

# Task 6: Statistical Analysis

Dongyuan Wu

2020-06-11

First of all, load data and packages what I need.

```
library(limma)

rawcount <- read.table("featurecounts.txt", header=TRUE)
rownames(rawcount) <- rawcount$GENEID
rawcount <- rawcount[, -1]
rawcount <- subset(rawcount, rowSums(rawcount) > 0) # remove genes with zeros in all samples
phenotype <- read.table("phenotype.txt", header=TRUE)
```

I used limma to do the statistical analysis. Genes with adjusted p-value smaller than 0.05 would be viewed as the differential expressed genes.

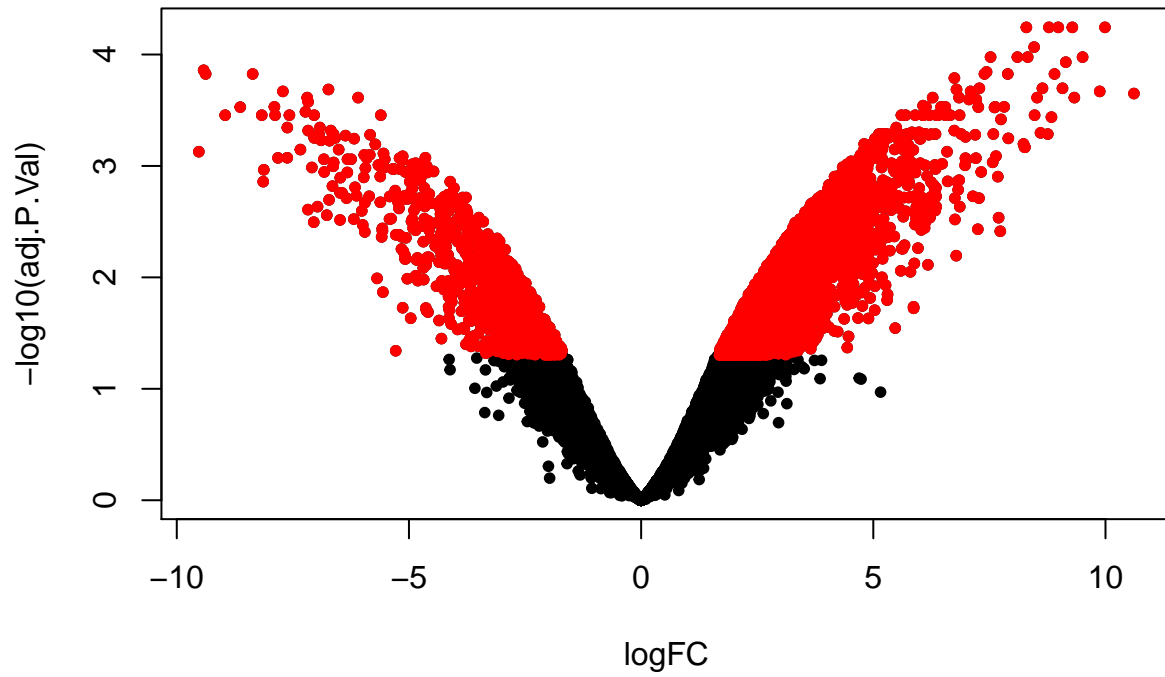
```
normcount <- log2(as.matrix(rawcount) + 1) # make log2 + 1 transformation

fit <- lmFit(normcount, model.matrix(~ phenotype$group))
fit <- eBayes(fit)
toptable <- topTable(fit, number=nrow(normcount))
deresult <- toptable[, c(1, 4, 5)]
deresult$gene <- rownames(deresult)
deresult <- deresult[, c(4, 1:3)]

write.table(deresult, "deresult.txt", quote=FALSE, sep="\t", row.names=FALSE)

with(deresult, plot(logFC, -log10(adj.P.Val), pch=20, main="Volcano plot"))
with(subset(deresult, adj.P.Val < 0.05), points(logFC, -log10(adj.P.Val), pch=20, col="red"))
```

## Volcano plot



The red points in volcano plot represent those differential expressed genes.

```
sum(deresult$adj.P.Val < 0.05)
```

```
## [1] 7094
```

As we can see, there are totally 7094 differential expressed genes detected.