DEG analysis using DESeq2 (TE)

Toshihiro Arae

General directory setting

```
wd <- here::here()
shared <- fs::path(fs::path_dir(wd), "shared")
```

Loading packages

```
library(magrittr)
library(ggplot2)
```

Load common R scripts

```
#source(fs::path(wd, "script_r", "MISC.R"))
#source(fs::path(here::here(), "script_r", "MISC_PALETTE.R"))
```

Directory setting

```
dir_output <- fs::path("analysis", "deseq2_te")
path_out <- function(...) fs::path(wd, dir_output, ...)
fs::dir_create(path_out())
```

Loading input files

```
dir_readcount <- fs::path("data_preproc", "readcount")

# data.frame of RNA-seq read count
rcdf_rna <-
  fs::path(wd, dir_readcount, "count_rna_exon", "count_by_gene.csv") %>%
  readr::read_csv(show_col_types = FALSE) %>%
  dplyr::select(Geneid, dplyr::starts_with("zt")) %>%
  ngsmisc::ds2_tbl_to_rcdf()
head(rcdf_rna)
```

```
          zt0_1_rna zt0_2_rna zt12_1_rna zt12_2_rna zt18_1_rna zt18_2_rna
AT1G01010        20        30         44         43         35         40
AT1G01020        81        82         81         66         91         88
AT1G03987         7         7          0          1          4          2
AT1G01030       167       205         55         62        121        119
AT1G01040       566       633        384        396        738        734
AT1G03993        87       111         68         68        130        134
          zt21_1_rna zt21_2_rna zt3_1_rna zt3_2_rna zt6_1_rna zt6_2_rna
AT1G01010         39         23        31        27        32        22
AT1G01020        109        109        72        80        66        79
AT1G03987          1          7         9         5         3         6
AT1G01030         88         87       114       121        93        81
AT1G01040        620        606       489       532       454       439
AT1G03993        105         84        89       100        73        63
```

```r
dim(rcdf_rna)
```

```
[1] 36917    12
```

```r
# data.frame of Ribo-seq (mORF) read count
rcdf_ribo <-
  fs::path(wd, dir_readcount, "count_ribo_central_cds_psite", "count_by_gene.csv") %>%
  readr::read_csv(show_col_types = FALSE) %>%
  dplyr::select(Geneid, dplyr::starts_with("zt")) %>%
  ngsmisc::ds2_tbl_to_rcdf()
head(rcdf_ribo)
```

```
          zt0_1_ribo zt0_2_ribo zt12_1_ribo zt12_2_ribo zt18_1_ribo zt18_2_ribo
AT1G01010         59         31          82          56          65          45
AT1G01020         74         56          35          32          55          50
AT1G01030        125         61          47          50          63          57
AT1G01040        136         55         334          77         175         177
AT1G01050       1123        597         616         314        1233        1276
AT1G01060       4257       2356           0           2         122         132
          zt21_1_ribo zt21_2_ribo zt3_1_ribo zt3_2_ribo zt6_1_ribo zt6_2_ribo
AT1G01010          31          41         48         25         19         28
AT1G01020          46          56         43         40         18         35
AT1G01030          43          74         55         45         20         38
AT1G01040         123         139        160         54        154         90
AT1G01050         887        1111        414        312        314        263
AT1G01060        1618        2120        168        102         12         18
```

```r
dim(rcdf_ribo)
```

```
[1] 27631    12
```

```r
common_rownames <- intersect(row.names(rcdf_rna), row.names(rcdf_ribo))
length(common_rownames)
```

```
[1] 27631
```

```r
rcdf_ribo <- rcdf_ribo[row.names(rcdf_ribo) %in% common_rownames,]
rcdf_ribo <- rcdf_ribo[order(row.names(rcdf_ribo)),]
head(rcdf_ribo)
```

```
          zt0_1_ribo zt0_2_ribo zt12_1_ribo zt12_2_ribo zt18_1_ribo zt18_2_ribo
AT1G01010         59         31          82          56          65          45
AT1G01020         74         56          35          32          55          50
AT1G01030        125         61          47          50          63          57
AT1G01040        136         55         334          77         175         177
AT1G01050       1123        597         616         314        1233        1276
AT1G01060       4257       2356           0           2         122         132
          zt21_1_ribo zt21_2_ribo zt3_1_ribo zt3_2_ribo zt6_1_ribo zt6_2_ribo
AT1G01010          31          41         48         25         19         28
AT1G01020          46          56         43         40         18         35
AT1G01030          43          74         55         45         20         38
AT1G01040         123         139        160         54        154         90
AT1G01050         887        1111        414        312        314        263
AT1G01060        1618        2120        168        102         12         18
```

```
nrow(rcdf_ribo)
```

```
[1] 27631
```

```
rcdf_rna <- rcdf_rna[row.names(rcdf_rna) %in% common_rownames,]
rcdf_rna <- rcdf_rna[order(row.names(rcdf_rna)),]
identical(row.names(rcdf_rna), row.names(rcdf_ribo))
```

```
[1] TRUE
```

```
# setting sample information
zt_lev <- paste0("zt", c(0, 3, 6, 12, 18, 21))
rna_condition <-
  rep(paste0("zt", c(0, 12, 18, 21, 3, 6)), each = 2) %>%
  forcats::fct_relevel(zt_lev)
ribo_condition <- rna_condition
```

Data pre-processing

```
coldata <- data.frame(
  zt =
    c(colnames(rcdf_rna), colnames(rcdf_ribo)) %>%
    stringr::str_extract("zt\\d+") %>%
    forcats::fct_relevel(zt_lev),
  type =
    forcats::fct_inorder(c("rna", "ribo")) %>%
    rep(each = 12)
)

dds <-
  ngsmisc::ds2_rcdf_to_dds(
    rcdf = cbind(rcdf_rna, rcdf_ribo),
    coldata = coldata,
    design = ~ zt + type + zt:type
  )

sf_default_rna <-
  readRDS(fs::path(wd, "analysis", "deseq2_rna", "sf_default_rna.rds"))
sf_default_ribo <-
  readRDS(fs::path(wd, "analysis", "deseq2_ribo", "sf_default_ribo.rds"))
dds %>% ngsmisc::ds2_dds_get_sizefactor()
```

```
   zt0_1_rna    zt0_2_rna   zt12_1_rna   zt12_2_rna   zt18_1_rna   zt18_2_rna
   0.8168349    0.9102019    1.0909891    1.0749195    0.9724547    0.9610146
  zt21_1_rna   zt21_2_rna    zt3_1_rna    zt3_2_rna    zt6_1_rna    zt6_2_rna
   0.8772633    0.8984975    0.9142426    0.9129412    0.8926448    0.8789882
   zt0_1_ribo   zt0_2_ribo  zt12_1_ribo  zt12_2_ribo  zt18_1_ribo  zt18_2_ribo
   1.5141551    0.9670998    1.4913948    0.9683269    1.2424976    1.2322901
  zt21_1_ribo  zt21_2_ribo   zt3_1_ribo   zt3_2_ribo   zt6_1_ribo   zt6_2_ribo
   1.0614554    1.2865899    1.1623312    0.8626377    0.8034826    0.8800929
```

```
dds <-
  dds %>%
  ngsmisc::ds2_dds_estimate_sizefactor() %>%
  ngsmisc::ds2_dds_estimate_disp() %>%
  ngsmisc::ds2_dds_test_nbinomLRT(reduced = ~ zt + type)
```

```
gene-wise dispersion estimates
```

```
mean-dispersion relationship
```

```
final dispersion estimates
```

```
ddr <- DESeq2::results(dds, alpha = .01)
ddr %>% DESeq2::summary()
```

```
out of 25046 with nonzero total read count
adjusted p-value < 0.01
LFC > 0 (up)       : 3845, 15%
LFC < 0 (down)     : 3798, 15%
outliers [1]       : 0, 0%
low counts [2]     : 5796, 23%
(mean count < 7)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

```
ddr@metadata$filterThreshold
```

```
30.32961%
 6.689895
```

```
ddr %>% ngsmisc::ds2_ddr_plot_independent_filtering()
```
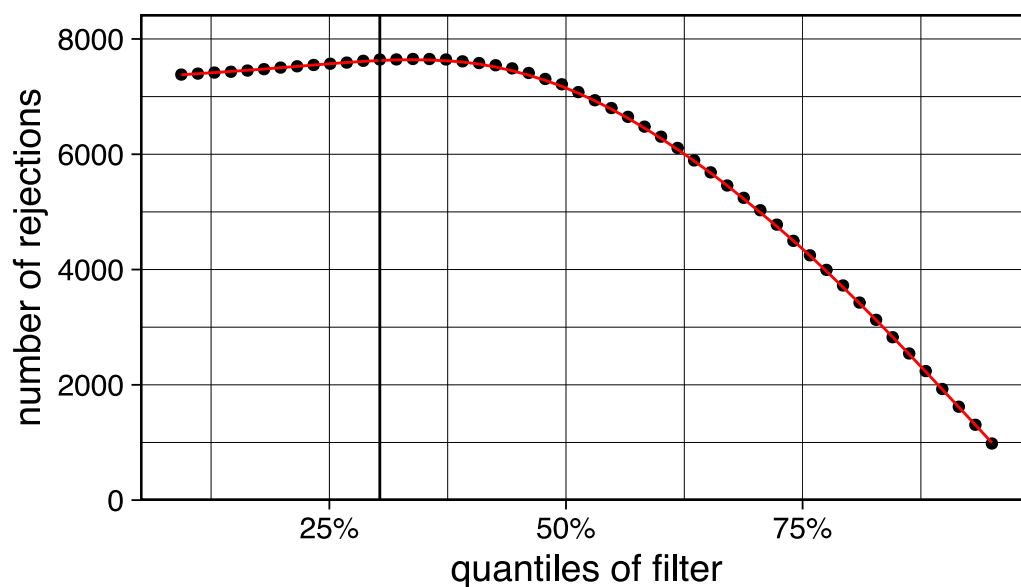


```
tbl_out <-
  dds %>%
  ngsmisc::ds2_dds_to_tbl() %>%
  dplyr::rowwise() %>%
  dplyr::mutate(
    l2fc_amp =
```

```
      range(dplyr::c_across(ztzt3.typeribo:ztzt21.typeribo)) %>%
      {.[2] - .[1]}
  ) %>%
  dplyr::ungroup()
tbl_out %>% dplyr::filter(abs(l2fc_amp) >= 1) %>% dplyr::glimpse()
```

```
Rows: 12,822
Columns: 44
$ Geneid            <chr> "AT1G01030", "AT1G01040", "AT1G01060", "AT1G01070"…
$ baseMean          <dbl> 8.450193e+01, 3.579584e+02, 2.205541e+03, 9.173573…
$ baseVar           <dbl> 2.664843e+03, 6.952273e+04, 1.605164e+07, 4.743736…
$ allZero           <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F…
$ dispGeneEst       <dbl> 0.0049598899, 0.0750919192, 0.0000000100, 0.012213…
$ dispGeneIter      <dbl> 76, 9, 41, 6, 2, 8, 3, 2, 6, 2, 14, 13, 8, 33, 2, …
$ dispFit           <dbl> 0.033966949, 0.014688393, 0.009697922, 0.031976977…
$ dispersion        <dbl> 0.018422883, 0.053077919, 0.004788910, 0.021549153…
$ dispIter          <dbl> 10, 10, 7, 10, 7, 9, 9, 8, 9, 9, 6, 8, 10, 9, 8, 9…
$ dispOutlier       <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F…
$ dispMAP           <dbl> 0.018422883, 0.053077919, 0.004788910, 0.021549153…
$ Intercept         <dbl> 7.7479480, 9.4391781, 13.7377224, 6.3580122, 2.793…
$ zt_zt3_vs_zt0     <dbl> -0.74103133, -0.31299021, -3.74856063, -0.26175966…
$ zt_zt6_vs_zt0     <dbl> -1.13050369, -0.46183833, -7.67989086, -1.05322558…
$ zt_zt12_vs_zt0    <dbl> -1.99218213, -0.94658714, -12.26777717, -2.0141531…
$ zt_zt18_vs_zt0    <dbl> -0.7922634, 0.1332187, -4.7521798, -0.8183533, 1.3…
$ zt_zt21_vs_zt0    <dbl> -1.124985919, -0.007480801, -1.016869635, -0.09811…
$ type_ribo_vs_rna  <dbl> -1.5455258, -3.2299271, -2.3788402, 1.6737628, 1.2…
$ ztzt3.typeribo    <dbl> 0.16953214, 0.76204877, -0.55955642, -1.32577561, …
$ ztzt6.typeribo    <dbl> 0.02847930, 1.44925364, 0.47569118, -1.45922597, -…
$ ztzt12.typeribo   <dbl> 1.12724173, 1.99652374, 0.61217403, -0.37049547, -…
$ ztzt18.typeribo   <dbl> 0.189283724, 0.809709835, 0.074775443, 0.007066597…
$ ztzt21.typeribo   <dbl> 0.54927704, 0.60443208, 0.28975197, -0.48373709, -…
$ SE_Intercept      <dbl> 0.15745312, 0.23870130, 0.07121953, 0.19272858, 0.…
$ SE_zt_zt3_vs_zt0  <dbl> 0.2298274, 0.3380150, 0.1057202, 0.2760798, 0.6762…
$ SE_zt_zt6_vs_zt0  <dbl> 0.2364946, 0.3384475, 0.1664080, 0.2988189, 0.6811…
$ SE_zt_zt12_vs_zt0 <dbl> 0.2484940, 0.3389451, 0.5974624, 0.3269472, 0.6205…
$ SE_zt_zt18_vs_zt0 <dbl> 0.2294266, 0.3370899, 0.1103618, 0.2875986, 0.6333…
$ SE_zt_zt21_vs_zt0 <dbl> 0.2363407, 0.3375102, 0.1011415, 0.2736172, 0.6677…
$ SE_type_ribo_vs_rna <dbl> 0.2360363, 0.3521778, 0.1019345, 0.2508343, 0.6291…
$ SE_ztzt3.typeribo <dbl> 0.3519622, 0.4969006, 0.1711678, 0.3693171, 0.8775…
$ SE_ztzt6.typeribo <dbl> 0.3769433, 0.4952552, 0.3277424, 0.4055517, 0.9198…
$ SE_ztzt12.typeribo <dbl> 0.3650116, 0.4928818, 1.1861238, 0.4152425, 0.8296…
$ SE_ztzt18.typeribo <dbl> 0.3465216, 0.4917629, 0.1751449, 0.3695291, 0.8367…
$ SE_ztzt21.typeribo <dbl> 0.3521037, 0.4941522, 0.1452416, 0.3580138, 0.8854…
$ LRTStatistic      <dbl> 12.49621313, 19.79866059, 27.40139807, 25.65496889…
$ LRTPvalue         <dbl> 2.858612e-02, 1.363237e-03, 4.763985e-05, 1.040891…
$ fullBetaConv      <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TR…
$ reducedBetaConv   <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TR…
$ betaIter          <dbl> 3, 3, 4, 3, 3, 2, 3, 3, 2, 2, 3, 5, 3, 2, 13, 3, 4…
$ deviance          <dbl> 178.12983, 260.70902, 216.19387, 182.39256, 124.48…
$ maxCooks          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
$ padj              <dbl> 5.910780e-02, 4.078011e-03, 2.026272e-04, 4.117132…
$ l2fc_amp          <dbl> 1.098762, 1.392092, 1.171730, 1.466293, 1.275295, …
```

```
tbl_out <- tbl_out %>% dplyr::select(1:11, 18:23, 30:44)
coefs <-
  c("intercept", "zt3", "zt6", "zt12", "zt18", "zt21") %>%
  paste0("coef_", .)
ses <-
  c("intercept", "zt3", "zt6", "zt12", "zt18", "zt21") %>%
  paste0("se_", .)
```

```
colnames(tbl_out) <-
  c("AGI", "baseMean", "baseVar", "allZero",
    "dispGeneEst", "dispGeneIter", "dispFit",
    "dispersion", "dispIter", "dispOutlier", "dispMAP",
    coefs, ses, "stat_LRT", "pvalue", "fullBetaConv", "reducedBetaConv",
    "betaIter", "deviance", "maxCooks", "padj", "l2fc_amp")
readr::write_csv(tbl_out, path_out("deg_all.csv"))



AGI_filtered_rna <-
  readRDS(fs::path(wd, "analysis", "deseq2_rna", "AGI_filtered_rna.rds"))
AGI_filtered_ribo <-
  readRDS(fs::path(wd, "analysis", "deseq2_ribo", "AGI_filtered_ribo.rds"))
```

Sessioninfo

```
sessionInfo()
```

```
R version 4.2.1 (2022-06-23)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Ventura 13.1

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices datasets  utils     methods   base

other attached packages:
[1] ggplot2_3.4.2  magrittr_2.0.3

loaded via a namespace (and not attached):
 [1] MatrixGenerics_1.8.1       Biobase_2.56.0
 [3] httr_1.4.5                 splines_4.2.1
 [5] bit64_4.0.5                vroom_1.6.0
 [7] jsonlite_1.8.4             here_1.0.1
 [9] BiocManager_1.30.18        stats4_4.2.1
[11] blob_1.2.3                 renv_1.0.3
[13] GenomeInfoDbData_1.2.8     yaml_2.3.6
[15] pillar_1.9.0               RSQLite_2.2.18
[17] lattice_0.20-45            glue_1.6.2
[19] digest_0.6.31              RColorBrewer_1.1-3
[21] GenomicRanges_1.48.0       XVector_0.36.0
[23] colorspace_2.0-3           htmltools_0.5.3
[25] Matrix_1.6-4               DESeq2_1.36.0
[27] XML_3.99-0.11              pkgconfig_2.0.3
[29] genefilter_1.78.0          zlibbioc_1.42.0
[31] purrr_1.0.1                xtable_1.8-4
[33] scales_1.2.1               tzdb_0.3.0
[35] BiocParallel_1.30.4        tibble_3.2.1
[37] annotate_1.74.0            KEGGREST_1.36.3
[39] farver_2.1.1               generics_0.1.3
[41] IRanges_2.30.1             cachem_1.0.6
[43] withr_2.5.0                SummarizedExperiment_1.26.1
[45] BiocGenerics_0.42.0        cli_3.6.0
[47] survival_3.3-1             crayon_1.5.2
```

```
[49] memoise_2.0.1          evaluate_0.20
[51] fs_1.5.2               fansi_1.0.3
[53] forcats_1.0.0          tools_4.2.1
[55] hms_1.1.3              lifecycle_1.0.3
[57] matrixStats_0.62.0     stringr_1.5.0
[59] S4Vectors_0.34.0       locfit_1.5-9.6
[61] munsell_0.5.0          DelayedArray_0.22.0
[63] Biostrings_2.64.1      AnnotationDbi_1.58.0
[65] compiler_4.2.1         GenomeInfoDb_1.32.4
[67] rlang_1.1.0            grid_4.2.1
[69] RCurl_1.98-1.9         rstudioapi_0.14
[71] ngsmisc_0.4.0          labeling_0.4.2
[73] bitops_1.0-7           rmarkdown_2.24
[75] gtable_0.3.1           codetools_0.2-18
[77] DBI_1.1.3              R6_2.5.1
[79] knitr_1.42             dplyr_1.1.1
[81] fastmap_1.1.0          bit_4.0.5
[83] utf8_1.2.2             rprojroot_2.0.3
[85] readr_2.1.4            stringi_1.7.12
[87] parallel_4.2.1         Rcpp_1.0.11
[89] geneplotter_1.74.0     png_0.1-7
[91] vctrs_0.6.1            tidyselect_1.2.0
[93] xfun_0.40
```