fig3 TE and MS plots

Nikita Verheyden

2023-05-23

setup

dir

```
setwd("D:/Krueger_Lab/Publications/miR181_paper/Figure3")
```

packages

```
source("D:/Krueger_Lab/Publications/miR181_paper/Supporting_scripts/themes/theme_paper.R")
library(xlsx)
library(ggplot2)
library(ggrastr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

data

```
#ms
ms <- as.data.frame(read.xlsx("D:/Krueger_Lab/R/ECDF plots/Kreuger_analysis_ms.xlsx",
                              sheetName = "Analysis"))
#adjust upper case lower case of gene names of MS data:
#function
capFirst <- function(s) {
  paste(toupper(substring(s, 1, 1)), substring(s, 2), sep = "")
}
```

```
ms$Gene.Symbol <- tolower(ms$Gene.Symbol)
ms$Gene.Symbol <- capFirst(ms$Gene.Symbol)
names(ms)[names(ms) == 'Gene.Symbol'] <- 'GeneName'
head(ms$GeneName)
```

```
## [1] "Ckb"     "Gnb4"    "Ccm2"    "Rnpep"   "Aldh1b1" "Macf1"
```

```
#Ribo profiling
RNA <- read.csv("D:/Krueger_Lab/Publications/miR181_paper/Figure3/RNA_masterframe.csv")
RPF <- read.csv("D:/Krueger_Lab/Publications/miR181_paper/Figure3/RPF_masterframe.csv")

#Translational efficiency
TEframe <- read.csv("D:/Krueger_Lab/Publications/miR181_paper_v07122022/Supporting scipts/deltaTE/TE_m23
head(TEframe)
```

```
##   X  baseMean log2FoldChange     lfcSE        stat      pvalue      padj
## 1 1 2354.3220    -0.07560025 0.1336485 -0.56566464 0.571621774 0.9277760
## 2 2  579.2399    -0.06784576 0.1617420 -0.41946900 0.674873409 0.9512329
## 3 3   20.3333    -0.02453968 0.5874974 -0.04176985 0.966682171 0.9964083
## 4 4 1154.1664    -0.05314227 0.1359683 -0.39084299 0.695913297 0.9542063
## 5 5  635.5639     0.36967194 0.1364473  2.70926532 0.006743239 0.2886809
## 6 6  378.2069     0.10567650 0.1653495  0.63910971 0.522751566 0.9143531
##                     Gene gene_symbol
## 1  ENSMUSG00000000001.4       Gnai3
## 2 ENSMUSG00000000028.15       Cdc45
## 3 ENSMUSG00000000037.17       Scml2
## 4  ENSMUSG00000000056.7        Narf
## 5  ENSMUSG00000000078.7        Klf6
## 6 ENSMUSG00000000085.16       Scmh1
```

colours

```
#colours
farbeneg <- "#b4b4b4"
farbe1 <- "#0073C2FF"
farbe2 <- "#EFC000FF"
farbe3 <- "#CD534CFF"
farbe4 <- "#7AA6DCFF"
farbe5 <- "#868686FF"
farbe6 <- "#003C67FF"
farbe7 <- "#8F7700FF"
farbe8 <- "#3B3B3BFF"
farbe9 <- "#A73030FF"
farbe10 <- "#4A6990FF"
farbe11 <- "#FF6F00FF"
farbe12 <- "#C71000FF"
farbe13 <- "#008EA0FF"
farbe14 <- "#8A4198FF"
farbe15 <- "#5A9599FF"
farbe16 <- "#FF6348FF"

RNApcol <- "#b56504"
RNAncol <- "#027d73"
RPFpcol <- "#c4c404"
```

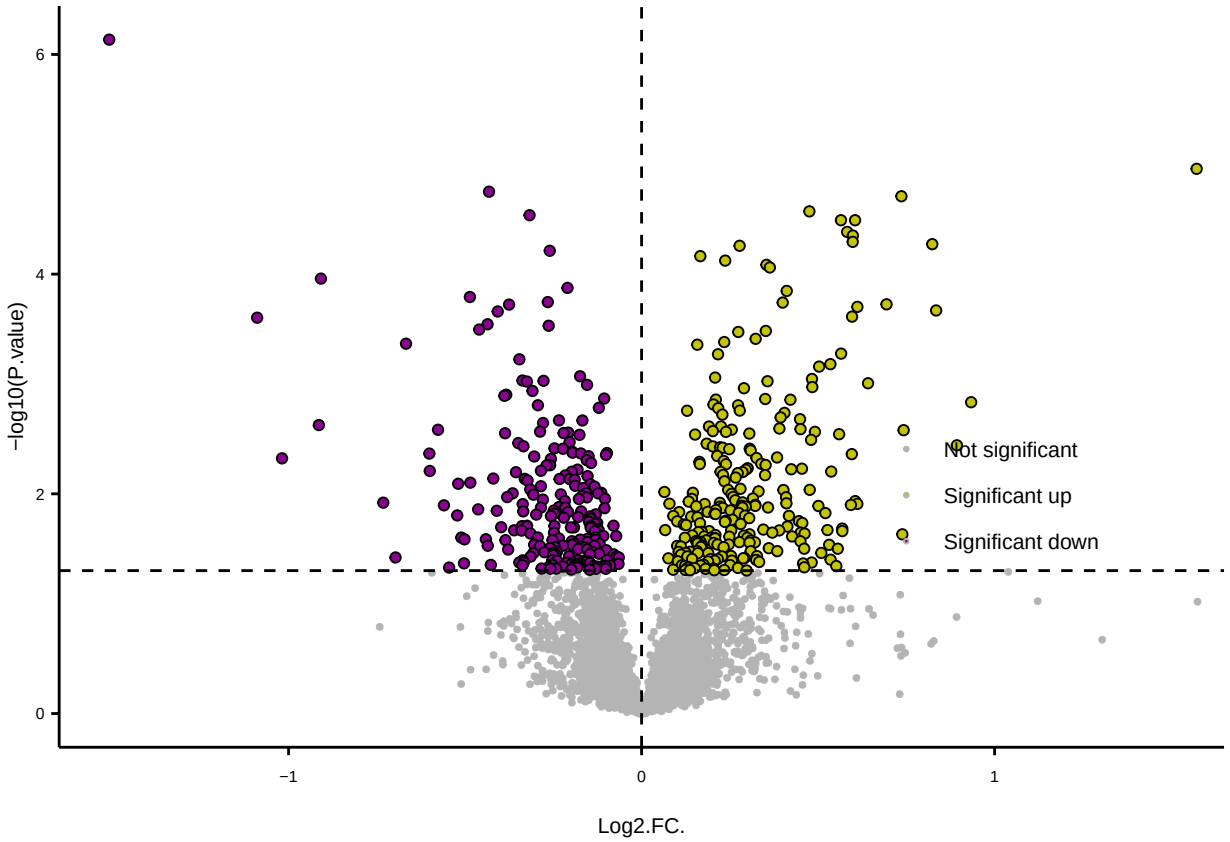
```
RPFncol <- "#8d0391"
```

ms volcano

```
pms <- 0.05
lfccutoffmax <- 1.5

# #make columns with identifiers
msco <- ms
# msco$lfccutoff <- "in"
# msco$lfccutoff[msco$Log2.FC. > lfccutoffmax] <- "high"
# msco$lfccutoff[msco$Log2.FC. < -lfccutoffmax] <- "low"
#
# #implement cutoff
# msco$Log2.FC.[msco$Log2.FC. > lfccutoffmax] <- lfccutoffmax
# msco$Log2.FC.[msco$Log2.FC. < -lfccutoffmax] <- -lfccutoffmax



#add significance identifier
msco$significance <- "Not significant"
msco$significance[msco$P.value <= pms & msco$Log2.FC. > 0] <- "Significant up"
msco$significance[msco$P.value <= pms & msco$Log2.FC. < 0] <- "Significant down"



#plot
mscovolcano <- ggplot(msco, aes(y=-log10(P.value), x=Log2.FC.,  fill=factor(significance, levels = c("N
  scale_fill_manual(values=c(farbeneg, RPFpcol, RPFncol)) +
  geom_point(data = msco[msco$significance == "Significant up",], shape=21, colour="black", fill=RPFpco
  geom_point(data = msco[msco$significance == "Significant down",], shape=21, colour="black", fill=RPFn
  # geom_point(data=msco[msco$significance == "Significant up" & msco$lfccutoff =="high",], aes(y=-log1
  # geom_point(data=msco[msco$significance == "Significant down" & msco$lfccutoff =="low",], aes(y=-log
  geom_hline(yintercept = -log10(pms), linetype="dashed") +
  geom_vline(xintercept = 0, linetype="dashed") +
  coord_cartesian(xlim = c(-lfccutoffmax, lfccutoffmax))+
  theme_paper()
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
mscovolcano
```




```
#export
pdf("Volcanoplot_MS.pdf", width = 2, height = 2)
mscovolcano
dev.off()
```

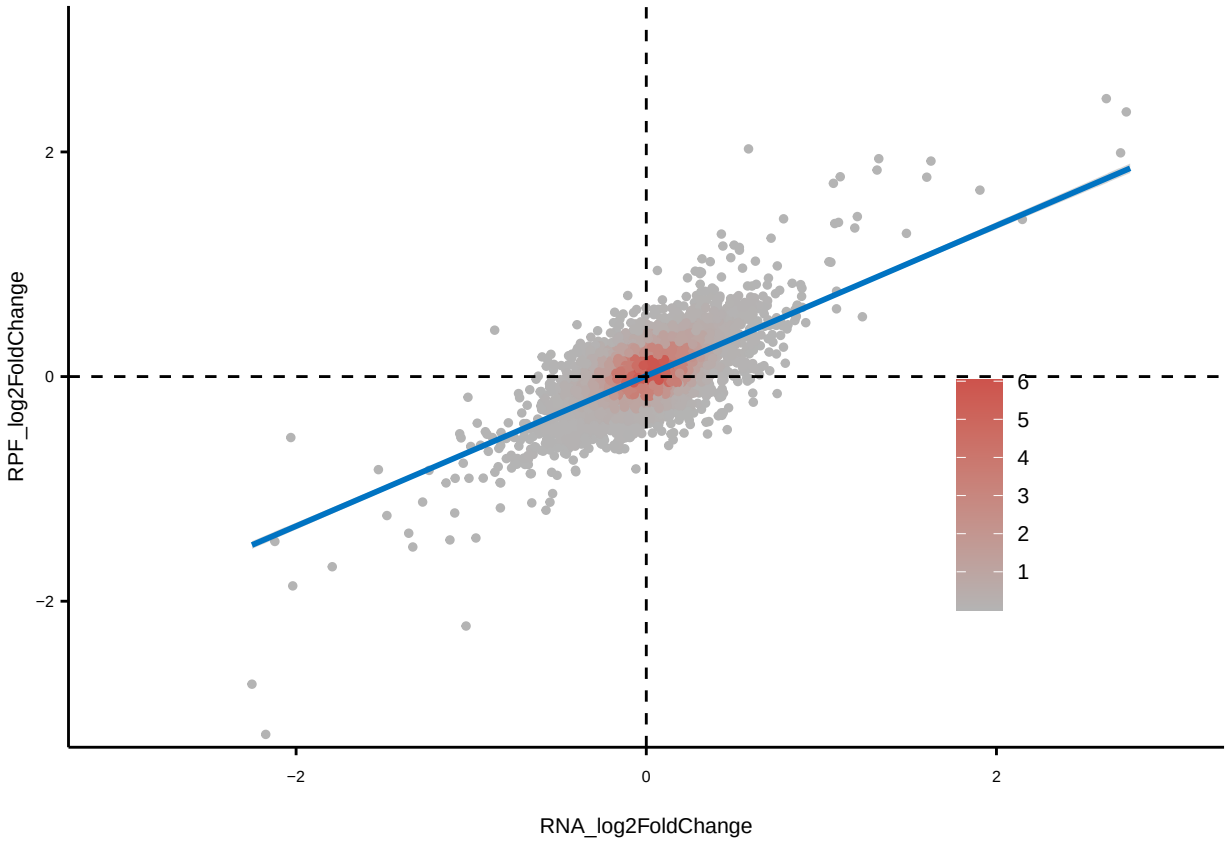
```
## pdf
##   2
```

TE crossover

```
names(RNA)[4] <- "RNA_log2FoldChange"
names(RPF)[4] <- "RPF_log2FoldChange"

ribo <- left_join(RPF[!is.na(RPF$gene_symbol),], RNA[!is.na(RNA$gene_symbol),], by="Gene")

#density function
get_density <- function(x, y, ...) {
  dens <- MASS::kde2d(x, y, ...)
  ix <- findInterval(x, dens$x)
  iy <- findInterval(y, dens$y)
  ii <- cbind(ix, iy)
  return(dens$z[ii])
}

set.seed(1)
```

```r
dat <- data.frame(
  x = c(
    rnorm(1e4, mean = 0, sd = 0.1),
    rnorm(1e3, mean = 0, sd = 0.1)
  ),
  y = c(
    rnorm(1e4, mean = 0, sd = 0.1),
    rnorm(1e3, mean = 0.1, sd = 0.2)
  )
)

#implement density
ribo$density <- get_density(ribo$RNA_log2FoldChange, ribo$RPF_log2FoldChange, n=100)

#colour brewer


#plot
riboplot <- ggplot(ribo, aes(x=RNA_log2FoldChange, y=RPF_log2FoldChange, colour=density)) +
  geom_point(size=1) +
  geom_smooth(method=lm, colour=farbe1) +
  geom_hline(yintercept = 0, linetype="dashed") +
  geom_vline(xintercept = 0, linetype="dashed") +
  coord_cartesian(ylim = c(-3,3), xlim = c(-3,3)) +
  scale_color_gradient(low = farbeneg, high = farbe3) +
  theme_paper()


riboplot
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
#export
pdf("Crossdots_TE.pdf", width = 2, height = 2)
riboplot
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
dev.off()
```

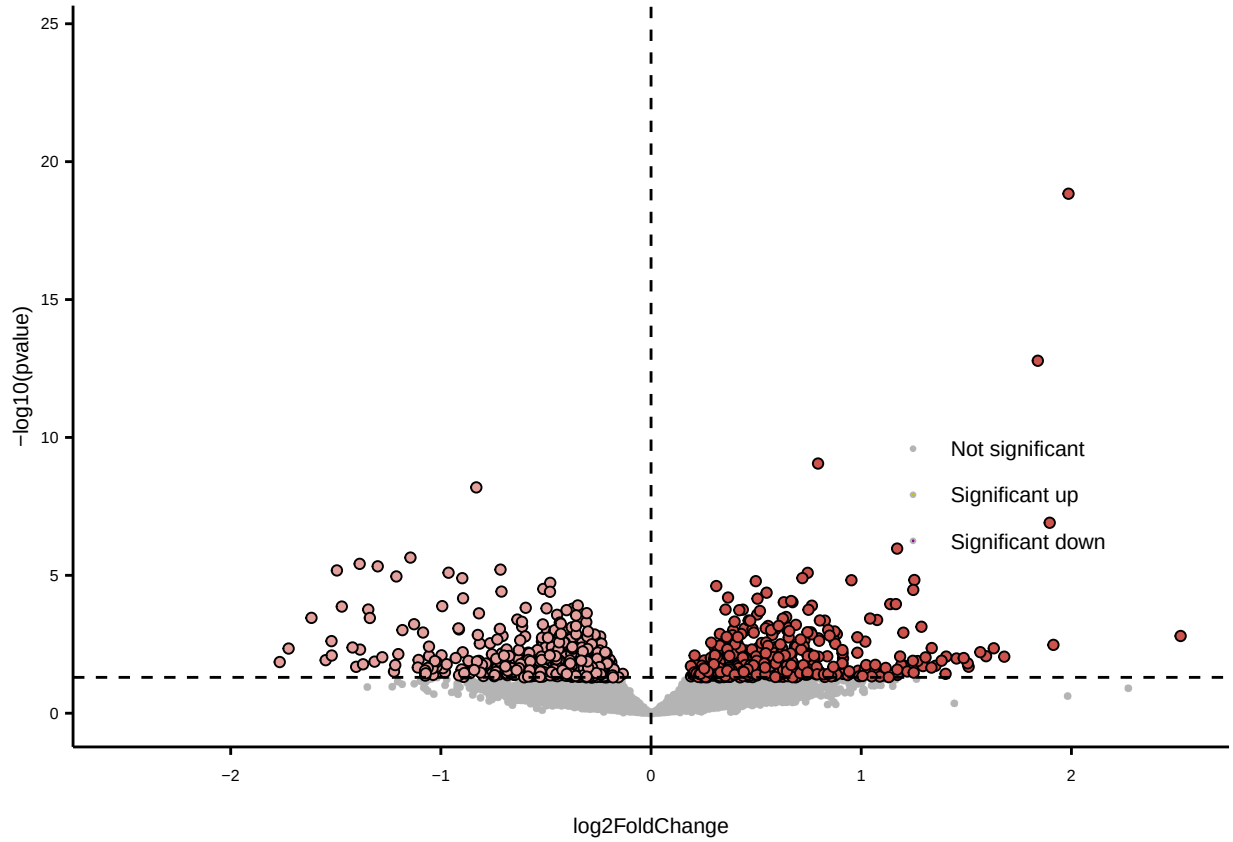
```
## pdf
##   2
```

#Te volcano

```
head(TEframe)
```

```
##   X  baseMean log2FoldChange     lfcSE        stat      pvalue      padj
## 1 1 2354.3220    -0.07560025 0.1336485 -0.56566464 0.571621774 0.9277760
## 2 2  579.2399    -0.06784576 0.1617420 -0.41946900 0.674873409 0.9512329
## 3 3   20.3333    -0.02453968 0.5874974 -0.04176985 0.966682171 0.9964083
## 4 4 1154.1664    -0.05314227 0.1359683 -0.39084299 0.695913297 0.9542063
## 5 5  635.5639     0.36967194 0.1364473  2.70926532 0.006743239 0.2886809
## 6 6  378.2069     0.10567650 0.1653495  0.63910971 0.522751566 0.9143531
##                   Gene gene_symbol
## 1  ENSMUSG00000000001.4       Gnai3
## 2 ENSMUSG00000000028.15       Cdc45
## 3 ENSMUSG00000000037.17       Scml2
## 4  ENSMUSG00000000056.7        Narf
## 5  ENSMUSG00000000078.7        Klf6
## 6 ENSMUSG00000000085.16       Scmh1
```

```
telfccutoff <- 2.5

#add significance identifier
TEco <- TEframe
TEco$significance <- "Not significant"
TEco$significance[TEco$pvalue <= pms & TEco$log2FoldChange > 0] <- "Significant up"
TEco$significance[TEco$pvalue <= pms & TEco$log2FoldChange < 0] <- "Significant down"



#plot
TEcovolcano <- ggplot(TEco, aes(y=-log10(pvalue), x=log2FoldChange,  fill=factor(significance, levels =
  scale_fill_manual(values=c(farbeneg, RPFpcol, RPFncol)) +
  geom_point(data = TEco[TEco$significance == "Significant up",], shape=21, colour="black", fill="#CD53
  geom_point(data = TEco[TEco$significance == "Significant down",], shape=21, colour="black", fill="#E3
  geom_hline(yintercept = -log10(pms), linetype="dashed") +
  geom_vline(xintercept = 0, linetype="dashed") +
  coord_cartesian(xlim = c(-telfccutoff, telfccutoff))+
  theme_paper()

TEcovolcano
```




```
#export
pdf("Volcanoplot_TE.pdf", width = 2, height = 2)
TEcovolcano
dev.off()
```

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

Session info

```
sessionInfo()
```

```
## R version 4.2.3 (2023-03-15 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Germany.utf8  LC_CTYPE=German_Germany.utf8
## [3] LC_MONETARY=German_Germany.utf8 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] MASS_7.3-58.2 dplyr_1.1.2   ggrastr_1.0.1 ggplot2_3.4.2 xlsx_0.6.5
##
## loaded via a namespace (and not attached):
##  [1] beeswarm_0.4.0   tidyselect_1.2.0 xfun_0.39        purrr_1.0.1
##  [5] lattice_0.20-45  splines_4.2.3    rJava_1.0-6      carData_3.0-5
##  [9] colorspace_2.1-0 vctrs_0.6.2      generics_0.1.3   htmltools_0.5.4
## [13] yaml_2.3.7       mgcv_1.8-42      utf8_1.2.3       rlang_1.1.0
## [17] pillar_1.9.0     ggpubr_0.6.0     glue_1.6.2       withr_2.5.0
## [21] lifecycle_1.0.3  munsell_0.5.0    ggsignif_0.6.4   gtable_0.3.3
## [25] evaluate_0.21    labeling_0.4.2   knitr_1.42       fastmap_1.1.1
## [29] Cairo_1.6-0      vipor_0.4.5      fansi_1.0.4      xlsxjars_0.6.1
## [33] highr_0.10       broom_1.0.4      scales_1.2.1     backports_1.4.1
## [37] abind_1.4-5      farver_2.1.1     digest_0.6.31    rstatix_0.7.2
## [41] grid_4.2.3       cli_3.6.0        tools_4.2.3      magrittr_2.0.3
## [45] tibble_3.2.1     crayon_1.5.2     tidyr_1.3.0      car_3.1-2
## [49] pkgconfig_2.0.3  Matrix_1.5-3     ggbeeswarm_0.7.2 rmarkdown_2.21
## [53] rstudioapi_0.14  R6_2.5.1         nlme_3.1-162     compiler_4.2.3
```