

Comparative Transcriptomic Analysis of Endothelial Cells Exposed to Unidirectional Versus Bidirectional Flow

Negar Vahdani

2024-06-01

Table of contents

1 Clean Memory, set working directory and load libraries	2
2 Import the data and prepare the metadata	2
3 Make the Deseq2 object and filter the low expressed genes	4
4 Run Desq2 analysis	5
5 Plot PCA	6
6 Save the differential expressed genes (DEGs) in each comparison	11
7 Add ENSMBL gene ids to the results	12
8 Plot upregulated and downregulated genes in volcano plots	14
9 GO analysis for DEGs	20
9.1 GO plots for selected pathways	27
10 Plot Heatmap for figure 2	33
11 Plot Heatmap for revision- selected pathways for 24h time point	35
12 Plot Heatmap for revision- selected pathways for 4h time point -	42
13 Save the workspace	48

1 Clean Memory, set working directory and load libraries

```
#Clear Memory and set the working directory
rm(list=ls())
gc()
setwd("/Users/negarvahdani/Deseq2")
dir.create("results", recursive = TRUE)

#load libraries
library(dplyr)
library(readxl)
library(DESeq2)
library(biomaRt)
library(ggplot2)
library(ggrepel)
library(ComplexHeatmap)
library(clusterProfiler)
library(tibble)
library(tidyverse)
library(matrixStats)
```

2 Import the data and prepare the metadata

```
#import data
counts110_trimmed <-
  read.delim("/Users/negarvahdani/gene_counts_ref110_trimmed.txt",
            sep = "\t", stringsAsFactor = FALSE, header= TRUE,
            fill = TRUE, row.names = 1)

#remove the columns and the rows of data containing sequencing information
counts110_trimmed_colremove <- counts110_trimmed %>% select(-c(1:5))

#remove the prefix and suffix of the the column name
colnames(counts110_trimmed_colremove) <-
  gsub("X.data.users.nvahdani.flow_project.bamsort110.output_trimmed.",
        "", colnames(counts110_trimmed_colremove))
colnames(counts110_trimmed_colremove) <-
  sub(".sorted.bam", "", colnames(counts110_trimmed_colremove))
```

```
#check the column names
head(counts110_trimmed_colremove)
```

	e3_24h_p	e3_24h_t	e3_4h_p	e3_4h_t	e4_24h_p	e4_24h_t	e4_4h_p
ENSG00000279928	0	0	0	0	0	0	0
ENSG00000228037	0	0	0	0	0	0	6
ENSG00000142611	0	1	0	0	0	0	0
ENSG00000284616	0	0	0	0	0	0	0
ENSG00000157911	292	447	115	495	296	386	377
ENSG00000269896	0	21	18	11	7	6	14
	e4_4h_t	e6_24h_p	e6_24h_t	e6_4h_p	e6_4h_t		
ENSG00000279928	0	2	0	0	0		
ENSG00000228037	0	0	0	0	0		
ENSG00000142611	0	0	0	5	0		
ENSG00000284616	0	0	0	0	0		
ENSG00000157911	407	377	467	383	388		
ENSG00000269896	23	14	12	6	4		

```
colnames(counts110_trimmed_colremove)
```

```
[1] "e3_24h_p" "e3_24h_t" "e3_4h_p" "e3_4h_t" "e4_24h_p" "e4_24h_t"
[7] "e4_4h_p" "e4_4h_t" "e6_24h_p" "e6_24h_t" "e6_4h_p" "e6_4h_t"
```

```
#load the metadata and check the column names
new_metadata <- read.delim("/Users/negarvahdani/new_metadata.txt",
                           sep = "\t", stringsAsFactor = FALSE, header= TRUE,
                           fill = TRUE, row.names = 1)

# check if the column names are the same as the row names of the metadata
colnames(counts110_trimmed_colremove) %in% rownames(new_metadata)
```

```
[1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
# check if the column names are the same as the row names of the metadata
identical(colnames(counts110_trimmed_colremove), rownames(new_metadata))
```

```
[1] TRUE
```

```

#create a new column with the flow profile and time
new_metadata$flow_time <- paste0(new_metadata$flow_profile, "_",
                                new_metadata$time)

#convert the columns to a factor
new_metadata$flow_time <- factor(new_metadata$flow_time)
new_metadata$flow_profile <- factor(new_metadata$flow_profile)
head(new_metadata)

```

	experiment	time	flow_profile	flow_time
e3_24h_p	3	24	unidirectional	unidirectional_24
e3_24h_t	3	24	bidirectional	bidirectional_24
e3_4h_p	3	4	unidirectional	unidirectional_4
e3_4h_t	3	4	bidirectional	bidirectional_4
e4_24h_p	4	24	unidirectional	unidirectional_24
e4_24h_t	4	24	bidirectional	bidirectional_24

3 Make the Deseq2 object and filter the low expressed genes

```

#make DESeq2 object
dds_flow_time <-
  DESeqDataSetFromMatrix(countData = counts110_trimmed_colremove,
                        colData = new_metadata,
                        design= ~ flow_time)

#remove the low expressed genes and check the no. of the removed columns
#set threshold and no. of samples
threshold <- 5
min_samples <- 3

#calculate the sum of counts across all samples for each gene
gene_counts <- rowSums(counts(dds_flow_time))

#create a logical vector indicating whether each gene meets the criteria
keep_genes <- gene_counts >= threshold

#subset the DESeqDataSet
dds_filtered <- dds_flow_time[keep_genes,]

#DESeq2 analysis
dds_flow_time <- DESeq(dds_flow_time)

```

```
estimating size factors  
  
estimating dispersions  
  
gene-wise dispersion estimates  
  
mean-dispersion relationship  
  
final dispersion estimates  
  
fitting model and testing
```

```
#check the lists the coefficients  
print(resultsNames(dds_flow_time))
```

```
[1] "Intercept"  
[2] "flow_time_bidirectional_4_vs_bidirectional_24"  
[3] "flow_time_unidirectional_24_vs_bidirectional_24"  
[4] "flow_time_unidirectional_4_vs_bidirectional_24"
```

4 Run Desq2 analysis

```
#extract the results for the comparison of the flow profile and time  
#tilter 24h vs pump 24h  
res_uni_bidirectional_24 <-  
  lfcShrink(dds_flow_time,parallel = TRUE,  
            contrast=c("flow_time","bidirectional_24","unidirectional_24"),  
            type="ashr")
```

using 'ashr' for LFC shrinkage. If used in published research, please cite:
Stephens, M. (2016) False discovery rates: a new deal. Biostatistics, 18:2.
<https://doi.org/10.1093/biostatistics/kxw041>

```
#convert the results to a data frame  
res_uni_bidirectional_24 <- as.data.frame(res_uni_bidirectional_24 )  
head(res_uni_bidirectional_24) #check the first rows
```

	baseMean	log2FoldChange	lfcSE	pvalue	padj
ENSG00000279928	0.1400055	-0.012117556	0.3593873	0.6798435	NA
ENSG00000228037	0.4547908	0.000000000	0.3580922	1.0000000	NA
ENSG00000142611	0.4967602	0.006178389	0.3584660	0.8319448	NA
ENSG00000284616	0.0000000	0.000000000	0.3692715	NA	NA
ENSG00000157911	354.5700642	0.158482844	0.1867622	0.1225302	0.4983356
ENSG00000269896	11.8541232	0.058353799	0.3059987	0.4419554	0.8178249

```
#tilter 4h vs pump 4h
res_uni_bidirectional_4 <-
  lfcShrink(dds_flow_time,parallel = TRUE,
             contrast=c("flow_time", "bidirectional_4","unidirectional_4"),
             type="ashr")
```

using 'ashr' for LFC shrinkage. If used in published research, please cite:
 Stephens, M. (2016) False discovery rates: a new deal. *Biostatistics*, 18:2.
<https://doi.org/10.1093/biostatistics/kxw041>

```
#convert the results to a data frame and check the first rows
res_uni_bidirectional_4 <- as.data.frame(res_uni_bidirectional_4 )
head(res_uni_bidirectional_4)
```

	baseMean	log2FoldChange	lfcSE	pvalue	padj
ENSG00000279928	0.1400055	0.000000000	0.2344532	1.0000000	NA
ENSG00000228037	0.4547908	-0.009906163	0.2362416	0.4399685	NA
ENSG00000142611	0.4967602	-0.009524368	0.2361023	0.4571412	NA
ENSG00000284616	0.0000000	0.000000000	0.2380879	NA	NA
ENSG00000157911	354.5700642	0.114227151	0.1988798	0.0852675	0.5495606
ENSG00000269896	11.8541232	-0.018984809	0.2063680	0.6158550	0.9694326

5 Plot PCA

```
## PCA function

PCA <- function(mat,color_pca="",shape_pca= "", label_pca= "",
                 save_plot= "no", name_of_plot= "PCA", comp1=1,
                 comp2=2, pdf_width=12, pdf_height=12){
  #Get the differential expressed values from the comparrison interested,
```

```

#extract the normalized values from the assay of vsd and plot them.
#Giving condition and group from your design table

#1. Extract the counts.
dt <- mat

#2. Perform pca
pca_dt <- prcomp(t(dt))
cat("PCA running...\n")
# Sys.sleep(0.2)

#3. Extract percentVar data.
percentVar_dt <- pca_dt$sdev^2/sum(pca_dt$sdev^2)
cat("Percents calculated...\n")
# Sys.sleep(0.2)

#4. Create the new dataframe to plot.
dt_f <- data.frame(PC1=pca_dt$x[,comp1],
                     PC2=pca_dt$x[,comp2],
                     color_pca=color_pca,
                     shape_pca=shape_pca,
                     label_pca= label_pca)
cat("Data frame built...\n")
# Sys.sleep(0.2)

#5. Plot it
cat("Plotting...\n")
# Sys.sleep(0.2)
print(save_plot)
require(ggplot2)
require(ggrepel)
if (save_plot== "no") {
  pca_p <- ggplot(data = dt_f, aes_string(x = paste0("PC1"),
                                             y = paste0("PC2"),
                                             color = "color_pca",
                                             shape= "shape_pca",
                                             label="label_pca")) +
    geom_point(size = 5) +
    geom_text_repel(size= 3, max.overlaps = 50,
                   box.padding = 1.5, point.padding = 0.5, force = 50) +
    xlab(paste0("PC", comp1,": ",
                round(percentVar_dt[comp1] * 100), "% variance")) +
}

```

```

      ylab(paste0("PC",comp2,": ",
                  round(percentVar_dt[comp2] * 100), "% variance")) +
# coord_fixed()+
NULL
}
if (save_plot== "yes"){

  cat("Saving plot as: ",paste0(name_of_plot,"...\n"))
  pca_p <- ggplot(data = dt_f, aes_string(x = paste0("PC",comp1),
                                             y = paste0("PC",comp2),
                                             color = "color_pca",
                                             shape= "shape_pca",
                                             label="label_pca")) +
  geom_text_repel(size= 3, max.overlaps = 50,
                  box.padding = 1.5,
                  point.padding = 0.5,force = 50) +
  geom_point(size = 5) +
  xlab(paste0("PC", comp1,": ", round(percentVar_dt[comp1] * 100),
              "% variance")) +
  ylab(paste0("PC",comp2,": ", round(percentVar_dt[comp2] * 100),
              "% variance")) +
# coord_fixed()+
NULL
print(pca_p)
dev.copy(pdf, paste0(name_of_plot,".pdf"),
        width = pdf_width,height = pdf_height)
dev.off()
}
# Sys.sleep(0.2)
cat("Done")
print(pca_p)

#return(pca_p)
}

#run the PCA function
PCA(counts110_trimmed_colremove, color_pca = factor(new_metadata$time),
    shape_pca = factor(new_metadata$flow_profile), save_plot = "no",
    name_of_plot = "PCA_nolog")

```

PCA running...
Percents calculated...

```
Data frame built...
```

```
Plotting...
```

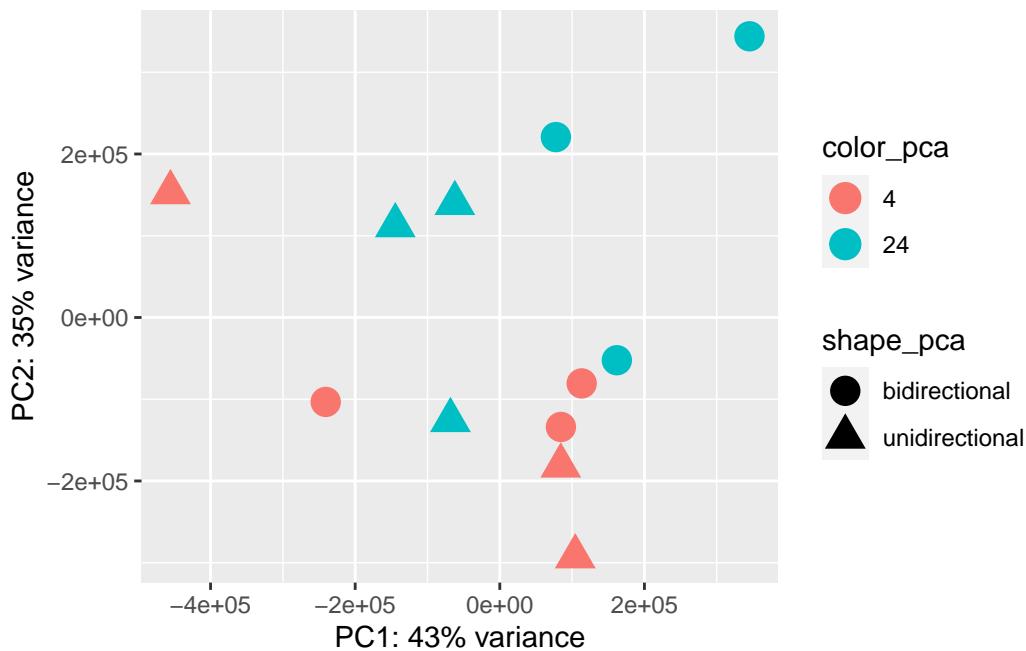
```
[1] "no"
```

```
Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
```

```
i Please use tidy evaluation idioms with `aes()``.
```

```
i See also `vignette("ggplot2-in-packages")` for more information.
```

```
Done
```



```
#PCA for each timepoint- revisions request
```

```
counts110_trimmed_colremove_4h <- counts110_trimmed_colremove[,c(3,4,7,8,11,12)]  
counts110_trimmed_colremove_24h <- counts110_trimmed_colremove[,c(1,2,5,6,9,10)]
```

```
PCA(counts110_trimmed_colremove_4h,  
    shape_pca = factor(new_metadata$flow_profile), save_plot = "no",  
    name_of_plot = "PCA_nolog")
```

```
PCA running...
```

```
Percents calculated...
```

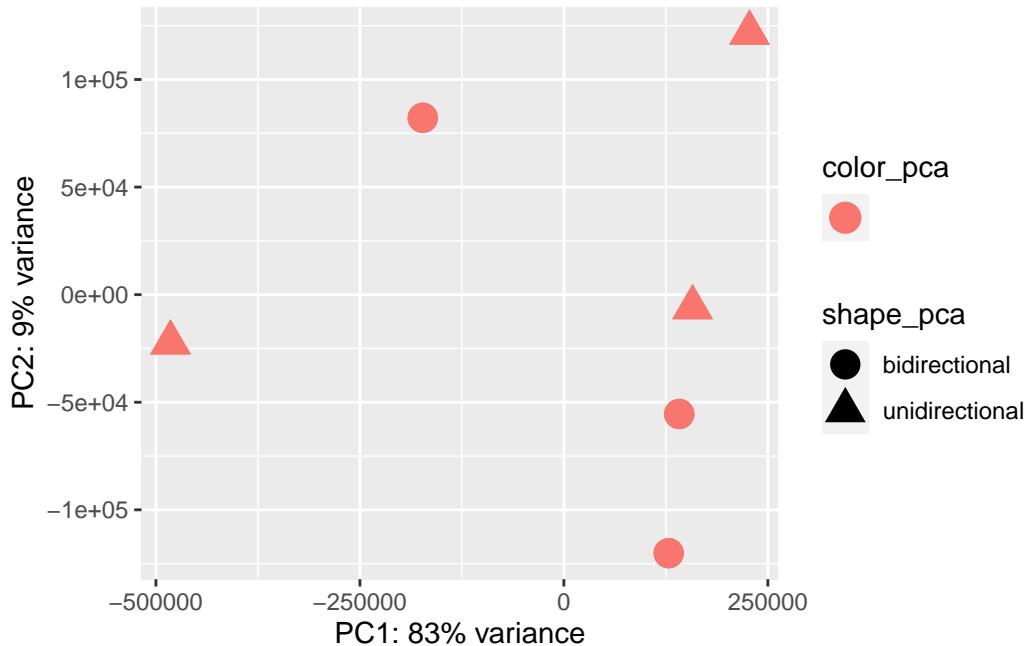
```
Warning in data.frame(PC1 = pca_dt$x[, comp1], PC2 = pca_dt$x[, comp2], : row  
names were found from a short variable and have been discarded
```

```
Data frame built...
```

```
Plotting...
```

```
[1] "no"
```

```
Done
```



```
PCA(counts110_trimmed_colremove_24h,  
shape_pca = factor(new_metadata$flow_profile), save_plot = "no",  
name_of_plot = "PCA_nolog")
```

```
PCA running...
```

```
Percents calculated...
```

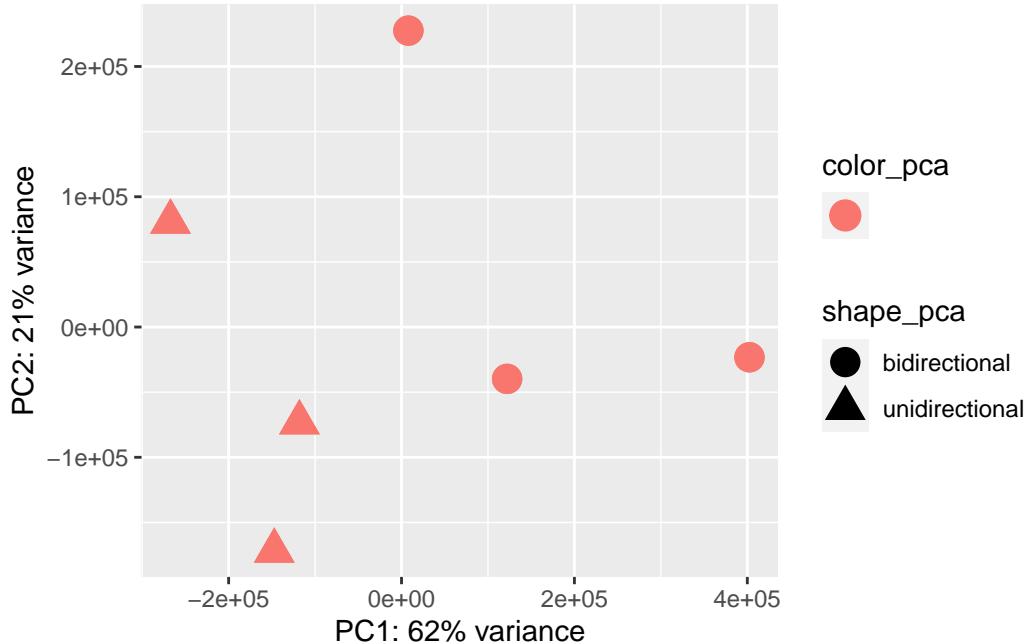
```
Warning in data.frame(PC1 = pca_dt$x[, comp1], PC2 = pca_dt$x[, comp2], : row  
names were found from a short variable and have been discarded
```

```
Data frame built...
```

```
Plotting...
```

```
[1] "no"
```

```
Done
```



6 Save the differential expressed genes (DEGs) in each comparison

```
#put 1 instead of the NAs because it causes problems in the volcano plot
res_uni_bidirectional_24$padj[is.na(res_uni_bidirectional_24$padj)] <- 1
res_uni_bidirectional_4$padj[is.na(res_uni_bidirectional_4$padj)] <- 1

#save the differential expressed genes in each comparison
res_uni_bidirectional_24_padj0.05 <- subset(res_uni_bidirectional_24,
                                              padj < 0.05)
write.csv(as.data.frame(res_uni_bidirectional_24_padj0.05),
          file="unidirectional_vs_bidirectional_24_0.05_nopb.csv")
head(res_uni_bidirectional_24_padj0.05)
```

	baseMean	log2FoldChange	lfcSE	pvalue	padj
ENSG00000130766	745.104354	-0.6042284	0.3080410	5.534883e-04	1.943932e-02
ENSG00000092853	1317.929292	1.4715801	0.3163299	2.375180e-07	3.497268e-05
ENSG00000236098	9.162552	-0.7006854	0.9759735	4.108929e-04	1.554073e-02
ENSG00000023909	2084.137196	-1.1834854	0.2824740	7.122420e-07	9.086322e-05
ENSG00000198198	1489.214895	-0.4057124	0.1879218	1.424821e-03	3.655861e-02
ENSG00000117525	308.979163	-2.6614545	0.4370338	2.789364e-12	1.561144e-09

```

res_uni_bidirectional_4_padj0.05 <- subset(res_uni_bidirectional_4,
                                             padj < 0.05)
write.csv(as.data.frame(res_uni_bidirectional_4_padj0.05),
          file="unidirectional_vs_bidirectional_4_0.05_nopb.csv")
head(res_uni_bidirectional_4_padj0.05)

```

	baseMean	log2FoldChange	lfcSE	pvalue	padj
ENSG00000132881	57.76324	0.8568570	0.5505004	2.173710e-04	1.487556e-02
ENSG00000162407	1978.62537	-0.5549861	0.2070021	3.110924e-04	1.865012e-02
ENSG00000127124	368.14958	-1.2764537	0.3464764	2.095402e-07	7.600025e-05
ENSG00000023909	2084.13720	-1.4496330	0.2912663	3.083979e-10	6.114980e-07
ENSG00000143315	275.00192	0.7384807	0.3265591	2.064510e-04	1.426281e-02
ENSG00000132763	117.11548	0.6430754	0.4443382	8.166160e-04	3.859109e-02

7 Add ENSMBL gene ids to the results

```

#convert gene ids to names using biomart
symbol_to_ensmbl_human <- function(){
  require(biomartR) # load biomart package
  ensembl <- useMart(biomart = "ensembl", dataset = "hsapiens_gene_ensembl",
                      host = "https://www.ensembl.org")
  # use the ensembl mart
  # query biomart
  annotations <- getBM(attributes = c("ensembl_gene_id","hgnc_symbol"),
                        # get the ensembl gene id and the hgnc symbol
                        mart = ensembl)# the ensembl mart
  return(annotations) # return the results
}

#get the gene ids and convert to data frame
gene_id <- symbol_to_ensmbl_human()
gene_id_df <- as.data.frame(gene_id)
head(gene_id_df)

```

	ensembl_gene_id	hgnc_symbol
1	ENSG00000210049	MT-TF
2	ENSG00000211459	MT-RNR1
3	ENSG00000210077	MT-TV
4	ENSG00000210082	MT-RNR2

```

5 ENSG00000209082      MT-TL1
6 ENSG00000198888      MT-ND1

#add the ensembl gene id as a column, merge with the data and save the csv
res_uni_bidirectional_24$ensembl_gene_id <- rownames(res_uni_bidirectional_24)
res_uni_bidirectional_4$ensembl_gene_id <- rownames(res_uni_bidirectional_4)
res_uni_bidirectional_24merged <- merge(res_uni_bidirectional_24,
                                         gene_id_df, by.x = "ensembl_gene_id",
                                         all.x=TRUE)
res_uni_bidirectional_4hmerged <- merge(res_uni_bidirectional_4,
                                         gene_id_df, by.x = "ensembl_gene_id",
                                         all.x=TRUE)

res_uni_bidirectional_24merged_padj0.05 <-
  subset(res_uni_bidirectional_24merged, padj < 0.05)
write.csv(as.data.frame(res_uni_bidirectional_24merged_padj0.05),
          file="unidirectional_vs_bidirectional_24_0.05_nopb_withsymbol.csv")
head(res_uni_bidirectional_24merged_padj0.05)

  ensembl_gene_id  baseMean log2FoldChange      lfcSE      pvalue
9   ENSG00000001084    502.9601     -0.5804792 0.3502721 1.356347e-03
13  ENSG00000001497    857.7391      0.5631373 0.2315451 1.766710e-04
40   ENSG00000003402   7649.1939     -0.4074632 0.1312337 1.380003e-04
82   ENSG00000005108    624.6087      1.1664631 0.2512031 1.196775e-07
85   ENSG00000005187    167.0716      2.2175175 0.3117089 4.535044e-14
137  ENSG00000006459    851.4950     -0.7895024 0.3706497 2.368339e-04

  padj hgnc_symbol
9    3.550911e-02       GCLC
13   8.356567e-03       LAS1L
40   6.840872e-03       CFLAR
82   1.977528e-05      THSD7A
85   4.628413e-11      ACSM3
137  1.040271e-02      KDM7A

res_uni_bidirectional_4hmerged_padj0.05 <-
  subset(res_uni_bidirectional_4hmerged, padj < 0.05)
write.csv(as.data.frame(res_uni_bidirectional_4hmerged_padj0.05),
          , file="unidirectional_vs_bidirectional_4_0.05_nopb_withsymbol.csv")
head(res_uni_bidirectional_4hmerged_padj0.05)

  ensembl_gene_id  baseMean log2FoldChange      lfcSE      pvalue

```

20	ENSG00000002016	168.1894	0.8707179	0.3757943	8.314925e-05
67	ENSG00000004866	604.4218	-0.5734751	0.2106566	2.636006e-04
135	ENSG00000006451	8337.3423	0.4658322	0.2073872	9.509839e-04
137	ENSG00000006459	851.4950	-1.0084843	0.3069049	3.948719e-06
152	ENSG00000006704	666.1753	-0.7424482	0.3083939	1.627811e-04
219	ENSG00000008405	751.4543	-0.9217799	0.2374767	1.045986e-06
		padj hgnc_symbol			
20	0.0072670439	RAD52			
67	0.0169412009	ST7			
135	0.0423870125	RALA			
137	0.0006739766	KDM7A			
152	0.0119192070	GTF2IRD1			
219	0.0002529193	CRY1			

8 Plot upregulated and downregulated genes in volcano plots

```
draw_volcano<- function(fileinput, title) {

  require(ggplot2)
  require(ggrepel)
  require(clusterProfiler)
  require(tidyverse)

  # read input file
  # drawing plots
  ggplot(data =fileinput , aes(x = log2FoldChange, y = -log10(padj))) +
    # draw lines
    geom_hline(yintercept = -log10(0.05), linetype = "dashed",
               col = "chartreuse4") +
    geom_hline(yintercept = -log10(0.05), linetype = "dotted",
               col = "darkgoldenrod") +
    geom_vline(xintercept = 0, linetype = "dashed")+
    # draw points
    geom_point(x = fileinput$log2FoldChange, y = -log10(fileinput$padj),
               alpha = 0.5,size = 2,color="grey51") +
    # draw coloured points
    geom_point(data = fileinput[which(fileinput$padj < 0.05 &
                                      fileinput$log2FoldChange < -0.585),],
               aes(x=log2FoldChange, y = -log10(padj)), shape = 21,
```

```

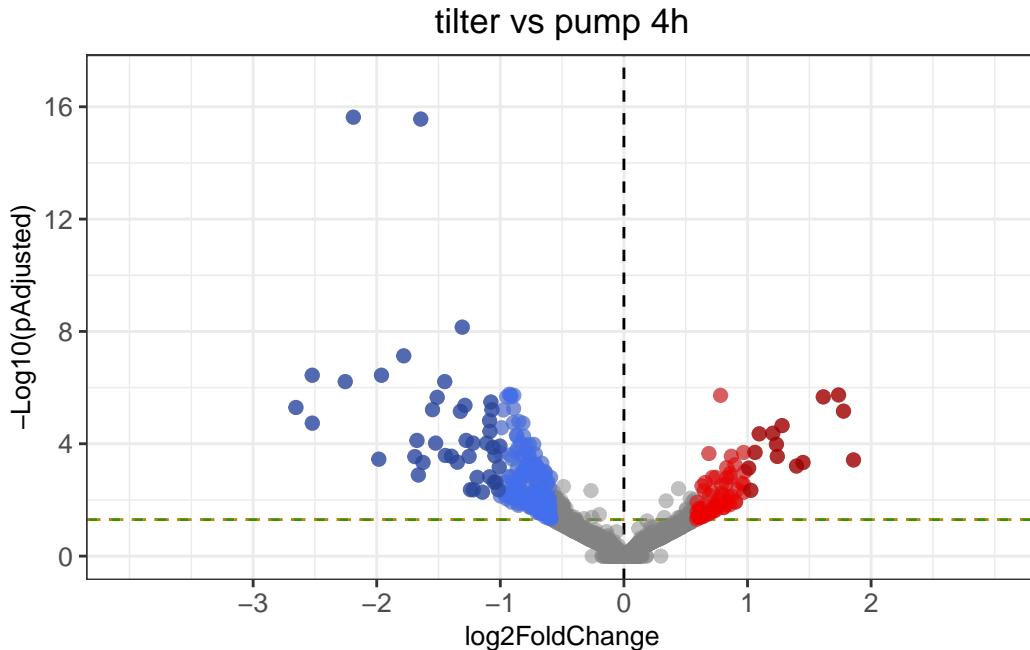
        color = "royalblue2", fill = "royalblue2",
        alpha = 0.5, size = 2) +
geom_point(data = fileinput[which(fileinput$padj < 0.05 &
                                 fileinput$log2FoldChange <= - 1),],
            aes(x=log2FoldChange, y = -log10(padj)),
            shape = 21, color = "royalblue4", fill = "royalblue4",
            alpha = 0.5, size = 2) +
geom_point(data = fileinput[which(fileinput$padj < 0.05 &
                                 fileinput$log2FoldChange > 0.585),],
            aes(x=log2FoldChange, y = -log10(padj)),
            shape = 21, color = "red2", fill = "red2",
            alpha = 0.5, size = 2) +
geom_point(data = fileinput[which(fileinput$padj < 0.05 &
                                 fileinput$log2FoldChange >= 1),],
            aes(x=log2FoldChange, y = -log10(padj)),
            shape = 21, color = "red4", fill = "red4",
            alpha = 0.5, size = 2) +
# x axis scale
scale_x_continuous(breaks =
                    seq(round(min(fileinput$log2FoldChange)- 0.5),
                        round(max(fileinput$log2FoldChange)+ 0.5),by = 1),
                    limits = c(round(min(fileinput$log2FoldChange)-1),
                               round(max(fileinput$log2FoldChange)+1))) +
xlab("log2FoldChange") + #ylab("-Log10(p.value)") +

scale_y_continuous(breaks = seq(0,round(-log10(min(fileinput$padj))+1),
                                by = 4),
                   limits = c(0,round(-log10(min(fileinput$padj))+1))) +
ylab("-Log10(pAdjusted)") + # ylab("-log10(p.value)")+

# set title
ggtitle(title) +
# x and y axis limits
# black and white theme
theme_bw() +
# center title
theme(plot.title = element_text(hjust = 0.5),
      axis.text = element_text(size = 10),
      axis.title.x = element_text(size = 10),
      axis.title.y = element_text(size = 10))
}

```

```
##volcano plot : tiltter vs pump at 4h
a = draw_volcano(res_uni_bidirectional_4hmerged, "tiltter vs pump 4h")
a
```



```
# add labels: make a list of top 25 genes and bottom 25 genes
options(ggrepel.max.overlaps=50) # set the maximum number of overlaps
label_set_top <- res_uni_bidirectional_4hmerged %>%
  arrange(desc(log2FoldChange)) %>% head(25) %>% pull(hgnc_symbol)
label_set_top
```

```
[1] ""
[7] "SH2D3C"    "BMP4"      "GIMAP1"    "SOX7"      "CAVIN2"    "GIMAP6"
[13] ""          "DHODH"     "IMP3"      "NUDT18"    "HTR1B"     "PTP4A2P2"
[19] "GIMAP8"    "TMEM144"   "LPP-AS2"   "EHD3"      "SPHK2"     "RAD52"
[25] "MAP3K12"
```

```
label_set_bottom <- res_uni_bidirectional_4hmerged %>%
  arrange(desc(log2FoldChange)) %>% tail(25) %>% pull(hgnc_symbol)
label_set_bottom
```

```
[1] "PITPNc1"    "HIVEP3"     "TXNRD1"     "GABARAPL1"  "CHSY3"      "SMARCD3"
```

```

[7] "NOG"          "KLF4"        "GCLM"        "EGFR"        "EGR1"        "NR4A1"
[13] "GLI2"         "MOAP1"       "NR4A3"       "HEY1"        "F2RL3"       "ZFPM1"
[19] "ATF3"         "DLX2"        "GLIS3"       "OSGIN1"      "HMOX1"       ""
[25] "IL11"

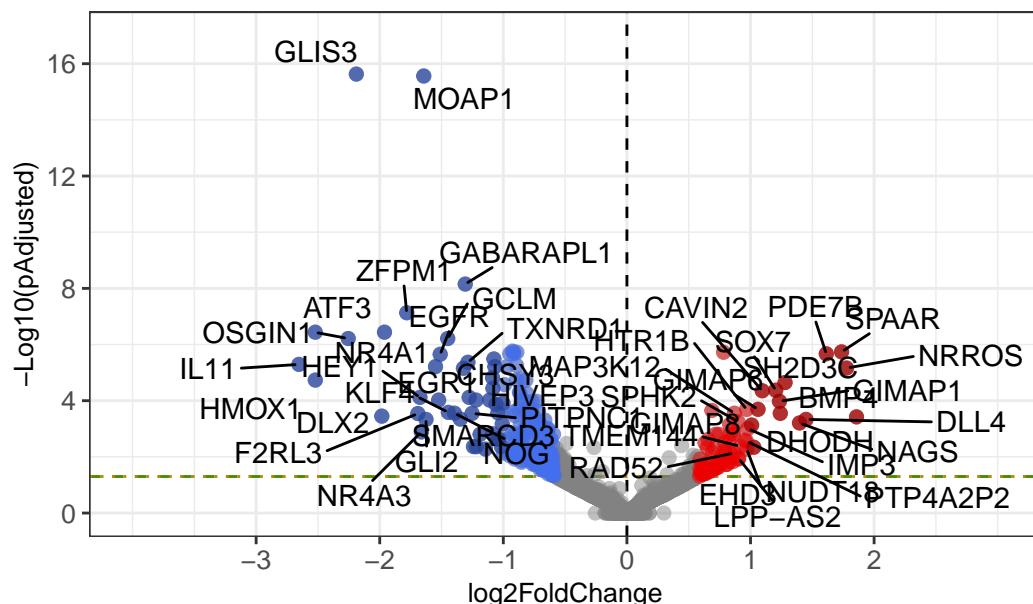
# combine the top and bottom genes altered genes
label_set <- c(label_set_bottom, label_set_top)
label_set

[1] "PITPNC1"     "HIVEP3"      "TXNRD1"      "GABARAPL1"   "CHSY3"       "SMARCD3"
[7] "NOG"          "KLF4"        "GCLM"        "EGFR"        "EGR1"        "NR4A1"
[13] "GLI2"         "MOAP1"       "NR4A3"       "HEY1"        "F2RL3"       "ZFPM1"
[19] "ATF3"         "DLX2"        "GLIS3"       "OSGIN1"      "HMOX1"       ""
[25] "IL11"         ""            "NRROS"       "SPAAR"       "PDE7B"       "DLL4"
[31] "NAGS"         "SH2D3C"      "BMP4"        "GIMAP1"      "SOX7"        "CAVIN2"
[37] "GIMAP6"       ""            "DHODH"       "IMP3"        "NUDT18"      "HTR1B"
[43] "PTP4A2P2"    "GIMAP8"      "TMEM144"     "LPP-AS2"     "EHD3"        "SPHK2"
[49] "RAD52"        "MAP3K12"    
```

```

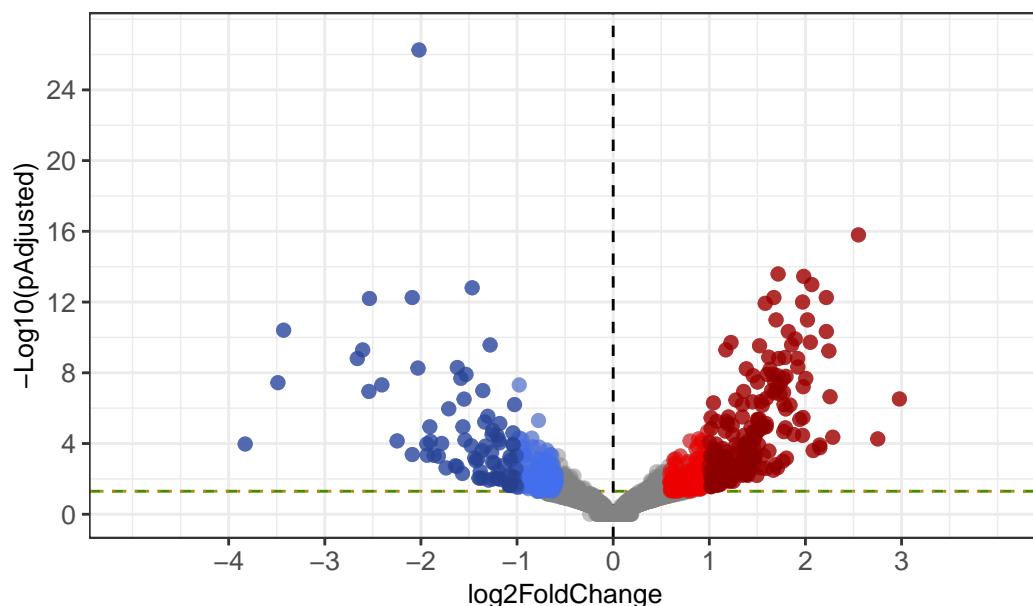
# add the labels to the plot
a2= a +
  geom_text_repel(data = res_uni_bidirectional_4hmerged
                  [res_uni_bidirectional_4hmerged$hgnc_symbol %in% label_set,],
                  aes(label=res_uni_bidirectional_4hmerged
                      [res_uni_bidirectional_4hmerged$hgnc_symbol %in%
                       label_set,]$hgnc_symbol),
  )
a2
```

tilter vs pump 4h



```
##volcano plot : tilter vs pump at 24h
b = draw_volcano(res_uni_bidirectional_24merged,"tilter vs pump 24h")
b
```

tilter vs pump 24h



```
# add labels: make a list of top 25 genes and bottom 25 genes
options(ggrepel.max.overlaps=50) # set the maximum number of overlaps
label_set_top <- res_uni_bidirectional_24merged %>%
  arrange(desc(log2FoldChange)) %>% head(25) %>% pull(hgnc_symbol)
label_set_top
```

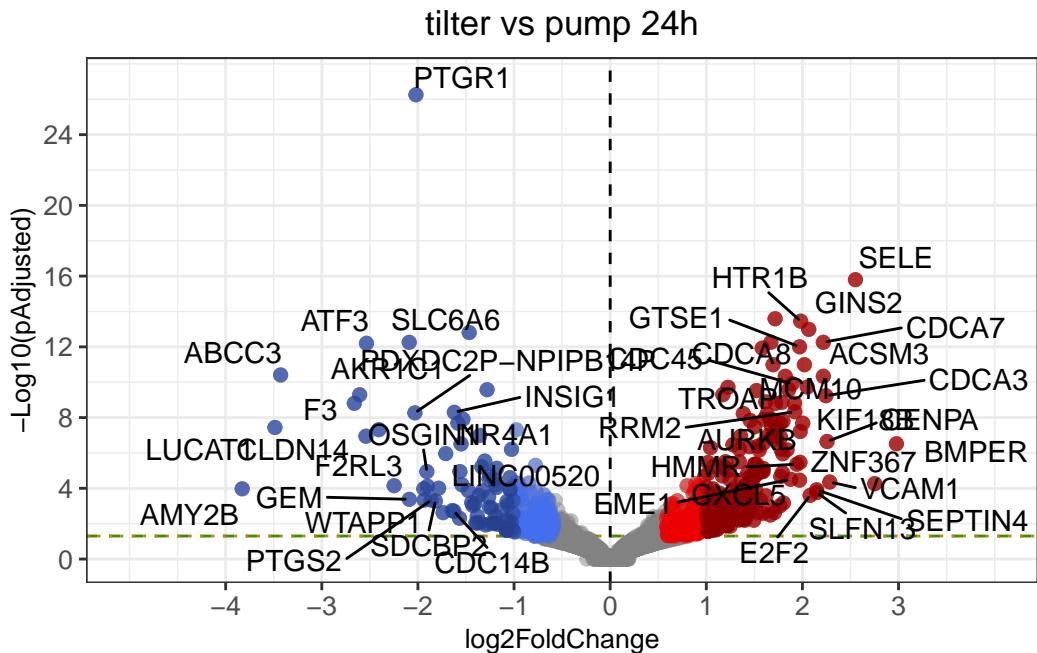
```
[1] "BMPER"      NA        "SELE"      "VCAM1"     "CENPA"     "CDCA3"     "ACSM3"
[8] "CDCA7"      "SLFN13"    "SEPTIN4"   "E2F2"      "GINS2"     "MCM10"     "CDCA8"
[15] "KIF18B"     "HTR1B"     "AURKB"     "ZNF367"    "GTSE1"     "CXCL5"     "HMMR"
[22] "RRM2"       "TROAP"     "CDC45"     "EME1"
```

```
label_set_bottom <- res_uni_bidirectional_24merged %>%
  arrange(desc(log2FoldChange)) %>% tail(25) %>% pull(hgnc_symbol)
label_set_bottom
```

```
[1] "INSIG1"      "CDC14B"     ""          "NR4A1"
[5] ""           "LINC00520"   "SDCBP2"    "PTGS2"
[9] "F2RL3"       "OSGIN1"     "WTAPP1"    ""
[13] "PTGR1"       "PDXDC2P-NPIPB14P" "GEM"       "SLC6A6"
[17] ""           ""           "ATF3"      "CLDN14"
[21] "AKR1C1"     "F3"         "ABCC3"    "LUCAT1"
[25] "AMY2B"
```

```
# combine the top and bottom genes altered genes
label_set <- c(label_set_bottom, label_set_top)
# add the labels to the plot
b2= b +
  geom_text_repel(data = res_uni_bidirectional_24merged
                  [res_uni_bidirectional_24merged$hgnc_symbol %in% label_set,],
                  aes(label=res_uni_bidirectional_24merged
                      [res_uni_bidirectional_24merged$hgnc_symbol %in%
                       label_set,]$hgnc_symbol),
  )
b2
```

Warning: Removed 1219 rows containing missing values (`geom_text_repel()`).



9 GO analysis for DEGs

```

##gene ontology

GO_function <- function(gene_list, pval = 0.05, onto= "MF",
                         prefix="", org="human"){
  require(clusterProfiler)
  require(ReactomePA)
  if(org=="mouse"){
    orgdb <- "org.Mm.eg.db"
    org_reactome <- "mouse"
  }else if(org=="human"){
    orgdb <- "org.Hs.eg.db"
    org_reactome <- "human"
  }else{
    message("Please enter a valid organism (mouse or human)")
  }

  if(onto %in% c("MF", "CC", "BP")){
    ...
  }
}

```

```

compGO <- enrichGO(gene = gene_list, pvalueCutoff = pval,
                    keyType = "SYMBOL",
                    pAdjustMethod = "BH", OrgDb = orgdb, ont = onto)
} else if(onto=="reactome"){
  gene_list <- bitr(gene_list, fromType = "SYMBOL",
                     toType = "ENTREZID", OrgDb = orgdb) #SYMBOL
  gene_list <- gene_list$ENTREZID
  compGO <- enrichPathway(gene = gene_list, pvalueCutoff = 0.05,
                           organism= org_reactome,      readable = TRUE)

} else{
  message("Please enter a valid GO term")
}

if(is.null(compGO)){
  message(paste0("No GO:", onto, " obtained"))
  message(paste0("*****"))
  message(paste0("\n"))

} else {

  compGO_df <- as.data.frame(compGO)
  compGO_df$GeneRatio_decimal <- compGO_df$GeneRatio
  compGO_df$GeneRatio_decimal <- sapply(compGO_df$GeneRatio_decimal,
                                         function(x)
                                           (eval(parse(text
                                                       = as.character(x)))))

  compGO_df$BgRatio_decimal <- compGO_df$BgRatio
  compGO_df$BgRatio_decimal <- sapply(compGO_df$BgRatio_decimal,
                                         function(x)
                                           (eval(parse(text = as.character(x)))))

  compGO_df <- compGO_df %>% tidyverse::separate_rows(geneID,
                                                         sep = "/",
                                                         convert = FALSE) %>%
    arrange(desc(GeneRatio_decimal))
  compGO_df %>% head

  if(nrow(compGO_df)==0){
    message(paste0("No GO:", onto, " obtained"))
    message(paste0("*****"))
  }
}

```

```

    message(paste0("\n"))

} else{

  write.csv(compGO_df, paste0(prefix,"_GO_",onto, "_pathways.csv"))

  full_name= switch(onto,
                     MF= "Molecular Function",
                     CC= "Cellular Components",
                     BP= "Biological Pathways"
  )

  print(dotplot(compGO, showCategory = 15,
                title = paste0("GO Pathway Enrichment Analysis \n",
                               full_name),
                font.size = 6))
  dev.copy(
    pdf,
    file = paste0(prefix,"_GO_",onto, "_pathways.pdf"),
    width = 12,
    height = 16
  )
  dev.off()

}

message(paste0("Pathway analysis GO:",onto, " done"))
message(paste0("*****"))
message(paste0("\n"))

}

}

}

# run the GO function : tilter vs pump at 4h
dir.create("results/GO_analysis/analysis_withoutpb", recursive = TRUE)

```

Warning in dir.create("results/GO_analysis/analysis_withoutpb", recursive =

```
TRUE): 'results/GO_analysis/analysis_withoutpb' already exists
```

```
GO_function((res_uni_bidirectional_4hmerged %>% filter(padj<0.05)
             %>% pull(hgnc_symbol)), onto="MF",
             prefix="results/GO_analysis/pump_tilter_4h" )
```

Loading required package: ReactomePA

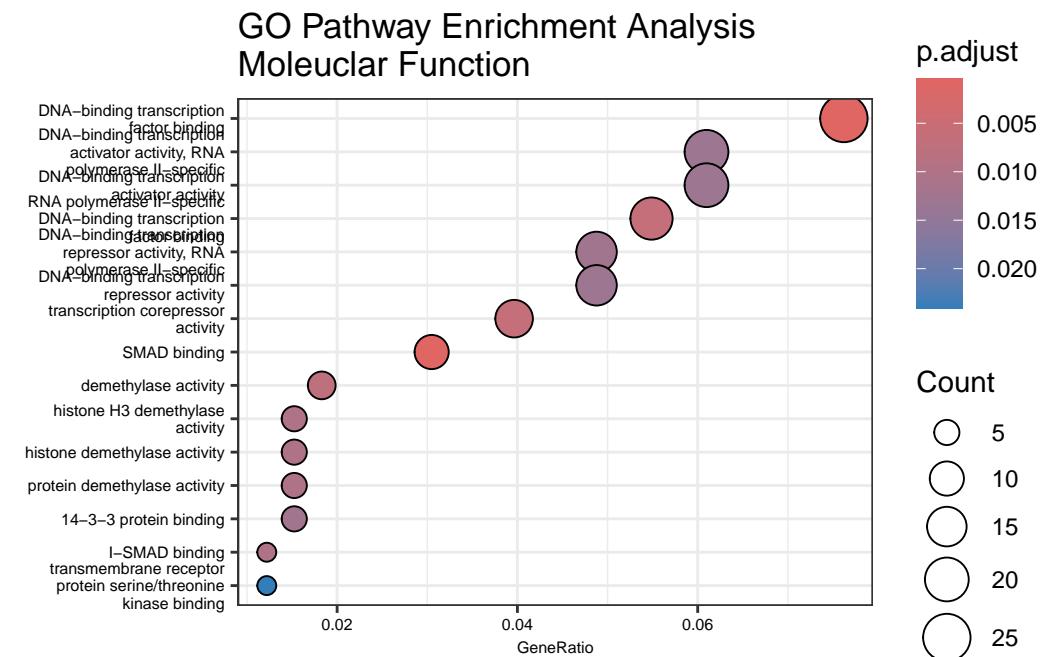
ReactomePA v1.46.0 For help: <https://yulab-smu.top/biomedical-knowledge-mining-book/>

If you use ReactomePA in published research, please cite:

Guangchuang Yu, Qing-Yu He. ReactomePA: an R/Bioconductor package for reactome pathway analysis. Bioinformatics, 2018, 34(18), 3193-3195.

Pathway analysis GO:MF done

```
*****
```



```

GO_function((res_uni_bidirectional_4hmerged %>% filter(padj<0.05)
             %>% pull(hgnc_symbol)), onto="CC",
             prefix="results/GO_analysis/pump_tilter_4h", org = "human" )

```

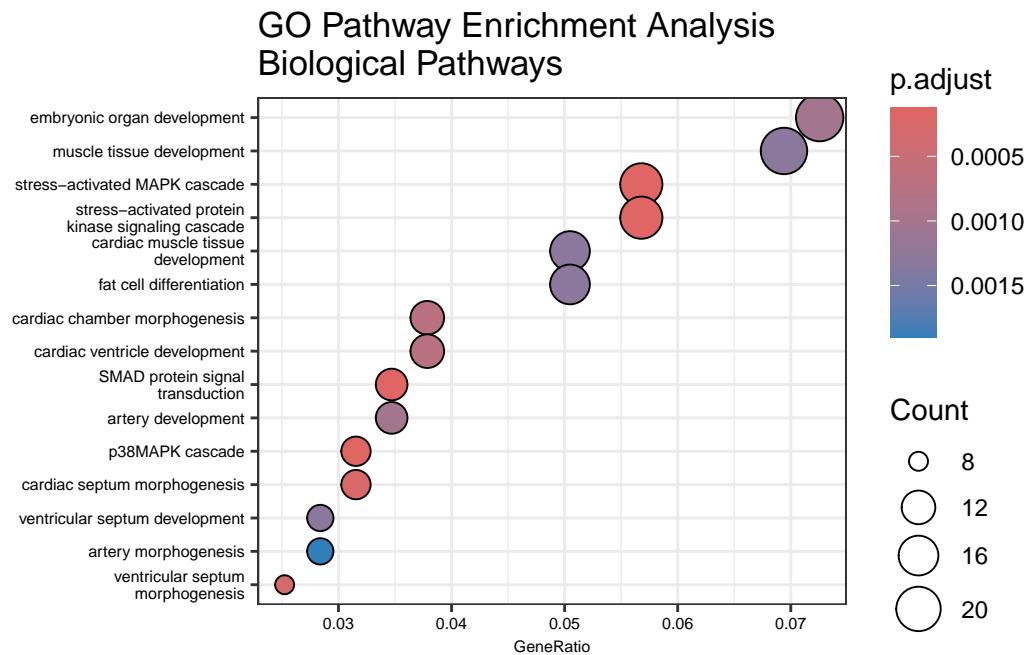
No GO:CC obtained

```

GO_function((res_uni_bidirectional_4hmerged %>% filter(padj<0.05)
             %>% pull(hgnc_symbol)), onto="BP",
             prefix="results/GO_analysis/pump_tilter_4h" )

```

Pathway analysis GO:BP done

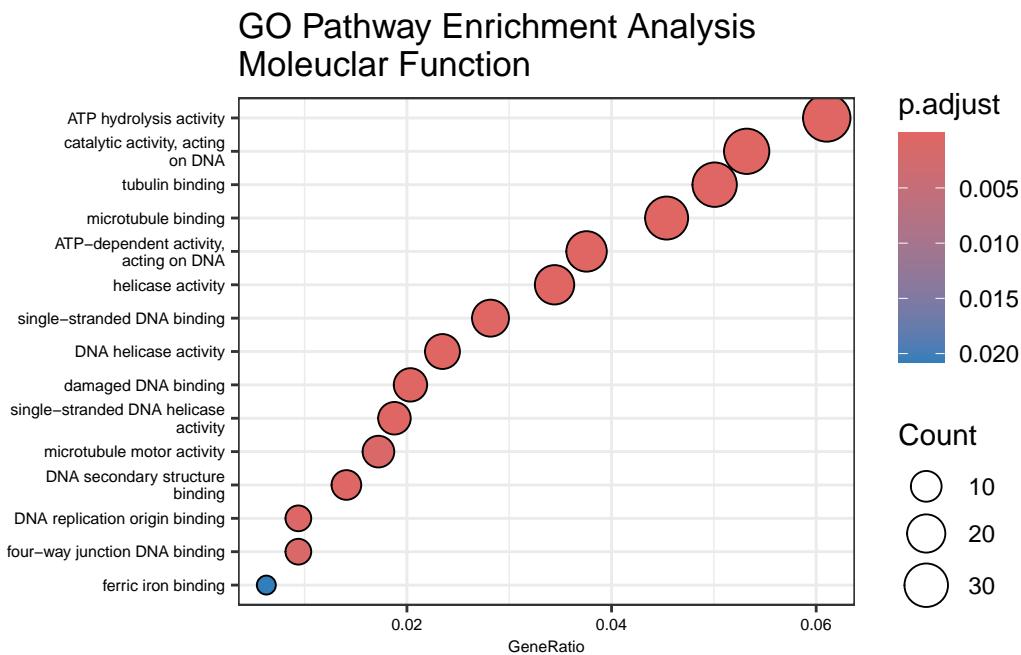


```

##volcano plot : tilter vs pump at 24h
GO_function((res_uni_bidirectional_24merged %>% filter(padj<0.05)
             %>% pull(hgnc_symbol)), onto="MF",
             prefix="results/GO_analysis/analysis_withoutpb/pump_tilter_24h" )

```

Pathway analysis GO:MF done

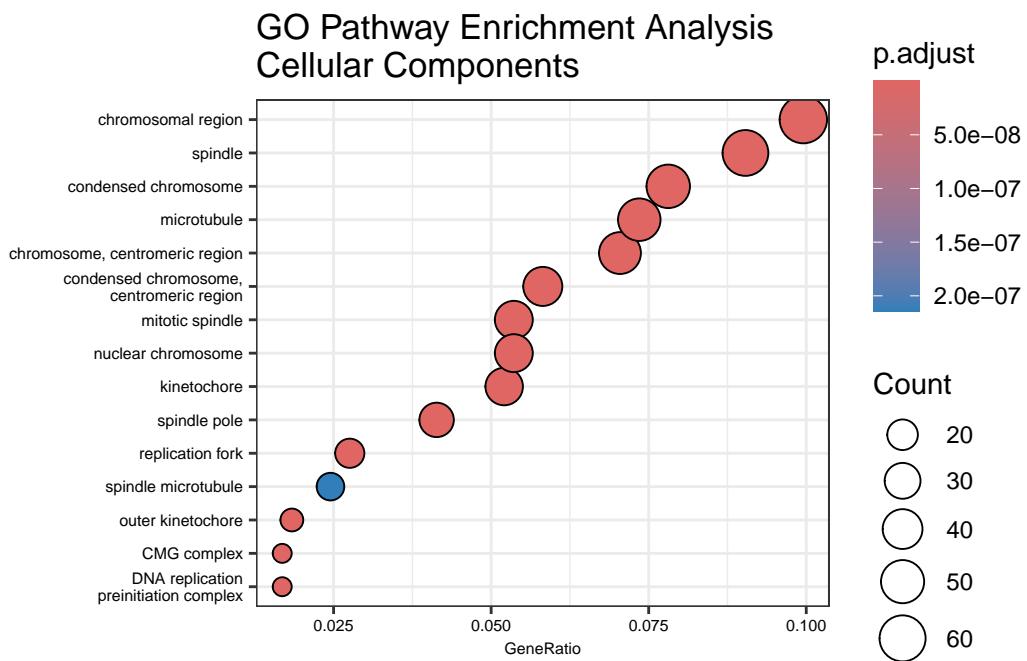


```

GO_function((res_uni_bidirectional_24merged %>% filter(padj<0.05)
             %>% pull(hgnc_symbol)), onto="CC",
             prefix="results/GO_analysis/analysis_withoutpb/pump_tilter_24h",
             org = "human" )

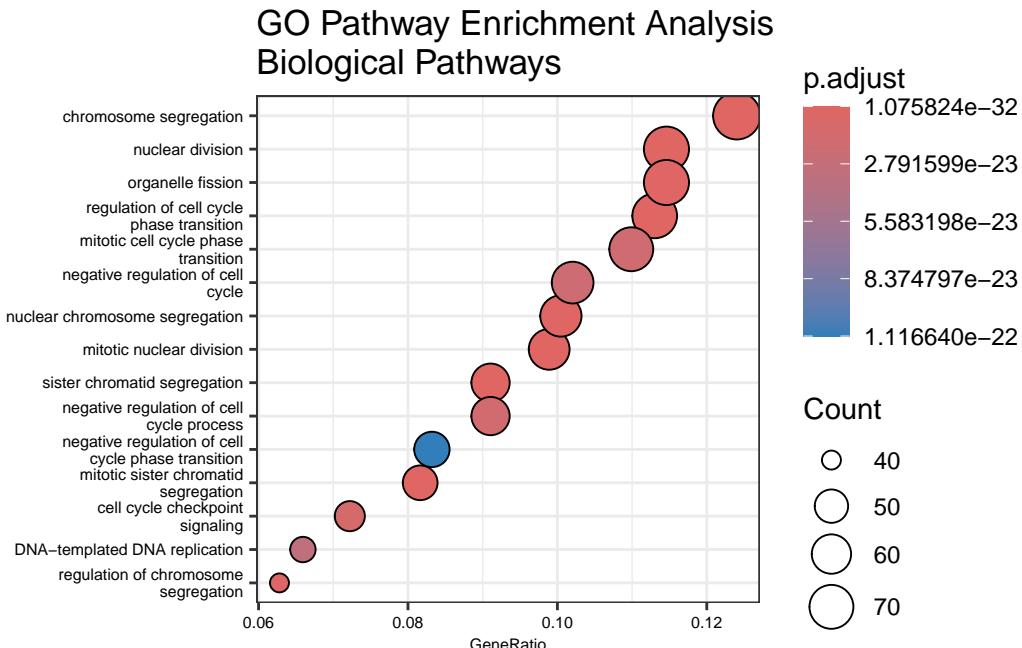
```

Pathway analysis GO:CC done



```
GO_function((res_uni_bidirectional_24merged %>% filter(padj<0.05)
            %>% pull(hgnc_symbol)), onto="BP",
            prefix="results/GO_analysis/analysis_withoutpb/pump_tilter_24h" )
```

Pathway analysis GO:BP done



9.1 GO plots for selected pathways

```
# GO analysis for the selected pathways for tilter and pump at 24h time point
# read the GO analysis results obtained in the previous step
GO_res_uni_bidirectional_24_ <-
  read.csv("/Users/negarvahdani/Deseq2/pump_tilfter_24h_GO_BP_pathways.csv")
head(GO_res_uni_bidirectional_24_)
```

X	ID	Description	GeneRatio	BgRatio	pvalue
1	1	GO:0007059 chromosome segregation	79/637	424/18870	2.302705e-36
2	2	GO:0007059 chromosome segregation	79/637	424/18870	2.302705e-36
3	3	GO:0007059 chromosome segregation	79/637	424/18870	2.302705e-36
4	4	GO:0007059 chromosome segregation	79/637	424/18870	2.302705e-36
5	5	GO:0007059 chromosome segregation	79/637	424/18870	2.302705e-36
6	6	GO:0007059 chromosome segregation	79/637	424/18870	2.302705e-36
p.adjust		qvalue	geneID	Count	GeneRatio_decimal
1	1.075824e-32	9.555014e-33	NCAPD2	79	0.1240188
2	1.075824e-32	9.555014e-33	DPF1	79	0.1240188
3	1.075824e-32	9.555014e-33	BRCA1	79	0.1240188
4	1.075824e-32	9.555014e-33	TACC3	79	0.1240188
5	1.075824e-32	9.555014e-33	RAD18	79	0.1240188

```
6 1.075824e-32 9.555014e-33 TRIP13    79      0.1240188      0.02246953
```

```
# make a list of the select the pathways of interest
selected_pathways_24h <- c('ERK1 and ERK2 cascade',
                           'epithelial cell migration','leukocyte migration',
                           'sprouting angiogenesis','regeneration',
                           'regulation of microtubule cytoskeleton organization',
                           'substrate adhesion-dependent cell spreading',
                           'chemokine production',
                           'microtubule depolymerization',
                           'positive regulation of vasculature development')

# check the selected pathways
selected_pathways_24h
```

```
[1] "ERK1 and ERK2 cascade"
[2] "epithelial cell migration"
[3] "leukocyte migration"
[4] "sprouting angiogenesis"
[5] "regeneration"
[6] "regulation of microtubule cytoskeleton organization"
[7] "substrate adhesion-dependent cell spreading"
[8] "chemokine production"
[9] "microtubule depolymerization"
[10] "positive regulation of vasculature development"
```

```
# subset the GO analysis results to the selected pathways
GO_res_uni_bidirectional_24_ <-
  GO_res_uni_bidirectional_24_[GO_res_uni_bidirectional_24_$Description
                            %in% selected_pathways_24h,]

# arrange the results by GeneRatio_decimal
GO_res_uni_bidirectional_24_ <- GO_res_uni_bidirectional_24_ %>%
  arrange((GeneRatio_decimal))
# set the levels of the Description column
GO_res_uni_bidirectional_24_$Description <-
  factor(GO_res_uni_bidirectional_24_$Description,
         levels=unique(GO_res_uni_bidirectional_24_$Description))

# check the dimensions of the data and the first rows
GO_res_uni_bidirectional_24_ %>% dim
```

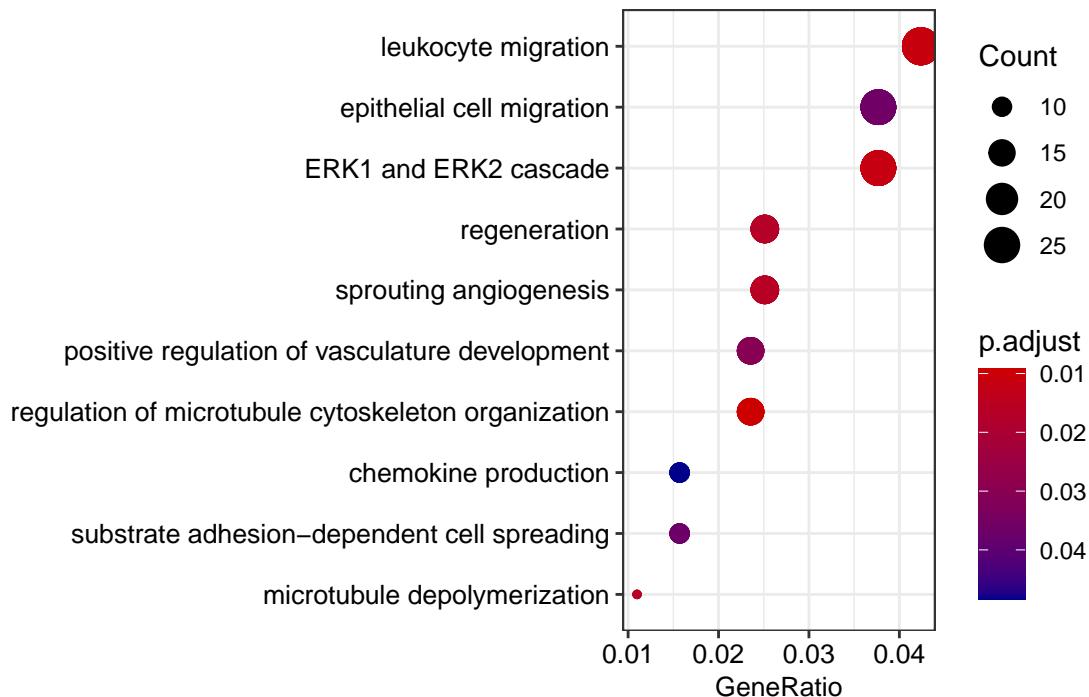
```
[1] 164 12
```

```
GO_res_uni_bidirectional_24_ %>% head
```

X	ID	Description	GeneRatio	BgRatio	pvalue
1	4776	GO:0007019 microtubule depolymerization	7/637	45/18870	0.0007183456
2	4777	GO:0007019 microtubule depolymerization	7/637	45/18870	0.0007183456
3	4778	GO:0007019 microtubule depolymerization	7/637	45/18870	0.0007183456
4	4779	GO:0007019 microtubule depolymerization	7/637	45/18870	0.0007183456
5	4780	GO:0007019 microtubule depolymerization	7/637	45/18870	0.0007183456
6	4781	GO:0007019 microtubule depolymerization	7/637	45/18870	0.0007183456
p.adjust	qvalue	geneID	Count	GeneRatio_decimal	BgRatio_decimal
1	0.01574125	0.01398072	NAV3	7	0.01098901
2	0.01574125	0.01398072	TPX2	7	0.01098901
3	0.01574125	0.01398072	STMN1	7	0.01098901
4	0.01574125	0.01398072	KIF18A	7	0.01098901
5	0.01574125	0.01398072	KIF2C	7	0.01098901
6	0.01574125	0.01398072	NCKAP5	7	0.01098901

```
# set the maximum number of overlaps
options(ggrepel.max.overlaps = 50)
ggplot(data = (GO_res_uni_bidirectional_24_ ) , # make ggplot object
        aes(x = GeneRatio_decimal, y = Description,
            # x axis is GeneRatio_decimal, y axis is Description
            color = p.adjust, size = Count)) +
# color by p.adjust, size of dot by Count
geom_point(alpha = 0.5) + # add points with alpha 0.5
scale_color_gradient(low = "red3", high = "blue4",
                      guide=guide_colourbar(reverse = TRUE)) +
# set color gradient to blue to red
theme_bw() + # set theme to black and white
labs(title = paste0("Tilter vs Pump 24h Selected Biological Processes"),
      x = "GeneRatio", y = "") + # set title and axis labels
theme(plot.title = element_text(hjust = 0.5),
      axis.text = element_text(size = 10),
      # set theme for plot title and axis text
      axis.title.x = element_text(size = 10),
      axis.title.y = element_text(size = 10),
      axis.text.x=element_text(colour="black"),
      axis.text.y=element_text(colour="black")) +
# change color of Counts in legend to black
guides(size = guide_legend(override.aes = list(color = "black", alpha=1)))
```

Tilter vs Pump 24h Selected Biological Process



```
# GO analysis for the selected pathways for tilter and pump at 4h time point
# read the GO analysis results obtained in the previous step
GO_res_uni_bidirectional_4_ <-
  read.csv("/Users/negarvahdani/Deseq2/pump_tilter_4h_GO_BP_pathways.csv")
#make a list of the select the pathways of interest
selected_pathways_4h <-
  c('transmembrane receptor protein serine/threonine kinase signaling pathway',
    'ERK1 and ERK2 cascade','p38MAPK cascade','stress-activated MAPK cascade',
    'BMP signaling pathway','SMAD protein signal transduction','JNK cascade',
    'canonical NF-kappaB signal transduction')
# check the selected pathways
selected_pathways_4h
```

```
[1] "transmembrane receptor protein serine/threonine kinase signaling pathway"
[2] "ERK1 and ERK2 cascade"
[3] "p38MAPK cascade"
[4] "stress-activated MAPK cascade"
[5] "BMP signaling pathway"
[6] "SMAD protein signal transduction"
[7] "JNK cascade"
```

```
[8] "canonical NF-kappaB signal transduction"

# subset the GO analysis results to the selected pathways
GO_res_uni_bidirectional_4_ <-
  GO_res_uni_bidirectional_4_[GO_res_uni_bidirectional_4_$Description
                            %in% selected_pathways_4h,]
GO_res_uni_bidirectional_4_ <- GO_res_uni_bidirectional_4_ %>%
  arrange((GeneRatio_decimal)) # arrange the results by GeneRatio_decimal
# set the levels of the Description column
GO_res_uni_bidirectional_4_$Description <-
  factor(GO_res_uni_bidirectional_4_$Description,
         levels=unique(GO_res_uni_bidirectional_4_$Description))
GO_res_uni_bidirectional_4_ %>% dim # check the dimensions of the data

[1] 90 12

GO_res_uni_bidirectional_4_ %>% head # check the first rows

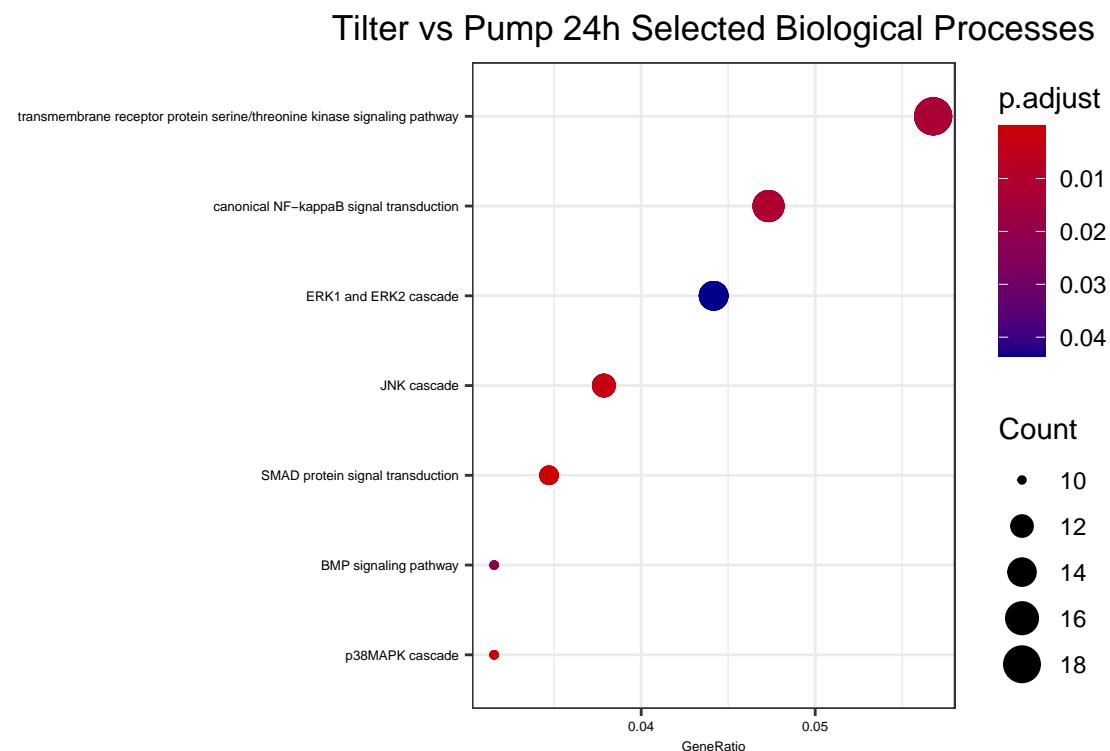
      X           ID Description GeneRatio   BgRatio      pvalue    p.adjust
1 931 GO:0038066 p38MAPK cascade     10/317 57/18870 3.335848e-08 0.0001227592
2 932 GO:0038066 p38MAPK cascade     10/317 57/18870 3.335848e-08 0.0001227592
3 933 GO:0038066 p38MAPK cascade     10/317 57/18870 3.335848e-08 0.0001227592
4 934 GO:0038066 p38MAPK cascade     10/317 57/18870 3.335848e-08 0.0001227592
5 935 GO:0038066 p38MAPK cascade     10/317 57/18870 3.335848e-08 0.0001227592
6 936 GO:0038066 p38MAPK cascade     10/317 57/18870 3.335848e-08 0.0001227592
      qvalue geneID Count GeneRatio_decimal BgRatio_decimal
1 0.0001045349 MAP2K3    10       0.03154574    0.003020668
2 0.0001045349 PHLPP1    10       0.03154574    0.003020668
3 0.0001045349 BMP4     10       0.03154574    0.003020668
4 0.0001045349 ZFP36    10       0.03154574    0.003020668
5 0.0001045349 DAB2IP    10       0.03154574    0.003020668
6 0.0001045349 MFHAS1   10       0.03154574    0.003020668

# create a dotplot for the GO terms with GeneRatio on x axis decreasing
# gene ratio , and GO term on y axis, color by p.adjust size of dot by Count
# set the maximum number of overlaps
options(ggrepel.max.overlaps = 50)
ggplot(data = (GO_res_uni_bidirectional_4_ ) , # make ggplot object
       aes(x = GeneRatio_decimal, y = Description,
            # x axis is GeneRatio_decimal, y axis is Description
            color = p.adjust, size = Count)) +
```

```

# color by p.adjust, size of dot by Count
geom_point(alpha = 0.5) + # add points with alpha 0.5
scale_color_gradient(low = "red3", high = "blue4",
                      guide=guide_colourbar(reverse = TRUE)) +
# set color gradient to blue to red
theme_bw() + # set theme to black and white
labs(title = paste0("Tilter vs Pump 24h Selected Biological Processes"),
      x = "GeneRatio", y = "") + # set title and axis labels
theme(plot.title = element_text(hjust = 0.5),
      axis.text = element_text(size = 5),
      # set theme for plot title and axis text
      axis.title.x = element_text(size = 5),
      axis.title.y = element_text(size = 5),
      axis.text.x=element_text(colour="black"),
      axis.text.y=element_text(colour="black")) +
# change color of Counts in legend to black
guides(size = guide_legend(override.aes = list(color = "black", alpha=1)))

```



10 Plot Heatmap for figure 2

```

# create a variance stabilized data
vst_data <- vst(dds_flow_time, blind = FALSE)
# create a matrix of the variance stabilized data
heatmap_matrix <- assay(vst_data)
degs <- rbind(res_uni_bidirectional_4_padj0.05,
               res_uni_bidirectional_24_padj0.05)
degs %>% head %>% print

```

	baseMean	log2FoldChange	lfcSE	pvalue	padj
ENSG00000132881	57.76324	0.8568570	0.5505004	2.173710e-04	1.487556e-02
ENSG00000162407	1978.62537	-0.5549861	0.2070021	3.110924e-04	1.865012e-02
ENSG00000127124	368.14958	-1.2764537	0.3464764	2.095402e-07	7.600025e-05
ENSG00000023909	2084.13720	-1.4496330	0.2912663	3.083979e-10	6.114980e-07
ENSG00000143315	275.00192	0.7384807	0.3265591	2.064510e-04	1.426281e-02
ENSG00000132763	117.11548	0.6430754	0.4443382	8.166160e-04	3.859109e-02

```

heatmap_matrix_degs <- heatmap_matrix[rownames(heatmap_matrix)
                                       %in% rownames(degs), ]

# Step 2: Calculate z-scores for the DEGs
heatmap_matrix_degs_z <- t(scale(t(heatmap_matrix_degs)))
# subset the heatmap matrix to the selected columns
heatmap_matrix_all_subset_col <-
  heatmap_matrix_degs_z[, c("e3_4h_t", "e4_4h_t", "e6_4h_t", "e3_4h_p",
                            "e4_4h_p", "e6_4h_p", "e3_24h_t", "e4_24h_t", "e6_24h_t",
                            "e3_24h_p", "e4_24h_p", "e6_24h_p")]

colnames(heatmap_matrix_all_subset_col) # check the column names

```

```

[1] "e3_4h_t"   "e4_4h_t"   "e6_4h_t"   "e3_4h_p"   "e4_4h_p"   "e6_4h_p"
[7] "e3_24h_t"  "e4_24h_t"  "e6_24h_t"  "e3_24h_p"  "e4_24h_p"  "e6_24h_p"

```

```

# create the heatmap annotation
ha = HeatmapAnnotation(
  Genotype = rep(c("4h_t", "4h_p", "24h_t", "24h_p"), each=3),
  col = list(Genotype = c(
    "4h_t"="#DE9F9D", "4h_p"="#B5C7DC", "24h_t" ="#E57A77", "24h_p"="#6391C5"
  ))

```

```

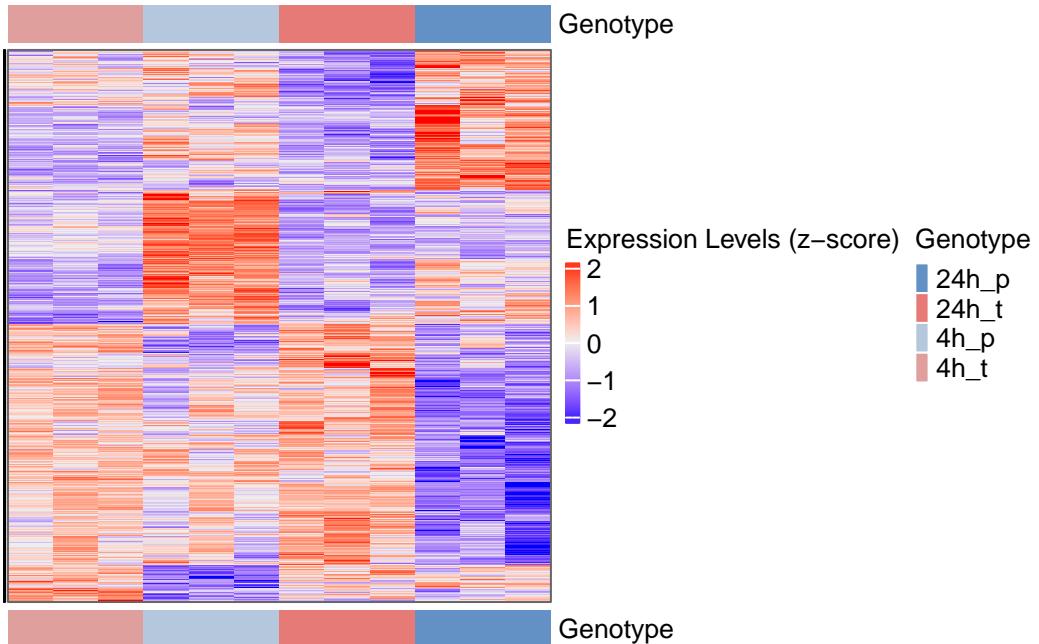
    )

),

# change font size of the annotation
annotation_name_gp = grid::gpar(fontsize = 10),
annotation_legend_param = list(
  # title = "Expression Levels",
  title_gp = gpar(fontsize = 10),
  labels_gp = gpar(fontsize = 10),
  grid_width = unit(5,"point"),
  grid_height = unit(5,"point")
)
)

Heatmap(heatmap_matrix_all_subset_col, name="Expression Levels",
        cluster_columns = FALSE,
        cluster_rows = TRUE,
        clustering_distance_columns = "pearson",
        clustering_method_columns = "complete",
        row_dend_width = unit(0, "cm"),
        show_column_names = FALSE,
        show_row_names = FALSE,
        top_annotation = ha,
        gap=unit(2, "point"),
        border= TRUE,
        border_gp = gpar(col = "grey40"),
        row_title_gp = gpar(fontsize = 5), # change row split font size
        row_title_rot = 0,
        bottom_annotation = ha,
        heatmap_legend_param = list(
          title = "Expression Levels (z-score)",
          title_gp = gpar(fontsize = 10),
          labels_gp = gpar(fontsize = 10),
          grid_width = unit(5,"point"),
          grid_height = unit(5,"point")
        ),
        row_names_gp = grid::gpar(fontsize = 14)
)

```



11 Plot Heatmap for revision- selected pathways for 24h time point

```
# create a variance stabilized data
vst_data <- vst(dds_flow_time, blind = FALSE)
# create a matrix of the variance stabilized data
heatmap_matrix <- assay(vst_data)
heatmap_matrix <- as.data.frame(heatmap_matrix)
#add the row names as a new column
heatmap_matrix$ensembl_gene_id <- rownames(heatmap_matrix)
# merge the data with the gene names
heatmap_matrix_merged <- merge(heatmap_matrix, gene_id_df,
                                by.x = "ensembl_gene_id", all.x=TRUE)

head(heatmap_matrix_merged) %>% print
```

	ensembl_gene_id	e3_24h_p	e3_24h_t	e3_4h_p	e3_4h_t	e4_24h_p	e4_24h_t
1	ENSG000000000003	9.202584	9.332839	8.870857	9.265021	9.239380	9.679730
2	ENSG000000000005	6.607224	6.607224	6.607224	6.607224	6.607224	6.607224
3	ENSG00000000419	10.398201	10.389236	10.677285	10.170131	10.034811	10.395053
4	ENSG00000000457	8.334541	8.423656	8.113262	8.356905	8.520805	8.088652

```

5 ENSG00000000460 8.261697 8.481267 8.613920 8.444441 8.203979 8.308101
6 ENSG00000000938 6.607224 6.920301 7.009329 6.835800 6.607224 6.607224
  e4_4h_p   e4_4h_t e6_24h_p e6_24h_t e6_4h_p   e6_4h_t hgnc_symbol
1 9.406244 9.317567 9.376216 8.981418 9.314274 9.362230      TSPAN6
2 6.607224 6.607224 6.607224 6.607224 6.607224 6.607224      TNMD
3 10.163925 10.123110 9.982052 9.596050 9.502058 9.444631      DPM1
4 8.482199 8.462258 8.350069 8.168159 8.530794 8.512267      SCYL3
5 8.620917 8.828837 7.807006 8.692401 8.129087 8.339573      FIRRM
6 6.848306 6.607224 7.007689 6.607224 6.994157 6.607224      FGR

```

```

heatmap_matrix_merged_subset24 <-
  heatmap_matrix_merged[, c("e3_24h_t", "e4_24h_t", "e6_24h_t",
                           "e3_24h_p", "e4_24h_p", "e6_24h_p",
                           "hgnc_symbol", "ensembl_gene_id")]
head(heatmap_matrix_merged_subset24) %>% print

```

	e3_24h_t	e4_24h_t	e6_24h_t	e3_24h_p	e4_24h_p	e6_24h_p	hgnc_symbol
1	9.332839	9.679730	8.981418	9.202584	9.239380	9.376216	TSPAN6
2	6.607224	6.607224	6.607224	6.607224	6.607224	6.607224	TNMD
3	10.389236	10.395053	9.596050	10.398201	10.034811	9.982052	DPM1
4	8.423656	8.088652	8.168159	8.334541	8.520805	8.350069	SCYL3
5	8.481267	8.308101	8.692401	8.261697	8.203979	7.807006	FIRRM
6	6.920301	6.607224	6.607224	6.607224	6.607224	7.007689	FGR
	ensembl_gene_id						
1	ENSG00000000003						
2	ENSG00000000005						
3	ENSG00000000419						
4	ENSG00000000457						
5	ENSG00000000460						
6	ENSG00000000938						

```
dim(heatmap_matrix_merged_subset24)
```

```
[1] 62758     8
```

```

# remove the NAs and the empty strings

heatmap_matrix_merged_subset24_cleaned <-
  heatmap_matrix_merged_subset24[!is.na(heatmap_matrix_merged_subset24$hgnc_symbol), ]
heatmap_matrix_merged_subset24_cleaned <-

```

```

  heatmap_matrix_merged_subset24_cleaned %>% filter(hgnc_symbol != "")
heatmap_matrix_merged_subset24_cleaned %>% filter(hgnc_symbol == "") %>% dim

[1] 0 8

dim(heatmap_matrix_merged_subset24_cleaned)

[1] 41182      8

sum(is.na(heatmap_matrix_merged_subset24_cleaned$hgnc_symbol))

[1] 0

# check for duplicates
duplicate <-
  heatmap_matrix_merged_subset24_cleaned[
    duplicated(heatmap_matrix_merged_subset24_cleaned$hgnc_symbol)
      | duplicated(
        heatmap_matrix_merged_subset24_cleaned$hgnc_symbol,
        fromLast=TRUE),]

df_filtered <- heatmap_matrix_merged_subset24_cleaned %>%
  group_by(hgnc_symbol) %>%
  rowwise() %>%
  filter(n_distinct(c_across(c("e3_24h_t","e4_24h_t","e6_24h_t",
    "e3_24h_p","e4_24h_p","e6_24h_p")))) > 1) %>%
  ungroup()

df_duplicated <-
  df_filtered[duplicated(df_filtered$hgnc_symbol)
    | duplicated(df_filtered$hgnc_symbol, fromLast=TRUE),]

# List of Ensembl Gene IDs to exclude from the DataFrame
gene_ids_to_exclude <- c("ENSG00000161912", "ENSG00000291140",
  "ENSG00000292339", "ENSG00000124549",
  "ENSG00000205771", "ENSG00000154874",
  "ENSG00000292362", "ENSG00000250959",
  "ENSG00000258137", "ENSG00000292358",
  "ENSG00000206149", "ENSG00000258724",
  "ENSG00000168255", "ENSG00000292327",

```

```

"ENSG00000205763", "ENSG00000189136",
"ENSG00000259658", "ENSG00000259429",
"ENSG00000124333", "ENSG00000292372",
"ENSG00000214717", "ENSG00000197083",
"ENSG00000198155", "ENSG00000228903",
"ENSG00000214331")

# Remove rows based on the column ensembl_gene_id
df_filtered <-
  df_filtered[!df_filtered$ensembl_gene_id %in% gene_ids_to_exclude, ]
df_duplicated_after <-
  df_filtered[duplicated(df_filtered$hgnc_symbol)
  |duplicated(df_filtered$hgnc_symbol, fromLast=TRUE),]

df_filtered <- as.data.frame(df_filtered)

#set row names using the hgnc_symbol column
rownames(df_filtered) <- df_filtered$hgnc_symbol
df_filtered$hgnc_symbol <- NULL

degs_24 <- res_uni_bidirectional_24merged_padj0.05
dim(degs_24)

```

[1] 761 7

```

degs_cleaned <- degs_24[!duplicated(degs_24$hgnc_symbol),]#
dim(degs_cleaned)

```

[1] 717 7

```

#remove the missing values in degs_cleaed
degs_cleaned <- degs_cleaned[!is.na(degs_cleaned$hgnc_symbol), ]
rownames(degs_cleaned) <- degs_cleaned$hgnc_symbol

heatmap_matrix_degs_24 <- df_filtered[rownames(df_filtered)
                                         %in% rownames(degs_cleaned), ]
head(heatmap_matrix_degs_24) %>% print

```

e3_24h_t e4_24h_t e6_24h_t e3_24h_p e4_24h_p e6_24h_p

GCLC	9.155070	9.408026	8.742333	10.120618	9.705010	9.654014
LAS1L	10.139252	10.229997	10.315455	9.368936	9.584177	9.828903
CFLAR	12.606448	12.565054	12.452735	12.958229	13.141769	13.016731
THSD7A	9.604160	9.995055	9.629444	8.660360	9.136361	8.668289
ACSM3	9.039761	8.983977	8.430656	7.758601	7.509860	7.606982
KDM7A	8.922242	9.272116	8.599267	10.113301	9.404068	9.640206
	ensembl_gene_id					
GCLC	ENSG000000001084					
LAS1L	ENSG000000001497					
CFLAR	ENSG000000003402					
THSD7A	ENSG000000005108					
ACSM3	ENSG000000005187					
KDM7A	ENSG000000006459					

```

genes_to_annotate <- c('SELE', 'KITLG', 'STK10', 'TGFB2', 'HMOX1', 'FLT1',
                      'IL1A', 'PGF', 'TNFSF18', 'EDNRB', 'IL33', 'CH25H',
                      'LYST', 'ADAM17', 'VCAM1', 'CXCL5', 'ITGA2', 'LGALS9',
                      'SELP', 'ANO6', 'GBA1', 'CALR', 'CSF1', 'SELL',
                      'HMGB1', 'CCL23', 'CCL16', 'CFLAR', 'CEACAM1', 'NOX4',
                      'FAM83D', 'HMGCR', 'GBP1', 'DUSP4', 'BMP4', 'SPRY2',
                      'ACKR3', 'MFHAS1', 'CNKSR3', 'ATF3', 'NEK10', 'BMPER',
                      'CDK1', 'TNFAIP8L3', 'IQGAP3', 'MIR126', 'ANLN', 'PTGS2',
                      'ANGPT2', 'RGCC', 'SASH1', 'PIK3R3', 'PROX1', 'NR4A1',
                      'EFNB2', 'MACF1', 'KLF4', 'HAS2', 'NANOS1', 'FGF16',
                      'E2F2', 'RECK', 'E2F8', 'TSPAN18', 'E2F7', 'AURKA',
                      'PCNA', 'POSTN', 'LPIN1', 'NREP', 'ATIC', 'CCNA2', 'ADM',
                      'NNMT', 'TYMS', 'DHFR', 'SPAAR', 'TACC3', 'NAV3', 'SPAG5',
                      'MAP2', 'TPX2', 'STMN1', 'KIF18A', 'STIL', 'HAUS8',
                      'PSRC1', 'SKA1', 'SKA3', 'PLK1', 'FAM107A', 'NUP62',
                      'BRCA1', 'HIPK2', 'F3', 'AGO2', 'ADM2', 'ST6GAL1',
                      'UNC13D', 'MERTK', 'AKIP1', 'PEAK1', 'S100A10', 'OAS1',
                      'OAS3', 'IL1RL1', 'KIF2C', 'NCKAP5', 'KIF18B')

row_annotation <-
  rowAnnotation(mark = anno_mark(at = which(rownames(heatmap_matrix_degs_24)
                                             %in% genes_to_annotate),
                                    labels =
                                    rownames(heatmap_matrix_degs_24)
                                    [rownames(heatmap_matrix_degs_24)
                                     %in% genes_to_annotate],
                                    labels_gp
                                    = gpar(fontsize = 10, fontface = "bold"),

```

```

padding = unit(2, "mm")))

#z score normalization
heatmap_matrix_degs_24 <-
  subset(heatmap_matrix_degs_24,
         select = -ensembl_gene_id)
heatmap_matrix_degs_24 <- as.matrix(heatmap_matrix_degs_24)
heatmap_matrix_degs_24 <- t(scale(t(heatmap_matrix_degs_24)))

# check how many NAs are there
sum(is.na(heatmap_matrix_degs))

```

[1] 0

```

ha = HeatmapAnnotation(
  Genotype = rep(c("24h_t", "24h_p"), each=3),
  col = list(Genotype = c(
    "24h_t" = "#E57A77", "24h_p" = "#6391C5"
  )),
  # change font size of the annotation
  annotation_name_gp = grid::gpar(fontsize = 10),
  annotation_legend_param = list(
    # title = "Expression Levels",
    title_gp = gpar(fontsize = 10),
    labels_gp = gpar(fontsize = 10),
    grid_width = unit(5, "point"),
    grid_height = unit(5, "point")
  )
)

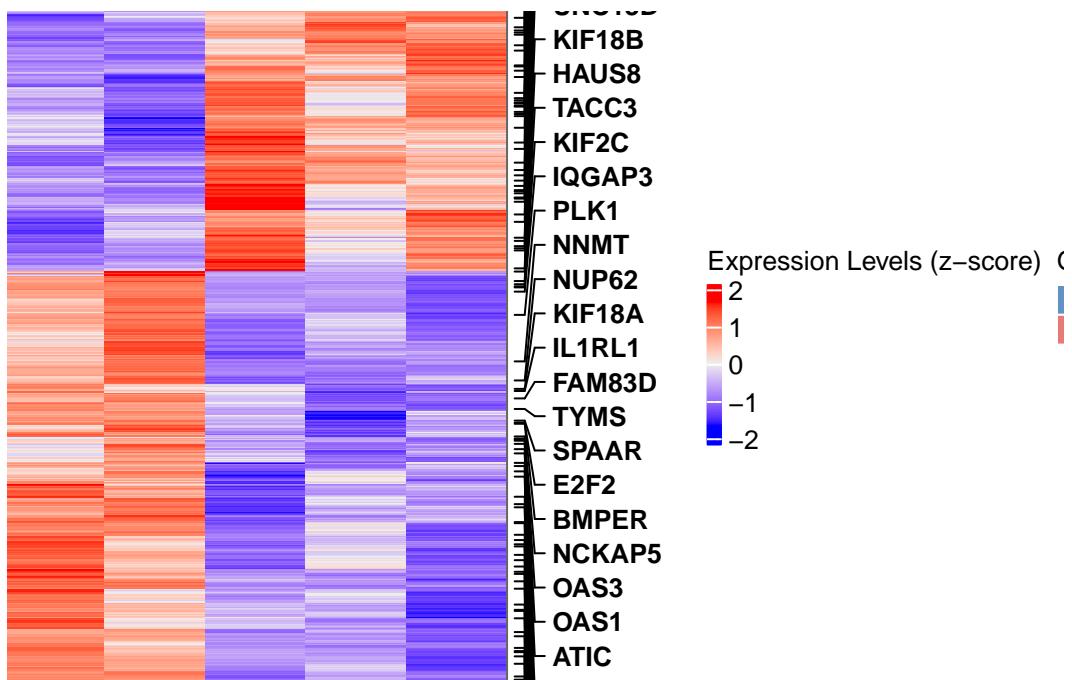
Heatmap(heatmap_matrix_degs_24, name="Expression Levels",
        cluster_columns = FALSE,
        cluster_rows = TRUE,
        clustering_distance_columns = "pearson",
        clustering_method_columns = "complete",
        row_dend_width = unit(0, "cm"),
        show_column_names = FALSE,
        show_row_names = FALSE,
        top_annotation = ha,

```

```

gap=unit(2, "point"),
border= TRUE,
border_gp = gpar(col = "grey40"),
row_title_gp = gpar(fontsize = 5),
row_title_rot = 0,
bottom_annotation = ha,
right_annotation = row_annotation,
width = unit(8, "cm"),
height = unit(10, "cm"),
heatmap_legend_param = list(
  title = "Expression Levels (z-score)",
  title_gp = gpar(fontsize = 10),
  labels_gp = gpar(fontsize = 10),
  grid_width    = unit(5,"point"),
  grid_height   = unit(5,"point")
),
row_names_gp = grid::gpar(fontsize = 14)
)

```



12 Plot Heatmap for revision- selected pathways for 4h time point -

```
# create a variance stabilized data
vst_data <- vst(dds_flow_time, blind = FALSE)
# create a matrix of the variance stabilized data
heatmap_matrix <- assay(vst_data)
heatmap_matrix <- as.data.frame(heatmap_matrix)
#add the row names as a new column
heatmap_matrix$ensembl_gene_id <- rownames(heatmap_matrix)
# merge the data with the gene names
heatmap_matrix_merged <- merge(heatmap_matrix, gene_id_df,
                                by.x = "ensembl_gene_id", all.x=TRUE)

head(heatmap_matrix_merged) %>% print
```

	ensembl_gene_id	e3_24h_p	e3_24h_t	e3_4h_p	e3_4h_t	e4_24h_p	e4_24h_t
1	ENSG000000000003	9.202584	9.332839	8.870857	9.265021	9.239380	9.679730
2	ENSG000000000005	6.607224	6.607224	6.607224	6.607224	6.607224	6.607224
3	ENSG00000000419	10.398201	10.389236	10.677285	10.170131	10.034811	10.395053
4	ENSG00000000457	8.334541	8.423656	8.113262	8.356905	8.520805	8.088652
5	ENSG00000000460	8.261697	8.481267	8.613920	8.444441	8.203979	8.308101
6	ENSG00000000938	6.607224	6.920301	7.009329	6.835800	6.607224	6.607224
	ensembl_gene_id	e3_24h_p	e3_24h_t	e3_4h_p	e3_4h_t	e4_24h_p	e4_24h_t
1	ENSG000000000003	9.202584	9.332839	8.870857	9.265021	9.239380	9.679730
2	ENSG000000000005	6.607224	6.607224	6.607224	6.607224	6.607224	6.607224
3	ENSG00000000419	10.398201	10.389236	10.677285	10.170131	10.034811	10.395053
4	ENSG00000000457	8.334541	8.423656	8.113262	8.356905	8.520805	8.088652
5	ENSG00000000460	8.261697	8.481267	8.613920	8.444441	8.203979	8.308101
6	ENSG00000000938	6.607224	6.920301	7.009329	6.835800	6.607224	6.607224
	ensembl_gene_id	e3_24h_p	e3_24h_t	e3_4h_p	e3_4h_t	e4_24h_p	e4_24h_t
1	ENSG000000000003	9.202584	9.332839	8.870857	9.265021	9.239380	9.679730
2	ENSG000000000005	6.607224	6.607224	6.607224	6.607224	6.607224	6.607224
3	ENSG00000000419	10.398201	10.389236	10.677285	10.170131	10.034811	10.395053
4	ENSG00000000457	8.334541	8.423656	8.113262	8.356905	8.520805	8.088652
5	ENSG00000000460	8.261697	8.481267	8.613920	8.444441	8.203979	8.308101
6	ENSG00000000938	6.607224	6.920301	7.009329	6.835800	6.607224	6.607224

```
# subset the heatmap matrix to the selected columns
heatmap_matrix_merged_subset4 <-
  heatmap_matrix_merged[, c("e3_4h_t", "e4_4h_t", "e6_4h_t",
                           "e3_4h_p", "e4_4h_p", "e6_4h_p",
                           "hgnc_symbol", "ensembl_gene_id")]
head(heatmap_matrix_merged_subset4) %>% print
```

	e3_4h_t	e4_4h_t	e6_4h_t	e3_4h_p	e4_4h_p	e6_4h_p	hgnc_symbol
1	9.265021	9.317567	9.362230	8.870857	9.406244	9.314274	TSPAN6
2	6.607224	6.607224	6.607224	6.607224	6.607224	6.607224	TNMD

```
3 10.170131 10.123110 9.444631 10.677285 10.163925 9.502058      DPM1
4 8.356905  8.462258 8.512267  8.113262  8.482199 8.530794      SCYL3
5 8.444441  8.828837 8.339573  8.613920  8.620917 8.129087      FIRRM
6 6.835800  6.607224 6.607224  7.009329  6.848306 6.994157      FGR
  ensembl_gene_id
1 ENSG000000000003
2 ENSG000000000005
3 ENSG000000000419
4 ENSG000000000457
5 ENSG000000000460
6 ENSG000000000938
```

```
dim(heatmap_matrix_merged_subset4)
```

```
[1] 62758     8
```

```
heatmap_matrix_merged_subset4_cleaned <-
  heatmap_matrix_merged_subset4[!is.na(heatmap_matrix_merged_subset4$hgnc_symbol)
                               , ]
heatmap_matrix_merged_subset4_cleaned <-
  heatmap_matrix_merged_subset4_cleaned %>% filter(hgnc_symbol != "")
dim(heatmap_matrix_merged_subset4_cleaned)
```

```
[1] 41182     8
```

```
sum(is.na(heatmap_matrix_merged_subset4_cleaned$hgnc_symbol))
```

```
[1] 0
```

```
heatmap_matrix_merged_subset4_cleaned %>% filter(hgnc_symbol == "") %>% dim
```

```
[1] 0 8
```

```
duplicate_4 <-
  heatmap_matrix_merged_subset4_cleaned[
    duplicated(heatmap_matrix_merged_subset4_cleaned$hgnc_symbol)
    | duplicated(
      heatmap_matrix_merged_subset4_cleaned$hgnc_symbol,
```

```

fromLast=TRUE),]

df_filtered_4 <- heatmap_matrix_merged_subset4_cleaned %>%
  group_by(hgnc_symbol) %>%
  rowwise() %>%
  filter(n_distinct(c_across(c("e3_4h_t","e4_4h_t","e6_4h_t",
    "e3_4h_p","e4_4h_p","e6_4h_p")))) > 1) %>%
  ungroup()
df_filtered_4 %>% dim

```

[1] 21426 8

```

df_duplicated_4 <-
  df_filtered_4[duplicated(df_filtered_4$hgnc_symbol)
    |duplicated(df_filtered_4$hgnc_symbol,
      fromLast=TRUE),]

# List of Ensembl Gene IDs to exclude from the DataFrame
gene_ids_to_exclude_4 <- c("ENSG00000161912", "ENSG00000291140",
  "ENSG00000292339", "ENSG00000124549",
  "ENSG00000205771", "ENSG00000154874",
  "ENSG00000292362", "ENSG00000250959",
  "ENSG00000258137", "ENSG00000292358",
  "ENSG00000206149", "ENSG00000258724",
  "ENSG00000168255", "ENSG00000292327",
  "ENSG00000205763", "ENSG00000189136",
  "ENSG00000259658", "ENSG00000259429",
  "ENSG00000124333", "ENSG00000292372",
  "ENSG00000214717", "ENSG00000197083",
  "ENSG00000198155", "ENSG00000225492",
  "ENSG00000238105", "ENSG00000243753",
  "ENSG00000236871", "ENSG00000292342",
  "ENSG00000182378", "ENSG00000292334",
  "ENSG00000228903", "ENSG00000214331")

heatmap_matrix_merged_subset4_cleaned <-
  heatmap_matrix_merged_subset4_cleaned[!heatmap_matrix_merged_subset4_cleaned$ensembl_gene_id
    %in% gene_ids_to_exclude,]

df_filtered_4 <-
  df_filtered_4[!df_filtered_4$ensembl_gene_id
    %in% gene_ids_to_exclude_4,]

```

```

df_duplicated_after_4 <-
  df_filtered_4[duplicated(df_filtered_4$hgnc_symbol)
    |duplicated(df_filtered_4$hgnc_symbol,
      fromLast=TRUE),]
# Convert tibble to a data frame
df_filtered_4 <- as.data.frame(df_filtered_4)

# Set row names using the hgnc_symbol column
rownames(df_filtered_4) <- df_filtered_4$hgnc_symbol

# Optionally, you can remove the hgnc_symbol column if you don't need it anymore
df_filtered_4$hgnc_symbol <- NULL

degs_4 <- res_uni_bidirectional_4hmerged_padj0.05
dim(degs_4)

```

[1] 365 7

```

degs_cleaned_4 <- degs_4[!duplicated(degs_4$hgnc_symbol),]
dim(degs_cleaned_4)

```

[1] 355 7

```

degs_cleaned_4 <- na.omit(degs_cleaned_4)
rownames(degs_cleaned_4) <- degs_cleaned_4$hgnc_symbol

heatmap_matrix_degs_4_ano <-
  df_filtered_4[rownames(df_filtered_4) %in% rownames(degs_cleaned_4), ]
heatmap_matrix_degs_4_ano <- as.data.frame(heatmap_matrix_degs_4_ano)

genes_to_annotate_4 <- c("TAB2", "SPHK2", "HMOX1", "NFKB1",
  "TRIM25", "EDNRB", "DAB2IP", "PELI2",
  "BRD4", "SLC20A1", "CXXC5",
  "C18orf32", "PER1", "ROR1", "PELI1",
  "RAPGEF2", "BTN2A2", "BMP4", "LIF",
  "SPRY2", "MAP3K12", "ACKR3", "EGFR",
  "MFHAS1", "EPHB1", "ATF3", "BMPER",
  "PDGFA", "HIPK2", "PHLPP1", "SH2D3C",

```

```

    "SH3RF1", "SH3RF3", "SMAD7", "TGFBR1",
    "ZMIZ1", "PMEPA1", "SMAD6", "SKI",
    "LDLRAD4", "BMPR2", "MAP2K3", "ZFP36",
    "MAPKAPK2", "SPHK1", "RELL1", "EGR1",
    "NOG", "SMURF1", "INTS9", "APPL2",
    "NRROS")
}

row_annotation_4 <-
  rowAnnotation(mark = anno_mark(at = which(rownames(heatmap_matrix_degs_4_ano)
                                             %in% genes_to_annotate_4),
                                   labels =
                                             rownames(heatmap_matrix_degs_4_ano)
                                             [rownames(heatmap_matrix_degs_4_ano)
                                              %in% genes_to_annotate_4],
                                   labels_gp =
                                             gpar(fontsize = 10, fontface = "bold"),
                                   padding = unit(2, "mm"))))

heatmap_matrix_degs_4_ano <-
  subset(heatmap_matrix_degs_4_ano,
         select = -ensembl_gene_id)
heatmap_matrix_degs_4_ano <-
  heatmap_matrix_degs_4_ano[, !names(heatmap_matrix_degs_4_ano)
                            %in% "hgnc_symbol", drop = FALSE]

heatmap_matrix_degs_4_ano <- na.omit(heatmap_matrix_degs_4_ano)
heatmap_matrix_degs_4_ano <- as.matrix(heatmap_matrix_degs_4_ano)
heatmap_matrix_degs_4_ano <- t(scale(t(heatmap_matrix_degs_4_ano)))

# check how many NAs are there
sum(is.na(heatmap_matrix_degs_4_ano))

```

[1] 0

```

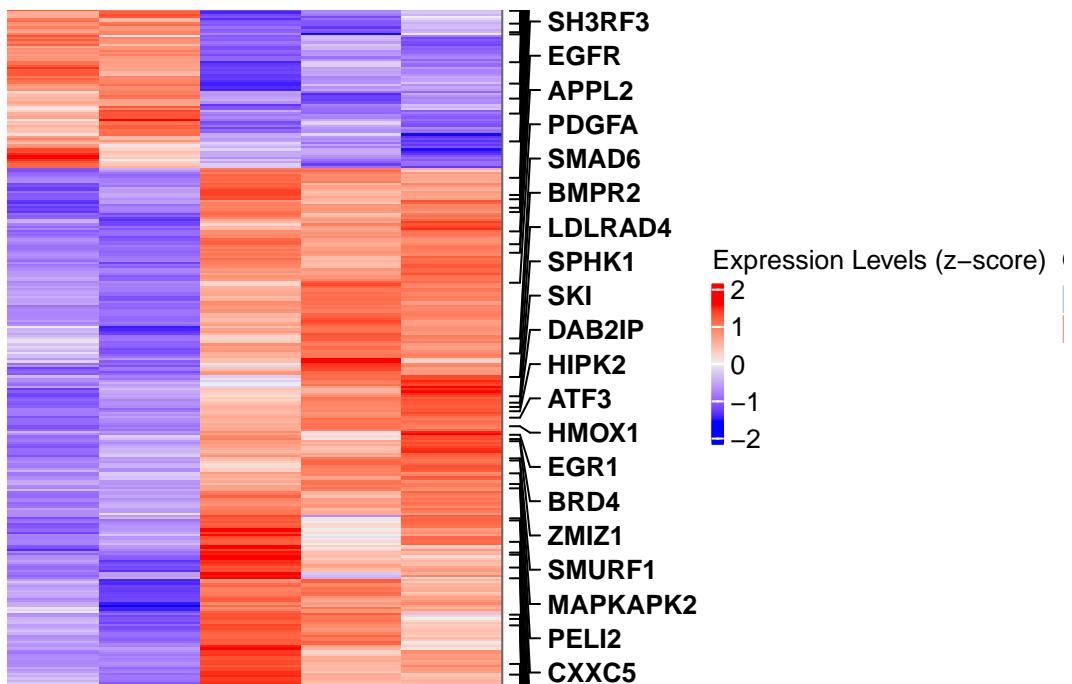
ha = HeatmapAnnotation(
  Genotype = rep(c("4h_p", "4h_t"), each=3),
  col = list(Genotype = c(
    "4h_p"="#B5C7DC", "4h_t"="#DE9F9D"
  )),
  )
,
```

```

# change font size of the annotation
annotation_name_gp = grid::gpar(fontsize = 10),
annotation_legend_param = list(
  # title = "Expression Levels",
  title_gp = gpar(fontsize = 10),
  labels_gp = gpar(fontsize = 10),
  grid_width = unit(5,"point"),
  grid_height = unit(5,"point")
)
)

Heatmap(heatmap_matrix_degs_4_ano, name="Expression Levels",
        cluster_columns = FALSE,
        cluster_rows = TRUE,
        clustering_distance_columns = "pearson",
        clustering_method_columns = "complete",
        row_dend_width = unit(0, "cm"),
        show_column_names = FALSE,
        show_row_names = FALSE,
        top_annotation = ha,
        gap=unit(1, "point"),
        border= TRUE,
        border_gp = gpar(col = "grey40"),
        row_title_gp = gpar(fontsize = 5), # change row split font size
        row_title_rot = 0,
        bottom_annotation = ha,
        right_annotation = row_annotation_4,
        width = unit(8, "cm"),
        height = unit(10, "cm"),
        heatmap_legend_param = list(
          title = "Expression Levels (z-score)",
          title_gp = gpar(fontsize = 10),
          labels_gp = gpar(fontsize = 10),
          grid_width    = unit(4,"point"),
          grid_height   = unit(4,"point")
        ),
        row_names_gp = grid::gpar(fontsize = 1)
)

```



13 Save the workspace

```
#save.image("workspace/Deseq_cleanedup_analysis_160224.Rdata")
```