Cystic Fibrosis: DEG Analysis

Zahra Ahmadi Fakhr

Report

First, I obtained samples for Cystic Fibrosis from the GEO database. Then, I performed some data preprocessing by removing rows with a sum of counts less than 10. After that, I created the DESeq2 object and ran the function. Finally, I filtered the end results to include only genes that were **upregulated** or **downregulated** by at least 1-fold (2 times) and had a p-adjusted value of less than 0.05. This was done to identify genes that were differentially expressed in the disease group compared to the control group. The process was carried out to compare the disease group versus the control group. Eventually, I found 855 differentially expressed genes between the healthy and disease groups. These findings can be used for further analyses to generate biologically meaningful insights into this rare genetic disorder, Cystic Fibrosis.

For this analysis, I used the DESeq2 and ggplot libraries:

```
suppressMessages(library(DESeq2))
suppressMessages(library(ggplot2))
```

Read the counts matrix

```
counts_data <- read.csv('counts_data.csv', row.names = 1)

status <- factor(c("Control", "Control", "Control",
                   "Control", "Disease", "Disease",
                   "Disease", "Disease"))

coldata <- data.frame(row.names = colnames(counts_data), status)

# Create a DESeq2 dataset
dds <- DESeqDataSetFromMatrix(countData = counts_data,
                              colData = coldata,
                              design = ~ status)
```

Pre-processing

```
filt_dds <- rowSums(counts(dds)) >= 10
dds <- dds[filt_dds, ]
dds <- DESeq(dds)
```

```
## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates
```

```
## fitting model and testing
r <- results(dds)
```

Filtering

Here, I filtered the identified genes based on log2foldchange and p-adjusted values:

```
r <- subset(r, abs(log2FoldChange) > 1 & padj < 0.05)
r <- as.data.frame(r)
```

Top results

```
top_ones <- r[order(r$padj)[1:3], ]
top_ones
```

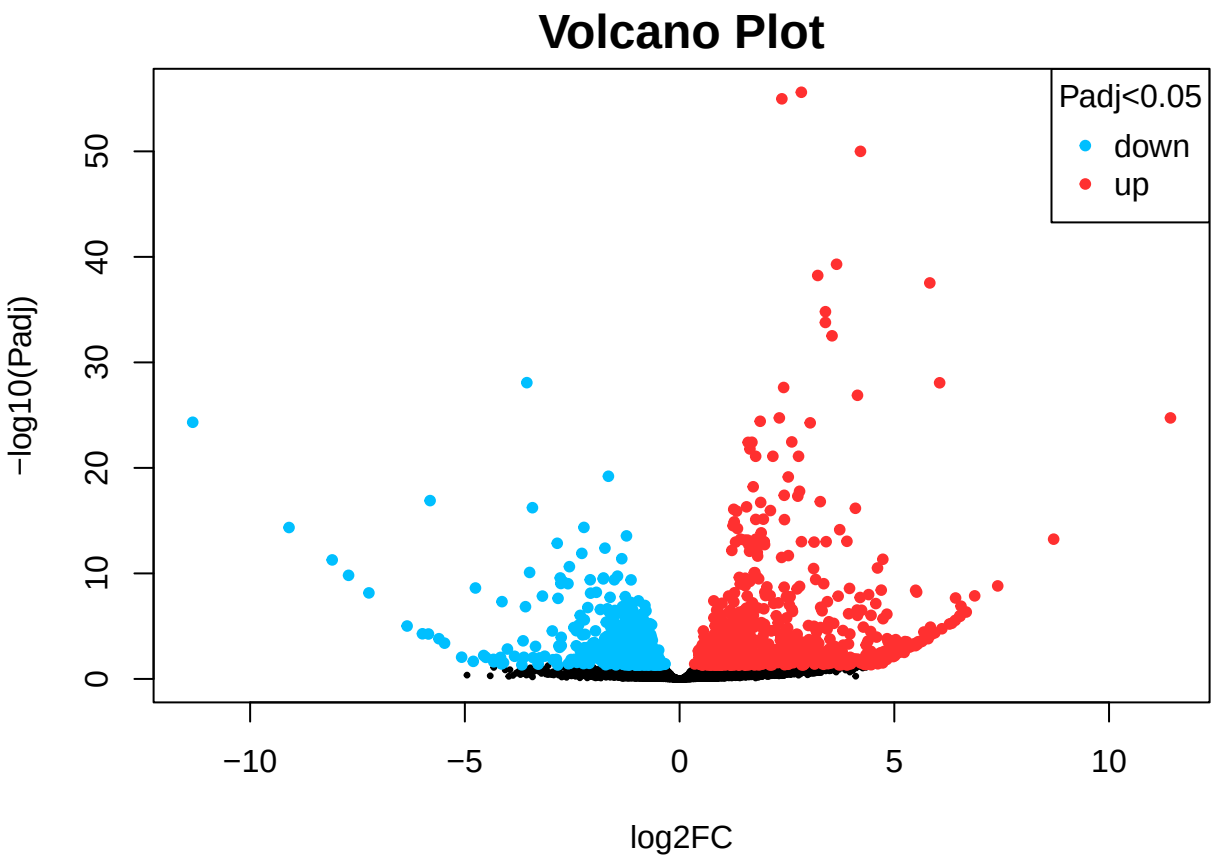
```
##                   baseMean log2FoldChange     lfcSE     stat       pvalue
## ENSG00000120885 2076.2218       2.835158 0.1728799 16.39958 1.925701e-60
## ENSG00000159403  384.2563       2.381344 0.1463656 16.26983 1.616029e-59
## ENSG00000175274  164.3117       4.211767 0.2712685 15.52619 2.306796e-54
##                         padj
## ENSG00000120885 2.508225e-56
## ENSG00000159403 1.052439e-55
## ENSG00000175274 1.001534e-50
```

Plots

To visualize the results, I created volcano and MD (mean-difference) plots using the ggplot2 package. Here, you can see my findings based on these two plots. In the volcano plot, you can observe a relatively higher number of upregulated genes in these samples, compared to downregulated genes.

Volcano plot

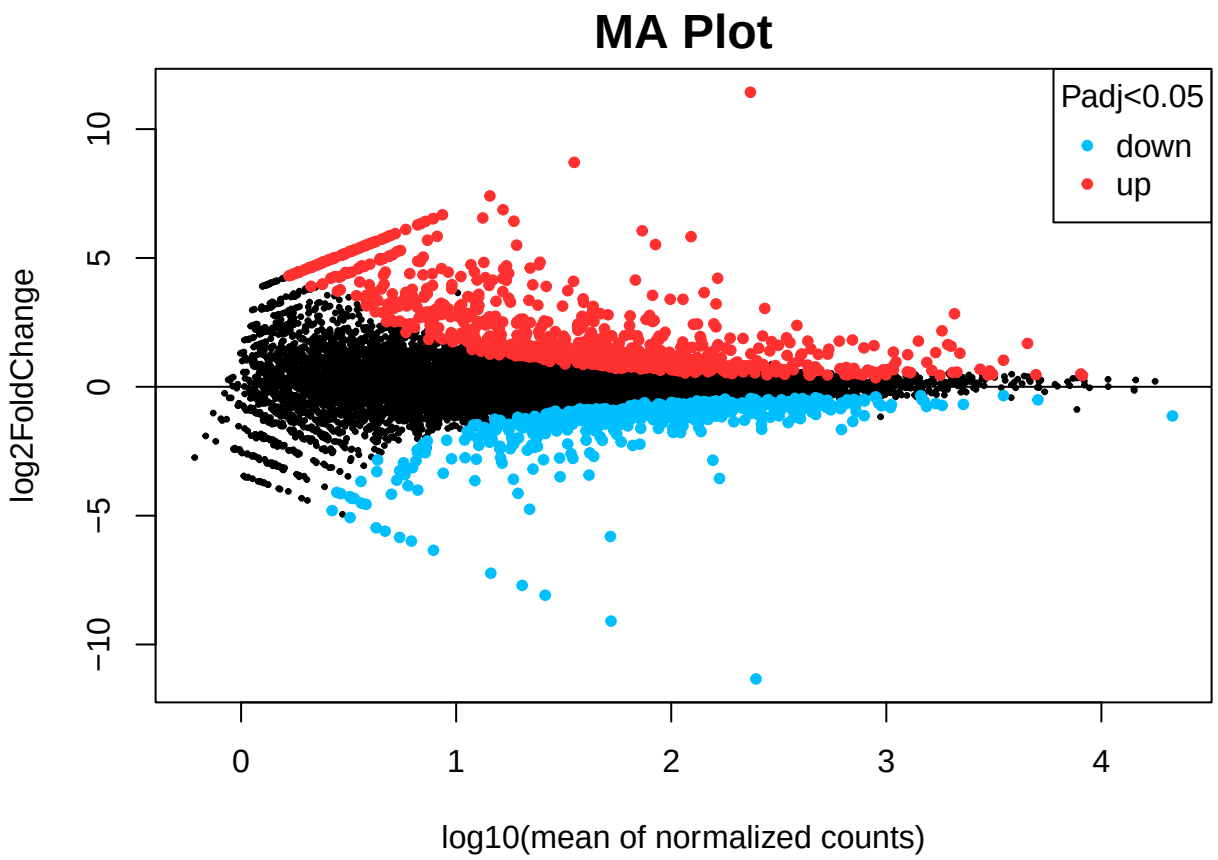
```
# Redefine the results for plotting
r <- results(dds)

# Define a color palette
old.pal <- palette(c("#00BFFF", "#FF3030")) # low-hi colors

# Volcano plot
par(mar=c(4,4,2,1), cex.main=1.5)
plot(r$log2FoldChange, -log10(r$padj), main="Volcano Plot",
     xlab="log2FC", ylab="-log10(Padj)", pch=20, cex=0.5)
with(subset(r, padj<0.05 & abs(log2FoldChange) >= 0),
     points(log2FoldChange, -log10(padj), pch=20, col=(sign(log2FoldChange) + 3)/2, cex=1))
legend("topright", title=paste("Padj<", 0.05, sep=""), legend=c("down", "up"), pch=20,col=1:2)
```



In the MD plot, you can see the comparison between the number of upregulated and downregulated genes in contrast to the genes that were not found to be differentially expressed based on the statistical tests conducted by the DESeq2 package.

MA plot

```
par(mar=c(4,4,2,1), cex.main=1.5)
plot(log10(r$baseMean), r$log2FoldChange, main="MA Plot",
     xlab="log10(mean of normalized counts)", ylab="log2FoldChange", pch=20, cex=0.5)
with(subset(r, padj<0.05 & abs(log2FoldChange) >= 0),
     points(log10(baseMean), log2FoldChange, pch=20, col=(sign(log2FoldChange) + 3)/2, cex=1))
legend("topright", title=paste("Padj<", 0.05, sep=""), legend=c("down", "up"), pch=20,col=1:2)
abline(h=0)
```




```
palette(old.pal)
```