生成 Differential_Exp_TPM.txt

> 姓名：李群
>
> 老师：吴飞珍
>
> 学号：20111510021
>
> 邮箱：20111510021@fudan.edu.cn
>
> 时间：2020年12月11日

这部分主要涉及R语言的代码

提前将TPM文件和差异表达文件拷贝到`/public/home/st28/report1/rawData/RAnalysis`文件夹下

```
#处理TPM文件
setwd('/public/home/st28/report1/rawData/RAnalysis')                #设置运行初始话文件夹
rm(list=ls())                                                       #删除环境中已有变量
# 加载、提取数据并赋予标签
DEG=read.table("gene_exp.diff",header = T)                          #导入差异表达文件
DEG=DEG[,c(3,10,12)]                                                #提取gene，logFC和p值
DEG=DEG[is.finite(DEG$log2.fold_change.),]                          #排除NA，NaN，无穷大，返回TRUE
DEG=DEG[abs(DEG$log2.fold_change.)>log2(1.5) & DEG$p_value<0.05,] #保留FC大于1.5，p值小于0.05的数据
names(DEG)=c("genes","foldchange","pvalue")                         #修改列名为genes，foldchange和pvalue
DEG$regulation="down"
DEG$regulation[DEG$foldchange<0]="up"                               #增加一列regulation，根据foldchange赋予up和down类别
                                                                    #计算差异表达的时候我是KT vs WT所以此处令FC<0为up
#处理差异表达文件
dd=read.table("Expression_TPM.xls",header = T)
DEG1=DEG[order(abs(DEG$foldchange),decreasing = T),]                #选取前40进行操作，方便上交作业进行查看
DEG1=DEG1[1:40,]                                                    #其他操作与此一致
dd1=dd[dd$gene_id %in% DEG1$genes,]
row.names(dd1)=dd1$gene_id
dd1$gene_id=NULL
```

```
head(dd1)
#           SRR8734708 SRR8734712 SRR8734718 SRR8734722
#    Acta1   66.151933 1.94048393 194.288144 1.77682715
#    Acta2   64.656162 1.19114537 246.012581 0.61783471
#    Actc1    4.700736 0.10098259  35.848814 0.31898462
#    Antxr2   1.981463 0.07358183   8.479333 0.05007741
#    Bmp1    28.094466 2.15089616 152.900672 1.51706758
#    Bmp2     3.513508 0.11223066  29.559151 0.06167556
names(dd1)=c("WT1","KO1","WT2","KO2")
head(dd1)
#                 WT1        KO1        WT2        KO2
#    Acta1  66.151933 1.94048393 194.288144 1.77682715
#    Acta2  64.656162 1.19114537 246.012581 0.61783471
#    Actc1   4.700736 0.10098259  35.848814 0.31898462
#    Antxr2  1.981463 0.07358183   8.479333 0.05007741
#    Bmp1   28.094466 2.15089616 152.900672 1.51706758
#    Bmp2    3.513508 0.11223066  29.559151 0.06167556

#合并两个数据
dd1$genes=rownames(dd1)                                    #给dd1增加一列genes,
与DEG合并，使用函数merge
data = merge(DEG,dd1,by="genes")
head(data)
#   genes foldchange  pvalue regulation        WT1       KO1        WT2
#1  Acta1   -7.32175 0.00005         up 1.50212528 66.047068 1.26174673
#2  Acta2   -7.39843 0.00005         up 1.32037326 64.998085 0.54098342
#3 Antxr2   -6.35281 0.04320         up 0.07254987  1.945530 0.04995185
#4   Bmp1   -5.58273 0.00005         up 2.16840139 27.834534 1.44127886
#5   Bmp2   -7.47658 0.01990         up 0.12769611  3.495013 0.05273834
#6   Cd44   -6.96206 0.00005         up 0.17185940  6.531589 0.21590387
#         KO2
#1 194.923310
#2 245.141102
#3   8.382594
#4 151.218199
#5  29.183706
#6  43.466384
#保存为老师需要的格式
data1 = data[,-4]                                          #删除regulation这一列
colnames(data1) = c("gene_id", "log2_foldchange", "p_value",
"set1_WT_D0","set1_NonoKO_D0","set2_WT_D0","set2_NonoKO_D0" )

head(data1)
#  gene_id log2_foldchange p_value set1_WT_D0 set1_NonoKO_D0 set2_WT_D0
#1   Acta1        -7.32175 0.00005 1.50212528      66.047068 1.26174673
#2   Acta2        -7.39843 0.00005 1.32037326      64.998085 0.54098342
#3  Antxr2        -6.35281 0.04320 0.07254987       1.945530 0.04995185
#4    Bmp1        -5.58273 0.00005 2.16840139      27.834534 1.44127886
#5    Bmp2        -7.47658 0.01990 0.12769611       3.495013 0.05273834
```

```
#6    Cd44        -6.96206 0.00005 0.17185940       6.531589 0.21590387
#  set2_NonoKO_D0
#1     194.923310
#2     245.141102
#3       8.382594
#4     151.218199
#5      29.183706
#6      43.466384
#保存文件
write.table(data1,file = "Differential_Exp_TPM.txt",sep="\t",quote =
F,row.names = F)
#查看文件
(base) [st28@ibs report3]$ head Differential_Exp_TPM.txt
#gene_id  log2_foldchange p_value set1_WT_D0  set1_NonoKO_D0  set2_WT_D0
set2_NonoKO_D0
#Acta1  -7.32175  5e-05 1.50212528356821  66.0470684201909  1.26174673301969
194.923309634562
#Acta2  -7.39843  5e-05 1.32037325924688  64.9980851876817  0.540983418370514
245.141102004707
#Antxr2 -6.35281  0.0432  0.0725498676467781  1.94553021726517
0.0499518527924758  8.38259384722471
#Bmp1 -5.58273  5e-05 2.16840139031331  27.8345335836616  1.4412788612435
151.218199220567
#Bmp2 -7.47658  0.0199  0.127696113267556 3.49501257121745  0.0527383369409183
29.183706239403
#Cd44 -6.96206  5e-05 0.171859398669249 6.53158901756655  0.215903874361817
43.4663838576435
#Cdk14  -6.50128  0.00425 0.0725735448360725  4.15407726801075
0.0750425985311999  9.64175337705559
#Cdx2 -6.38677  2e-04 0.269148969480038 13.4213160720506  0.461386419972748
49.9243063717431
#Col12a1  -6.66436  0.00325 0.0367899126905428  1.40551766627632
0.0304987367698262  5.62416386329396
```

生成的文件`Differential_Exp_TPM.txt`在附件当中，也可在github中进行下载 链接