

TCGA甲基化数据分析

发表于2018年8月12日由daizao



```
cd /home/train/Liuping/raw_data
mkdir ProstateCancer_methylation
cd ProstateCancer_methylation/

#合并下载好的数据
nohup ./Merge.pl metadata.cart.2018-08-01.json > /home/train/Liuping/raw_data/TCGA_Methylation_ProstateCancer/methylayio
#注意记录下normal '和 tumor'的个数
# normal count: 50
# tumor count: 503

mkdir ../result
cp geneMethy.txt ../result/
cd ../result/

#差异甲基化
source("http://bioconductor.org/biocLite.R")
biocLite("limma")
library(limma)
setwd("D:/TCGA_methylation")
inputFile <- "geneMethy.txt"
normalNum=50
tumorNum=503
fdrFilter=0.05
logFCfilter=1

outTab <- data.frame()
grade <- c(rep(1,normalNum),rep(2,tumorNum))
Type <- c(rep("Normal",normalNum),rep("Tumor",tumorNum))
rt <- read.table(inputFile,sep="\t",header=T,check.names=F)
rt <- as.matrix(rt)
rownames(rt) <- rt[,1]
exp <- rt[,2:ncol(rt)] #行名可被直接移植
dimnames <- list(rownames(exp),colnames(exp))
data <- matrix(as.numeric(as.matrix(exp)),nrow=nrow(exp),dimnames=dimnames)
data <- avereps(data)
data <- data[rowMeans(data)>0,]

data <- normalizeBetweenArrays(data)
normalData <- cbind(id=row.names(data),data) #为了让id号写入文件
write.table(normalData,file="normalizeMethy.txt",sep="\t",row.names=F,quote=F)

for (i in row.names(data)){
  rt <- rbind(expression=data[i,],grade=grade)
  rt <- as.matrix(t(rt))
  wilcoxTest <- wilcox.test(expression ~ grade, data=rt)

  normalGeneMeans <- mean(data[i,1:normalNum])
  tumorGeneMeans <- mean(data[i,(normalNum+1):ncol(data)])
  logFC <- log2(tumorGeneMeans) - log2(normalGeneMeans)
  pvalue <- wilcoxTest$p.value
  normalMed <- median(data[i,1:normalNum])
  tumorMed <- median(data[i,(normalNum+1):ncol(data)])
  diffMed <- tumorMed - normalMed
  if((( logFC>0) & ( diffMed >0)) | (( logFC < 0 ) & ( diffMed <0 ))){
    outTab <- rbind(outTab,cbind(gene=i,normalMean=normalGeneMeans,TumorMean=tumorGeneMeans,logFC=logFC,pValue=pvalue))
  }
}
pValue <- outTab[, "pValue"]
fdr <- p.adjust(as.numeric(as.vector(pValue)),method="fdr")
outTab <- cbind(outTab,fdr=fdr)
write.table(outTab,file="allGene.xls",sep="\t",row.names=F,quote=F)
```

```

outDiff <- outTab[(abs(as.numeric(outTab$logFC))) > logFCfilter & as.numeric(as.vector(outTab$fdr)) < fdrFilter]
write.table(outDiff,file="diff.xls",sep="\t",row.names=F,quote=F)
heatmap <- data[as.vector(outDiff[,1]),]
heatmap <- cbind(id=row.names(heatmap),heatmap)
write.table(heatmap,file="heatmap.txt",sep="\t",row.names=F,quote=F)

#pheatmap图
install.packages("pheatmap")
setwd("D:/TCGA_methylation/pheatmap")
library(pheatmap)

rt <- read.table("heatmap.txt",sep="\t",header=T,row.names=1,check.names=F)
Type <- c(rep("normal",50),rep("tumor",503))
names(Type) <- colnames(rt)
Type <- as.data.frame(Type)

tiff(file="heatmap.tiff",width=45,height=70,units="cm",compression="lzw",bg="white",res=300)
pheatmap(rt,annotation=Type,color=colorRampPalette(c("green","black","red"))(50),cluster_cols=F,fontsize_row=5,fontsize_dev.off())

#甲基化与转录组数据融合
setwd("D:/TCGA_methylation/MethyGeneMerge")
methy <- read.table(file="normalizeMethy.txt",row.names=1,header=T,sep="\t",check.names=F)
rna <- read.table(file="normalizeExp.txt",row.name=1,header=T,sep="\t",check.names=F)

colnames(methy) <- gsub("(.*?)-(.*?)-(.*?)-(.*?)-.*", "\\1-\\2-\\3-\\4",colnames(methy))
colnames(rna) <- gsub("(.*?)-(.*?)-(.*?)-(.*?)-.*", "\\1-\\2-\\3-\\4",colnames(rna))
sameSample <- intersect(colnames(methy),colnames(rna))
rownames(methy) <- paste(rownames(methy),"methy",sep="|")
rownames(rna) <- paste(rownames(rna),"exp",sep="|")
merge <- rbind(id=sameSample,methy[,sameSample],rna[,sameSample])
write.table(merge,file="merge.txt",sep="\t",quote=F,col.names=F)

#基因甲基化与转录组数据相关关系
setwd("D:/TCGA_methylation/methycor")
inputFile <- "merge.txt"
methyGene <- "NCAPD3|methy"
expGene <- "NCAPD3|exp"
rt <- read.table(inputFile,sep="\t",header=T,check.names=F,row.names=1)
i <- rt[methyGene,]
j <- rt[expGene,]
x <- as.numeric(i)
y <- log2(as.numeric(j)+1)
corT <- cor.test(x,y)
methyGeneName <- unlist(strsplit(methyGene,"\\|"))[[1]]
expGeneName <- unlist(strsplit(expGene,"\\|"))[[1]]
z <- lm(y~x)
estimate <- corT$estimate
cor <- round(estimate,3)
pvalue <- corT$p.value #以小数的方式保留位数
pval <- signif(pvalue,4) #以科学计数法的方式保留位数
pval <- format(pval,scientific=T)

outTiff <- "cor.tiff"
tiff(file=outTiff,width=15,height=15,units="cm",compression="lzw",bg="white",res=300)
plot(x,y,type="p",pch=16,main=paste("Cor=",cor,"(p-value=",pval,")",sep=""),xlab=paste(methyGeneName,"methylation"),ylab=expGeneName,
lines(x,fitted(z),col=2)
dev.off()

#甲基化位点提取
nohup ./posget.pl metadata.cart.2018-08-01.json NCAPD3 > /home/train/Liuping/raw_data/TCGA_Methylation_ProstateCancer/NCAPD3/
#1 40635

#甲基化位点与gene表达合并
setwd("D:/TCGA_methylation/posmethygene")
methyFile="posMethy.txt"
expFile="normalizeExp.txt"
methy <- read.table(methyFile,row.names=1,header=T,sep="\t",check.names=F)
rna <- read.table(expFile,row.names=1,header=T,sep="\t",check.names=F)
colnames(methy) <- gsub("(.*?)-(.*?)-(.*?)-(.*?)-.*", "\\1-\\2-\\3-\\4",colnames(methy))
colnames(rna) <- gsub("(.*?)-(.*?)-(.*?)-(.*?)-.*", "\\1-\\2-\\3-\\4",colnames(rna))
sameSample <- intersect(colnames(methy),colnames(rna))
rownames(methy) <- paste(rownames(methy),"methy",sep="|")
rownames(rna) <- paste(rownames(rna),"exp",sep="|")
merge <- rbind(id=sameSample,methy[,sameSample],rna[,sameSample])

```

```
#甲基化位点与gene表达相关
setwd("D:/TCGA_methylation/cor_posmethygene")
inputFile <- "merge.txt"
posmethy <- "cg14934141lmethy"
expGene <- "NCAPD3lexp"
rt <- read.table(inputFile,sep="\t",header=T,check.names=F,row.names=1)
i <- rt[posmethy,]
j <- rt[expGene,]
x <- as.numeric(i)
y <- log2(as.numeric(j)+1)
corT <- cor.test(x,y)
posmethyid <- unlist(strsplit(posmethy,"\\|"))[[1]]
expgenename <- unlist(strsplit(expGene,"\\|"))[[1]]
z <- lm(y~x)
cor <- corT$estimate
cor <- round(cor,3)
pvalue <- corT$p.value
pval <- signif(pvalue,4)
pval <- format(pval,scientific=T)
outTiff <- "posCor.tiff"
tiff(file=outTiff,width=15,height=15,units="cm",compression="lzw",bg="white",res=300)
plot(x,y,type="p",pch=16,main=paste("Cor=",cor,"(p-value=",pval,")",sep=""),xlab=paste(posmethyid,"methylation"),ylab=paste(expgenename,"expression"),
lines(x,fitted(z),col=2)
dev.off()
```

[合并的脚本](#)，类似转录本的合并脚本，需要自己根据需要自行修改



欢迎添加下面这个群提问：

此条目发表在TCGA分类目录。将[固定链接](#)加入收藏夹。

《TCGA 甲基化数据分析》有103条回应



无聊说：
2018年9月12日 下午9:23

请问如何把TCGA下载的文件汇总为一个矩阵
[回复](#)



daizao说：
2018年9月13日 上午12:36

需要自己专门写一个脚本提取，你需要哪个癌症么？甲基化合并需要及计算机配置好点才行，我可以帮你合并好，发给你。
[回复](#)



无聊说：
2018年9月16日 上午11:07

我用你的代码进行分析，在做差异性分析的时候for语句特别慢，是电脑问题还是数据问题？

[回复](#)



daizao说：

2018年9月17日 上午3:23

R里面循环确实挺慢的

[回复](#)



Happy说：

2018年10月10日 下午8:27

能否帮我合并一个肿瘤甲基化数据，个人机内存太小实在不行，谢谢，我扣扣2953-52900

[回复](#)



daizao说：

2018年10月10日 下午11:13

可以

[回复](#)



王烨金鹏说：

2018年10月24日 下午5:18

你好，可否也帮我下一个甲基化的数据呢？十分感谢！QQ 30665-0707



daizao说：

2018年10月25日 下午3:57

最近忙自己的课题，过阵子吧～



zhuy说：

2018年12月4日 上午12:07

您好，能否在有空的时候帮忙合一下宫颈癌的甲基化的数据呢？个人电脑实在跑不动，感激不尽！QQ 1059512605



daizao说：

2018年12月4日 下午5:28

最近忙着毕业，不好意思啊



程芳说：

2019年4月25日 下午1:08

你好请问有时间帮我合并一下甲基化矩阵吗，非常感谢，还有我下载数据也有些障碍。。

[回复](#)



daizao说：

2019年4月25日 下午10:02

最近忙毕业课题，所以没有时间了哦，不好意思啦

[回复](#)



无聊说：

2018年9月16日 下午1:04

你的outtab是不是缺少了几列：logFC、pvalue、normalMed、tumorMed、diffMed

[回复](#)



daizao说：

2018年9月17日 上午3:23

格式可以自己制定，但是确实少写了logFC和pvalue，已添加，谢谢指正！

[回复](#)



无聊说:

2018年9月18日 上午12:12

想请教一下TCGA下载的甲基化文件如何注释ID? 我下载的450K文件每个ID对应多个基因名字, 应如何配对?

[回复](#)



[daizao](#)说:

2018年9月20日 下午2:58

下载的文件不是有ID和基因的对应关系么? 我是取第一个配对的基因

[回复](#)



吴锦杰说:

2018年9月22日 下午10:20

谢谢, 你的分析中NCAPD3是如何挑选的?

[回复](#)



[daizao](#)说:

2018年9月25日 上午12:51

实验室同学要这个基因

[回复](#)



蒋三说:

2018年11月23日 上午9:16

为什么只取第一个。我将gene_symbol去重后取所有的, 能多出5000个symbol

[回复](#)



[daizao](#)说:

2018年11月24日 上午1:03

为了方便我写程序, 你的方法能解决你的问题就好~

[回复](#)



snowy说:

2018年9月26日 下午10:33

甲基化位点提取矩阵, 如何获得, 谢谢!

[回复](#)



[daizao](#)说:

2018年9月27日 上午3:06

需要根据自己需要写脚本提取

[回复](#)



王说:

2018年11月10日 下午10:19

请问merge之后的文件是第一列为基因名还是探针号, 其他列为样本? 然后数据都是 β 值的数据吗? 还是什么样的格式? 我没有从TCGA上下载原始数据, 是下载的整合后的, 但是不知道该整理成什么样, 才能作为limma的输入文件

[回复](#)



[daizao](#)说:

2018年11月11日 下午2:07

是你说的那个样子

[回复](#)



求教说:

2018年11月10日 下午10:30

亲，合并后的矩阵是什么样的格式呢，基因和样本信息，及 β 值吗

[回复](#)



[daizao](#)说:

2018年11月11日 下午2:07

是的

[回复](#)



刘辉说:

2018年11月20日 上午12:43

你好，能否帮我合并下甲基化数据，我的电脑合并不起来

[回复](#)



[daizao](#)说:

2018年11月24日 上午12:24

最近忙着毕业，不好意思啊

[回复](#)



刘辉说:

2018年11月25日 下午6:06

没事，我后来合并好了，然后那个甲基化位点的提取，我提取后只有样本的ID，没有甲基化位点的，不知道能否发我个脚本，或者买也可以的，我的扣扣：1098210582，麻烦了！打扰了！感谢！

[回复](#)



刘辉说:

2018年11月25日 下午6:18

我后来捣鼓了好久，合并好了，但甲基化位点没能提取成功，提取出来只有ID号，没有甲基化位点，不知道能否给我个脚本，或者帮我修改下脚本，有偿的，不好意思，打扰了！感谢！我的扣扣：1098210582

[回复](#)



[daizao](#)说:

2018年12月3日 下午10:23

因为实验室原因这个脚本目前还不能给你，而且忙于毕业课题所以木有时间帮你看脚本~抱歉啦~

[回复](#)



刘辉说:

2018年12月8日 下午9:32

没事没事，如果后面有时间的话，方便的话能不能帮我合并下甲基化文件，因为我合并的时候老是有一列样品名全部为o



周说:

2018年12月17日 下午10:26

请问合并的矩阵作为输入文件是这个格式的吗？

GENE SYMBOL NORMAL_1 NORMAL_2 TUMOR_1 TUMOR_2

gene1 beta_value beta_value beta_value beta_value

gene2 beta_value beta_value beta_value beta_value

....

正常样本和癌症样本都有多个，往后排列？

[回复](#)



daizao 说:
2018年12月18日 下午9:41

尽量把正常的和癌症的分开放在一起

[回复](#)



小穆 说:
2019年2月22日 下午4:51

我在TCGA上下载甲基化数据，我想确定promoter的甲基化水平。在下载的数据中，每一个cg号探针都对应着一个 position to tss 。按理说，每个探针都对应一个甲基化位点，所以position to tss应该是一个数字，但为什么下载的数据里面position to tss里是好多个数字呢

[回复](#)



daizao 说:
2019年2月22日 下午10:41

因为一个基因可以有不同甲基化位点，所以甲基化位点与TSS的距离，所以可以有多个

[回复](#)



小穆 说:
2019年2月23日 下午2:49

还是有点不明白。一个cg号探针应该是只对应一个甲基化位点吧，所以一个位点对应一个tss应该是一个数字啊。

[回复](#)



daizao 说:
2019年2月24日 下午3:48

你说的没问题，但是一个基因对应着不同的转录本，而这些转录本的TSS并不一定是同一个，所以会出现多个数字（原始文件里面相同的gene_symbol有不同的Transcript_ID）
一个cg号 -> 相同的gene_symbol -> 多个转录本（同一个基因） -> 多个TSS

[回复](#)



小穆 说:
2019年3月7日 上午11:57

真的是十分感谢，找了很多资料都没有明白，谢谢你！



谭增琦 说:
2019年2月25日 下午2:39

非常详细的甲基化数据分析，非常有用！请问你现在有空吗，可否帮忙下载个BRCA甲基化数据呢，我这边网实在是不行，如果可以方便的话，请加我QQ544931709

[回复](#)



谭增琦 说:
2019年2月25日 下午4:20

非常细致用心的教程！受益匪浅，请问你现在还忙吗？可否麻烦你帮我下载个BRCA的甲基化数据呢？十分感谢！扣544931709

[回复](#)



daizao 说:
2019年2月26日 下午2:53

这段时间忙自己的课题，不好意思哦

[回复](#)



周瑶瑶 说:

你好，在运行代码

```
for (i in row.names(data)){  
rt <- rbind(expression=data[i,],grade=grade)  
rt <- as.matrix(t(rt))  
wilcoxTest <- wilcox.test(expression ~ grade, data=rt)
```

```
normalGeneMeans <- mean(data[i,1:normalNum])  
tumorGeneMeans <- mean(data[i,(normalNum+1):ncol(data)])  
logFC <- log2(tumorGeneMeans) – log2(normalGeneMeans)  
pvalue <- wilcoxTest$p.value  
normalMed <- median(data[i,1:normalNum])  
tumorMed <- median(data[i,(normalNum+1):ncol(data)])  
diffMed o) & ( diffMed >o)) | (( logFC < o ) & ( diffMed <o ))){  
outTab <-  
rbind(outTab,cbind(gene=i,normalMean=normalGeneMeans,TumorMean=tumorGeneMeans,logFC=logFC,pValue=pvalue))  
}}
```

的时候出现Error in data[i,]:subscript out of bounds.但是不知道哪里错了求解决

[回复](#)



daizao 说:

2019年3月10日 上午1:09

你的data是什么样的结构?

[回复](#)



周瑶瑶 说:

2019年3月11日 下午12:15

就是类似于这种的data,列名就是GA-IG-A3I8-11A-1 GA-L5-A43C-11A-1 GA-L5-A4OE-11A-1 GA-L5-A4OF-11A-1.....行名就是cg开头的.....
id GA-IG-A3I8-11A-1
cg000000029 0.343621638
cg000000108 NA
cg000000109 NA

[回复](#)



daizao 说:

2019年3月12日 上午10:27

grade定义了么?

[回复](#)



周瑶瑶 说:

2019年3月12日 上午11:33

定义了，normal是1，tumor是2



daizao 说:

2019年3月12日 下午7:44

你要定义成normal数是多少，tumor数是多少



周瑶瑶 说:

2019年3月13日 上午11:53

normal是16，turmor是186



daizao 说:

2019年3月13日 下午1:17

要不先别用循环，把i换成你的一个行名，一句一句试试，这样容易发现哪里出了问题。



周瑶瑶 说:

2019年3月13日 下午9:26

我用debug发现是rt <- rbind(expression=data[i,],grade=grade)这一行代码出现的Error in data[i,]:subscript out of bounds，报错是rbind用法错误，但是不知道哪里错了



daizao 说:

2019年3月14日 下午1:15

我觉得是数据格式问题，所以建议你先用第一行或者第二行的数据跑一下，我觉得可能是NA值。



周瑶瑶 说:

2019年3月17日 下午7:05

你好，我的问题解决了，就是不要i了，直接用row.names(data)代替就好了，所以很感谢您，但还是想咨询您一个问题，就是我得到了差异甲基化的基因，但是有将近两万个，怎么从这里面留下比较适合做GO和KEGG的基因呢？



daizao 说:

2019年3月17日 下午9:07

我一般全放进去做看富集结果，如果太多的差异基因我一般会用别的种类的数据来筛选(例如:snp,cnv，lnc靶基因等等等)



周瑶瑶 说:

2019年3月25日 下午7:02

您好！请问得到了差异性甲基化的基因怎么知道是高甲基化还是低甲基化？麻烦您了



daizao 说:

2019年3月26日 上午7:38

logFC的正负值



朱良宇 说:

2019年5月18日 下午3:44

rbind(expression=data[i,],grade=grade)是直接把这个i换了吗
rbind(expression=data[row.names(data),],grade=grade)这样吗？



daizao 说:

2019年5月18日 下午11:55

可是i就是行名额



daizao 说:

2019年5月19日 下午4:26

na替换成O，具体的得看数据了



果子 说:

2019年3月29日 上午12:38

你好，我跟你遇到一样的问题，将rbind中的i改成row.names(data)还是报错，请指教一下怎么改才对，多谢！

回复



橘子 说:

2019年4月11日 下午1:54

我也遇到了和您一样的问题，改了i之后还是报错，请问一下您怎么解决的呢



daizao 说:

2019年4月12日 下午2:42

我觉得是数据格式问题



朱良宇说:

2019年5月27日 下午8:49

您说的格式问题具体什么意思啊，我用他们数据，课程上的数据能跑代码，用我自己就不行了，

[回复](#)



daizao说:

2019年5月28日 下午8:44

我也不是很清楚额

[回复](#)



果子说:

2019年3月29日 上午12:27

我跟你一样的情况呢，metadata的数据比别人教程里下载的小很多，不晓得什么原因。

[回复](#)



abc说:

2019年4月11日 下午2:02

您好，我也遇到这个问题Error in data[i,]:subscript out of bounds，但是改了i之后还是报错，请问还应该怎么改呢

[回复](#)



daizao说:

2019年4月12日 下午2:44

很好奇你们的这个报错，我重复不出来你们的问题。。。。欢迎你们贴出文件，我来重复你们的问题

[回复](#)



gongyue说:

2019年4月10日 上午10:30

您好，打扰啦，想请问一下您的基因表达文件“normalizeExp”是怎么生成的呢

[回复](#)



daizao说:

2019年4月10日 下午2:44

转录组数据分析中矫正后的矩阵

[回复](#)



橘子说:

2019年4月11日 上午10:13

甲基化差异分析流程按照您的代码跑还是有一些问题可以请教一下吗，不知道怎么联系您，我的QQ2325933960,有偿，麻烦啦

[回复](#)



daizao说:

2019年4月14日 下午11:26

你可以加下群，我抽空看下你的文件是怎么回事

[回复](#)



泮说:

2019年4月16日 下午8:11

outDiff logFCfilter & as.numeric(as.vector(outTab\$fdr)) < fdrFilter),]

请问这里为什么需要挑选logFC的绝对值大于1的行呢？

[回复](#)



daizao 说:

2019年4月17日 下午12:53

差异倍数，这个看个人设置习惯

[回复](#)

宋晓帆 说:

2019年5月9日 下午3:41

您好，下载的数据里有NA值，这些值您是采取什么方法处理的呢？用什么算法填充一下吗？另外很感谢您的分享，对最近的学习有很大帮助。

[回复](#)



daizao 说:

2019年5月9日 下午6:45

建议统一改成o

[回复](#)



新大麦味茶 说:

2019年6月16日 下午4:05

请问有原始TCGA甲基化数据整合成矩阵的代码，可以分享一下吗，谢谢

[回复](#)



daizao 说:

2019年6月17日 下午3:56

目前因为实验室的原因，暂时不能公开，其实写起来不难的。

[回复](#)



le 说:

2019年6月18日 下午3:55

这是生信自学网的视频脚本吧，我好像用过

[回复](#)



daizao 说:

2019年6月18日 下午6:02

对，他们的代码和生信技能树的代码都挺好用。不过有的代码有问题，理解一下改改就好啦~

[回复](#)



xueshengxin 说:

2019年7月2日 下午4:08

请问怎样把TCGA下载下来的数据合并成矩阵呢？

[回复](#)



daizao 说:

2019年7月2日 下午9:30

你可以用正则去提取一目的个文件，再写循环合并，或者用json模块，写循环合并。

[回复](#)



FLL 说:

2019年7月26日 下午5:56

您好，我想问一下在第一步甲基化差异分析中，为什么没有把分组信息输入呢？

[回复](#)



daizao 说:

2019年7月28日 下午9:00

因为根据样本的id就可以知道是癌症还是正常样品啊

[回复](#)



FLL 说:

2019年7月29日 下午4:55

程序是会自动识别嘛？（还有我看到您另一篇是关于TCGA样品排序的，请问这个表达矩阵需要先排序再做差异分析嘛

[回复](#)



daizao 说:

2019年7月29日 下午9:20

一般排好序，才好做差异分析

[回复](#)



马妮籽 说:

2019年8月28日 上午10:58

真是一个好博主，回答问题这么仔细，你毕业了吗

[回复](#)



daizao 说:

2019年8月28日 下午2:14

还没

[回复](#)



shaoqing liu 说:

2019年9月27日 下午9:36

那么多位点，怎样确定一个基因的甲基化水平啊，是把各个位点的b值平均一下吗？

[回复](#)



daizao 说:

2019年9月28日 下午10:50

是的，我后来又重新写了脚本，发现计算平均值的结果会可靠一点

[回复](#)



于永春 说:

2019年10月3日 上午11:11

不知道博主毕业了没有，有没有时间，我最近也是在合并矩阵上出的问题，文件太大了，电脑运行不下来，停止运行后得出的文件不知道全不全，样本数倒是对，郁闷，可不可以帮个忙，谢谢！

[回复](#)



flxi 说:

2019年10月17日 下午2:35

ucsc xena上有合并号的矩阵可以直接下载，矩阵差异用什么做的？查了minfi,ChAMP，输入格式 都不对。

[回复](#)



daizao 说:

2019年10月17日 下午9:42

不是很清楚额，格式不对，那就了解下写代码整合就好了

[回复](#)



flxi说:
2019年10月18日 下午5:15

本来也是这么打算的，找到矩阵和分组数据的位置，但难度太大，几个软件都是用的甲基化原始数据.IDAT做的。代码封装在后面

[回复](#)



daizao说:
2019年10月18日 下午9:21

有难度，写出来才有进步～

[回复](#)



flxi说:
2019年10月17日 下午2:38

差异是怎么做的ttest吗? ucsc xena上有合并好的Illumina Human Methylation 450矩阵. 查了minfi、ChAMP这些计算甲基化数据的包，格式都不对

[回复](#)



daizao说:
2019年10月17日 下午9:45

差异看个人怎么做的，我这里用的wilcoxTest检验

[回复](#)



Jenson Zhou说:
2019年11月18日 上午9:35

博主，你这里对提取后的甲基化数据用normalizeBetweenArrays做标准化看上去有点问题。normalizeBetweenArrays函数是用于标准化不同array之间数据的，它默认一列数据是一个array，而你的data结构中一列是一个样本的各个基因的甲基化 β 值，那么执行完normalizeBwtweenArrays函数后，不同样本之间的各基因甲基化水平的都被标准化到同样的scale了，而这显然是不合理的，因为偏恶性的肿瘤组织中，其各基因整体的甲基化水平应该比偏良性的肿瘤组织高。

[回复](#)



daizao说:
2019年11月19日 上午1:43

但要是不做均一化，直接作比较，这也不妥吧

[回复](#)



刘杨说:
2019年11月20日 下午3:02

有前列腺TCGA甲基化基因矩阵吗，求分享

[回复](#)



daizao说:
2019年11月20日 下午9:48

这个真不好分享。。

[回复](#)



彭振说:
2019年12月5日 上午10:39

为什么有的甲基化位点会对应好几个基因啊?比如cg00000029对应的gene_symbol是RBL2;RBL2;RBL2，对于这种位点应该是删除还是做什么处理?

[回复](#)



彭振说:

2019年12月5日 上午10:46

为什么有的位点对应好几个gene_symbol? 比如cg00000029对应的gene_symbol为RBL2;RBL2;RBL2, 对于这类位点应该做什么处理呢?

回复



daizao 说:

2019年12月5日 下午8:55

我取的第一个, 因为有的甲基化位点在这个基因的附近或者在基因上, 你可以用promoter区域过滤一下, 然后再进行合并。

回复

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)

十年的生信 | **About**

苏公网安备 32011302320554号 自豪地采用WordPress。