Projreport20160627

hy

6/27/2016

# 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

# Why are there more non-promoter H3K4me3 peaks in K562 than in H1hesc?

In previous analysis, we have found the resluts, showing that cancer cell lines seem to have more H3K4me3 in non-promoter region than normal cell lines. next, we use K562 and H1hesc to explore the charaterisitic of these non-promoter peaks. There are two parts: the analysis in K562 and H1hesc, respectivly, and the analysis between K562 and H1hesc.

# The analysis in K562 and H1hesc, respectivly

in the section, we want to check whether the enhancer with H3K4me3 make their target gene overexpression?

Can the expression of different genes be comparable?

If you suppose that gene expressions depend on enhancer's activity, they are comparable.

In order to do that, we need to choose the genuine peaks by using M-value cutoff between -1 and 1. 33658 H3K4me1 18110 H3K27ac 26623 H3K4me3

we use H3K4me1 to define enhancer. According to the state of H3K27ac and H3K4me3, enhancer are classified into four types:Both as enhancer with H3K27ac and H3K4me3(10184), H3K27ac as enhancer with H3K27ac(836), H3K4me3 as enhancer with H3K4me3(6641), and None as enhancer with only H3K4me1(5682)

p1

let's see the expression pattern of target genes in the four ehnacer states.

p2 and p3

It seemed that the gene expression targeted by enhancer with Both state were highest, but the tendency were not as obvious as the expression pattern of genes classified by histone modification state in promoter region

p4 and p5

we also investigate the average DNase signal distribution in the four enhancer states.

p6

the result shows that the average DNase signals have this order:Both > H3K27ac > H3K4me3 > None

# The analysis between K562 and H1hesc

First, Let's see the charateristic of these peaks. ##1.most of the non-promoter H3K4me3 peaks are K562-specific because most of these peaks are K562-unique(M value > 1) after running MAnorm, H3K4me3 peaks are classed into three types: K562-unique(M>1),common(-1<=M<=1),H1hesc-unique(M<-1). we defined promoters as the region which is upstream and downstream 2000 bp from TSS. Let's look at the distribution of the three group H3K4me3 peaks in the whole genome.

p1

It turned out the total number of unique peaks are similar, but the K562-unique non-promoter H3K4me3 peaks are more than twice the H1hesc-unique non-promoter H3K4me3 peaks

## 2.they are enriched in K562-specific enhancer (K562-unique H3K4me1 and H3K27ac)

## 3.GATA factors are enriched in K562-specific enhancer