options(stringsAsFactors = F)

library(BiocManager)

BiocManager::install('GEOquery')

library(GEOquery)

gset=getGEO('GSE24206,destdir='.', ###GSE40839

AnnotGPL=F,

getGPL=F)

exprSet=exprs(gset[[1]])

pd1=pData(gset[[1]]

)

GPL <- read.delim("GPL570-55999.txt",stringsAsFactors=FALSE,skip = 16,sep = '\t' )

#GPL <- read.delim("GPL96.annot",stringsAsFactors=FALSE,skip = 27 )

library(ggpubr)

library(dplyr)

colnames(GPL)

probe<-GPL %>%

select("ID","GENE\_SYMBOL","ENTREZ\_GENE\_ID")

head(probe)

dim(probe)

library(tidyverse)

probe2<-apply(probe,1,function(x){

paste(x[1],

str\_split(x[2]," /// ",simplify = T),

sep = "|")

}

) %>%

unlist()

###

head(probe2)

length(probe2)

probe2<-as\_tibble(probe2)

probe2<-probe2 %>% separate(value,c("ID","GeneName"),sep = "\\|")

dim(probe2)

table(table(probe2$ID))

test2<-probe2 %>% count(ID) %>% filter(n==1) %>%

inner\_join(probe2,by="ID") %>%

select(-n)##remove "n" column

dim(test2)

head(test2)

probe2<-test2

exprSet <- exprSet[rownames(exprSet) %in% probe2$ID,]

identical(rownames(exprSet),ids$ID)

exprSet$Symbol=ids$GeneName

exprSet=exprSet%>%group\_by(Symbol)%>%summarise\_each(funs(mean))#去重

exprSet=as.data.frame(exprSet)

rownames(exprSet)=exprSet$Symbol

exprSet$Symbol=NULL

exprSet\_new=exprSet

metadata=pd1[,c(1,2)]

metadata$`phenotype:ch1`[1:6]=‘control’

metadata$`phenotype:ch1`[7:23]='IPF'

exprSet\_new=as.data.frame(t(exprSet\_new))

identical(rownames(exprSet\_new),rownames(metadata))

exprSet\_new$group=metadata$`phenotype:ch1`

exp <- rownames\_to\_column(as.data.frame(exprSet),var = " Symbol ")

exp\_L<- melt(exp)

colnames(exp\_L)<- c("Symbol","sample","value");table(exp\_L$sample)

exp\_L$group<- rep(Group,each=nrow(exp))

head(exp\_L)

library(ggplot2)

ggplot(exp\_L,aes(x=sample,y=value,fill=group))+geom\_boxplot()+

theme(axis.text.x = element\_text(angle=0, vjust=0.6))

ggplot(

data= exprSet\_new,

aes(x = group, y = COL20A1)) + ###COL27A1

geom\_point(position = 'jitter',

color= 'black',

size= 2,

alpha= 0.8) +

theme\_classic()+

ggtitle("GSE24206 COL20A1")+

geom\_violin(aes(fill = group),

color= 'grey',

scale= 'width',

linewidth= 0.8,

trim= TRUE,

alpha= 0.8)+

scale\_fill\_manual(values=c('#58A4C3','#E95C59'))+

geom\_boxplot(color = 'white',

width= 0.15,

size= 0.8,

fill= NA)+

stat\_compare\_means(comparisons = list( c('IPF’,’control’)

),label = "p.format",method = "wilcox.test",)