```
#!/usr/bin/env Rscript
#Usage: Rscript exactTest_edgeR.r rawGeneCountsFile startColumn endColumn
#Usage Ex: Rscript exactTest_edgeR.r daphnia_rawGeneCounts_htseq.csv 1 6
#R script to perform exact test DE analysis of raw gene counts using edgeR

#Load the edgeR library
library("edgeR")

#Retrieve inputs to the script
args = commandArgs(trailingOnly=TRUE)
#Test if there is one input argument
if (length(args)!=3) {
  stop("One file name and a range of columns must be supplied.n", call.=FALSE)
}

#Import gene count data
countsTable <- read.csv(file=args[1], row.names="gene")[ ,args[2]:args[3]]
head(countsTable)

#Add grouping factor
group <- factor(c(rep("ctrl",3),rep("treat",3)))
#Create DGE list object
list <- DGEList(counts=countsTable,group=group)

#Plot the library sizes before normalization
jpeg("plotBarsBefore.jpg")
barplot(list$samples$lib.size*1e-6, names=1:6, ylab="Library size (millions)")
dev.off()

#Draw a MDS plot to show the relative similarities of the samples
jpeg("plotMDSBefore.jpg")
plotMDS(list, col=rep(1:3, each=3))
dev.off()

#Draw a heatmap of individual RNA-seq samples
jpeg("plotHeatMapBefore.jpg")
logcpm <- cpm(list, log=TRUE)
heatmap(logcpm)
dev.off()

#Filter raw gene counts by expression levels
keep <- filterByExpr(list)
table(keep)
```

```
list <- list[keep, , keep.lib.sizes=FALSE]
#Calculate normalized factors
list <- calcNormFactors(list)
#Write normalized counts to file
normList <- cpm(list, normalized.lib.sizes=TRUE)
write.table(normList, file="stats_normalizedCounts.csv", sep=",", row.names=TRUE)
#View normalization factors
list$samples
dim(list)

#Plot the library sizes after normalization
jpeg("plotBarsAfter.jpg")
barplot(list$samples$lib.size*1e-6, names=1:6, ylab="Library size (millions)")
dev.off()

#Draw a MDS plot to show the relative similarities of the samples
jpeg("plotMDSAfter.jpg")
plotMDS(list, col=rep(1:3, each=3))
dev.off()

#Draw a heatmap of individual RNA-seq samples
jpeg("plotHeatMapAfter.jpg")
logcpm <- cpm(list, log=TRUE)
heatmap(logcpm)
dev.off()

#Produce a matrix of pseudo-counts
list <- estimateDisp(list)
list$common.dispersion
#View dispersion estimates and biological coefficient of variation
jpeg("plotBCV.jpg")
plotBCV(list)
dev.off()

#Perform an exact test for treat vs ctrl
tested <- exactTest(list, pair=c("ctrl", "treat"))
topTags(tested)
#Create results table of DE genes
resultsTbl <- topTags(tested, n=nrow(tested$table))$table
#Output resulting table
write.table(resultsTbl, file="stats_exactTest.csv", sep=",", row.names=TRUE)

#Look at the counts per million in individual samples for the top genes
```

```
o <- order(tested$table$PValue)
cpm(list)[o[1:10],]
#View the total number of differentially expressed genes at 5% FDR
summary(decideTests(tested))

#Make a MD plot of logFC against logcpm
jpeg("plotMDResults.jpg")
plotMD(tested)
abline(h=c(-1, 1), col="blue")
dev.off()

#Make a MA plot of the libraries of count data
jpeg("plotMAResults.jpg")
plotSmear(tested)
dev.off()
```