

deltaTE_figures_m23

Nikita Verheyden

2023-01-20

Setup

Directory

```
setwd("D:/Krueger_Lab/Publications/miR181_paper/Supporting_scripts/deltaTE")
```

Packages

```
library(DESeq2)
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

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

```
##
```

```
##      IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
##      colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
##      get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
##      match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
##      Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,  
##      table, tapply, union, unique, unsplit, which.max, which.min
```

```
##
```

```
## Attaching package: 'S4Vectors'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      expand.grid, I, unname
```

```
## Loading required package: IRanges
```

```
##
```

```
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:grDevices':
```

```
##
```

```

## windows
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Loading required package: SummarizedExperiment
## Loading required package: MatrixGenerics
## Loading required package: matrixStats
##
## Attaching package: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
##
## colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
## colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
## colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
## colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
## colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
## colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
## colWeightedMeans, colWeightedMedians, colWeightedSds,
## colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
## rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
## rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
## rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
## rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
## rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
## rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
## rowWeightedSds, rowWeightedVars
## Loading required package: Biobase
## Welcome to Bioconductor
##
## Vignettes contain introductory material; view with
## 'browseVignettes()'. To cite Bioconductor, see
## 'citation("Biobase")', and for packages 'citation("pkgname)".
##
## Attaching package: 'Biobase'
## The following object is masked from 'package:MatrixGenerics':
##
## rowMedians
## The following objects are masked from 'package:matrixStats':
##
## anyMissing, rowMedians
library(dplyr)

##
## Attaching package: 'dplyr'
## The following object is masked from 'package:Biobase':
##
## combine

```

```
## The following object is masked from 'package:matrixStats':
##
##     count

## The following objects are masked from 'package:GenomicRanges':
##
##     intersect, setdiff, union

## The following object is masked from 'package:GenomeInfoDb':
##
##     intersect

## The following objects are masked from 'package:IRanges':
##
##     collapse, desc, intersect, setdiff, slice, union

## The following objects are masked from 'package:S4Vectors':
##
##     first, intersect, rename, setdiff, setequal, union

## The following objects are masked from 'package:BiocGenerics':
##
##     combine, intersect, setdiff, union

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

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

Load data

```
ribo_counts = read.csv("D:/Krueger_Lab/Publications/miR181_paper/Figure3/RPFadjusted_countmatrix_15112020.csv")
rna_counts = read.csv("D:/Krueger_Lab/Publications/miR181_paper/Figure3/RNAadjusted_countmatrix_15112020.csv")

colnames(ribo_counts) <- c("Gene ID", "RPF_WT_1411", "RPF_WT_1601", "RPF_WT_1710", "RPF_KO_1411", "RPF_KO_1601", "RPF_KO_1710")
colnames(rna_counts) <- c("Gene ID", "RNA_WT_1411", "RNA_WT_1601", "RNA_WT_1710", "RNA_KO_1411", "RNA_KO_1601", "RNA_KO_1710")

head(ribo_counts)
```

```
##           Gene ID RPF_WT_1411 RPF_WT_1601 RPF_WT_1710 RPF_KO_1411
## 1 ENSMUSG00000000001.4      5647      5703      3586      3716
## 2 ENSMUSG00000000003.15         0         0         0         0
## 3 ENSMUSG00000000028.15     1672     1811     1098     1029
## 4 ENSMUSG00000000031.16         5         8         5         7
## 5 ENSMUSG00000000037.17        21        35        27        27
## 6 ENSMUSG00000000049.11         1         2         1         0
## RPF_KO_1601 RPF_KO_1710
## 1      5361      1326
## 2         0         0
## 3      1386      355
## 4         6         2
## 5        24         2
## 6         2         0
```

```
head(rna_counts)
```

```
##           Gene ID RNA_WT_1411 RNA_WT_1601 RNA_WT_1710 RNA_KO_1411
## 1  ENSMUSG00000000001.4      1300      1561      1284      1435
## 2  ENSMUSG00000000003.15         0         0         0         0
## 3  ENSMUSG00000000028.15      205      265      304      271
## 4  ENSMUSG00000000031.16       50       66      109       46
## 5  ENSMUSG00000000037.17       16       25       27       25
## 6  ENSMUSG00000000049.11         0         0         0         0
##   RNA_KO_1601 RNA_KO_1710
## 1         828        1914
## 2          0          0
## 3        149         241
## 4         25          17
## 5         11          20
## 6          1           0
```

```
#Load results tables form Ribo profiling normal analysis to pick the same genes
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")
```

Sort out genes that are cut of in the normal analysis

```
ribo_countsCUT = ribo_counts[ribo_counts$`Gene ID` %in% RNA$Gene & ribo_counts$`Gene ID` %in% RPF$Gene,]
rna_countsCUT = rna_counts[rna_counts$`Gene ID` %in% RNA$Gene & rna_counts$`Gene ID` %in% RPF$Gene,]
```

deltaTE analysis

make metadata table

```
sample_info = data.frame(as.factor(c(colnames(ribo_countsCUT[, -1]), colnames(rna_countsCUT[, -1]))), as.
colnames(sample_info) <- c("Sample ID", "Condition", "SeqType", "Batch")
sample_info
```

```
##      Sample ID Condition SeqType Batch
## 1  RPF_WT_1411      WT      RIBO  1411
## 2  RPF_WT_1601      WT      RIBO  1601
## 3  RPF_WT_1710      WT      RIBO  1710
## 4  RPF_KO_1411      KO      RIBO  1411
## 5  RPF_KO_1601      KO      RIBO  1601
## 6  RPF_KO_1710      KO      RIBO  1710
## 7  RNA_WT_1411      WT      RNA   1411
## 8  RNA_WT_1601      WT      RNA   1601
## 9  RNA_WT_1710      WT      RNA   1710
## 10 RNA_KO_1411      KO      RNA   1411
## 11 RNA_KO_1601      KO      RNA   1601
## 12 RNA_KO_1710      KO      RNA   1710
```

Count tables

```
rownames(ribo_countsCUT) <- ribo_countsCUT[, 1]
ribo_countsCUT <- ribo_countsCUT[, -1]
rownames(rna_countsCUT) <- rna_countsCUT[, 1]
rna_countsCUT <- rna_countsCUT[, -1]
```

```
head(ribo_countsCUT)
```

```
##               RPF_WT_1411 RPF_WT_1601 RPF_WT_1710 RPF_KO_1411
## ENSMUSG000000000001.4      5647      5703      3586      3716
## ENSMUSG000000000028.15     1672      1811      1098      1029
## ENSMUSG000000000037.17        21        35        27        27
## ENSMUSG000000000056.7     2391     2147     1252     1279
## ENSMUSG000000000078.7       994     1055       578     1003
## ENSMUSG000000000085.16       498       515       368       334
##               RPF_KO_1601 RPF_KO_1710
## ENSMUSG000000000001.4      5361      1326
## ENSMUSG000000000028.15     1386       355
## ENSMUSG000000000037.17        24         2
## ENSMUSG000000000056.7     2099       560
## ENSMUSG000000000078.7     1376       391
## ENSMUSG000000000085.16       471        99
```

```
head(rna_countsCUT)
```

```
##               RNA_WT_1411 RNA_WT_1601 RNA_WT_1710 RNA_KO_1411
## ENSMUSG000000000001.4     1300     1561     1284     1435
## ENSMUSG000000000028.15      205      265      304      271
## ENSMUSG000000000037.17       16       25       27       25
## ENSMUSG000000000056.7      925      962      937      924
## ENSMUSG000000000078.7      423      503      427      537
## ENSMUSG000000000085.16      406      471      445      379
##               RNA_KO_1601 RNA_KO_1710
## ENSMUSG000000000001.4       828     1914
## ENSMUSG000000000028.15       149      241
## ENSMUSG000000000037.17        11       20
## ENSMUSG000000000056.7       628     1218
## ENSMUSG000000000078.7       315      705
## ENSMUSG000000000085.16       228      463
```

```
str(sample_info)
```

```
## 'data.frame':   12 obs. of  4 variables:
## $ Sample ID: Factor w/ 12 levels "RNA_KO_1411",...: 10 11 12 7 8 9 4 5 6 1 ...
## $ Condition: Factor w/ 2 levels "KO","WT": 2 2 2 1 1 1 2 2 2 1 ...
## $ SeqType : Factor w/ 2 levels "RIBO","RNA": 1 1 1 1 1 1 2 2 2 2 ...
## $ Batch : Factor w/ 3 levels "1411","1601",...: 1 2 3 1 2 3 1 2 3 1 ...
```

Reset reference samples and run DEseq

```
ddsMat = DESeqDataSetFromMatrix(countData=cbind(ribo_countsCUT,rna_countsCUT), colData=sample_info, des
```

```
ddsMat$SeqType <- relevel(ddsMat$SeqType, ref = "RNA")
ddsMat$Condition <- relevel(ddsMat$Condition, ref = "WT")
```

```
#Run DESeq2:
```

```
ddsMat = DESeq(ddsMat)
```

```
## estimating size factors
```

```

## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing
#Obtain fold changes for TE:

resultsNames(ddsMat)

## [1] "Intercept"          "Condition_KO_vs_WT"
## [3] "SeqType_RIBO_vs_RNA" "ConditionK0.SeqTypeRIBO"

#this is the correct one
res = results(ddsMat, name="ConditionK0.SeqTypeRIBO")

head(res)

## log2 fold change (MLE): ConditionK0.SeqTypeRIBO
## Wald test p-value: ConditionK0.SeqTypeRIBO
## Dataframe with 6 rows and 6 columns
##
##          baseMean log2FoldChange    lfcSE      stat      pvalue
##          <numeric>      <numeric> <numeric> <numeric> <numeric>
## ENSMUSG000000000001.4  2354.3220   -0.0756003  0.133649 -0.5656646  0.57162177
## ENSMUSG000000000028.15  579.2399   -0.0678458  0.161742 -0.4194690  0.67487341
## ENSMUSG000000000037.17   20.3333   -0.0245397  0.587497 -0.0417698  0.96668217
## ENSMUSG000000000056.7  1154.1664   -0.0531423  0.135968 -0.3908430  0.69591330
## ENSMUSG000000000078.7   635.5639    0.3696719  0.136447  2.7092653  0.00674324
## ENSMUSG000000000085.16  378.2069    0.1056765  0.165350  0.6391097  0.52275157
##
##          padj
##          <numeric>
## ENSMUSG000000000001.4   0.927776
## ENSMUSG000000000028.15  0.951233
## ENSMUSG000000000037.17  0.996408
## ENSMUSG000000000056.7   0.954206
## ENSMUSG000000000078.7   0.288681
## ENSMUSG000000000085.16  0.914353

add gene names

res$Gene = rownames(res)
TEframe = left_join(as.data.frame(res), RNA[, c("Gene", "gene_symbol")], by = "Gene")
head(TEframe)

##          baseMean log2FoldChange    lfcSE      stat      pvalue      padj
## 1  2354.3220   -0.07560025  0.1336485 -0.56566464  0.571621774  0.9277760
## 2   579.2399   -0.06784576  0.1617420 -0.41946900  0.674873409  0.9512329
## 3    20.3333   -0.02453968  0.5874974 -0.04176985  0.966682171  0.9964083
## 4  1154.1664   -0.05314227  0.1359683 -0.39084299  0.695913297  0.9542063
## 5   635.5639    0.36967194  0.1364473  2.70926532  0.006743239  0.2886809
## 6   378.2069    0.10567650  0.1653495  0.63910971  0.522751566  0.9143531
##
##          Gene gene_symbol
## 1  ENSMUSG000000000001.4   Gnai3
## 2  ENSMUSG000000000028.15   Cdc45

```

```
## 3 ENSMUSG00000000037.17      Scml2
## 4  ENSMUSG00000000056.7       Narf
## 5  ENSMUSG00000000078.7       Klf6
## 6  ENSMUSG00000000085.16      Scmh1
```

write results

```
write.csv(TEframe, "TE_m23_genesInRNAandRPF.csv", quote=F)
```

```
sessionInfo()
```

```
## R version 4.2.3 (2023-03-15 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## 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] stats4      stats      graphics  grDevices  utils      datasets  methods
## [8] base
##
## other attached packages:
## [1] dplyr_1.1.1           DESeq2_1.38.3
## [3] SummarizedExperiment_1.28.0 Biobase_2.58.0
## [5] MatrixGenerics_1.10.0    matrixStats_0.63.0
## [7] GenomicRanges_1.50.2     GenomeInfoDb_1.34.9
## [9] IRanges_2.32.0          S4Vectors_0.36.2
## [11] BiocGenerics_0.44.0
##
## loaded via a namespace (and not attached):
## [1] locfit_1.5-9.7          Rcpp_1.0.10             lattice_0.20-45
## [4] png_0.1-8              Biostrings_2.66.0       digest_0.6.31
## [7] utf8_1.2.3             R6_2.5.1               RSQLite_2.3.0
## [10] evaluate_0.20          httr_1.4.5             ggplot2_3.4.1
## [13] pillar_1.9.0           zlibbioc_1.44.0         rlang_1.1.0
## [16] rstudioapi_0.14        annotate_1.76.0         blob_1.2.4
## [19] Matrix_1.5-3           rmarkdown_2.21         BiocParallel_1.32.6
## [22] geneplotter_1.76.0     RCurl_1.98-1.12        bit_4.0.5
## [25] munsell_0.5.0          DelayedArray_0.23.2     compiler_4.2.3
## [28] xfun_0.37              pkgconfig_2.0.3        htmltools_0.5.4
## [31] tidyselect_1.2.0       KEGGREST_1.38.0        tibble_3.2.1
## [34] GenomeInfoDbData_1.2.9 codetools_0.2-19       XML_3.99-0.14
## [37] fansi_1.0.4            crayon_1.5.2           bitops_1.0-7
## [40] grid_4.2.3            xtable_1.8-4           gtable_0.3.3
## [43] lifecycle_1.0.3       DBI_1.1.3              magrittr_2.0.3
## [46] scales_1.2.1          cli_3.6.0              cachem_1.0.7
## [49] XVector_0.38.0        generics_0.1.3         vctrs_0.6.1
## [52] RColorBrewer_1.1-3    tools_4.2.3           bit64_4.0.5
## [55] glue_1.6.2            parallel_4.2.3         fastmap_1.1.1
## [58] yaml_2.3.7            AnnotationDbi_1.60.2    colorspace_2.1-0
```

```
## [61] memoise_2.0.1      knitr_1.42
```