**Project:** Differential expression analysis in asthma cases

**Objective:** Identify upregulated and downregulated genes in asthmatic patients.

**Methods:**

The analysis in this project is developed in R studio using the DESeq2 package.

To perform the analysis first load the following packages.

library(DESeq2)

library(edgeR)

library(tidyverse)

library(reshape2)

library(dendextend)

library(magrittr)

Import the data in R studio to perform the differential expression analysis.

data\_count <- read.csv("https://github.com/HackBio-Internship/public\_datasets/blob/main/transcriptomics/asthma/airway\_scaledcounts.csv?raw=true")

metadata <- read.csv("https://raw.githubusercontent.com/HackBio-Internship/public\_datasets/main/transcriptomics/asthma/airway\_metadata.csv")

Run the DESeq2 package, stating which are the counts (countData = data\_count), metadata (colData = metadata) and the experimental desing (design = ~ celltype + dex).

dds <- DESeqDataSetFromMatrix(countData = data\_count,

colData = metadata,

design = ~ celltype + dex, tidy=T)

dds@colData$dex <- relevel(dds@colData$dex, ref = "control")

dds <- DESeq(dds)

Generate the results table, order it based on pvalue.

result <- results(dds, tidy = TRUE)

result[order(result$pvalue),] *# Order table based on pvalue*

Add a column with gene names

result$gene <- result$row

result <- as\_tibble(result) *# Change results to a tibble format to work with dplyr*

Add column to show which genes are significant and filter result based on pvalue

result <- mutate(result, sig=padj<0.01)

result <- result %>% mutate(sig=padj<0.01)

result %>%

filter(padj<0.01)

Plot the results

plot(result$log2FoldChange,-log10(result$padj),ylim =c(0,20), xlim =c(-5,5),col='black',pch=20)

*#Generate subsets of the data*

significant<-subset(result, padj< 0.01)

Up<-subset(result,log2FoldChange > 1.5 & padj < 0.05)

Down<-subset(result,log2FoldChange < -1.5 & padj < 0.05)

NotSignificant<-subset(result,log2FoldChange >-1.5 & log2FoldChange <1.5)

*#Change the color of the points in the plot*

points(low\_Pvalue$log2FoldChange,-log10(low\_Pvalue$padj),col='grey',pch=20)

points(Up$log2FoldChange,-log10(Up$padj),col='red',pch=20, labels=row)

points(Down$log2FoldChange,-log10(Down$padj),col='blue',pch=20)

Add labels to the most significant Up and downregulated genes

with(subset(result,padj<0.0001 & abs(log2FoldChange)>3),text(x=log2FoldChange, y= -log10(padj), col="red", labels=gene,cex=0.5 ))

Visualize the counts using a dendrogram and perform a PCA analysis.

*#Visualizing normalizing counts*

vsd <- vst(dds, blind =FALSE)

mat <- assay(vsd)

head(as.data.frame(mat))

dend = t(mat) %>%

scale %>%

dist %>%

hclust %>%

as.dendrogram

l = metadata$dex[metadata$id %in% labels(dend)]

dend %>%

set("labels", l) %>%

plot

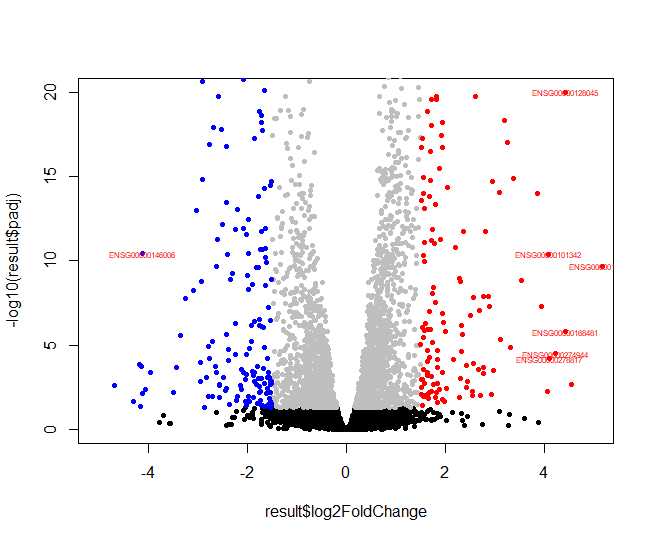
*#PCA*

mat.pca<-prcomp(t(mat)) #perform pca

summary(mat.pca)

**Results**

The following plot shows the up regulated genes in red and the down regulated in blue, along with names of some of the genes.



The dendrogram and PCA analysis are shown below.

