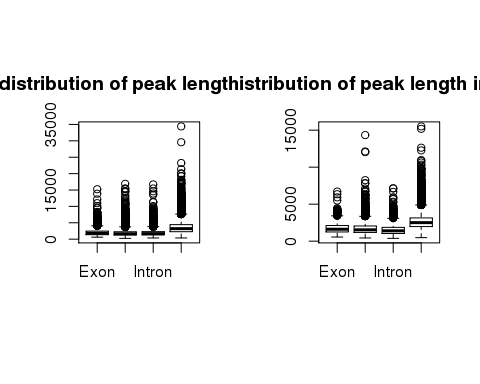
# Using K562 and H1hesc as an example

## First, let's check the H3K4me3 peak distribution in K562 and H1 cell lines

k562.me3<- read.table("/mnt/local-disk1/rsgeno2/MAmotif/macs.20151224/H3K4me3/K562.H3k4me3.Rep2.macs.20151224\_peaks.xls",header = TRUE)  
h1.me3<- read.table("/mnt/local-disk1/rsgeno2/MAmotif/macs.20151224/H3K4me3/H1.H3k4me3.Rep2.macs.20151224\_peaks.xls",header = TRUE)  
Peaks.Distribution<- function(peaks,refhg19=refhg19){  
 peaks$state<- "Intergenic"  
 tp.promoter<- overlap(peaks,refhg19$PROMOTER)  
 peaks$state[tp.promoter[[1]]>0]<- "Promoter"  
 tp.exon<- overlap(peaks[peaks$state=="Intergenic",],refhg19$EXON)  
 peaks$state[peaks$state=="Intergenic"][tp.exon[[1]]>0]<- "Exon"  
 tp.intron<- overlap(peaks[peaks$state=="Intergenic",],refhg19$INTRON)  
 peaks$state[peaks$state=="Intergenic"][tp.intron[[1]]>0]<- "Intron"  
 return(peaks)  
}  
k562.me3<- Peaks.Distribution(k562.me3,refhg19)  
h1.me3<- Peaks.Distribution(h1.me3,refhg19)  
op <- par(mfrow = c(1, 2), pty = "s")  
pie(table(k562.me3$state),main="K562 H3K4me3")  
pie(table(h1.me3$state),main = "H1 H3K4me3")



par(op)  
op <- par(mfrow = c(1, 2), pty = "s")  
boxplot(k562.me3$end-k562.me3$start~k562.me3$state,main="the distribution of peak length in K562")  
boxplot(h1.me3$end - h1.me3$start ~ h1.me3$state, main="the distribution of peak length in H1hesc")



par(op)

In K562 cells, intergenic peaks(or non-promoter peaks) are higher than intergenic peaks in H1hesc. **1.next, we can check other normal and cancer cell lines to inverestage whether this is common phenomenon between them.** **2.we can also look into other histone modification(such as H3K27ac, H3K4me1, and H3K27me3) change from normal cell lines to cancer cell lines.** the length of H3K4me3 peaks in promoter are longer than other peaks.

# Questions

1. why are intergenic peaks in K562 cells higher than intergenic peaks in H1hesc? 1.1 Do cancer cell lines have more H3K4me3 histone modification in non-promoter region than normal? 1.2 what about the distribution of other histone modification between cancer cell lines and normal cell lines?
2. what is the relationship between H3K4me3 intergenic peaks and enhancer? 2.1 what is the relationship between H3K4me3 intergenic peaks and the expression of target gene?

Fisrt, Let's try to study questions 1.1 and 1.2

## K562-unique non-promoter H3K4me3 peaks in Rep1 and Rep2

Top 20k peaks are cho

## The common target genes of K562-unique non-promoter peaks two H3K4me3 Rep

## The relationship between H3K27ac and H3K4me1 in K562-unique non-promoter peaks

## K562-unique enhancer H3K4me3 peaks in Rep1 and Rep2

## GO analysis for K562-unique target genes

## the relationship between K562-unique enhancer H3K4me3 and the expression of their target gene

# oncogenes

# Future work

The relationship between K562-unique enhancer H3K4me3 signal and the expression of their target genes The distribution H3K4me3, H3K4me1, and H3K27ac signal in K562-unique enhancer H3K4me3 peak regions and the promoter regions of their target genes The distribution of Pol II and eRNA signal in K562-unique enhancer H3K4me3 peak regions The enrichment of TF motifs in K562-unique enhancer H3K4me3 peak regions and the promoter regions of their target gene to explore how these enhancers interact with their target genes