

Proximity Labelling - Mutants vs WT Student t-test script

2022-07-27

```
library(reshape2)
library(ggplot2)
library(RColorBrewer)
library(data.table)

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:reshape2':
##
##      dcast, melt

library(tidyr)

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:reshape2':
##
##      smiths
```

1. Turbo

1.1 Data Preprocessing

```
turbo_id_preprocess_all <- function(raw_data, lfq=FALSE) {

  if (lfq==TRUE){

    preprocessed <- raw_data[,c("T..Majority.protein.IDs",
                                "T..Gene.names",
                                colnames(raw_data)[grepl("LFQ.intensity.T_",
                                                            colnames(raw_data))])]

  } else {

    preprocessed <- raw_data[,c("T..Majority.protein.IDs",
                                "T..Gene.names",
                                colnames(raw_data)[grepl("Intensity.T_",
                                                            colnames(raw_data))])]

  }

  preprocessed$PROTID_temp <- sapply(preprocessed$T..Majority.protein.IDs,
                                     function(protid) {
```

```

    unlist(strsplit(x = protid, split = ";"))[1]
  })
preprocessed$PROTID <- gsub(preprocessed$PROTID_temp,
                           pattern = "CON__",
                           replacement = "")
preprocessed$PROTID <- gsub(preprocessed$PROTID,
                           pattern = "REV__",
                           replacement = "")

preprocessed$PROTID_temp <- NULL
preprocessed$T..Majority.protein.IDs <- NULL

preprocessed$Gene.name <- preprocessed$T..Gene.names
preprocessed$T..Gene.names <- NULL

preprocessed <-
  preprocessed[,c("PROTID",
                  "Gene.name",
                  colnames(preprocessed)[1:(dim(preprocessed)[2]-2)])]

return(preprocessed)
}

```

Comparisons between CFTR mutants and WT condition were performed based on the LFQ (Label-free quantification) log2-transformed data.

```

turbo_lfq_log_pl <- read.table("../raw_data/Turbo_LFQintensities_log2_export.txt",
                              sep = "\t",
                              header = T)
turbo_lfq_log_pl_df <- turbo_id_preprocess_all(turbo_lfq_log_pl,
                                              lfq=TRUE)

colnames(turbo_lfq_log_pl_df)[-1] <- gsub(x = colnames(turbo_lfq_log_pl_df)[-1],
                                           pattern = "LFQ.intensity.T_",
                                           replacement = "")

```

1.2 Keep proteins that are present in all replicates of all conditions.

```

WT_samples <- which(grepl(pattern = "WT", colnames(turbo_lfq_log_pl_df)))
G551D_samples <- which(grepl(pattern = "G551D", colnames(turbo_lfq_log_pl_df)))
W1282X_samples <- which(grepl(pattern = "W1282X", colnames(turbo_lfq_log_pl_df)))

turbo_lfq_log_pl_df$filter <- apply(X = turbo_lfq_log_pl_df[,c(WT_samples,
                                                                G551D_samples,
                                                                W1282X_samples)],
                                   MARGIN = 1,
                                   function(x){return(any(is.na(x)))})

turbo_filtered <- turbo_lfq_log_pl_df[which(!turbo_lfq_log_pl_df$filter),]
turbo_filtered$filter <- NULL

```

1.3 Comparison G551D vs WT.

1.3.1 Student T-test for G551D vs WT.

```
turbo_filtered$t_test.G551D.pval <-
  sapply(1:nrow(turbo_filtered), function(iprot){
    return(t.test(as.numeric(turbo_filtered[iprot,WT_samples]),
                    as.numeric(turbo_filtered[iprot,G551D_samples]))$p.value)
  })
# Correction for multiple testing
turbo_filtered$t_test.G551D.padj <- p.adjust(turbo_filtered$t_test.G551D.pval,
                                             method = "BH")

#-log10(p.value) for the plot
turbo_filtered$t_test.G551D.log.pval <- -log10(turbo_filtered$t_test.G551D.pval)

# log2(FoldChange)
turbo_filtered$diff.mean.G551D.WT <- sapply(1:nrow(turbo_filtered),
                                             function(iprot){
        return(mean(as.numeric(turbo_filtered[iprot,G551D_samples])) -
               mean(as.numeric(turbo_filtered[iprot,WT_samples])))
      })

# Over- or under-expressed class for the volcano plot
turbo_filtered$G551D.class <- sapply(1:nrow(turbo_filtered), function(i_row){

  p.val <- turbo_filtered[i_row, "t_test.G551D.pval"]
  FC <- turbo_filtered[i_row, "diff.mean.G551D.WT"]

  if (p.val<0.1 & FC<0){
    return("DOWN")
  } else if (p.val<0.1 & FC>0){
    return("UP")
  } else {
    return("NO")
  }
})
turbo_filtered$G551D.class <- factor(turbo_filtered$G551D.class,
                                   levels = c("NO", "DOWN", "UP"))
```

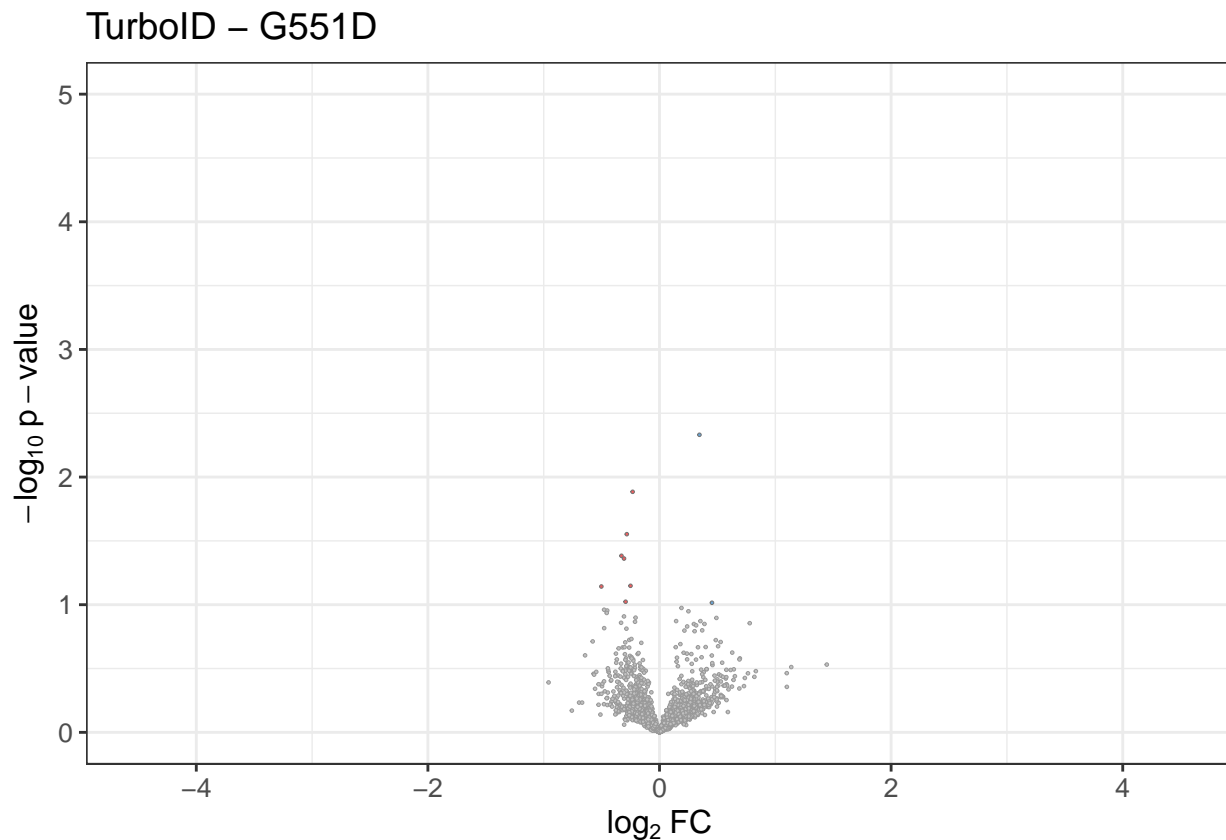
1.3.2 Volcano plot for G551D vs WT.

```
volcano_turbo_G551D <- ggplot(turbo_filtered, aes(x = diff.mean.G551D.WT,
                                                    y = t_test.G551D.log.pval,
                                                    color = G551D.class,
                                                    fill = G551D.class))+

  geom_point(pch=21, size=0.5, stroke=0.2)+
  ylim(0, 5)+
  scale_color_manual(values=c("#999999", "#737373", "#737373"))+
  scale_fill_manual(values=c("#bfbfbf", "#ed5e5f", "#69a3d2"))+
  theme_bw()+
  theme(legend.position="none",
        text=element_text(size = 12),
        title = element_text(size = 12))+
```

```
ggtitle("TurboID - G551D") +
  guides(fill="legend")+
  scale_x_continuous(breaks=c(-4, -2, 0, 2, 4), limits=c(-4.5,4.5))+
  labs(y=expression(-log[10]~p-value),
       x=expression(log[2]~FC))
```

volcano_turbo_G551D



1.4 Comparison W1282X vs WT.

1.4.1 Student T-test for W1282X vs WT.

```
turbo_filtered$t_test.W1282X.pval <- sapply(1:nrow(turbo_filtered), function(iprot){
  return(t.test(as.numeric(turbo_filtered[iprot,WT_samples]),
                as.numeric(turbo_filtered[iprot,W1282X_samples]))$p.value)
})

# Correction for multiple testing
turbo_filtered$t_test.W1282X.padj <- p.adjust(turbo_filtered$t_test.W1282X.pval,
                                             method = "BH")

#-log10(p.value) for the plot
turbo_filtered$t_test.W1282X.log.pval <- -log10(turbo_filtered$t_test.W1282X.pval)

# log2(FoldChange)
turbo_filtered$diff.mean.W1282X.WT <- sapply(1:nrow(turbo_filtered),
                                             function(iprot){
```

```

return(mean(as.numeric(turbo_filtered[iprot,W1282X_samples]))-
       mean(as.numeric(turbo_filtered[iprot,WT_samples]))))

# Over- or under-expressed class for plot
turbo_filtered$W1282X.class <- sapply(1:nrow(turbo_filtered), function(i_row){

  p.val <- turbo_filtered[i_row, "t_test.W1282X.pval"]
  FC <- turbo_filtered[i_row, "diff.mean.W1282X.WT"]

  if (p.val<0.1 & FC<0){
    return("DOWN")
  } else if (p.val<0.1 & FC>0){
    return("UP")
  } else {
    return("NO")
  }
})
turbo_filtered$W1282X.class <- factor(turbo_filtered$W1282X.class,
                                   levels = c("NO", "DOWN", "UP"))

```

1.4.2 Volcano plot for W1282X vs WT.

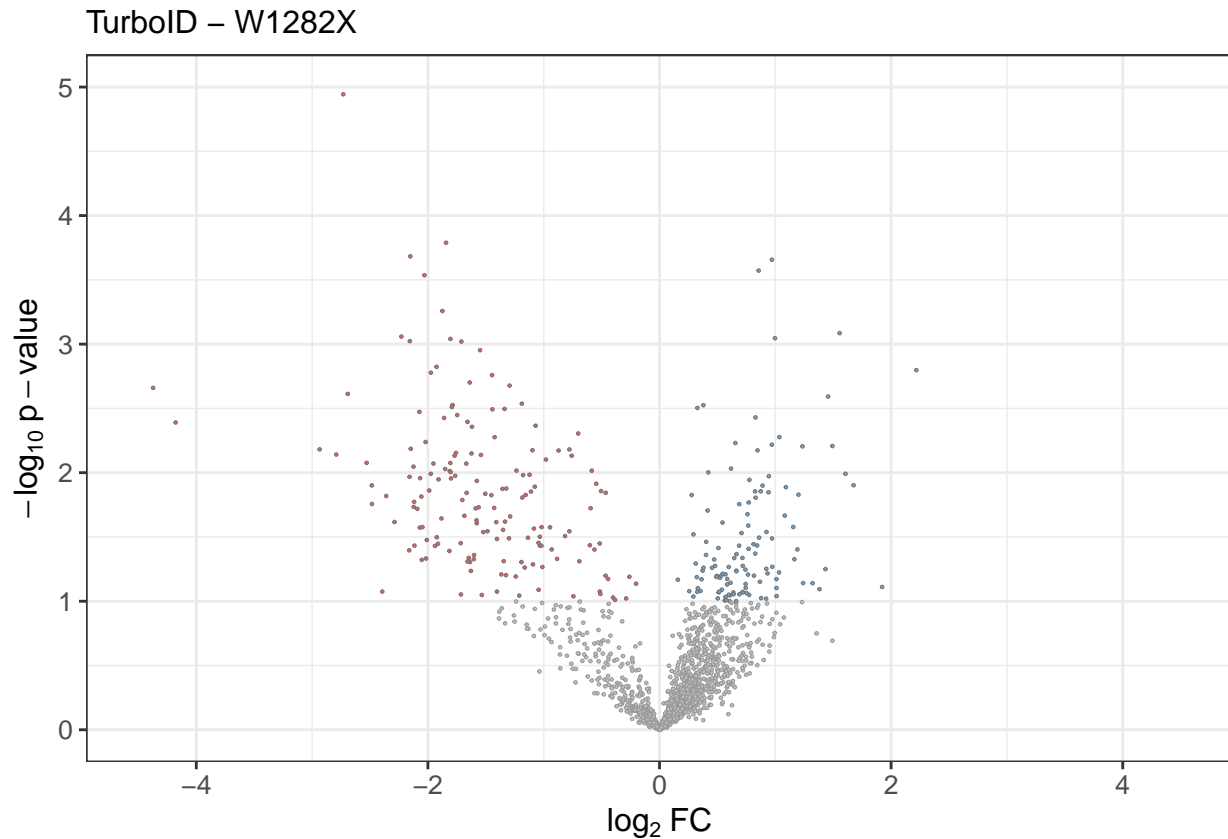
```

volcano_turbo_W1282X <- ggplot(turbo_filtered, aes(x = diff.mean.W1282X.WT,
                                                    y = t_test.W1282X.log.pval,
                                                    fill = W1282X.class,
                                                    color = W1282X.class))+

  geom_point(shape=21, size=0.5, stroke=0.2)+
  ylim(0, 5)+
  scale_color_manual(values=c("#999999", "#737373", "#737373"))+
  scale_fill_manual(values=c("#bfbfbf", "#ed5e5f", "#69a3d2"))+
  theme_bw()+
  theme(legend.position="none",
        text=element_text(size = 12),
        plot.title = element_text(size = 12),
        axis.title = element_text(size=12))+
  ggtitle("TurboID - W1282X") +
  guides(fill="legend")+
  scale_x_continuous(breaks=c(-4, -2, 0, 2, 4), limits=c(-4.5,4.5))+
  labs(y=expression(-log[10]~p-value),
       x=expression(log[2]~FC))

volcano_turbo_W1282X

```



2. APEX2

2.1 Data Preprocessing

```
apex_preprocess_all <- function(raw_data, lfq=FALSE) {

  if (lfq==TRUE){

    preprocessed <- raw_data[,c("T..Majority.protein.IDs",
                                "T..Gene.names",
                                colnames(raw_data)[grepl("LFQ.intensity.A_",
                                                            colnames(raw_data))])]

  } else {

    preprocessed <- raw_data[,c("T..Majority.protein.IDs",
                                "T..Gene.names",
                                colnames(raw_data)[grepl("Intensity.A_",
                                                            colnames(raw_data))])]

  }

  preprocessed$PROTID_temp <- sapply(preprocessed$T..Majority.protein.IDs,
                                    function(protid) {
                                      unlist(strsplit(x = protid,
                                                         split = ";"))[1])
                                    })

  preprocessed$PROTID <- gsub(preprocessed$PROTID_temp,
```

```

        pattern = "CON__",
        replacement = "")
preprocessed$PROTID <- gsub(preprocessed$PROTID,
        pattern = "REV__",
        replacement = "")

preprocessed$PROTID_temp <- NULL
preprocessed$T..Majority.protein.IDs <- NULL

preprocessed$Gene.name <- preprocessed$T..Gene.names
preprocessed$T..Gene.names <- NULL

preprocessed <-
  preprocessed[,c("PROTID",
    "Gene.name",
    colnames(preprocessed)[1:(dim(preprocessed)[2]-2)])]

return(preprocessed)
}

```

Comparisons between CFTR mutants and WT condition were performed based on the LFQ (Label-free quantification) log2-transformed data.

```

apex_lfq_log_pl <- read.table("../raw_data/Apex_LFQintensities_log2_export.txt",
        sep = "\t",
        header = T)
apex_lfq_log_pl_df <- apex_preprocess_all(apex_lfq_log_pl,
        lfq=TRUE)

colnames(apex_lfq_log_pl_df)[-1] <- gsub(x = colnames(apex_lfq_log_pl_df)[-1],
        pattern = "LFQ.intensity.A_",
        replacement = "")

```

2.2 Keep proteins that are present in all replicates of all conditions.

```

WT_samples <- which(grepl(pattern = "WT", colnames(apex_lfq_log_pl_df)))
G551D_samples <- which(grepl(pattern = "G551D", colnames(apex_lfq_log_pl_df)))
W1282X_samples <- which(grepl(pattern = "W1282X", colnames(apex_lfq_log_pl_df)))

apex_lfq_log_pl_df$filter <- apply(X = apex_lfq_log_pl_df[,c(WT_samples,
        G551D_samples,
        W1282X_samples)],
        MARGIN = 1,
        function(x){return(any(is.na(x)))})

apex_filtered <- apex_lfq_log_pl_df[which(!apex_lfq_log_pl_df$filter),]
apex_filtered$filter <- NULL

```

2.3 Comparison G551D vs WT.

2.3.1 Student T-test for G551D vs WT.

```
apex_filtered$t_test.G551D.pval <- sapply(1:nrow(apex_filtered), function(iprot){
  return(t.test(apex_filtered[iprot,WT_samples],
                apex_filtered[iprot,G551D_samples])$p.value)
})
# Correction for multiple testing
apex_filtered$t_test.G551D.padj <- p.adjust(apex_filtered$t_test.G551D.pval,
                                           method = "BH")

#-log10(p.value) for the plot
apex_filtered$t_test.G551D.log.pval <- -log10(apex_filtered$t_test.G551D.pval)

# log2(FoldChange)
apex_filtered$diff.mean.G551D.WT <- sapply(1:nrow(apex_filtered), function(iprot){
  return(mean(as.numeric(apex_filtered[iprot,G551D_samples])) -
         mean(as.numeric(apex_filtered[iprot,WT_samples]))))

# Over- or under-expressed class for the volcano plot
apex_filtered$G551D.class <- sapply(1:nrow(apex_filtered), function(i_row){

  p.val <- apex_filtered[i_row, "t_test.G551D.pval"]
  FC <- apex_filtered[i_row, "diff.mean.G551D.WT"]

  if (p.val<0.1 & FC<0){
    return("DOWN")
  } else if (p.val<0.1 & FC>0){
    return("UP")
  } else {
    return("NO")
  }
})
apex_filtered$G551D.class <- factor(apex_filtered$G551D.class,
                                  levels = c("NO", "DOWN", "UP"))
```

2.3.2 Volcano plot for G551D vs WT.

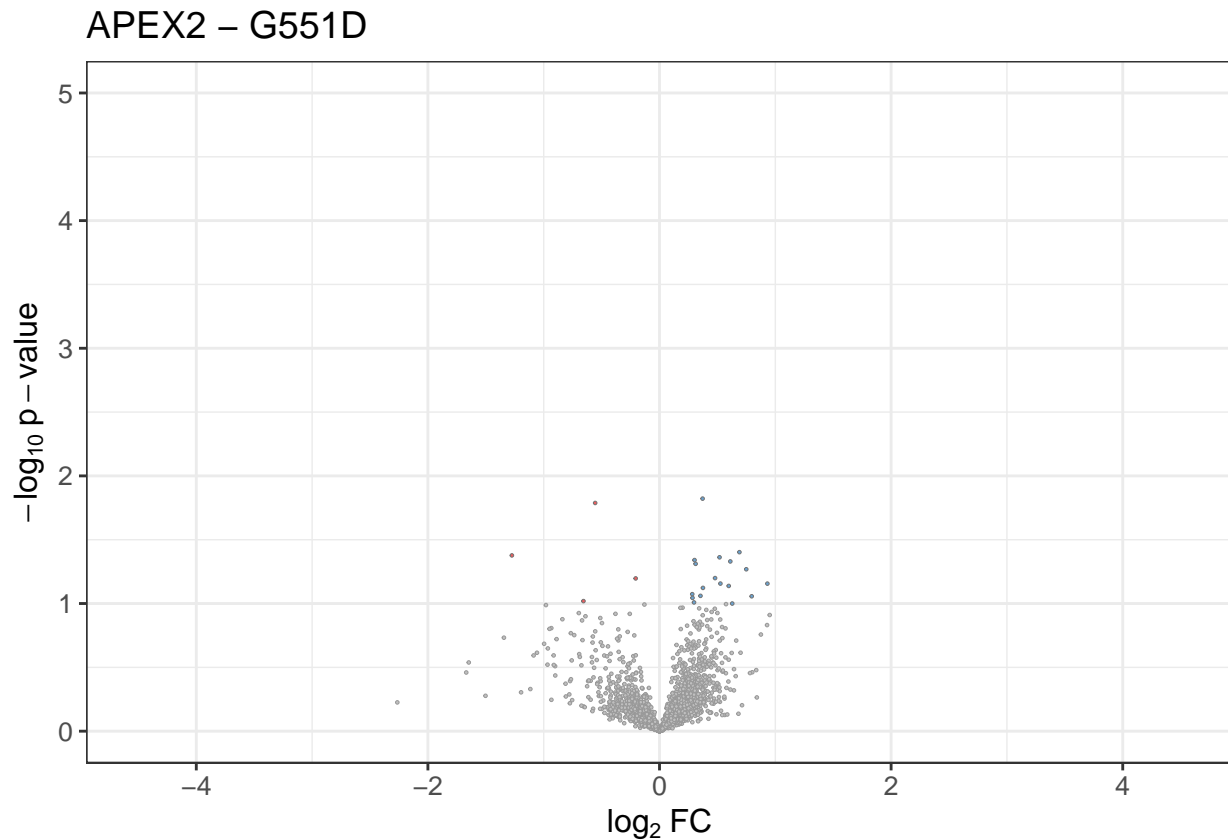
```
volcano_apex_G551D <- ggplot(apex_filtered, aes(x = diff.mean.G551D.WT,
                                                  y = t_test.G551D.log.pval,
                                                  color = G551D.class,
                                                  fill = G551D.class))+

  geom_point(pch=21, size=0.5, stroke=0.2)+
  ylim(0, 5)+
  scale_color_manual(values=c("#999999", "#737373", "#737373"))+
  scale_fill_manual(values=c("#bfbfbf", "#ed5e5f", "#69a3d2"))+
  theme_bw()+
  theme(legend.position="none",
        text=element_text(size = 12),
        title = element_text(size = 12))+
  ggtitle("APEX2 - G551D") +
  guides(fill="legend")+
  scale_x_continuous(breaks=c(-4, -2, 0, 2, 4), limits=c(-4.5,4.5))+
```



```
labs(y=expression(-log[10]~p-value),
     x=expression(log[2]~FC))
```

volcano_apex_G551D



2.4 Comparison W1282X vs WT.

2.4.1 Student T-test for W1282X vs WT.

```
apex_filtered$t_test.W1282X.pval <- sapply(1:nrow(apex_filtered), function(iprot){
  return(t.test(apex_filtered[iprot,WT_samples],
               apex_filtered[iprot,W1282X_samples])$p.value)
})

# Correction for multiple testing
apex_filtered$t_test.W1282X.padj <- p.adjust(apex_filtered$t_test.W1282X.pval,
                                           method = "BH")

#-log10(p.value) for the plot
apex_filtered$t_test.W1282X.log.pval <- -log10(apex_filtered$t_test.W1282X.pval)

# log2(FoldChange)
apex_filtered$diff.mean.W1282X.WT <- sapply(1:nrow(apex_filtered), function(iprot){
  return(mean(as.numeric(apex_filtered[iprot,W1282X_samples])) -
         mean(as.numeric(apex_filtered[iprot,WT_samples]))))})
```

```

# Over- or under-expressed class for plot
apex_filtered$W1282X.class <- sapply(1:nrow(apex_filtered), function(i_row){

  p.val <- apex_filtered[i_row, "t_test.W1282X.pval"]
  FC <- apex_filtered[i_row, "diff.mean.W1282X.WT"]

  if (p.val<0.1 & FC<0){
    return("DOWN")
  } else if (p.val<0.1 & FC>0){
    return("UP")
  } else {
    return("NO")
  }
})
apex_filtered$W1282X.class <- factor(apex_filtered$W1282X.class,
                                   levels = c("NO", "DOWN", "UP"))

```

2.4.2 Volcano plot for W1282X vs WT.

```

volcano_apex_W1282X <- ggplot(apex_filtered, aes(x = diff.mean.W1282X.WT,
                                                  y = t_test.W1282X.log.pval,
                                                  color = W1282X.class,
                                                  fill = W1282X.class))+

  geom_point(pch=21, size=0.5, stroke=0.2)+
  ylim(0, 5)+
  scale_color_manual(values=c("#999999", "#737373", "#737373"))+
  scale_fill_manual(values=c("#bfbfbf", "#ed5e5f", "#69a3d2"))+
  theme_bw()+
  theme(legend.position="none",
        text=element_text(size = 12),
        title = element_text(size = 12))+
  ggtitle("APEX2 - W1282X") +
  guides(fill="legend")+
  scale_x_continuous(breaks=c(-4, -2, 0, 2, 4), limits=c(-4.5,4.5))+
  labs(y=expression(-log[10]~p-value),
       x=expression(log[2]~FC))

volcano_apex_W1282X

```

