BIOS 7659 Homework 3

Tim Vigers

09 October 2020

# 1. T Statistics

Read in the data:

array <- read.table("./hw3data/hw3arraydata.txt")  
gene\_names <- read.table("./hw3data/hw3genenames.txt",  
 blank.lines.skip = FALSE)

## a) Fold Change

For each gene (row), find the mean expression among controls and among the knock out group. Then calculate fold change using :

fc <- apply(array,1,function(x){  
 control = mean(as.numeric(x[1:8]))  
 knockout = mean(as.numeric(x[9:16]))  
 return(control-knockout)  
})  
fc\_results <- as.data.frame(cbind(gene\_names,fc))  
colnames(fc\_results) <- c("Gene","log2FC")  
kable(head(fc\_results[order(abs(fc\_results$log2FC),decreasing = T),],10),  
 caption = "Top 10 genes with largest absolute value of fold change",  
 row.names = F)

Top 10 genes with largest absolute value of fold change

|  |  |
| --- | --- |
| Gene | log2FC |
| ApoAI,lipid-Img | 4.749247 |
| EST,HighlysimilartoA | 4.572826 |
| CATECHOLO-METHYLTRAN | 2.772249 |
| EST,WeaklysimilartoC | 1.540431 |
| ESTs,Highlysimilarto | 1.514718 |
| est | 1.466135 |
| similartoyeaststerol | 1.432454 |
| ApoCIII,lipid-Img | 1.398874 |
| psoriasis-associated | 1.256714 |
| Cy3RT | -1.193286 |

First, create the design matrix for limma:

design <- matrix(ncol = 2,nrow = ncol(array))  
colnames(design) <- c("Control","Knockout")  
rownames(design) <- colnames(array)  
design[,1] <- rep(1,nrow(design))  
design[,2] <- ifelse(grepl("k",rownames(design)),1,0)  
kable(design)

|  |  |  |
| --- | --- | --- |
|  | Control | Knockout |
| c1 | 1 | 0 |
| c2 | 1 | 0 |
| c3 | 1 | 0 |
| c4 | 1 | 0 |
| c5 | 1 | 0 |
| c6 | 1 | 0 |
| c7 | 1 | 0 |
| c8 | 1 | 0 |
| k1 | 1 | 1 |
| k2 | 1 | 1 |
| k3 | 1 | 1 |
| k4 | 1 | 1 |
| k5 | 1 | 1 |
| k6 | 1 | 1 |
| k7 | 1 | 1 |
| k8 | 1 | 1 |

Fit the model with limma:

fit <- lmFit(array, design)