```
#!/usr/bin/env Rscript
#Usage: Rscript glmQLF_edgeR.r rawGeneCountsFile experimentalDesign
#Usage Ex: Rscript glmQLF_edgeR.r daphnia_rawGeneCounts_htseq.csv
daphnia_experimentalDesign.csv
#R script to perform QL F-tests using generalized linear models in edgeR

#Install edgeR and statmod, this should only need to be done once
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install("edgeR")
install.packages("statmod")

#Load the edgeR and statmod libraries
library("edgeR")
library("statmod")

#Retrieve inputs to the script
args = commandArgs(trailingOnly=TRUE)

#Import gene count data and view the first few rows
countsTable <- read.csv(file=args[1], row.names="gene")
head(countsTable)
#Import grouping factor
targets <- read.csv(file=args[2], row.names="sample")

#Setup and view the grouping factors
group <- factor(paste(targets$treatment,targets$genotype,sep="."))
cbind(targets,Group=group)
#Create a DGE list object
list <- DGEList(counts=countsTable,group=group)
colnames(list) <- targets$sample

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

#Retain genes only if it is expressed at a minimum level
keep <- filterByExpr(list)
list <- list[keep, , keep.lib.sizes=FALSE]
#View a summary of the normalized counts
summary(keep)

#Use TMM normalization to eliminate composition biases
```

```
list <- calcNormFactors(list)
#Write normalized counts to file
normList <- cpm(list, normalized.lib.sizes=TRUE)
write.table(normList, file="glmQLF_normalizedCounts.csv", sep=",", row.names=TRUE)

#Verify TMM normalization using a MD plot
jpeg("glmQLF_plotNormalizedMD.jpg")
plotMD(cpm(list, log=TRUE), column=1)
abline(h=0, col="red", lty=2, lwd=2)
dev.off()

#Use a MDS plot to visualizes the differences between samples
#Set the point shapes and colors
points <- c(0,1,15,16)
colors <- rep(c("blue", "red"), 2)
#Write plot with legend to file
jpeg("glmQLF_plotMDS.jpg")
plotMDS(list, col=colors[group], pch=points[group])
legend("topleft", legend=levels(group), pch=points, col=colors, ncol=2)
dev.off()

#Setup the design matrix
design <- model.matrix(~ 0 + group)
colnames(design) <- levels(group)
#View the design matrix
design

#Generate the NB dispersion estimates
list <- estimateDisp(list, design, robust=TRUE)
#View the common dispersion
list$common.dispersion
#Visualize the dispersion estimates with a BCV plot
jpeg("glmQLF_plotBCV.jpg")
plotBCV(list)
dev.off()

#Estimate and view the QL dispersions
fit <- glmQLFit(list, design, robust=TRUE)
head(fit$coefficients)
#Plot to the QL dispersions and write to file
jpeg("glmQLF_plotQLDisp.jpg")
plotQLDisp(fit)
dev.off()
```

```
#Design a contrast to test overall effect of the first factor
con.UVvsVIS <- makeContrasts(UVvsVIS = (UV.Y023 + UV.Y05)/2
  - (VIS.Y023 + VIS.Y05)/2,
  levels=design)

#Look at genes expressed across all UV groups
test.anov.one <- glmQLFTest(fit, contrast=con.UVvsVIS)
summary(decideTests(test.anov.one))
#Write plot to file
jpeg("glmQLF_UVvsVIS_plotMD.jpg")
plotMD(test.anov.one)
abline(h=c(-1, 1), col="blue")
dev.off()
#Write top tags table of DE genes to file
tagsTblANOVA.one <- topTags(test.anov.one, n=nrow(test.anov.one$table))$table
tagsTblANOVA.one.keep <- tagsTblANOVA.one$FDR <= 0.05
tagsTblANOVA.one.out <- tagsTblANOVA.one[tagsTblANOVA.one.keep,]
write.table(tagsTblANOVA.one.out, file="glmQLF_UVvsVIS_topTags.csv", sep=",",
row.names=TRUE)

#Look at genes with significant expression across all UV groups
treat.anov.one <- glmTreat(fit, contrast=con.UVvsVIS, lfc=log2(1.2))
summary(decideTests(treat.anov.one))
#Write plot to file
jpeg("glmQLF_UVvsVIS_plotMD_filtered.jpg")
plotMD(treat.anov.one)
abline(h=c(-1, 1), col="blue")
dev.off()
#Write tags table of DE genes to file
tagsTblANOVA.one.filtered <- topTags(treat.anov.one, n=nrow(treat.anov.one$table))$table
tagsTblANOVA.one.filtered.keep <- tagsTblANOVA.one.filtered$FDR <= 0.05
tagsTblANOVA.one.filtered.out <- tagsTblANOVA.one.filtered[tagsTblANOVA.one.filtered.keep,]
write.table(tagsTblANOVA.one.filtered.out, file="glmQLF_UVvsVIS_topTags_filtered.csv",
sep=",", row.names=TRUE)

#Design a contrast to test overall effect of the first factor
con.TvsN <- makeContrasts(TvsN = (UV.Y05 + VIS.Y05)/2
  - (UV.Y023 + VIS.Y023)/2,
  levels=design)

#Look at genes expressed across all tolerant groups
test.anov.two <- glmQLFTest(fit, contrast=con.TvsN)
summary(decideTests(test.anov.two))
#Write plot to file
```

```
jpeg("glmQLF_TvsN_plotMD.jpg")
plotMD(test.anov.two)
abline(h=c(-1, 1), col="blue")
dev.off()
#Write tags table of DE genes to file
tagsTblANOVA.two <- topTags(test.anov.two, n=nrow(test.anov.two$table))$table
tagsTblANOVA.two.keep <- tagsTblANOVA.two$FDR <= 0.05
tagsTblANOVA.two.out <- tagsTblANOVA.two[tagsTblANOVA.two.keep,]
write.table(tagsTblANOVA.two.out, file="glmQLF_TvsN_topTags.csv", sep=",",
row.names=TRUE)

#Look at genes with significant expression across all tolerant groups
treat.anov.two <- glmTreat(fit, contrast=con.TvsN, lfc=log2(1.2))
summary(decideTests(treat.anov.two))
#Write plot to file
jpeg("glmQLF_TvsN_plotMD_filtered.jpg")
plotMD(treat.anov.two)
abline(h=c(-1, 1), col="blue")
dev.off()
#Write tags table of DE genes to file
tagsTblANOVA.two.filtered <- topTags(treat.anov.two, n=nrow(treat.anov.two$table))$table
tagsTblANOVA.two.filtered.keep <- tagsTblANOVA.two.filtered$FDR <= 0.05
tagsTblANOVA.two.filtered.out <- tagsTblANOVA.two.filtered[tagsTblANOVA.two.filtered.keep,]
write.table(tagsTblANOVA.two.filtered.out, file="glmQLF_TvsN_topTags_filtered.csv", sep=",",
row.names=TRUE)

#Test whether there is an interaction effect
con.Interaction <- makeContrasts(Interaction = ((UV.Y023 + UV.Y05)/2
  - (VIS.Y023 + VIS.Y05)/2)
  - ((UV.Y05 + VIS.Y05)/2
  - (UV.Y023 + VIS.Y023)/2),
  levels=design)

#Look at genes expressed across all tolerant groups
test.anov.Interaction <- glmQLFTest(fit, contrast=con.Interaction)
summary(decideTests(test.anov.Interaction))
#Write plot to file
jpeg("glmQLF_Interaction_plotMD.jpg")
plotMD(test.anov.Interaction)
abline(h=c(-1, 1), col="blue")
dev.off()
#Write tags table of DE genes to file
tagsTblANOVAInteraction <- topTags(test.anov.Interaction,
n=nrow(test.anov.Interaction$table))$table
```

```
tagsTblANOVAInteraction.keep <- tagsTblANOVAInteraction$FDR <= 0.05
tagsTblANOVAInteraction.out <- tagsTblANOVAInteraction[tagsTblANOVAInteraction.keep,]
write.table(tagsTblANOVAInteraction.out, file="glmQLF_Interaction_topTags.csv", sep=",",
row.names=TRUE)

#Look at genes with significant expression across all tolerant groups
treat.anov.Interaction <- glmTreat(fit, contrast=con.Interaction, lfc=log2(1.2))
summary(decideTests(treat.anov.Interaction))
#Write plot to file
jpeg("glmQLF_Interaction_plotMD_filtered.jpg")
plotMD(treat.anov.Interaction)
abline(h=c(-1, 1), col="blue")
dev.off()
#Write tags table of DE genes to file
tagsTblANOVAInteraction.filtered <- topTags(treat.anov.Interaction,
n=nrow(treat.anov.Interaction$table))$table
tagsTblANOVAInteraction.filtered.keep <- tagsTblANOVAInteraction.filtered$FDR <= 0.05
tagsTblANOVAInteraction.filtered.out <-
tagsTblANOVAInteraction.filtered[tagsTblANOVAInteraction.filtered.keep,]
write.table(tagsTblANOVAInteraction.filtered.out,
file="glmQLF_Interaction_topTags_filtered.csv", sep=",", row.names=TRUE)
```