

# Volcano Plot for Proximity Labeling Proteomics

2023-11-05

This R script generates a volcano plot to visualize differentially regulated proteins on a proximity-labeling proteomics (e.g TurboID) dataset, highlighting proteins based on significance (p-value) and regulation (up- or down-regulated).

Based on the following script: <https://github.com/kevinblighe/EnhancedVolcano>

Install necessary packages if not already installed

Install packages

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
```

```
## Bioconductor version '3.16' is out-of-date; the current release version '3.20'
##   is available with R version '4.4'; see https://bioconductor.org/install
```

```
BiocManager::install("EnhancedVolcano")
```

```
## Bioconductor version 3.16 (BiocManager 1.30.22), R 4.2.2 (2022-10-31)
```

```
## Warning: package(s) not installed when version(s) same as or greater than current; use
##   `force = TRUE` to re-install: 'EnhancedVolcano'
```

```
## Old packages: 'ape', 'aplot', 'askpass', 'backports', 'BH', 'BiocManager',
##   'bit', 'bit64', 'bitops', 'boot', 'brew', 'brio', 'broom', 'bslib', 'cachem',
##   'callr', 'cli', 'cluster', 'codetools', 'colorspace', 'commonmark',
##   'cowplot', 'cpp11', 'crayon', 'credentials', 'curl', 'data.table', 'DBI',
##   'dbplyr', 'dendextend', 'desc', 'digest', 'downlit', 'evaluate', 'fansi',
##   'farver', 'fastmap', 'foreign', 'fs', 'gert', 'ggforce', 'ggfun', 'ggh4x',
##   'ggnewscale', 'ggplot2', 'ggraph', 'ggrepel', 'gh', 'glue', 'graphlayouts',
##   'gtable', 'haven', 'highr', 'htmltools', 'htmlwidgets', 'httpuv', 'httr2',
##   'igraph', 'jsonlite', 'KernSmooth', 'knitr', 'later', 'lattice', 'markdown',
##   'mgcv', 'munsell', 'nlme', 'openssl', 'patchwork', 'pkgbuild', 'pkgdown',
##   'pkgload', 'plotly', 'polyclip', 'processx', 'profvis', 'progress',
##   'promises', 'ps', 'ragg', 'Rcpp', 'RcppArmadillo', 'RcppEigen', 'RCurl',
##   'readr', 'remotes', 'reprex', 'rlang', 'rmarkdown', 'roxygen2', 'rpart',
##   'RSQLite', 'rstudioapi', 'RUnit', 'rvest', 'sass', 'scales', 'scatterpie',
##   'seriation', 'shadowtext', 'shiny', 'stringi', 'survival', 'sys',
##   'systemfonts', 'testthat', 'textshaping', 'tidygraph', 'tidyr', 'tidyselect',
##   'tidytrees', 'timechange', 'tinytex', 'tweenr', 'usethis', 'uuid', 'vctrs',
##   'vegan', 'viridis', 'vroom', 'waldo', 'withr', 'xfun', 'XML', 'xml2',
##   'xopen', 'yaml', 'yulab.utils', 'zip'
```

Load data from a CSV file

```
mat<-read.csv("/Users/jeffreyreina/Documents/Salk/TurboID analysis/MDATurboUniprot.csv",header=TRUE,row
```

Ensure the log2 fold change column is numeric

```
mat$log2fc<-as.numeric(mat$log2FC)
```

Load EnhancedVolcano package for plotting

```
library(EnhancedVolcano)
```

```
## Loading required package: ggplot2
```

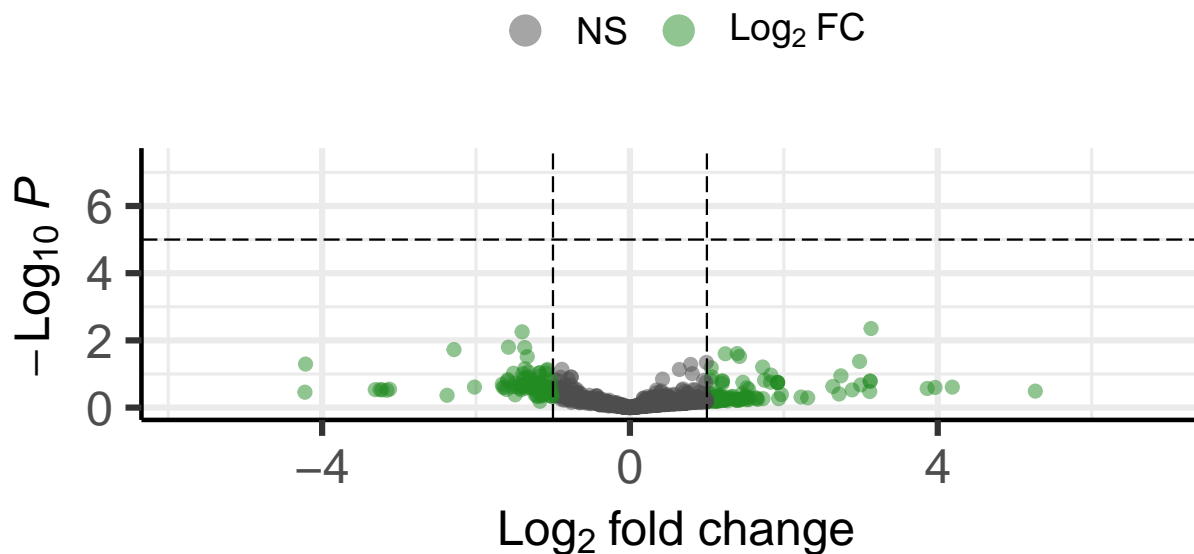
```
## Loading required package: ggrepel
```

Generate a basic volcano plot

```
EnhancedVolcano(mat,
  lab = rownames(mat),
  x = "log2fc",
  y = "Pval")
```

## Volcano plot

*EnhancedVolcano*



total = 1457 variables

Check the first few rows of the matrix

```
head(mat)
```

```
##      Gene_Name      FC      Pval  log2FC log10pval  log2fc
## Q9H008      LHPP 38.532596 0.323603098 5.268007 0.4899873 5.268007
## P01833      PIGR 18.241203 0.245907627 4.189129 0.6092280 4.189129
## P01877      IGHA2 15.620454 0.251158232 3.965364 0.6000526 3.965364
## P01876      IGHA1 14.555833 0.269833965 3.863525 0.5689034 3.863525
## Q14699      RFTN1  8.776471 0.004432621 3.133641 2.3533394 3.133641
## P59665      DEFA1  8.704354 0.163087103 3.121737 0.7875804 3.121737
```

Add a significance column based on p-value threshold

```
mat$significant<-ifelse (mat$Pval <= 0.05, "true", "false")
```

Calculate -log10 of the p-value for plotting

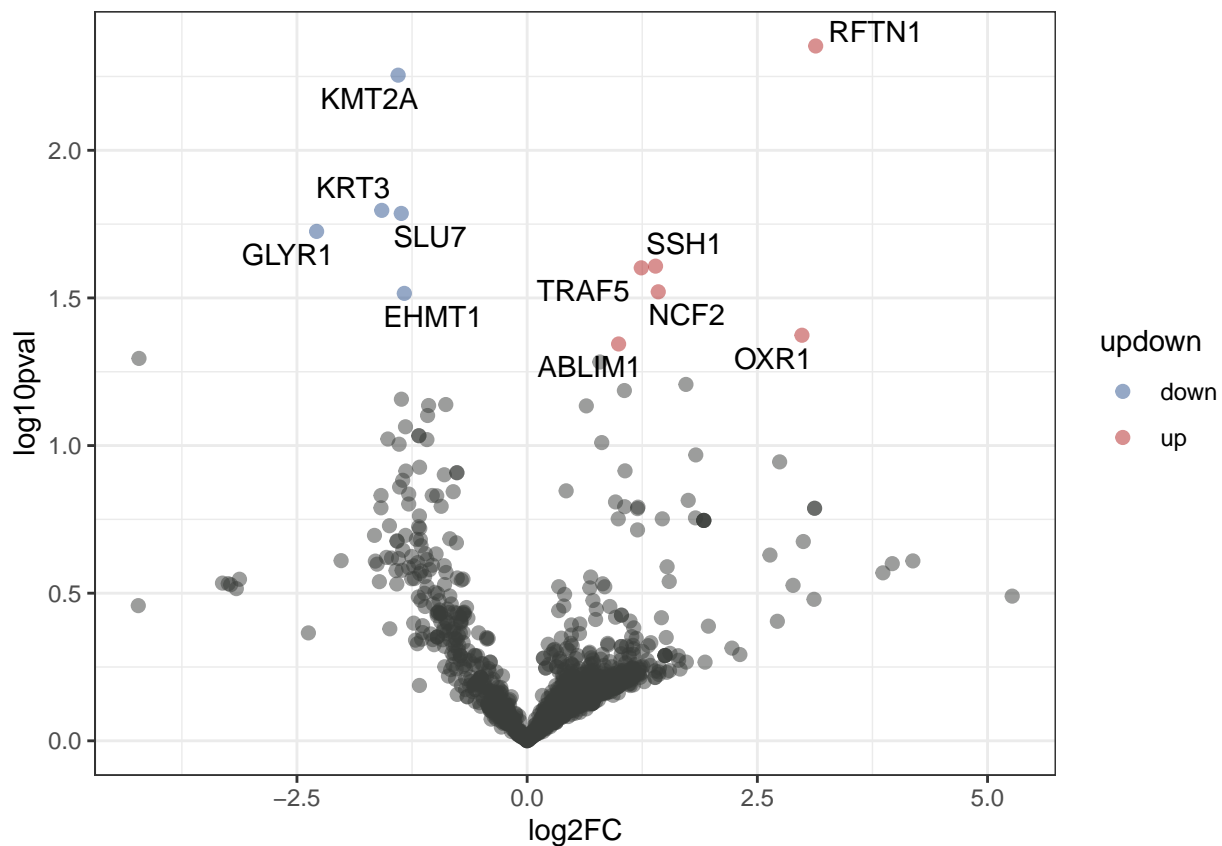
```
mat$log10pval<--log10(mat$Pval)
```

Identify upregulated and downregulated genes

```
mat$updown<-ifelse (mat$FC <= 1, "down", "up")
```

Create a volcano plot with significant proteins labeled and colored

```
ggplot(data=mat, aes(x = log2FC, y = log10pval)) +  
  geom_point(data = subset(mat, significant %in% c("true")), size = 2,  
    alpha = 0.5, aes(color = updown), shape=19) +  
  scale_color_manual(values = c("#2c538f", "firebrick")) + geom_point(data = subset(mat, significant %in% c("true")),  
    alpha = 0.5, colour = "#3a3d3a") +  
  theme_bw() +  
  geom_text_repel(data = subset(mat, significant %in% c("true")), aes(label = Gene_Name))
```



Save the plot as a PDF

```
pdf("volcanoTurboID.pdf", width=6, height=4.5)
```

```
ggplot(data=mat, aes(x = log2FC, y = log10pval)) +  
  geom_point(data = subset(mat, significant %in% c("true")), size = 2,  
    alpha = 0.5, aes(color = updown), shape=19) +  
  scale_color_manual(values = c("#2c538f", "firebrick")) + geom_point(data = subset(mat, significant %in% c("true")),  
    alpha = 0.5, colour = "#3a3d3a") +
```

```
theme_bw() +  
geom_text_repel(data = subset(mat, significant %in% c("true")), aes(label = Gene_Name))
```