# RNA-seq.R

Louie

2019-09-26

```
countdata <- read.table("WT_osdrm2_matrix.out", header=TRUE, row.names=1)
colnames(countdata) <- c("WT_1","WT_2","DRM2_1","DRN2_2")
condition <- factor(c("WT","WT","DRM2","DRM2"))
library(DESeq2)
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter,
##     Find, get, grep, grepl, intersect, is.unsorted, lapply, Map,
##     mapply, match, mget, order, paste, pmax, pmax.int, pmin,
##     pmin.int, Position, rank, rbind, Reduce, rownames, sapply,
##     setdiff, sort, table, tapply, union, unique, unsplit, which,
##     which.max, which.min
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:base':
##
##     expand.grid
```

```
## Loading required package: IRanges
```

```
##
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:grDevices':
##
##     windows
```

```
## Loading required package: GenomicRanges
```

```
## Loading required package: GenomeInfoDb
```

```
## Loading required package: SummarizedExperiment
```

```
## Loading required package: Biobase
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
## Loading required package: DelayedArray
```

```
## Loading required package: matrixStats
```

```
##
## Attaching package: 'matrixStats'
```

```
## The following objects are masked from 'package:Biobase':
##
```

```
##     anyMissing, rowMedians
```

```
## Loading required package: BiocParallel
```

```
##
## Attaching package: 'DelayedArray'
```

```
## The following objects are masked from 'package:matrixStats':
##
##     colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges
```

```
## The following objects are masked from 'package:base':
##
##     aperm, apply, rowsum
```

```
coldata <- data.frame(row.names=colnames(countdata), condition)
dds <- DESeqDataSetFromMatrix(countData=countdata, colData=coldata, design=~condition)
dds <- DESeq(dds)
```

```
## estimating size factors
```

```
## estimating dispersions
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

```
## fitting model and testing
```

```
resdata <- results(dds)
table(resdata$padj<0.05) # p<0.05的基因数
```

```
##
## FALSE  TRUE
## 19120  9515
```

```
res_padj <- resdata[order(resdata$padj), ]  ##按照padj列的值排序
names(resdata)[1] <- "Gene" # 将第一列的列名改为Gene
```

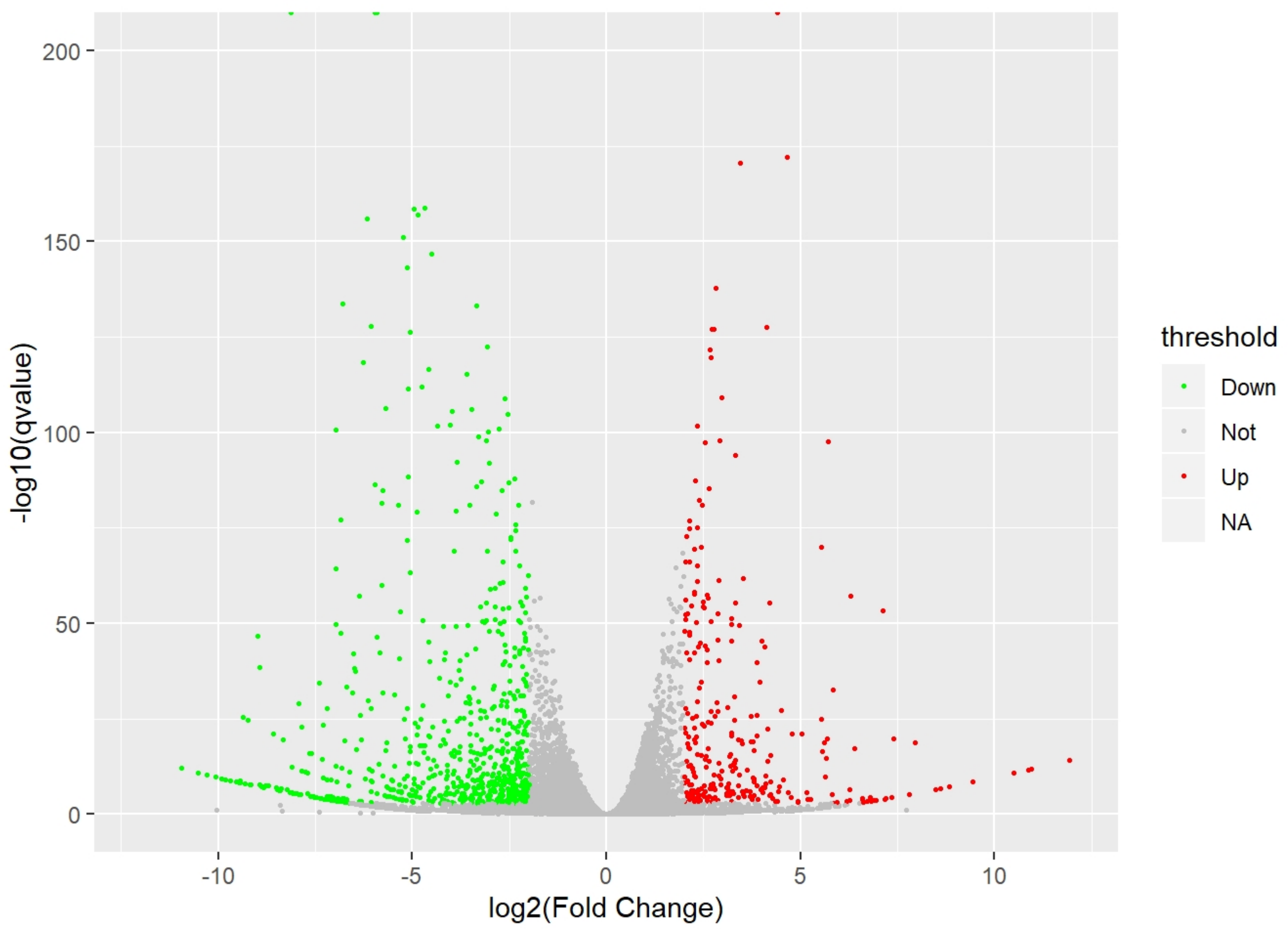
```
write.table(resdata, file="diffexpr_padj_results.csv",sep = "\t",row.names = F)
## 筛选差异基因
subset(resdata,pvalue < 0.001) -> diff ## 先筛选pvalue < 0.01的行
subset(diff,log2FoldChange < -2) -> down ## 筛选出下调的
subset(diff,log2FoldChange > 2) -> up ## 筛选出上调的
print(up)
```

```
## log2 fold change (MLE): condition WT vs DRM2
## Wald test p-value: condition WT vs DRM2
## DataFrame with 329 rows and 6 columns
##                              Gene   log2FoldChange             lfcSE
##                         <numeric>        <numeric>         <numeric>
## LOC_Os01g01910 155.135446663235 3.21244934669379 0.325355350254599
## LOC_Os01g02370 43.5155536314079 2.01626988614896 0.506933562373583
## LOC_Os01g02550 29.4412038062435   2.795196212752 0.654853656435659
## LOC_Os01g02560 62.7620161680862 2.21257978136606 0.443263822345301
## LOC_Os01g02570 38.6096342230744 3.25358268983719 0.614150096065336
## ...                          ...              ...               ...
## LOC_Os09g28500 84.6399898716598 2.77921889238728 0.398139714936849
## LOC_Os09g31130 2461.05166245426 2.33051894547913 0.124332828976063
## LOC_Os09g31514  888.02991913481  2.4869837214939 0.155790690732708
## LOC_Os09g33850 251.259628604867  2.5132266627841  0.23973744584058
## LOC_Os09g35940 93.8088539951849 3.28819794760727 0.408503203301435
##                              stat               pvalue                 padj
##                         <numeric>            <numeric>            <numeric>
## LOC_Os01g01910 9.87366380844812 5.41498102948654e-23 2.76889253177405e-21
## LOC_Os01g02370 3.97738487999948 6.96773368857235e-05 0.000430931002531899
## LOC_Os01g02550 4.26842880891303 1.96854631999382e-05 0.000138567659471541
## LOC_Os01g02560  4.9915640975601 5.98923002986468e-07 5.73392182899281e-06
## LOC_Os01g02570 5.29769955371147 1.17270714430967e-07 1.26242364952283e-06
## ...                          ...                  ...                  ...
## LOC_Os09g28500 6.98051158455292  2.9410710755249e-12 5.97287732252875e-11
## LOC_Os09g31130 18.7441962406229 2.15909498197209e-78 7.72821060109636e-76
## LOC_Os09g31514 15.9636221509592 2.29044022731366e-57 4.89453402306915e-55
## LOC_Os09g33850 10.4832461778013 1.03141467453491e-25 6.24409285524461e-24
## LOC_Os09g35940 8.04938106000825 8.32137826238537e-16 2.43393939268034e-14
```

```
rm(list=ls())

## 加载DESeq2中生成的resdata文件
resdata <- read.csv('diffexpr_padj_results.csv',header = T , sep = "\t")
threshold <- as.factor(ifelse(resdata$padj < 0.001 & abs(resdata$log2FoldChange) >= 2 ,ifelse(resdata$log2FoldChange >= 2 ,'Up','Down'),'Not'))
library('ggplot2')
ggplot(resdata,aes(x=log2FoldChange,y=-log10(padj),colour=threshold)) +
  xlab("log2(Fold Change)")+ylab("-log10(qvalue)") +
  geom_point(size = 0.5,alpha=1) +
  ylim(0,200) + xlim(-12,12) +
  scale_color_manual(values=c("green","grey", "red"))
```

```
## Warning: Removed 27359 rows containing missing values (geom_point).
```


-log10(qvalue)
log2(Fold Change)
threshold
Down
Not
Up
NA