HW3 - Stefan zdraljevic

This dataset was obtained from the GEO database - Reference series **GSE21784**.

The data consists of *C.elegans* expression data at three timepoints during the lifecycle (L4, adult day 6, adult day 15). Each timepoint is replicated in triplicate.

Comparisons between Benjamani and Hochberg will be made using L4 vs day 15 data

```
#  source("http://bioconductor.org/biocLite.R")
# biocLite("affy")
#biocLite("impute")
library(affy)
library(SAM)
library(multtest)
# biocLite("siggenes")
# biocLite("multtest")
library(siggenes)
library(samr)
library(impute)
library(dplyr)

setwd("~/Northwestern/Courses/Stats_for_Bioinformatics/Homework/HW2/GSE21784_RAW/")
```

Read in .CEL files and rename

```
setwd("~/Northwestern/Courses/Stats_for_Bioinformatics/Homework/HW2/GSE21784_RAW/")
expd <- ReadAffy()
enorm <- rma(expd)
```

```
## Warning: replacing previous import by 'utils::head' when loading 'celeganscdf'
## Warning: replacing previous import by 'utils::tail' when loading 'celeganscdf'
```

```
##
```

```
## Background correcting
## Normalizing
## Calculating Expression
```

```
write.exprs(enorm,"wormGrowth.txt")

exnorm <- read.table("wormGrowth.txt",sep="\t",skip=1)
exnorm1 <- exnorm
colnames(exnorm1) <- c("gene","L4_1","L4_2","L4_3", "Day6_1", "Day6_2","Day6_3","Day15_1","
Day15_2","Day15_3")
```

```
h3.cl <-c(rep(0,3),rep(1,3))

h3df <- exnorm1[,c(1,2:4,7:10)]

###calcuate the teststat
teststat <- mt.teststat(h3df[,2:7],test="t.equalvar",h3.cl)   ###default is t test.



rawp0<-2*(1-pt(abs(teststat),df=4))
procs <- c("Bonferroni", "Holm", "Hochberg", "SidakSS", "SidakSD","BH", "BY")
res <- mt.rawp2adjp(rawp0, procs)
```

```
####sort the adjusted p-value to original order
adjp <- data.frame(res$adjp[order(res$index), ])

###round the adjusted p-value to the second decimal places
adjp <- round(adjp, 12)
```

Filter p values based on adjusted p values for Benjamini and Hochberg procedure

```
temp <- filter(adjp, BH < .05)

# bnh <- data.frame(hoch = (adjp$BH))
# bnh <- arrange(bnh, hoch)
# bnh$index <- c(1:nrow(bnh))
# bnh$cutoff <- (bnh$index/nrow(bnh))*.05
# bnh$fin <- ifelse(bnh$hoch < bnh$cutoff,1,0)
# diffexp <- filter(bnh, fin == 1)

nrow(temp)
```

```
## [1] 4657
```

4657 differentially expressed genes using Benjamini and Hochberg @ FDF = .05

```
exmat <- exnorm[,c(2:4,7:10)]
h3.cl <-c(rep(1,3),rep(2,3))

exmat1 <- list(x=as.matrix(exmat), y = h3.cl, logged2=T)

samr.obj <- samr::samr(exmat1,  resp.type="Two class unpaired")
```

```
## perm= 1
## perm= 2
## perm= 3
## perm= 4
## perm= 5
## perm= 6
## perm= 7
## perm= 8
## perm= 9
## perm= 10
## perm= 11
## perm= 12
## perm= 13
## perm= 14
## perm= 15
## perm= 16
## perm= 17
## perm= 18
## perm= 19
## perm= 20
## perm= 21
## perm= 22
## perm= 23
## perm= 24
```

```
## perm= 25
## perm= 26
## perm= 27
## perm= 28
## perm= 29
## perm= 30
## perm= 31
## perm= 32
## perm= 33
## perm= 34
## perm= 35
## perm= 36
## perm= 37
## perm= 38
## perm= 39
## perm= 40
## perm= 41
## perm= 42
## perm= 43
## perm= 44
## perm= 45
## perm= 46
## perm= 47
## perm= 48
## perm= 49
## perm= 50
## perm= 51
## perm= 52
## perm= 53
## perm= 54
## perm= 55
## perm= 56
## perm= 57
## perm= 58
## perm= 59
## perm= 60
## perm= 61
## perm= 62
## perm= 63
## perm= 64
## perm= 65
## perm= 66
```

```
## perm= 67
## perm= 68
## perm= 69
## perm= 70
## perm= 71
## perm= 72
## perm= 73
## perm= 74
## perm= 75
## perm= 76
## perm= 77
## perm= 78
## perm= 79
## perm= 80
## perm= 81
## perm= 82
## perm= 83
## perm= 84
## perm= 85
## perm= 86
## perm= 87
## perm= 88
## perm= 89
## perm= 90
## perm= 91
## perm= 92
## perm= 93
## perm= 94
## perm= 95
## perm= 96
## perm= 97
## perm= 98
## perm= 99
## perm= 100
```

```
delta.table <- samr::samr.compute.delta.table(samr.obj, min.foldchange=0.1,nvals=100)
```

```
##
## Computing delta table
## 1
## 2
```

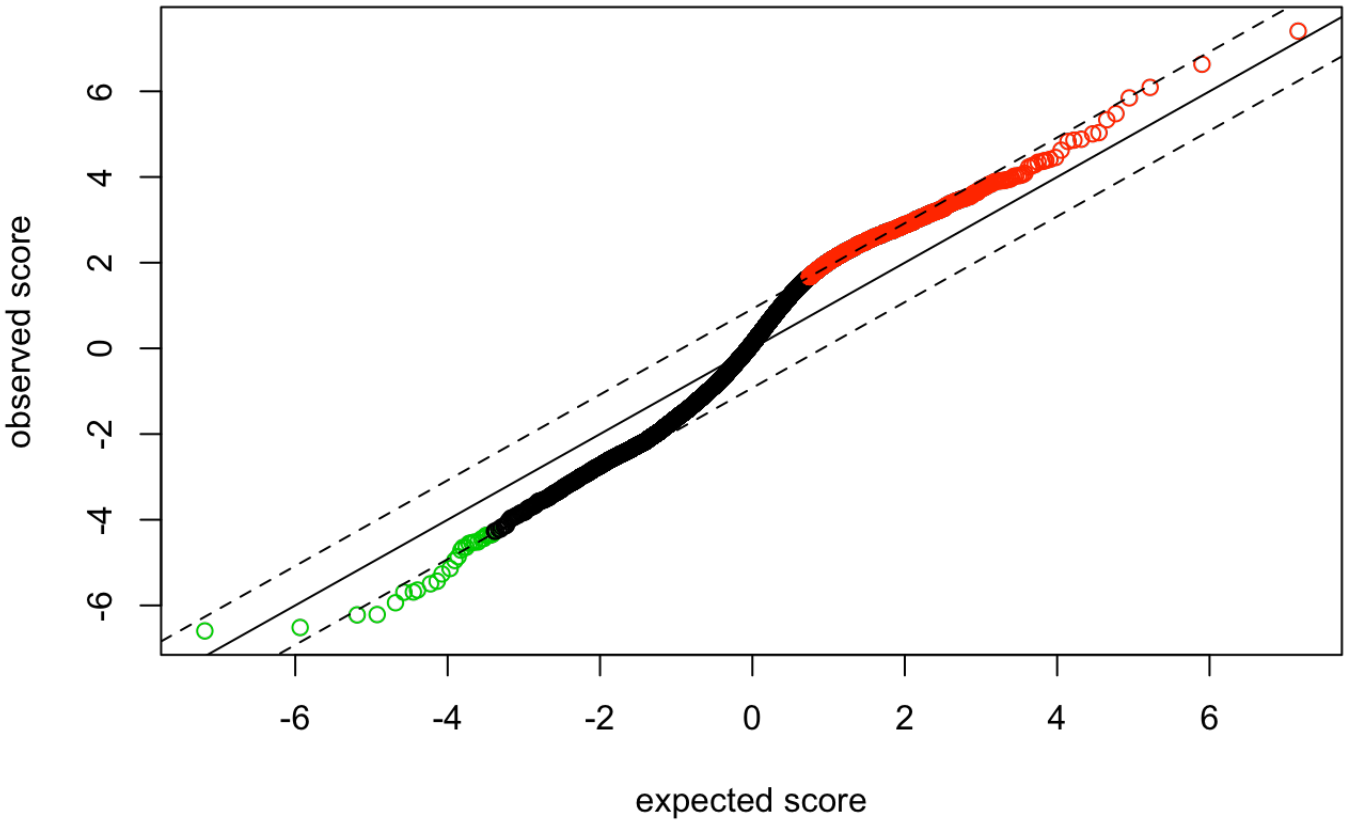
```
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
## 31
## 32
## 33
## 34
## 35
## 36
## 37
## 38
## 39
## 40
## 41
## 42
## 43
## 44
```

```
## 45
## 46
## 47
## 48
## 49
## 50
## 51
## 52
## 53
## 54
## 55
## 56
## 57
## 58
## 59
## 60
## 61
## 62
## 63
## 64
## 65
## 66
## 67
## 68
## 69
## 70
## 71
## 72
## 73
## 74
## 75
## 76
## 77
## 78
## 79
## 80
## 81
## 82
## 83
## 84
## 85
## 86
```

```
## 87
## 88
## 89
## 90
## 91
## 92
## 93
## 94
## 95
## 96
## 97
## 98
## 99
## 100
```

```
siggenes.table <- samr::samr.compute.siggenes.table(samr.obj, del=0.9212618985, exmat1, del
ta.table)

samr::samr.plot(samr.obj,del=0.9212618985)
```




```
nrow(siggenes.table[[1]])+nrow(siggenes.table[[2]]) # genes up + genes down
```

```
## [1] 5069
```

differentiall expressed genes using SAM method (above)

The difference between the two methods is a result of the way significance is calculated. Permutation in the case of SAM and ranked p values in the case of Benjamani and Hochberg. SAM uses non-parametric statistics to identify differentially expressed genes (DEGs), while the BH method assumes independence of genes (which is not always true in biology) and a parametric distribution.