Proteomics: Differentially expressed protein analysis in kidney organoids in prolonged culture

Hassan Saei

2024-12-12

Global proteomics analysis of kidney organoids - Rachel Lennon's lab - Manchester

Loading libraries

```
library(tidyverse)
library(limma)
library(QFeatures)
library(msqrob2)
library(plotly)
library(gridExtra)
library(proteusLabelFree)
library(dplyr)
library(msqrob2gui)
library(ExploreModelMatrix)
library(VennDiagram)
library(ggrepel)
library(ggplot2)
library("ggVennDiagram")
library(clusterProfiler)
library(enrichplot)
organism = "org.Hs.eg.db"
#BiocManager::install(organism, character.only = TRUE)
library(organism, character.only = TRUE)
#launchMsqrob2App()
```

Preprocessing MaxQuant output file

```
setwd("0_ECM/")
peptidesFile <- "peptides.txt"

pep <- read.table(peptidesFile, sep = "\t", header = T)

# All dataset
pep <- pep %>%
  mutate(Proteins = ifelse(Proteins == "", NA, Proteins)) %>%
  mutate(Intensity.1C3_d37_ECM_R1 = ifelse(Intensity.1C3_d37_ECM_R1 == "", NA, Intensity.1C3_d37_ECM_R1
  filter(Proteins != "NA") %>%
  filter(Intensity.1C3_d37_ECM_R1 != "NA")
```

```r
# ECM
pep <- pep[,-c(87:91, 95,96,97)]
# F1
#pep <- pep[,-c(83:88, 92:94)]

pep <- pep %>%
  mutate(Proteins = sapply(str_split(Proteins, ";"), function(x) paste(x[1], collapse = ";")))

#names(pep)[names(pep) == "Intensity.2G9_d37_ECM_R1"] <- "Intensity.ECM_d37.R1"

count_blanks <- function(x) {
  sum(x == "")
}

blank_counts <- sapply(pep, count_blanks)

ecols <- grep("Intensity\\.", names(pep))

pe <- readQFeatures(assayData = pep, fnames = 1,
                    ecol = ecols, name = "peptideRaw", sep = "\t")
```

```
## Checking arguments.
```

```
## Warning in .checkWarnEcol(quantCols, ecol): 'ecol' is deprecated, use
## 'quantCols' instead.
```

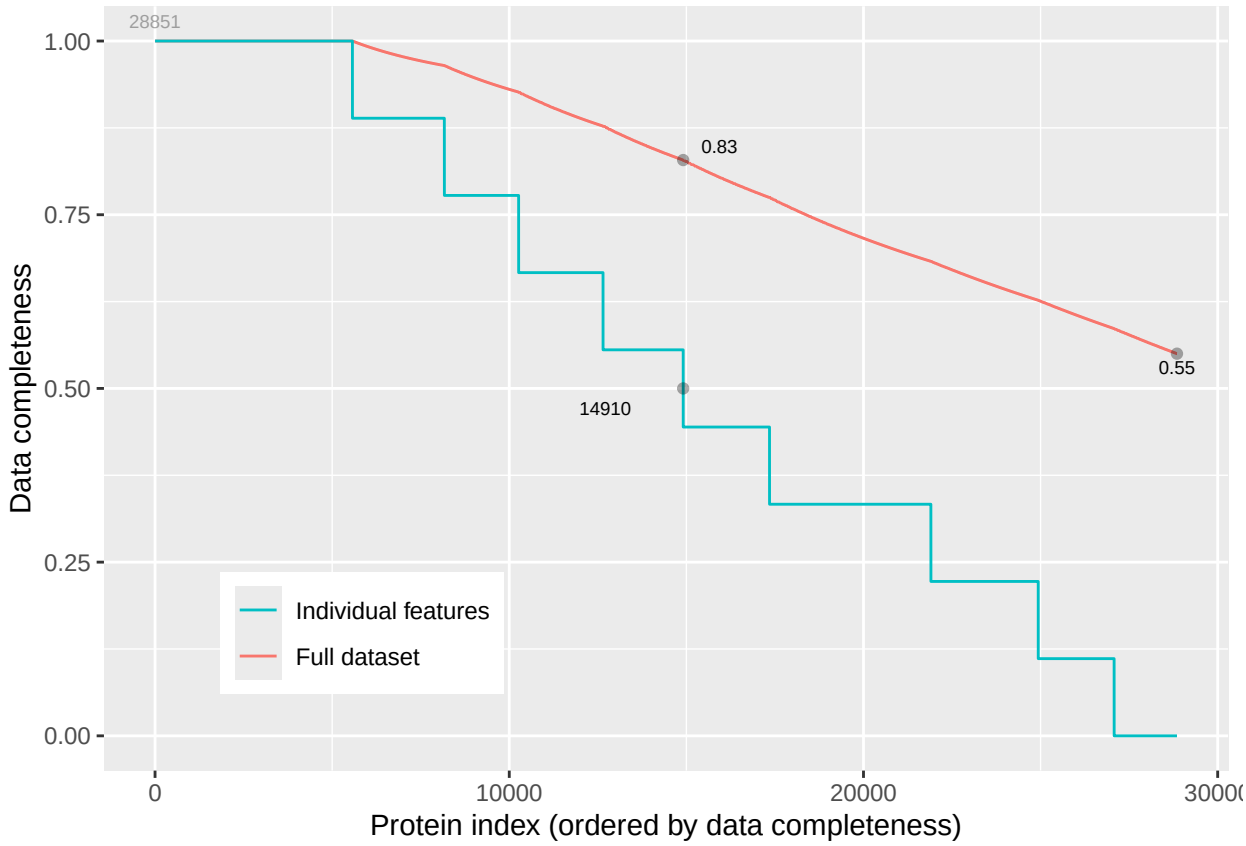
```
## Loading data as a 'SummarizedExperiment' object.
```

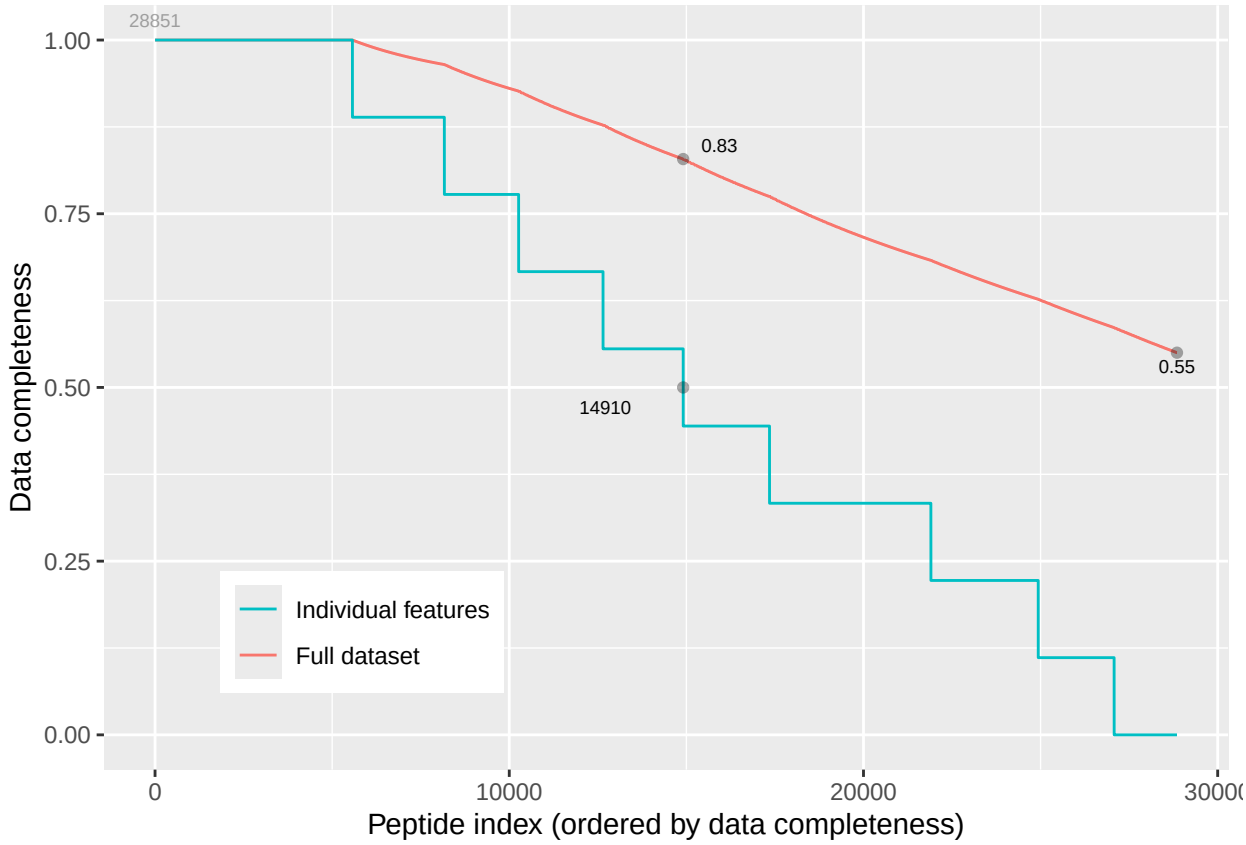
```
## Formatting sample annotations (colData).
```

```
## Formatting data as a 'QFeatures' object.
```

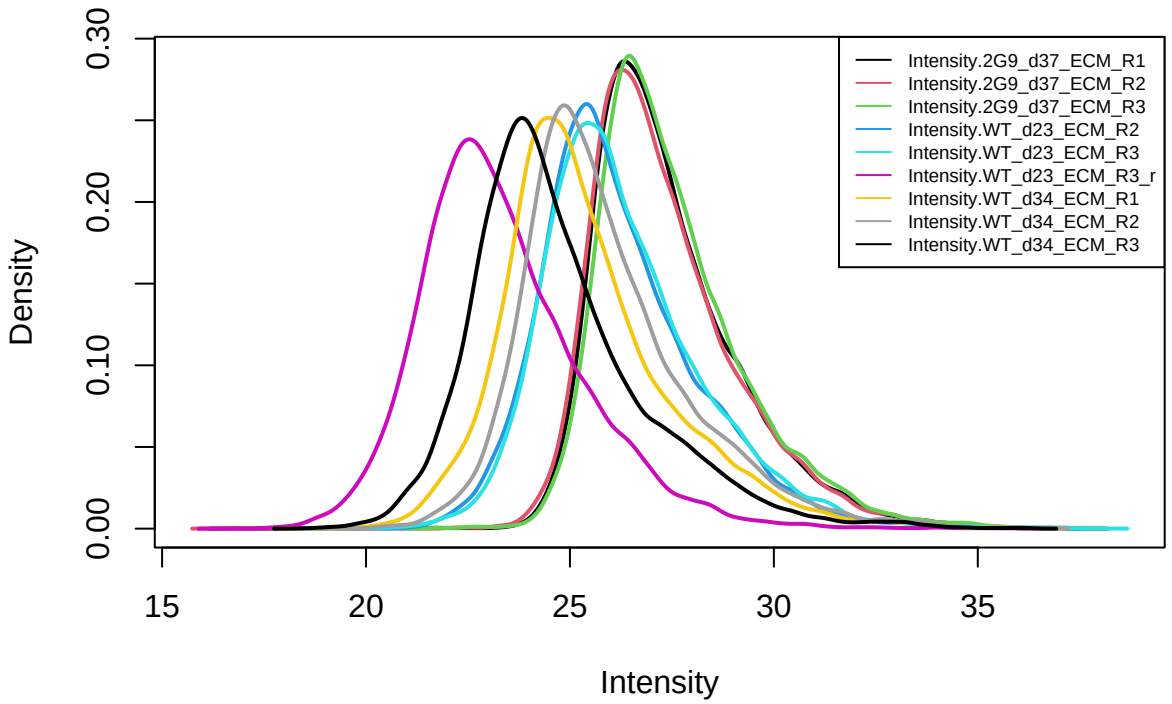
```r
# Zero values (missingness)

rowData(pe[["peptideRaw"]])$nNonZero <- rowSums(assay(pe[["peptideRaw"]]) > 0)
pe <- zeroIsNA(pe, "peptideRaw")

MSnbase::plotNA(assay(pe[["peptideRaw"]])) +
  xlab("Peptide index (ordered by data completeness)")
```

```
## Warning in fun(libname, pkgname): mzR has been built against a different Rcpp version (1.0.11)
## than is installed on your system (1.0.13.1). This might lead to errors
## when loading mzR. If you encounter such issues, please send a report,
## including the output of sessionInfo() to the Bioc support forum at
## https://support.bioconductor.org/. For details see also
## https://github.com/sneumann/mzR/wiki/mzR-Rcpp-compiler-linker-issue.
```


28851
1.00
0.75
0.50
0.25
0.00
Data completeness
0.83
0.55
14910
Individual features
Full dataset
0
10000
20000
30000
Protein index (ordered by data completeness)


28851
1.00
0.75
0.50
0.25
0.00
Data completeness
0.83
0.55
14910
Individual features
Full dataset
0
10000
20000
30000
Peptide index (ordered by data completeness)

```
# Log transformation

pe <- logTransform(pe, base = 2, i = "peptideRaw", name = "peptideLog")
limma::plotDensities(assay(pe[["peptideLog"]]), legend = FALSE)

legend("topright", legend = colnames(assay(pe[["peptideLog"]])), col = 1:ncol(assay(pe[["peptideLog"]])))
```




```
# Handeling ovrlapping protein groups

Protein_filter <- rowData(pe[["peptideLog"]])$Proteins %in% smallestUniqueGroups(rowData(pe[["peptideLog"]])$Proteins)
pe <- pe[Protein_filter,,]

pe <- filterFeatures(pe, ~ Reverse != "+")
```

```
## 'Reverse' found in 2 out of 2 assay(s)
```

```
pe <- filterFeatures(pe, ~ Potential.contaminant != "+")
```

```
## 'Potential.contaminant' found in 2 out of 2 assay(s)
```

```
pe <- filterFeatures(pe, ~ nNonZero >= 2)
```

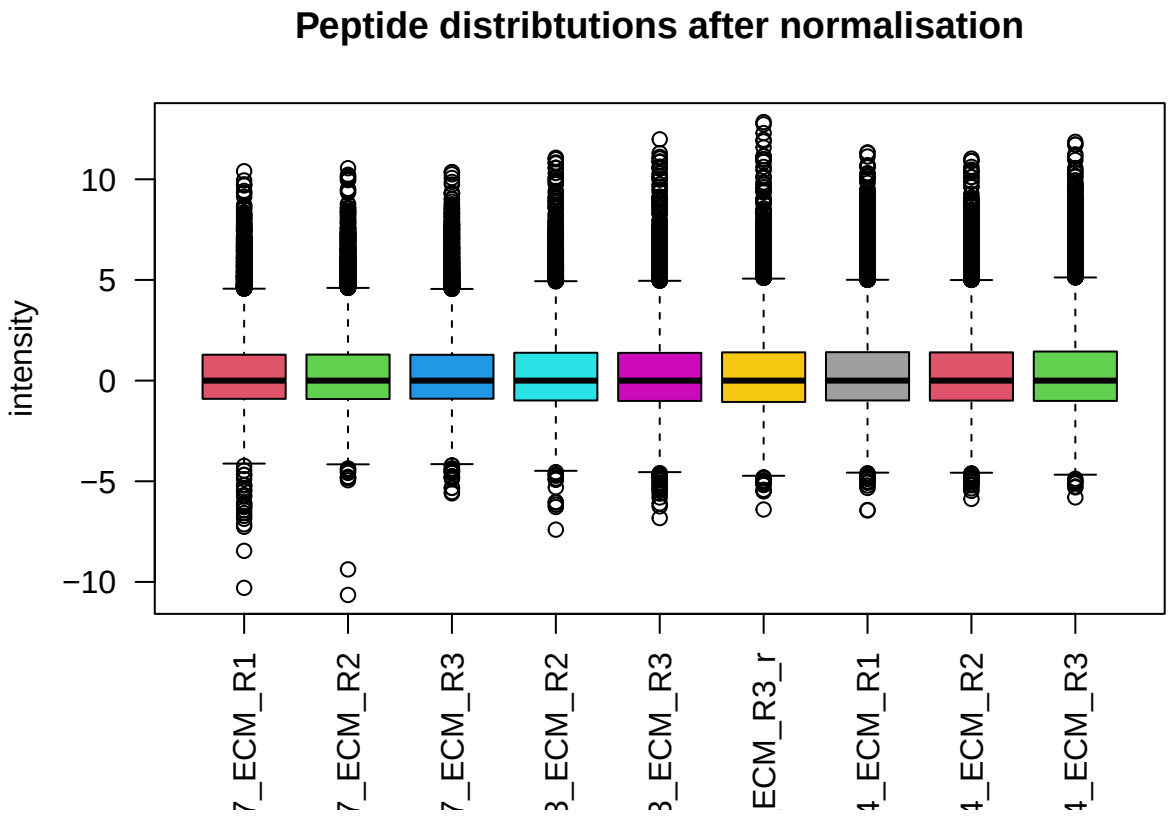
```
## 'nNonZero' found in 2 out of 2 assay(s)
```

```
nrow(pe[["peptideLog"]])
```

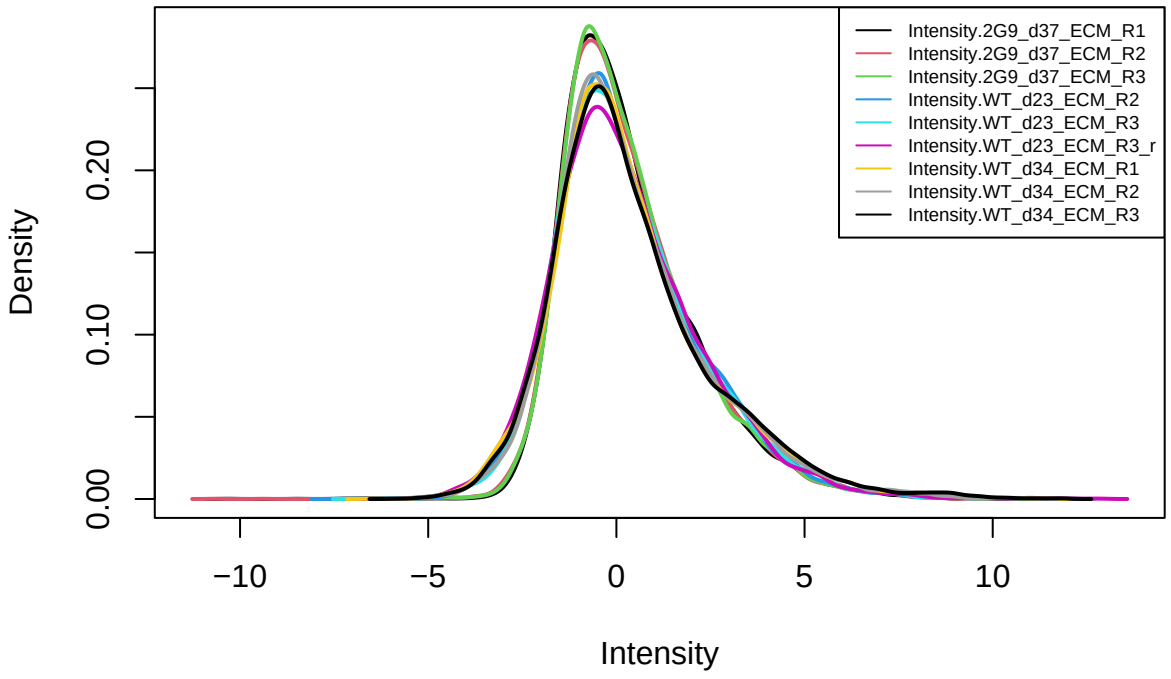
```
## [1] 24795
```

```
# Median normalization
pe <- normalize(pe,
                i = "peptideLog",
                name = "peptideNorm",
                method = "center.median"
)
```

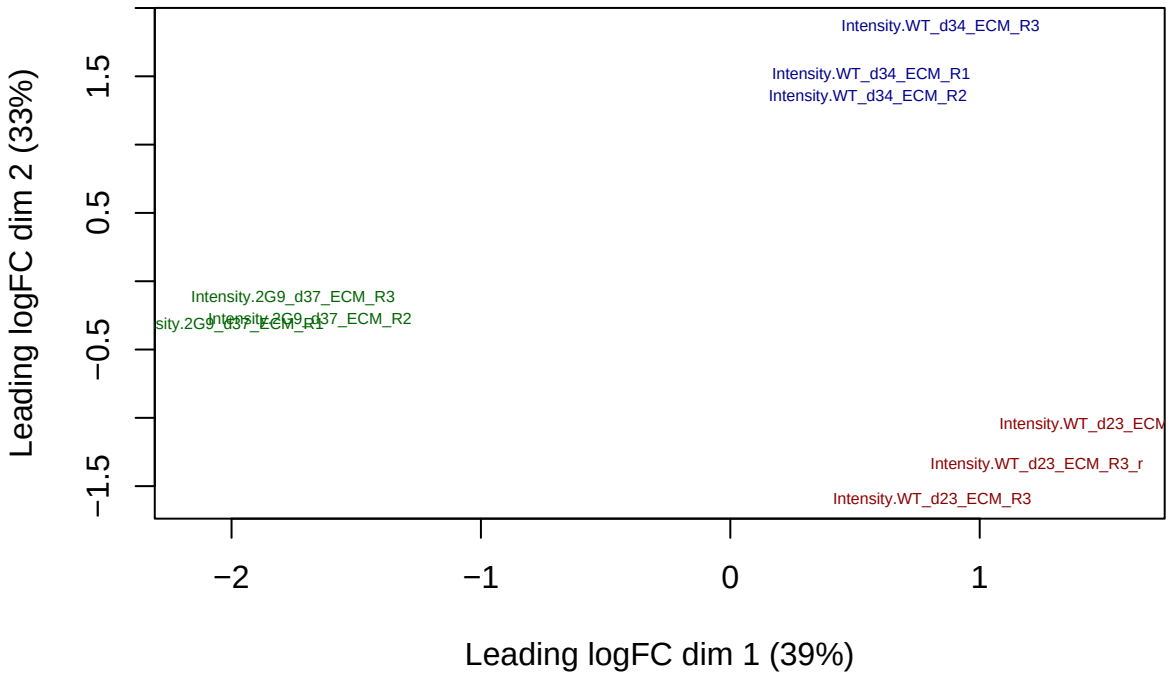
```
boxplot(assay(pe[["peptideNorm"]]),
        col = palette()[-1],
        main = "Peptide distribtutions after normalisation", ylab = "intensity", las=2)
```




```
limma::plotDensities(assay(pe[["peptideNorm"]]), legend = FALSE)
legend("topright", legend = colnames(assay(pe[["peptideLog"]])), col = 1:ncol(assay(pe[["peptideLog"]])))
```

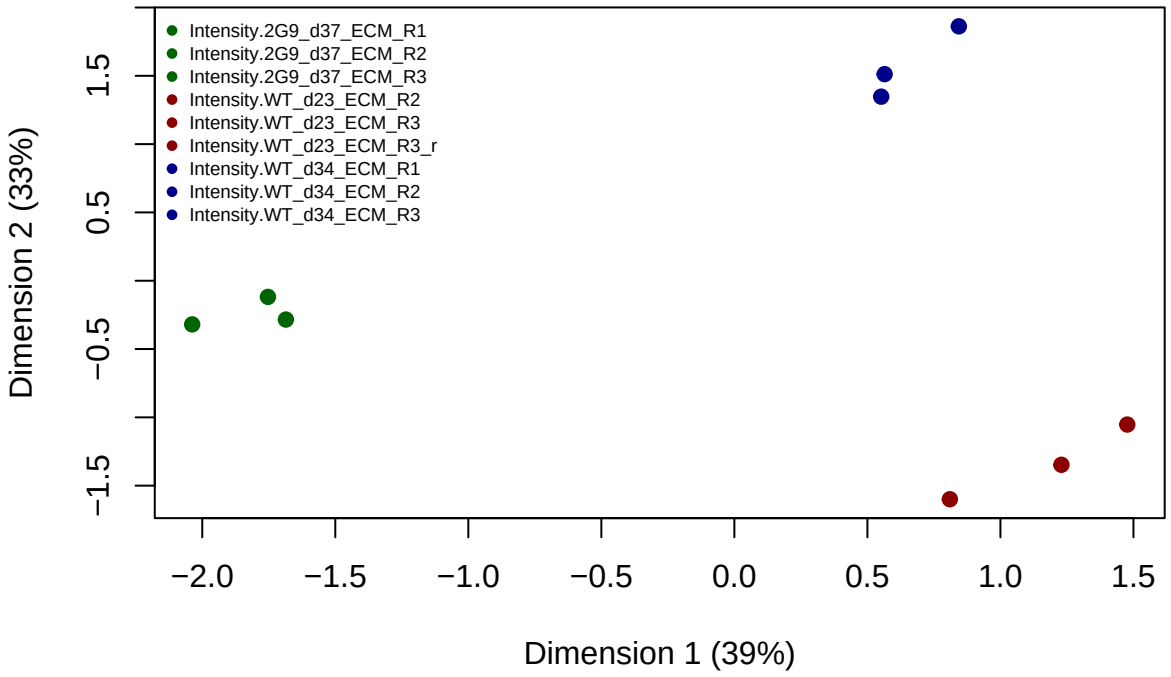


```
group <- factor(c(rep("Day38", 3), rep("Day22", 3), rep("Day34", 3)))
colors <- c("darkred", "darkblue", "darkgreen")[group]

limma::plotMDS(assay(pe[["peptideNorm"]]), col = colors, cex = 0.5)
```

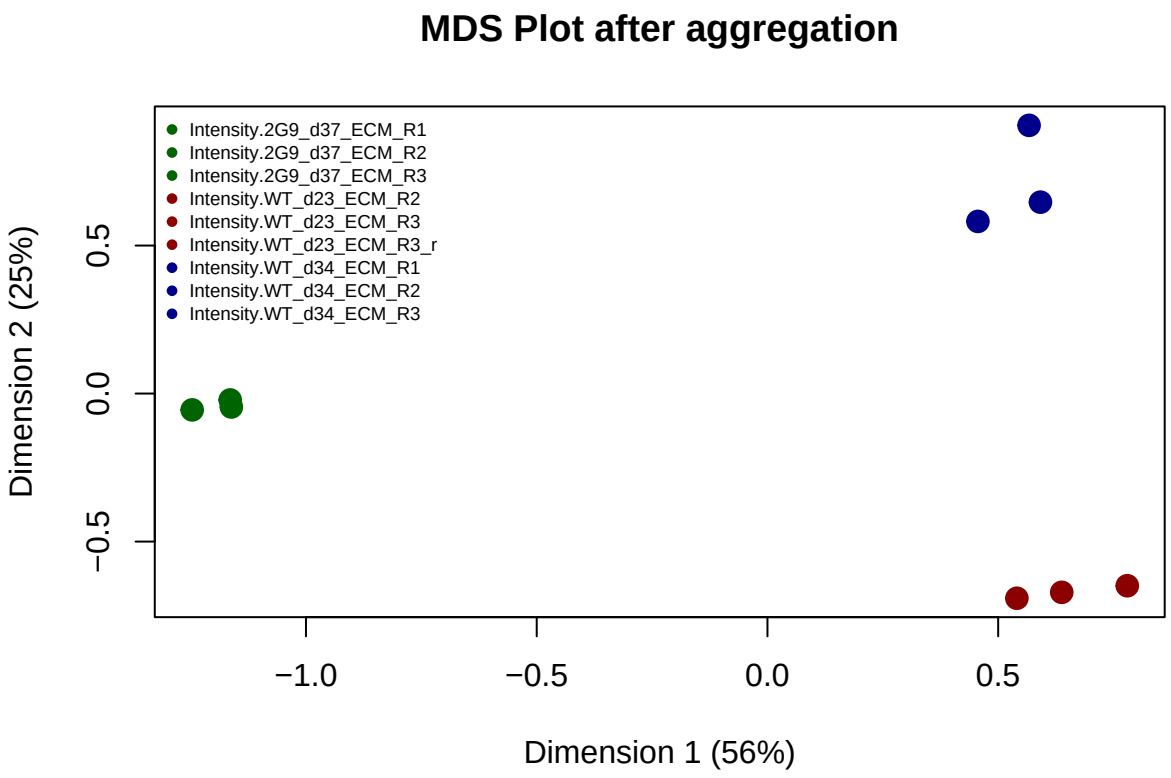


Generate MDS plot

```
# Beautifull MDS plot

mds <- plotMDS(assay(pe[["peptideNorm"]]), plot = FALSE)

plot(mds$x, mds$y, col = colors, pch = 19, cex = 1, xlab = "Dimension 1 (39%)", ylab = "Dimension 2 (33
legend("topleft", legend = colnames(assay(pe[["peptideLog"]])), col = colors, pch = 19, cex = 0.6, bty =
```




```
# Summarization to protein level
pe <- aggregateFeatures(pe, i = "peptideNorm", na.rm = TRUE, name = "protein", fcol= "Proteins")

mds <- plotMDS(assay(pe[["protein"]]), plot = FALSE)
plot(mds$x, mds$y, col = colors, pch = 19, cex = 1.5, xlab = "Dimension 1 (56%)", ylab = "Dimension 2 (
legend("topleft", legend = colnames(assay(pe[["peptideLog"]])), col = colors, pch = 19, cex = 0.6, bty =
```

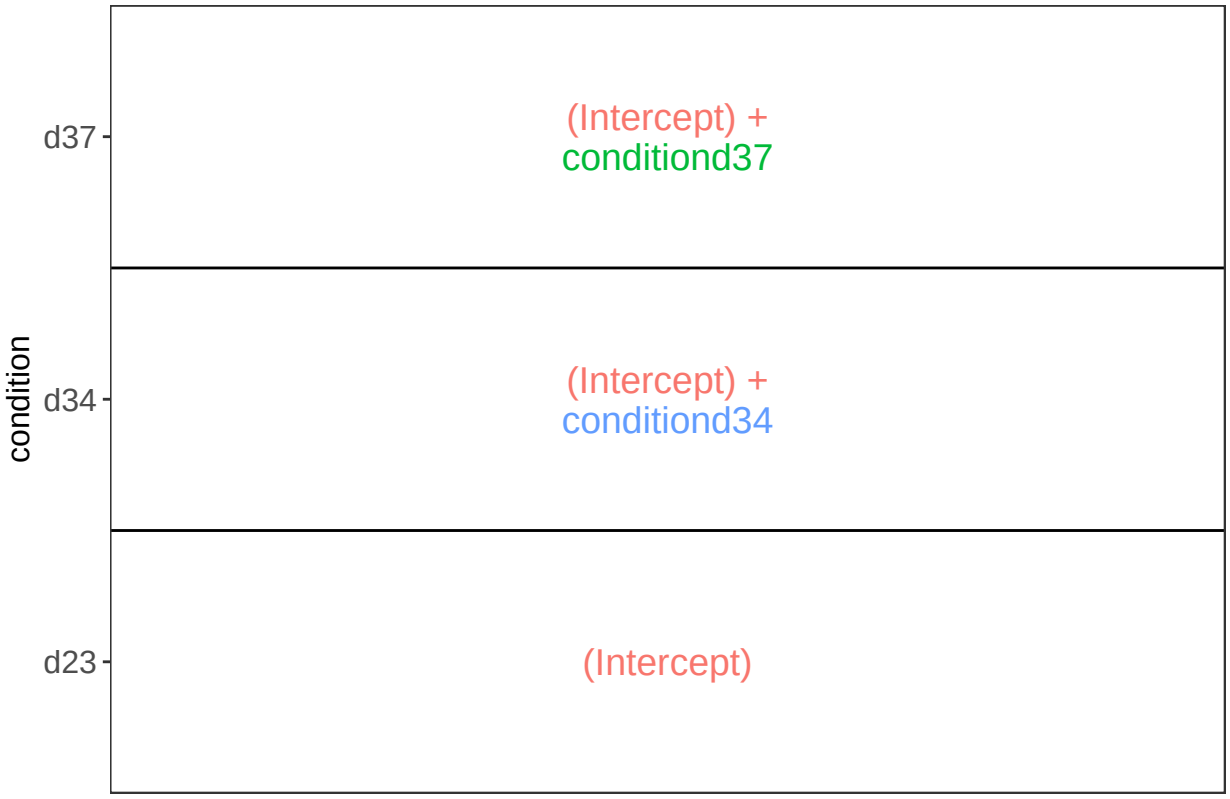


Estimations

```
condition <- read.table("0_ECM/design_A1.txt", sep = "\t", header = T)
#condition <- read.table("F1/design_A1.txt", sep = "\t", header = T)

colData(pe)$condition <- as.factor(condition$condition)

pe <- msqrob(object = pe, i = "protein", formula = ~condition, overwrite=TRUE)

getCoef(rowData(pe[["protein"]])$msqrobModels[[1]])
```

Visualizing designs and conditions

```
VisualizeDesign(colData(pe),~condition)$plotlist[[1]]
```




```
# Test all pairwise comparisons
contrasts1 <- makeContrast("conditiond37=0", parameterNames = c("conditiond37"))
contrasts2 <- makeContrast("conditiond34=0", parameterNames = c("conditiond34"))

pe1 <- hypothesisTest(object = pe, i = "protein", contrast = contrasts1, overwrite = TRUE)
pe2 <- hypothesisTest(object = pe, i = "protein", contrast = contrasts2, overwrite = TRUE)
```

Extracting differentially expressed proteins comparing two time points

```
DEP_d34d22 <- rowData(pe2[["protein"]])$"conditiond34"
DEP_d34d22 <- rowData(pe2[["protein"]])$"conditiond34" %>%
  filter(adjPval < 0.05)

DEP_d38d22 <- rowData(pe1[["protein"]])$"conditiond37"
DEP_d38d22 <- rowData(pe1[["protein"]])$"conditiond37" %>%
  filter(adjPval < 0.05)

match <- pep[,c("Proteins", "Gene.names")]

DEP_d34d22$Proteins <- rownames(DEP_d34d22)
DEP_d38d22$Proteins <- rownames(DEP_d38d22)

DEP_d34d22 <- merge(DEP_d34d22, match, by="Proteins")
DEP_d38d22 <- merge(DEP_d38d22, match, by="Proteins")

DEP_d34d22 <- DEP_d34d22 %>%
```

```
  mutate(Gene.names = sapply(str_split(Gene.names, ";"), `[`, 1))

DEP_d38d22 <- DEP_d38d22 %>%
  mutate(Gene.names = sapply(str_split(Gene.names, ";"), `[`, 1))

DEP_d34d22 <- DEP_d34d22 %>%
  distinct(Gene.names, .keep_all = TRUE)

DEP_d38d22 <- DEP_d38d22 %>%
  distinct(Gene.names, .keep_all = TRUE)


colnames(DEP_d34d22) <- c("ProteinIDs", "logFC", "se", "df", "t", "pval", "adjPval", "Symbol")
colnames(DEP_d38d22) <- c("ProteinIDs", "logFC", "se", "df", "t", "pval", "adjPval", "Symbol")
```

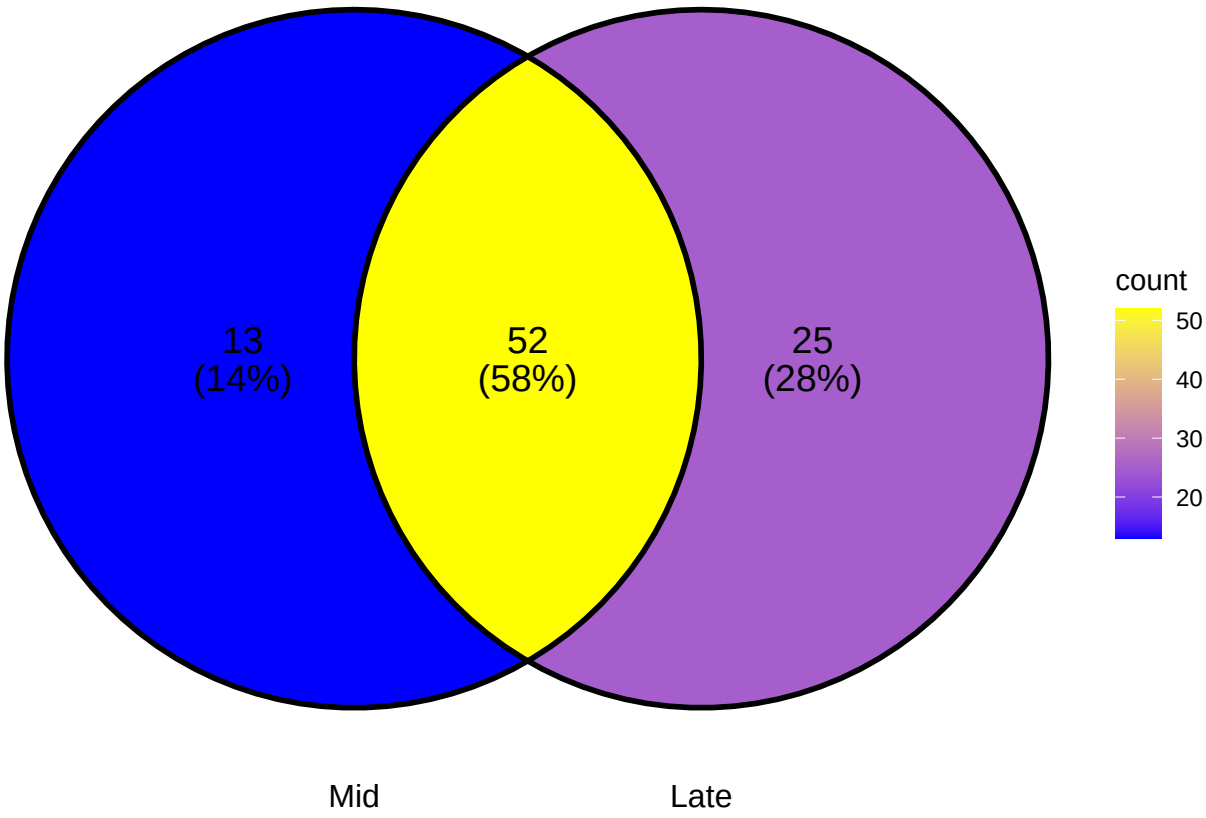
matching matrisome associated proteins

```
NABA <- read.table("../Bulk_RNASeq/Listes_Res_MultiTests/NABA_MATRISOME.v2023.2.txt", sep = "\t", heade

DEP_d34d22 <- DEP_d34d22 %>%
  mutate(Matrix = ifelse(Symbol %in% NABA$Symbol, "Matrix", "Other"))

DEP_d38d22 <- DEP_d38d22 %>%
  mutate(Matrix = ifelse(Symbol %in% NABA$Symbol, "Matrix", "Other"))

#write.table(DEP_d34d22, file = "Proteomics_d34vsd22_F1_DEP.txt", sep = "\t", row.names = F)
#write.table(DEP_d38d22, file = "Proteomics_d38vsd22_F1_DEP.txt", sep = "\t", row.names = F)
```

Venn diagram for shared ECM proteins

```
matrix_d34 <- DEP_d34d22$Matrix == "Matrix"
matrix_d38 <- DEP_d38d22$Matrix == "Matrix"

pro_d34 <- DEP_d34d22 %>%
  filter(matrix_d34) %>%
  pull(Symbol)

pro_d38 <- DEP_d38d22 %>%
  filter(matrix_d38) %>%
  pull(Symbol)

list = list(pro_d34, pro_d38)
ggVennDiagram(list, label_alpha = 0, label_size = 5,
              category.names = c("Mid", "Late")) + ggplot2::scale_fill_gradient(low="blue",high = "yell
```

```
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```



Enrichment analysis of DEPs

```
# Enrichment
original_gene_list <- DEP_d38d22$Symbol

gene_sets_hallmark <- msigdbr::msigdbr(species = "Homo sapiens", category = "H")
gene_sets_GO_BP <- msigdbr::msigdbr(species = "Homo sapiens", category = "C5", subcategory = "BP")
gene_sets_GO_CC <- msigdbr::msigdbr(species = "Homo sapiens", category = "C5", subcategory = "CC")
gene_sets_GO_MF <- msigdbr::msigdbr(species = "Homo sapiens", category = "C5", subcategory = "MF")
gene_sets_KEGG <- msigdbr::msigdbr(species = "Homo sapiens", category = "C2", subcategory = "KEGG")
gene_sets_Reactome <- msigdbr::msigdbr(species = "Homo sapiens", category = "C2", subcategory = "REACTO

gene_sets_hallmark <- gene_sets_hallmark %>%
  dplyr::select(gs_name, gene_symbol)

gene_sets_GO_BP <- gene_sets_GO_BP %>%
  dplyr::select(gs_name, gene_symbol)

gene_sets_GO_CC <- gene_sets_GO_CC %>%
  dplyr::select(gs_name, gene_symbol)

gene_sets_GO_MF <- gene_sets_GO_MF %>%
  dplyr::select(gs_name, gene_symbol)

gene_sets_KEGG <- gene_sets_KEGG %>%
  dplyr::select(gs_name, gene_symbol)
```

```r
gene_sets_Reactome <- gene_sets_Reactome %>%
  dplyr::select(gs_name, gene_symbol)
```

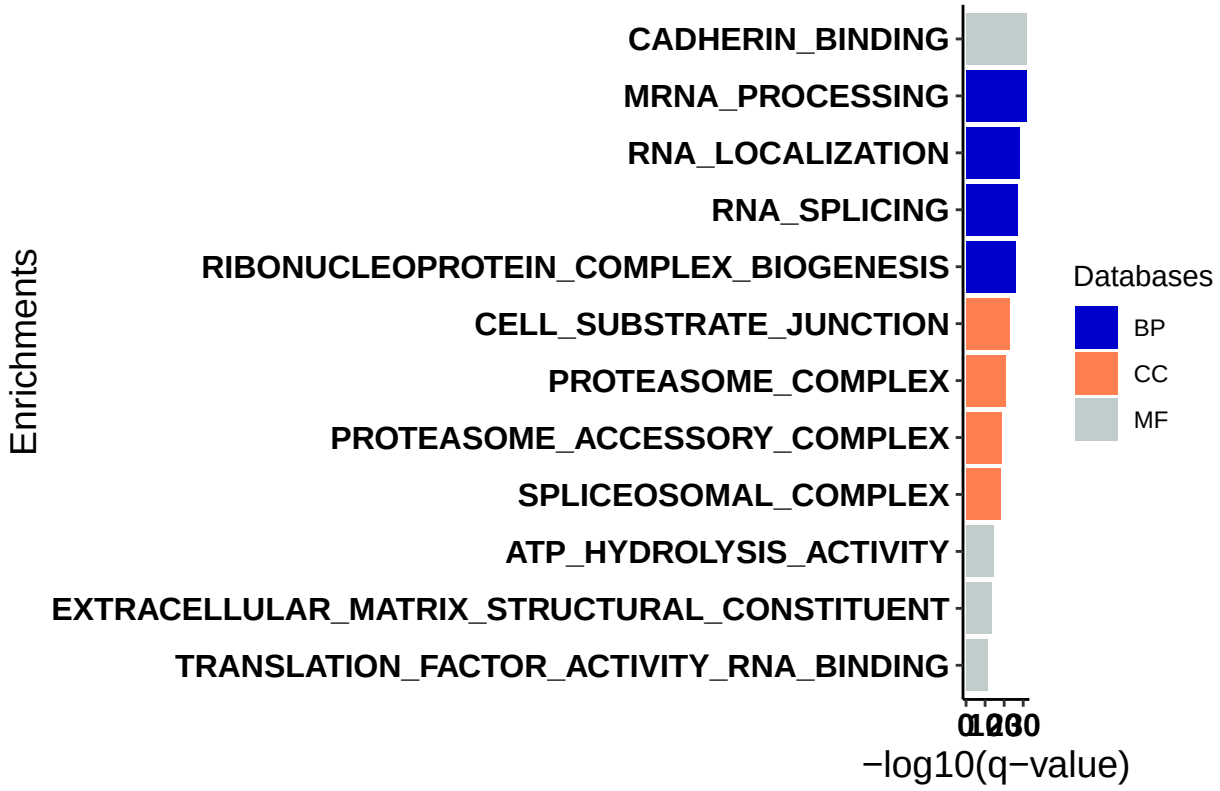
Enrichr for enrichemnt analysis

```r
CC <- as.data.frame(enricher(gene = original_gene_list,
                             TERM2GENE = gene_sets_GO_CC, pAdjustMethod = "BH", pvalueCutoff = 0.05, qva
BP <- as.data.frame(enricher(gene = original_gene_list, TERM2GENE = gene_sets_GO_BP,
                             pAdjustMethod = "BH", pvalueCutoff = 0.05, qvalueCutoff = 0.05))
MF <- as.data.frame(enricher(gene = original_gene_list, TERM2GENE = gene_sets_GO_MF,
                             pAdjustMethod = "BH", pvalueCutoff = 0.05, qvalueCutoff = 0.05))

# Select the first three columns
df_combined <- rbind(CC, BP, MF)
df_combined$log10_Adjusted <- -log10(df_combined$qvalue)
df_combined <- df_combined %>%
  mutate(ONTOLOGY = case_when(
    grepl("^GOMF_", Description) ~ "MF",
    grepl("^GOCC_", Description) ~ "CC",
    grepl("^GOBP_", Description) ~ "BP",
    TRUE ~ NA_character_
  ),
  Description = gsub("GOMF_|GOCC_|GOBP_", "", Description))
write.table(df_combined, file = "GO_d38vsd22_F1.txt", sep = "\t")

df_BP <- BP[1:4,]
df_CC <- CC[1:4,]
df_MF <- MF[1:4,]

# Combine the dataframes back if needed
df_combined <- rbind(df_BP, df_CC, df_MF)
df_combined$ID <- gsub("GOMF_|GOCC_|GOBP_", "", df_combined$ID)
df_combined$log10_Adjusted <- -log10(df_combined$qvalue)
df_combined <- df_combined %>%
  mutate(ONTOLOGY = case_when(
    grepl("^GOMF_", Description) ~ "MF",
    grepl("^GOCC_", Description) ~ "CC",
    grepl("^GOBP_", Description) ~ "BP",
    TRUE ~ NA_character_
  ),
  Description = gsub("GOMF_|GOCC_|GOBP_", "", Description))
```

barplot for enrichment results

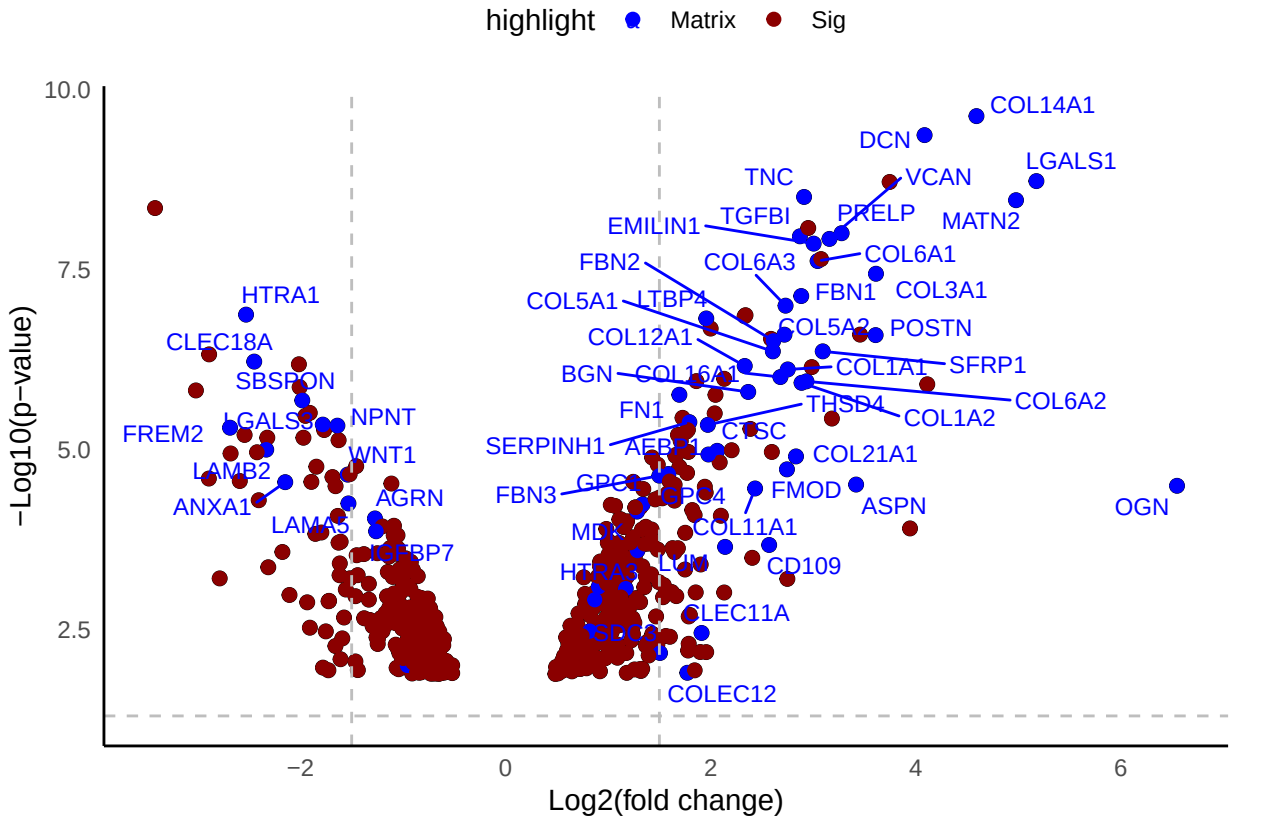
```r
ggplot(df_combined, aes(x = log10_Adjusted, y = reorder(ID, log10_Adjusted), fill = ONTOLOGY)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("BP" = "blue3", "CC" = "coral", "MF" = "azure3")) +
  labs(
    title = "",
    x = "-log10(q-value)",
    y = "Enrichments",
```

```
    fill = "Databases"
  ) +
  theme_classic() +
  theme(
    axis.text.y = element_text(size = 12, face = "bold", colour = "black"),
    axis.text.x = element_text(size = 12, face = "bold", colour = "black"),
    axis.title.x = element_text(size = 14),
    axis.title.y = element_text(size = 14),
    plot.title = element_text(size = 16, face = "bold"),
  )
```

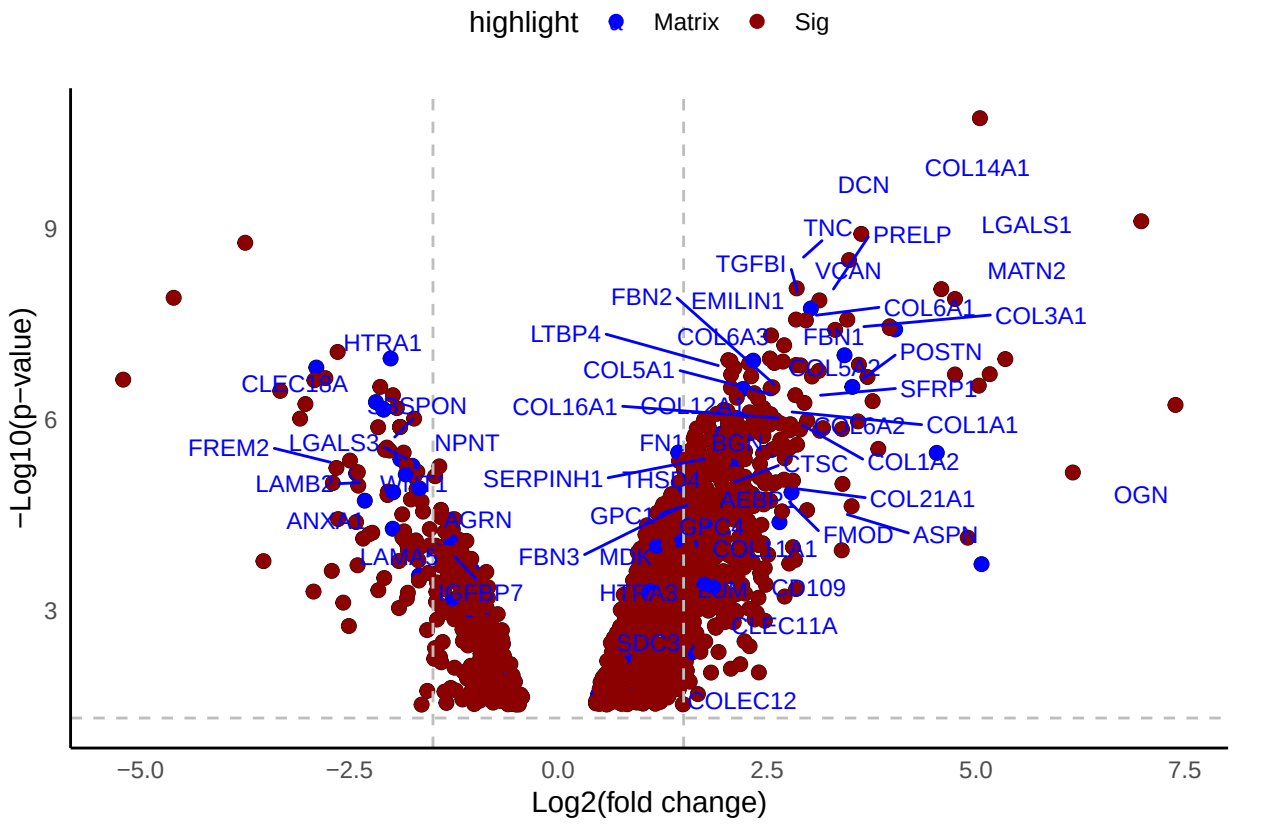



```
# Volcano plot
DEP_d34d22 <- DEP_d34d22 %>%
  mutate(
    highlight = case_when(
      pval < 0.05 & grepl("Matrix", Matrix) ~ "Matrix",
      pval < 0.05 ~ "Sig",
      TRUE ~ "Not_sig"
    )
  )

DEP_d38d22 <- DEP_d38d22 %>%
  mutate(
    highlight = case_when(
      pval < 0.05 & grepl("Matrix", Matrix) ~ "Matrix",
      pval < 0.05 ~ "Sig",
      TRUE ~ "Not_sig"
    )
  )
```

```r
color_map <- c(
  "Sig" = "darkred",
  "Not_sig" = "black",
  "Matrix" = "blue"
)

volcano1 <- ggplot(DEP_d34d22,
  aes(x = logFC, y = -log10(pval), colour= highlight)) +
  geom_point(cex = 2.0, stroke = 0.6, color = "black") +
  geom_point(aes(fill = highlight), cex = 2.0, stroke = 0.5) +
  theme_minimal() +
  labs(x="Log2(fold change)",
       y="-Log10(p-value)") +
  scale_color_manual(values=color_map) +
  theme(legend.position="top", axis.line = element_line(color = "black"),
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
  geom_hline(yintercept=-log10(0.05), linetype="dashed", color = "gray") +
  geom_vline(xintercept=c(-1.5, 1.5), linetype="dashed", color = "gray")

volcano1 <- volcano1 +
  geom_text_repel(data = subset(DEP_d34d22, highlight %in% "Matrix" & abs(logFC) > 1.2),
                  aes(label = Symbol),
                  size = 3,
                  max.overlaps = 20)

volcano2 <- ggplot(DEP_d38d22,
  aes(x = logFC, y = -log10(pval), colour= highlight)) +
  geom_point(cex = 2.0, stroke = 0.6, color = "black") +
  geom_point(aes(fill = highlight), cex = 2.0, stroke = 0.5) +
  theme_minimal() +
  labs(x="Log2(fold change)",
       y="-Log10(p-value)") +
  scale_color_manual(values=color_map) +
  theme(legend.position="top", axis.line = element_line(color = "black"),
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
  geom_hline(yintercept=-log10(0.05), linetype="dashed", color = "gray") +
  geom_vline(xintercept=c(-1.5, 1.5), linetype="dashed", color = "gray")

volcano2 <- volcano2 +
  geom_text_repel(data = subset(DEP_d34d22, highlight %in% "Matrix" & abs(logFC) > 1.2),
                  aes(label = Symbol),
                  size = 3,
                  max.overlaps = 20)

volcano1
```


highlight
Matrix
Sig
10.0
7.5
5.0
2.5
−Log10(p−value)
−2
0
2
4
6
Log2(fold change)
COL14A1
DCN
LGALS1
TNC
VCAN
TGFBI
PRELP
MATN2
EMILIN1
FBN2
COL6A3
COL6A1
FBN1
COL3A1
HTRA1
COL5A1
LTBP4
COL5A2
POSTN
CLEC18A
COL12A1
COL1A1
SFRP1
SBSPON
BGN
COL16A1
COL6A2
FN1
THSD4
COL1A2
FREM2
LGALS3
NPNT
CTSC
SERPINH1
AEBP1
COL21A1
WNT1
LAMB2
GPC1
FMOD
FBN3
GPC4
ASPN
OGN
ANXA1
AGRN
COL11A1
LAMA5
MDK
IGFBP7
HTRA3
LUM
CD109
CLEC11A
SDC3
COLEC12

```
volcano2
```


highlight
Matrix
Sig
9
6
3
−Log10(p−value)
−5.0
−2.5
0.0
2.5
5.0
7.5
Log2(fold change)
COL14A1
DCN
TNC
PRELP
LGALS1
TGFBI
VCAN
MATN2
FBN2
EMILIN1
COL6A1
COL3A1
LTBP4
COL6A3
FBN1
HTRA1
POSTN
COL5A1
SFRP1
CLEC18A
COL16A1
COL6A2
COL1A1
SBSPON
FREM2
LGALS3
NPNT
FN1
BGN
CTSC
COL1A2
SERPINH1
THSD4
COL21A1
OGN
LAMB2
AEBP1
ANXA1
GPC1
GPC4
FMOD
ASPN
AGRN
FBN3
MDK
COL11A1
LAMA5
IGFBP7
HTRA3
LUM
CD109
CLEC11A
SDC3
COLEC12