Analysis of Dataset 2 - Expression profiling by High Throughput Sequencing (GSE216609)

2023-12-22

Data Exploration and Preprocessing

```
# Set the CRAN mirror
options(repos = c(CRAN = "https://cran.rstudio.com"))
## Install packages
BiocManager::install("edgeR", update = F)
BiocManager::install("DESeq2", update = F)
BiocManager::install("limma", update = F)
BiocManager::install("tximport", update = F)
BiocManager::install("tximeta", update = F)
BiocManager::install("biomaRt", update = F)

## Load packages
library(edgeR)
library(DESeq2)
library(limma)
library(tximport)
library(tximeta)
library(biomaRt)

## Set working directory
setwd("C:/Users/maxim/OneDrive - UGent/1 Master Bio-informatics/Semester 1/Applied High Throughput Analy
```

Transcript annotation is loaded to be merged with the kallisto results.

```
# Get annotation data
grch38 <-  useEnsembl(biomart="ensembl", dataset="hsapiens_gene_ensembl")

data <- getBM(attributes = c('ensembl_gene_id', 'ensembl_transcript_id', 'external_gene_name', 'entrezg

tx2gene <- dplyr::select(data, ensembl_transcript_id, ensembl_gene_id, external_gene_name, entrezgene_i
tx2gene <- dplyr::rename(tx2gene, TXNAME = ensembl_transcript_id)
tx2gene <- dplyr::rename(tx2gene, GENEID = ensembl_gene_id)
tx2gene <- dplyr::rename(tx2gene, GENENAME = external_gene_name)
tx2gene <- dplyr::rename(tx2gene, ENTREZID = entrezgene_id)

head(tx2gene)
```

```
##            TXNAME          GENEID GENENAME ENTREZID transcript_length
## 1 ENST00000387314 ENSG00000210049    MT-TF       NA                71
```

```
## 2 ENST00000389680 ENSG00000211459  MT-RNR1       NA               954
## 3 ENST00000387342 ENSG00000210077    MT-TV       NA                69
## 4 ENST00000387347 ENSG00000210082  MT-RNR2       NA              1559
## 5 ENST00000386347 ENSG00000209082   MT-TL1       NA                75
## 6 ENST00000361390 ENSG00000198888   MT-ND1     4535               956
```

Loading Kallisto results.

```
# Get file locations
files <- list.files("kallisto_quant", pattern=".tsv")
files <- files[grep("abundance.tsv",files)]
samples <- unlist(strsplit(files,"_"))[c(1:length(files))*2-1]
files <- paste0("Kallisto_quant/", files)
names(files) <- samples

# Load RNAseq data; normalization for transcript length and library size
txi <- tximport(files, type = "kallisto", tx2gene = tx2gene, countsFromAbundance = "lengthScaledTPM")
```

```
## Note: importing `abundance.h5` is typically faster than `abundance.tsv`

## reading in files with read_tsv

## 1 2 3 4 5 6 7
## removing duplicated transcript rows from tx2gene
## summarizing abundance
## summarizing counts
## summarizing length
```

```
head(txi$counts)
```

```
##                 SRR22047906 SRR22047907 SRR22047908 SRR22047909 SRR22047910
## ENSG00000000003 1052.211543  1250.95892   212.80567  880.519668  736.136729
## ENSG00000000005    6.641032     0.00000     0.00000    3.403567    2.541978
## ENSG00000000419  872.767962   541.57261   244.54158  627.076292  513.961841
## ENSG00000000457  456.750580   391.26787    44.17282  372.299427  257.964547
## ENSG00000000460  130.558501   118.58838    21.56224  126.243089   99.377983
## ENSG00000000938   28.003540    20.13517    10.01190   17.755812   34.721916
##                 SRR22047911 SRR22047912
## ENSG00000000003  393.573330  918.848861
## ENSG00000000005    0.000000    1.771098
## ENSG00000000419  261.945095  523.609305
## ENSG00000000457   98.319452  265.386694
## ENSG00000000460   69.471070  116.579223
## ENSG00000000938    4.046165   25.588438
```

```
dim(txi$counts)
```

```
## [1] 62754     7
```

Checking for duplicate rows

```
sum(duplicated(rownames(txi$counts)))
```

```
## [1] 0
```

Merging metatdata with the results. There are no confounders present in the study.

```
sdrf_rnaseq <- read.delim("SraRunTable.txt", sep=',')
print(sdrf_rnaseq[,c("Run","cell_type","disease_state","sex")])
```

```
##           Run             cell_type                    disease_state    sex
## 1 SRR22047906 Cumulus granule cells polycystic ovary syndrome (PCOS) female
## 2 SRR22047907 Cumulus granule cells polycystic ovary syndrome (PCOS) female
## 3 SRR22047908 Cumulus granule cells polycystic ovary syndrome (PCOS) female
## 4 SRR22047909 Cumulus granule cells                          control female
## 5 SRR22047910 Cumulus granule cells                          control female
## 6 SRR22047911 Cumulus granule cells                          control female
## 7 SRR22047912 Cumulus granule cells                          control female
```

Preprocessing

```
# Creating a DGEList object for use in edgeR
y <- DGEList(txi$counts)

# Filtering
condition <- factor(sdrf_rnaseq$disease_state)
y$samples$group <- condition
keep <- filterByExpr(y, group = y$samples$group)
y <- y[keep, , keep.lib.sizes=FALSE]
summary(keep)
```

```
##    Mode   FALSE    TRUE
## logical   39341   23413
```

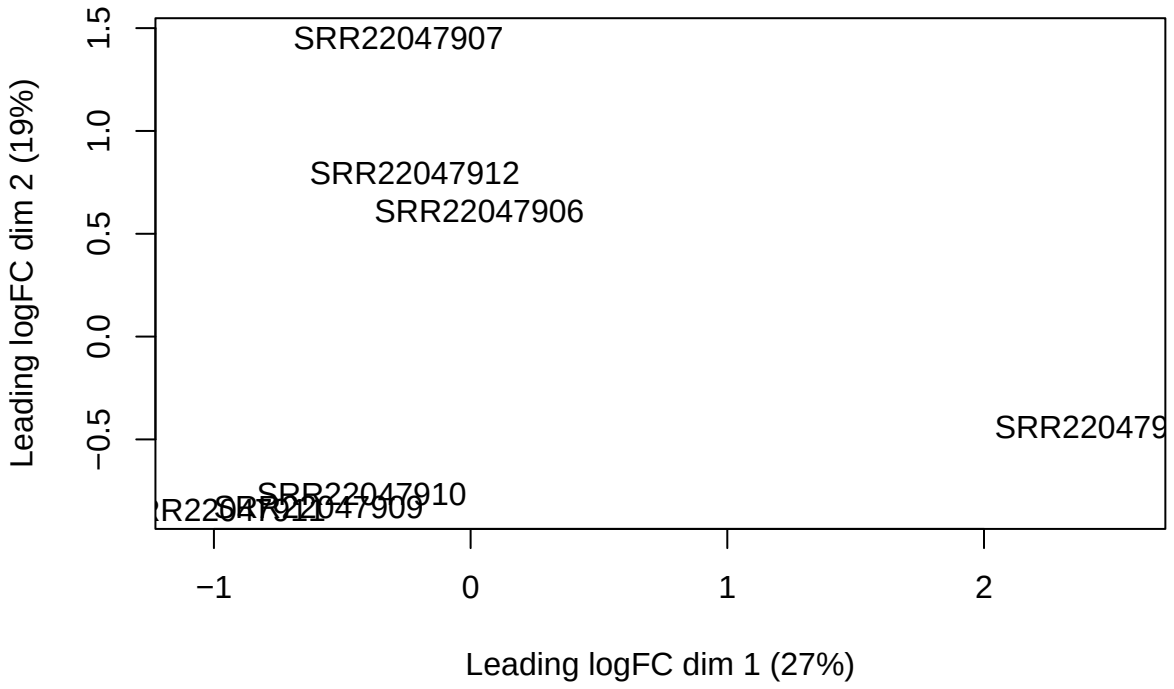
Calculating TMM normalisation factors

```
y <- calcNormFactors(y)
y$samples
```

```
##                                      group lib.size norm.factors
## SRR22047906 polycystic ovary syndrome (PCOS) 29601074    0.9846408
## SRR22047907 polycystic ovary syndrome (PCOS) 29812363    0.9012685
## SRR22047908 polycystic ovary syndrome (PCOS)  7300093    0.9584984
## SRR22047909                          control 21483794    1.1296602
## SRR22047910                          control 21954210    1.0700831
## SRR22047911                          control  9864893    1.0409527
## SRR22047912                          control 22999052    0.9342869
```

PCA on the normalized count data to visualize the overall grouping of samples

```
plotMDS(y)
```



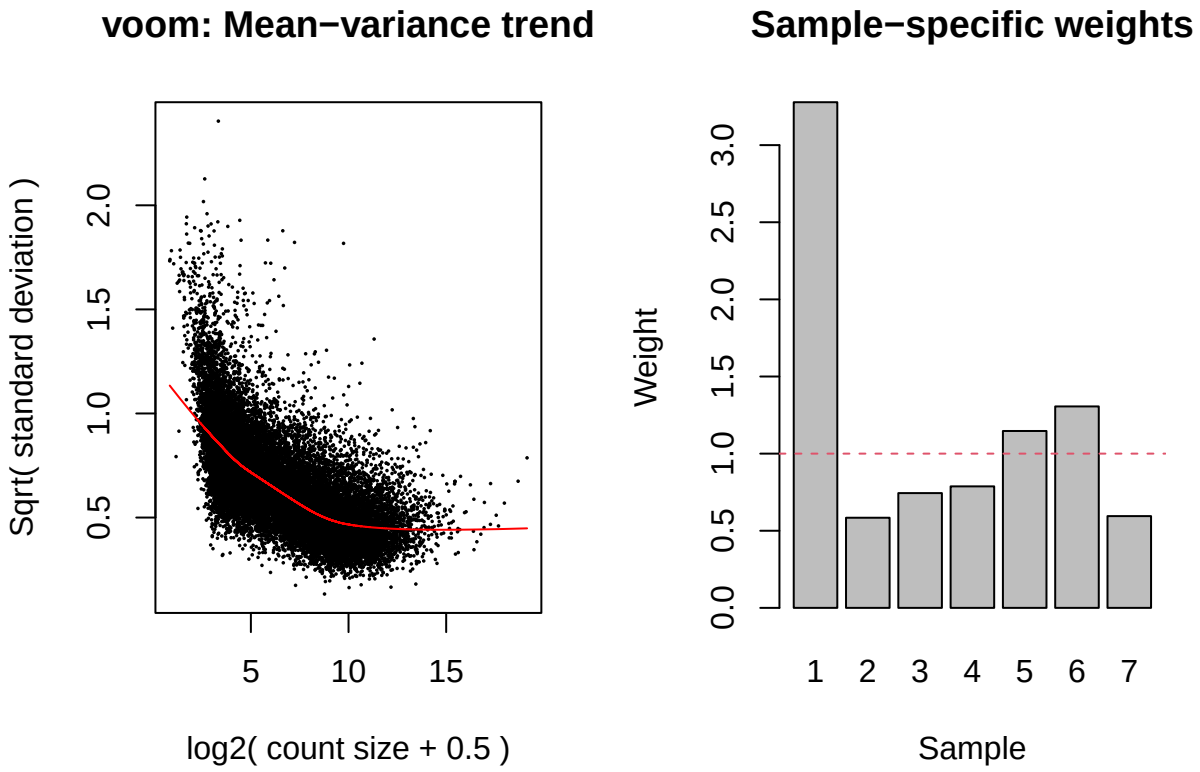
Making design

```
design <- model.matrix(~0 + condition)
colnames(design) <- c('control', 'PCOS')
design
```

```
##   control PCOS
## 1       0    1
## 2       0    1
## 3       0    1
## 4       1    0
## 5       1    0
## 6       1    0
## 7       1    0
## attr(,"assign")
## [1] 1 1
## attr(,"contrasts")
## attr(,"contrasts")$condition
## [1] "contr.treatment"
```

Differential expression analysis

```
v <- voomWithQualityWeights(y, design, plot=TRUE)
```




```
v
```

```
## An object of class "EList"
## $targets
##                                                 group lib.size norm.factors
## SRR22047906 polycystic ovary syndrome (PCOS) 29146425    0.9846408
## SRR22047907 polycystic ovary syndrome (PCOS) 26868943    0.9012685
## SRR22047908 polycystic ovary syndrome (PCOS)  6997128    0.9584984
## SRR22047909                             control 24269387    1.1296602
## SRR22047910                             control 23492829    1.0700831
## SRR22047911                             control 10268887    1.0409527
## SRR22047912                             control 21487714    0.9342869
##             sample.weights
## SRR22047906      3.2781846
## SRR22047907      0.5843890
## SRR22047908      0.7440253
## SRR22047909      0.7875824
## SRR22047910      1.1466648
## SRR22047911      1.3059924
## SRR22047912      0.5948448
```

```
##
## $E
##                 SRR22047906 SRR22047907 SRR22047908 SRR22047909 SRR22047910
## ENSG00000000003  5.17464739    5.541528   4.9300155   5.1819646   4.9706609
## ENSG00000000419  4.90503353    4.334475   5.1301197   4.6925812   4.4527716
## ENSG00000000457  3.97159412    3.865988   2.6745624   3.9411900   3.4596740
## ENSG00000000460  2.16882005    2.148021   1.6567441   2.3846975   2.0879461
## ENSG00000000938 -0.03217785   -0.380834   0.5871884  -0.4107819   0.5842529
##                 SRR22047911 SRR22047912
## ENSG00000000003    5.262112   5.4190284
## ENSG00000000419    4.675664   4.6082837
## ENSG00000000457    3.266515   3.6292276
## ENSG00000000460    2.768479   2.4459011
## ENSG00000000938   -1.175558   0.2798985
## 23408 more rows ...
##
## $weights
##          [,1]      [,2]     [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 69.57113 12.196885 8.972914 16.058557 23.201440 19.355502 11.755155
## [2,] 66.29527 11.566012 7.903963 14.229895 20.474474 15.581191 10.256884
## [3,] 49.84318  8.532434 5.011532 10.004077 14.306465 10.008602  7.050336
## [4,] 25.84258  4.393119 2.708867  6.430209  9.185076  6.565923  4.522387
## [5,] 11.91839  2.037497 1.084924  2.317397  3.306881  2.154823  1.619975
## 23408 more rows ...
##
## $design
##   control PCOS
## 1       0    1
## 2       0    1
## 3       0    1
## 4       1    0
## 5       1    0
## 6       1    0
## 7       1    0
## attr(,"assign")
## [1] 1 1
## attr(,"contrasts")
## attr(,"contrasts")$condition
## [1] "contr.treatment"
```

```
fit_limma <- lmFit(v, design)
cont.matrix <- makeContrasts(PCOS - control, levels = design)
fit2 <- contrasts.fit(fit_limma, cont.matrix)
fit_limma2 <- eBayes(fit2)
res_limma <- topTable(fit_limma2, coef=1, adjust.method="BH", number=Inf)
head(res_limma)
```

```
##                     logFC  AveExpr         t      P.Value  adj.P.Val        B
## ENSG00000159674  1.077184 6.072442  8.676081 2.673478e-06 0.03323942 5.143870
## ENSG00000166257 -1.503025 5.676478 -8.262216 4.312968e-06 0.03323942 4.672593
## ENSG00000263499 -1.794698 1.981602 -7.887603 6.753776e-06 0.03323942 4.160554
## ENSG00000114771  1.541361 6.397131  7.800815 7.509756e-06 0.03323942 4.092922
## ENSG00000118137  1.221194 5.673698  7.558263 1.014781e-05 0.03394154 3.803867
## ENSG00000184956 -3.218921 1.520790 -7.698638 8.518196e-06 0.03323942 3.777795
```

```
topGenes <- res_limma[res_limma$adj.P.Val < 0.05 & abs(res_limma$logFC) < 1.5, ]
topGenes
```

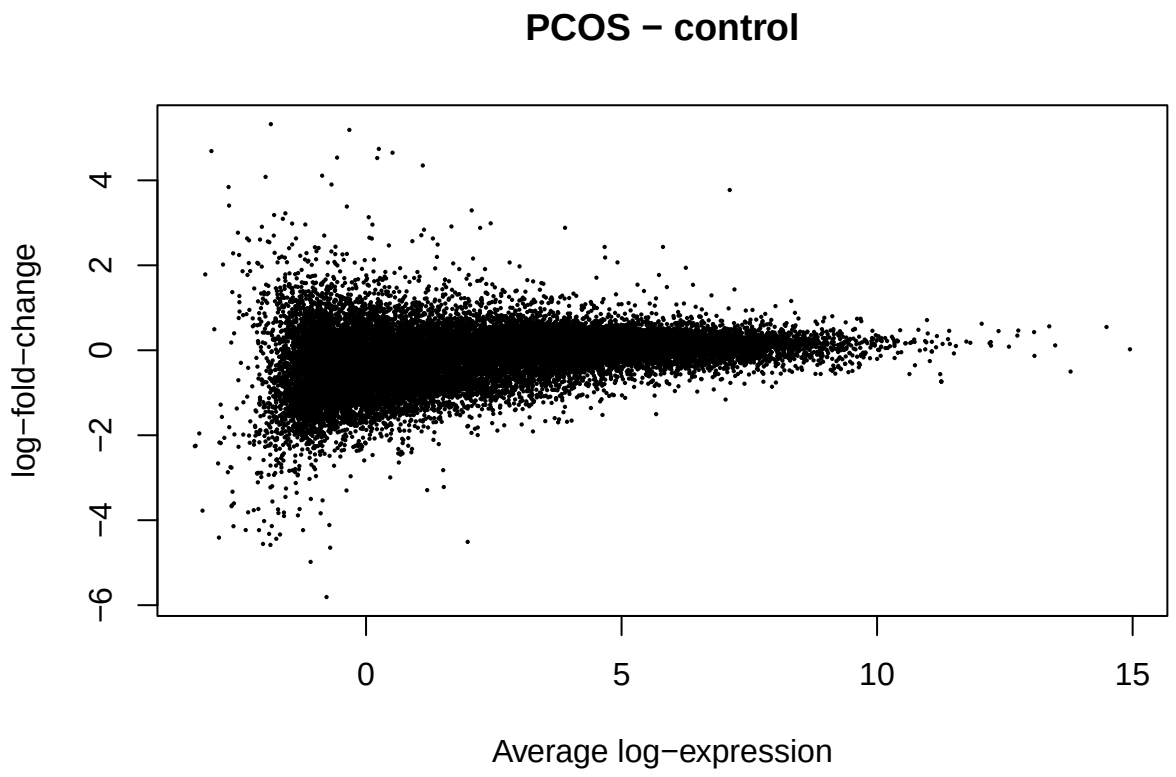
```
##                     logFC  AveExpr        t      P.Value  adj.P.Val        B
## ENSG00000159674 1.077184 6.072442 8.676081 2.673478e-06 0.03323942 5.143870
## ENSG00000118137 1.221194 5.673698 7.558263 1.014781e-05 0.03394154 3.803867
```

MA plot

```
jpeg("res_limmaWQW_PCOS_MA.png")
limma::plotMA(fit_limma2)
dev.off()
```

```
## pdf
##   2
```

```
limma::plotMA(fit_limma2)
```

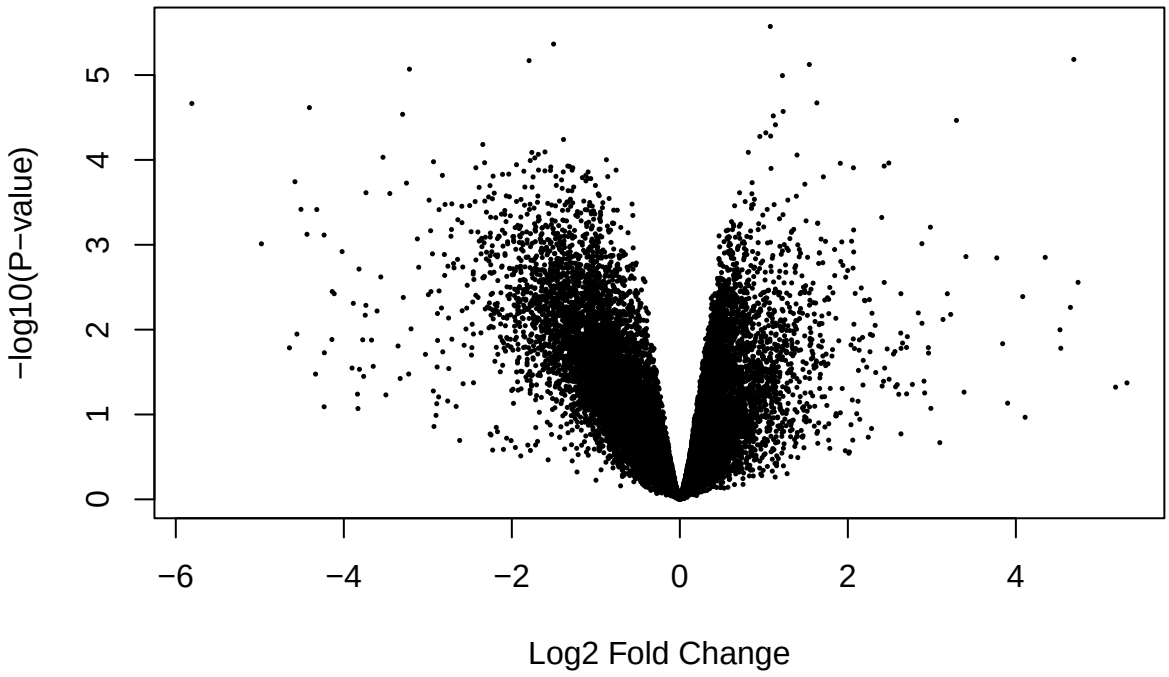


Volcano plot

```
jpeg("res_limmaWQW_PCOS_volcano.png")
limma::volcanoplot(fit_limma2, coef=1)
dev.off()
```

```
## pdf
##   2
```

```
limma::volcanoplot(fit_limma2, coef=1)
```



Gene Set Analysis

Gene annotation

As the package WebGestaltR is used, which requires EntrezIDs or Ensembl IDs to perform an Over Representation Analysis, annotation of the genes is performed first.

```
# Extract gene names from tx2gene
gene_names <- tx2gene$GENENAME

# Match gene names based on gene IDs
res_limma$GENENAME <- gene_names[match(rownames(res_limma), tx2gene$GENEID)]

# Match gene IDs based on gene names
gene_ids <- tx2gene$GENEID[match(res_limma$GENENAME, tx2gene$GENENAME)]
transcript_ids <- tx2gene$TXNAME[match(res_limma$GENENAME, tx2gene$GENENAME)]
transcript_lengths <- tx2gene$transcript_length[match(res_limma$GENENAME, tx2gene$GENENAME)]
```

```r
# Add ENSEMBL_GENE_ID and ENSEMBL_TRANSCRIPT_ID to res_limma
res_limma$ensembl_gene_id <- gene_ids
res_limma$ensembl_transcript_id <- transcript_ids

# Add transcript to res_limma
res_limma$transcript_length <- transcript_lengths

# Display the annotated result
head(res_limma)
```

```
##                     logFC  AveExpr         t      P.Value  adj.P.Val        B
## ENSG00000159674  1.077184 6.072442  8.676081 2.673478e-06 0.03323942 5.143870
## ENSG00000166257 -1.503025 5.676478 -8.262216 4.312968e-06 0.03323942 4.672593
## ENSG00000263499 -1.794698 1.981602 -7.887603 6.753776e-06 0.03323942 4.160554
## ENSG00000114771  1.541361 6.397131  7.800815 7.509756e-06 0.03323942 4.092922
## ENSG00000118137  1.221194 5.673698  7.558263 1.014781e-05 0.03394154 3.803867
## ENSG00000184956 -3.218921 1.520790 -7.698638 8.518196e-06 0.03323942 3.777795
##                 GENENAME ensembl_gene_id ensembl_transcript_id
## ENSG00000159674    SPON2 ENSG00000159674       ENST00000290902
## ENSG00000166257    SCN3B ENSG00000166257       ENST00000527125
## ENSG00000263499          ENSG00000276197       ENST00000619317
## ENSG00000114771    AADAC ENSG00000114771       ENST00000232892
## ENSG00000118137    APOA1 ENSG00000118137       ENST00000375320
## ENSG00000184956     MUC6 ENSG00000277518       ENST00000627256
##                 transcript_length
## ENSG00000159674              1579
## ENSG00000166257              3289
## ENSG00000263499                90
## ENSG00000114771              1563
## ENSG00000118137               940
## ENSG00000184956             14829
```

```r
# Sort results based on Ensembl gene ID
res_sorted <- res_limma[order(res_limma$ensembl_gene_id), ]
head(res_sorted)
```

```
##                      logFC     AveExpr          t    P.Value adj.P.Val
## ENSG00000000003  0.02582857  5.21142237  0.1740604 0.86492942 0.9262918
## ENSG00000000419  0.25598714  4.68556118  1.5837213 0.14111243 0.3348548
## ENSG00000000457  0.29655225  3.54410726  1.2412850 0.23991048 0.4432976
## ENSG00000000460 -0.26341447  2.23722980 -1.1770797 0.26361263 0.4681404
## ENSG00000000938  0.08225589 -0.07828741  0.1996374 0.84534849 0.9152313
## ENSG00000000971 -0.29478317  6.80004550 -1.8261521 0.09462508 0.2715956
##                        B GENENAME ensembl_gene_id ensembl_transcript_id
## ENSG00000000003 -6.888845   TSPAN6 ENSG00000000003       ENST00000373020
## ENSG00000000419 -5.632861     DPM1 ENSG00000000419       ENST00000466152
## ENSG00000000457 -5.904985    SCYL3 ENSG00000000457       ENST00000367771
## ENSG00000000460 -5.727274 C1orf112 ENSG00000000460       ENST00000498289
## ENSG00000000938 -5.950640      FGR ENSG00000000938       ENST00000374005
## ENSG00000000971 -5.431845      CFH ENSG00000000971       ENST00000359637
##                 transcript_length
## ENSG00000000003              3768
## ENSG00000000419              1097
```

```
## ENSG00000000457              6308
## ENSG00000000460              3849
## ENSG00000000938              2637
## ENSG00000000971              1756
```

```
# Check dimensions and overlap
cat("Overlap of Ensembl gene IDs between results and annotation data:",
    sum(res_sorted$ensembl_gene_id %in% tx2gene$GENEID), "\n")
```

```
## Overlap of Ensembl gene IDs between results and annotation data: 23413
```

```
cat("Dimensions of sorted results:", dim(res_sorted), "\n")
```

```
## Dimensions of sorted results: 23413 10
```

```
cat("Dimensions of annotation data:", dim(tx2gene), "\n")
```

```
## Dimensions of annotation data: 286849 5
```

```
# Add annotation to sorted results
res_sorted$gene_symbol <- tx2gene$GENENAME[match(res_sorted$ensembl_gene_id, tx2gene$GENEID)]
res_sorted$EntrezID <- tx2gene$ENTREZID[match(res_sorted$ensembl_gene_id, tx2gene$GENEID)]
res_sorted$transcript_length <- tx2gene$transcript_length[match(res_sorted$ensembl_gene_id, tx2gene$GENE
```

```
# Sort on p-values
res_annot <- res_sorted[order(res_sorted$P.Value), ]
head(res_annot)
```

```
##                     logFC   AveExpr         t      P.Value  adj.P.Val        B
## ENSG00000159674  1.077184  6.072442  8.676081 2.673478e-06 0.03323942 5.143870
## ENSG00000166257 -1.503025  5.676478 -8.262216 4.312968e-06 0.03323942 4.672593
## ENSG00000201059  4.689397 -3.025543  7.913281 6.546112e-06 0.03323942 1.983946
## ENSG00000263499 -1.794698  1.981602 -7.887603 6.753776e-06 0.03323942 4.160554
## ENSG00000114771  1.541361  6.397131  7.800815 7.509756e-06 0.03323942 4.092922
## ENSG00000184956 -3.218921  1.520790 -7.698638 8.518196e-06 0.03323942 3.777795
##                   GENENAME ensembl_gene_id ensembl_transcript_id
## ENSG00000159674     SPON2 ENSG00000159674       ENST00000290902
## ENSG00000166257     SCN3B ENSG00000166257       ENST00000527125
## ENSG00000201059 RNA5SP336 ENSG00000201059       ENST00000364189
## ENSG00000263499           ENSG00000276197       ENST00000619317
## ENSG00000114771     AADAC ENSG00000114771       ENST00000232892
## ENSG00000184956      MUC6 ENSG00000277518       ENST00000627256
##                 transcript_length gene_symbol EntrezID
## ENSG00000159674              1579       SPON2    10417
## ENSG00000166257              3289       SCN3B    55800
## ENSG00000201059               117   RNA5SP336       NA
## ENSG00000263499                90                   NA
## ENSG00000114771              1563       AADAC       13
## ENSG00000184956             14829        MUC6     4588
```

Over Representation Analysis

A filtering step is performed to only have entries with EntrezIDs

```
res_annot <- res_annot[complete.cases(res_annot$EntrezID), ]
```

The top 300 genes are selected based on p-value and logFC, these genes are most likely to be differently expressed.

```
logFC_interest <- res_annot[abs(res_annot$logFC) > 1.5, ]
logFC_interest_sorted <- logFC_interest[sort(logFC_interest$P.Value, index.return = T)$ix, ]
top_genes <- head(logFC_interest_sorted, 300)
head(top_genes)
```

```
##                      logFC    AveExpr         t      P.Value  adj.P.Val        B
## ENSG00000166257 -1.503025  5.6764776 -8.262216 4.312968e-06 0.03323942 4.672593
## ENSG00000114771  1.541361  6.3971307  7.800815 7.509756e-06 0.03323942 4.092922
## ENSG00000184956 -3.218921  1.5207896 -7.698638 8.518196e-06 0.03323942 3.777795
## ENSG00000100867  1.631042  2.1767636  6.983778 2.128518e-05 0.05448599 3.144624
## ENSG00000188039 -3.300016 -0.3819401 -6.752018 2.903122e-05 0.05448599 2.094728
## ENSG00000251655  3.292497  2.0670291  6.630956 3.423412e-05 0.05725167 2.681677
##                  GENENAME ensembl_gene_id ensembl_transcript_id
## ENSG00000166257    SCN3B ENSG00000166257       ENST00000527125
## ENSG00000114771    AADAC ENSG00000114771       ENST00000232892
## ENSG00000184956     MUC6 ENSG00000277518       ENST00000627256
## ENSG00000100867    DHRS2 ENSG00000100867       ENST00000553896
## ENSG00000188039     NWD1 ENSG00000188039       ENST00000524140
## ENSG00000251655     PRB1 ENSG00000282673       ENST00000632933
##                  transcript_length gene_symbol EntrezID
## ENSG00000166257              3289       SCN3B    55800
## ENSG00000114771              1563       AADAC       13
## ENSG00000184956             14829        MUC6     4588
## ENSG00000100867              1352       DHRS2    10202
## ENSG00000188039              7764        NWD1   284434
## ENSG00000251655               710        PRB1     5542
```

From these genes, we select the upregulated and downregulated genes. We store the entrez ids of these genes in txt files for further use in the Over Representation Analysis. The reference gene set is equal to all genes in the array.

```
# Upregulated genes
upreg_genes <- top_genes$EntrezID[top_genes$logFC > 0]
upreg_genes <- upreg_genes[!is.na(upreg_genes)]
cat("Number of downregulated genes:", length(upreg_genes), "\n")
```

```
## Number of downregulated genes: 82
```

```
# Downregulated genes
downreg_genes <- top_genes$EntrezID[top_genes$logFC < 0]
downreg_genes <- downreg_genes[!is.na(downreg_genes)]
cat("Number of downregulated genes:", length(downreg_genes), "\n")
```

```
## Number of downregulated genes: 218
```

```
# Reference genes
ref_genes <- res_annot$EntrezID[!is.na(res_annot$EntrezID)]
cat("Number of reference genes:", length(ref_genes), "\n")
```

```
## Number of reference genes: 16661
```

KEGG

```
# Perform KEGG pathway analysis on the upregulated genes
kegg_upreg <- limma::kegga(de = upreg_genes, universe = ref_genes, trend = res_annot$transcript_length,

# Sort results and calculate FDR
kegg_upreg <- kegg_upreg[sort(kegg_upreg$P.DE, index.return=T)$ix, ]
kegg_upreg$P.DE.adj <- p.adjust(kegg_upreg$P.DE, n=nrow(kegg_upreg), method="BH")
head(kegg_upreg)
```

```
##                                                    Pathway   N DE
## hsa04610                  Complement and coagulation cascades  63  4
## hsa05418               Fluid shear stress and atherosclerosis 123  5
## hsa04933 AGE-RAGE signaling pathway in diabetic complications  93  4
## hsa05143                              African trypanosomiasis  25  2
## hsa04080              Neuroactive ligand-receptor interaction 178  4
## hsa04670                 Leukocyte transendothelial migration  97  3
##                  P.DE   P.DE.adj
## hsa04610 0.0002629346 0.06310054
## hsa05418 0.0003544974 0.06310054
## hsa04933 0.0011523055 0.13674025
## hsa05143 0.0067042979 0.59668252
## hsa04080 0.0116465139 0.62649668
## hsa04670 0.0122439464 0.62649668
```

```
# Perform KEGG pathway analysis on the downregulated genes
kegg_downreg <- limma::kegga(de = downreg_genes, universe = ref_genes, trend = res_annot$transcript_leng

# Print or further process the results
# Sort results and calculate FDR
kegg_downreg <- kegg_downreg[sort(kegg_downreg$P.DE, index.return=T)$ix, ]
kegg_downreg$P.DE.adj <- p.adjust(kegg_downreg$P.DE, n=nrow(kegg_downreg), method="BH")
head(kegg_downreg)
```

```
##                                              Pathway   N DE        P.DE
## hsa04080      Neuroactive ligand-receptor interaction 178  8 0.002449781
## hsa04024                          cAMP signaling pathway 171  7 0.007441897
## hsa04672 Intestinal immune network for IgA production  32  3 0.008346555
## hsa05033                             Nicotine addiction  22  2 0.033312446
## hsa04740                          Olfactory transduction  94  4 0.035207778
## hsa04728                            Dopaminergic synapse 109  4 0.055352538
##           P.DE.adj
```

```
## hsa04080 0.8721219
## hsa04024 0.9904579
## hsa04672 0.9904579
## hsa05033 1.0000000
## hsa04740 1.0000000
## hsa04728 1.0000000
```

Gene Ontology: Molecular Function (MF)

```r
# Perform Gene Ontology: Molecular Function Analysis on the upregulated genes
goana_upreg <- limma::goana(de = upreg_genes, universe = ref_genes, trend = res_annot$transcript_length

# Print or further process the results
goana_upreg <- goana_upreg[goana_upreg$Ont == "MF", ]
goana_upreg <- goana_upreg[sort(goana_upreg$P.DE, index.return=T)$ix, ]
goana_upreg$P.DE.adj <- p.adjust(goana_upreg$P.DE, n=nrow(goana_upreg), method="BH")
head(goana_upreg)
```

```
##                                                 Term Ont   N DE         P.DE
## GO:0005201 extracellular matrix structural constituent  MF 139  6 6.397903e-05
## GO:0030545       signaling receptor regulator activity  MF 277  8 6.620640e-05
## GO:0098631             cell adhesion mediator activity  MF  55  4 1.552614e-04
## GO:0048018                    receptor ligand activity  MF 251  7 2.396536e-04
## GO:0030546       signaling receptor activator activity  MF 255  7 2.638633e-04
## GO:0008131              primary amine oxidase activity  MF   6  2 3.563991e-04
##             P.DE.adj
## GO:0005201 0.1590609
## GO:0030545 0.1590609
## GO:0098631 0.2486770
## GO:0048018 0.2535726
## GO:0030546 0.2535726
## GO:0008131 0.2854163
```

```r
# Perform Gene Ontology: Molecular Function Analysis using WebGestaltR on the downregulated genes
goana_downreg <- limma::goana(de = downreg_genes, universe = ref_genes, trend = res_annot$transcript_le

# Print or further process the results
goana_downreg <- goana_downreg[goana_downreg$Ont == "MF", ]
goana_downreg <- goana_downreg[sort(goana_downreg$P.DE, index.return=T)$ix, ]
goana_downreg$P.DE.adj <- p.adjust(goana_downreg$P.DE, n=nrow(goana_downreg), method="BH")
head(goana_downreg)
```

```
##                                           Term Ont   N DE         P.DE
## GO:0140345 phosphatidylcholine flippase activity  MF   5  3 2.185915e-05
## GO:1990782       protein tyrosine kinase binding  MF 100  7 3.467238e-04
## GO:0140333 glycerophospholipid flippase activity  MF  12  3 4.492516e-04
## GO:0140351    glycosylceramide flippase activity  MF   3  2 5.097840e-04
## GO:0030594    neurotransmitter receptor activity  MF  57  5 8.936658e-04
## GO:0140327                     flippase activity  MF  15  3 9.024984e-04
##             P.DE.adj
## GO:0140345 0.1050332
```

```
## GO:1990782 0.6123780
## GO:0140333 0.6123780
## GO:0140351 0.6123780
## GO:0030594 0.6223595
## GO:0140327 0.6223595
```

Writing out Data for comparison

Results of the analysis of Dataset 1 are saved for comparison between dataset analysis. The results of limma analysis are saved in txt file for further use.

```
write.table(res_annot, sep= "\t", file="Dataset2_limma_results.txt")
```

The results of gene set analysis are saved in a txt file for further use.

```
write.table(kegg_upreg, sep= "\t", file="Dataset2_PathwayAnalysis_upreg_results.txt")
write.table(kegg_downreg, sep= "\t", file="Dataset2_PathwayAnalysis_downreg_results.txt")
write.table(goana_upreg, sep= "\t", file="Dataset2_MolecularFunction_upreg_results.txt")
write.table(goana_downreg, sep= "\t", file="Dataset2_MolecularFunction_downreg_results.txt")
```