Differential Gene Expression Practice

2023-10-27

ACESO Genomics and TIDREC collaboration on the SCOPE project

Author: Pavol Genzor

Setup your environment

```r
## clean
ls()
```

```
## character(0)
```

```r
rm(list = ls())
gc()
```

```
##          used (Mb) gc trigger (Mb) limit (Mb) max used (Mb)
## Ncells 477397 25.5    1029291   55       NA   669400 35.8
## Vcells 897300  6.9    8388608   64    256000 1851660 14.2
```

```r
## directories
getwd()
```

```
## [1] "/Users/a_PGenzor/Documents/GITHUB/ahjf_scope/markdowns"
```

```r
setwd("/Users/a_PGenzor/Documents/GITHUB/ahjf_scope")
FIG.DIR <- "/Users/a_PGenzor/Documents/GITHUB/ahjf_scope/results/figures/"
TAB.TIR <- "/Users/a_PGenzor/Documents/GITHUB/ahjf_scope/results/tables/"
SES.DIR <- "/Users/a_PGenzor/Documents/GITHUB/ahjf_scope/sessions/"

## versioning and seed
version.date <- "26OCT23"
aSeed = 2610
set.seed(aSeed)

## libraries
library(data.table)
library(ggplot2)
library(scales)
library(ggrepel)
library(ggplot2)
library(ggpubr)
```

Load the data

```r
## metadata address
meta_path <- "/Users/a_PGenzor/Documents/GITHUB/ahjf_scope/data/batch1/aceso_scope_batch1_data/A_Encoded
meta_path
```

```
## [1] "/Users/a_PGenzor/Documents/GITHUB/ahjf_scope/data/batch1/aceso_scope_batch1_data/A_Encoded/batch1
```

```r
## load into data table
#?fread

meta.dt <- fread(input = meta_path)
str(meta.dt)
```

```
## Classes 'data.table' and 'data.frame':  13 obs. of  5 variables:
##  $ pid         : chr  "s2" "s3" "s36" "s47" ...
##  $ biofire     : chr  "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)" "Severe Acute 
##  $ sex         : chr  "Female" "Female" "Female" "Male" ...
##  $ age         : int  44 25 36 36 42 22 42 42 42 37 ...
##  $ comorbidities: chr  "" "" "" "Diabetes mellitus, Smoking" ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```r
##
## Load raw count data
##

rc_path <- "/Users/a_PGenzor/Documents/GITHUB/ahjf_scope/data/batch1/aceso_scope_batch1_data/A_Encoded/g
rc.dt <- fread(rc_path)
rc.dt[,1:5]
```

```
##                               gene_id s2_t0 s3_t0 s47_t0 s51_t0
##     1: ENSG00000272438.1|ENSG00000272438     0     0      0      0
##     2: ENSG00000230699.2|ENSG00000230699     0     0      0     10
##     3: ENSG00000241180.1|ENSG00000241180     0     0      0      0
##     4:        ENSG00000223764.2|LINC02593     7     0      5     11
##     5:          ENSG00000187961.15|KLHL17   142    78    101    197
##    ---
## 61902: ENSG00000278510.1|ENSG00000278510     0     0      0      0
## 61903: ENSG00000277836.1|ENSG00000277836     0     0      0      0
## 61904: ENSG00000276017.1|ENSG00000276017     0     0      0      0
## 61905: ENSG00000277196.4|ENSG00000277196     0     0      0      2
## 61906:            ENSG00000277374.1|U1     0     0      0      0
```

```r
str(rc.dt)
```

```
## Classes 'data.table' and 'data.frame':  61906 obs. of  20 variables:
##  $ gene_id: chr  "ENSG00000272438.1|ENSG00000272438" "ENSG00000230699.2|ENSG00000230699" "ENSG0000024
##  $ s2_t0  : int  0 0 0 7 142 17 30 0 324 265 ...
##  $ s3_t0  : int  0 0 0 0 78 33 8 21 208 25 ...
##  $ s47_t0 : int  0 0 0 5 101 27 13 0 242 20 ...
##  $ s51_t0 : int  0 10 0 11 197 0 9 0 451 157 ...
##  $ s58_t0 : int  0 0 0 6 152 29 0 0 350 14 ...
##  $ s78_t0 : int  0 0 0 5 201 33 0 14 332 56 ...
##  $ s81_t0 : int  0 0 0 0 187 22 0 0 344 34 ...
##  $ s94_t0 : int  0 25 0 9 113 17 0 0 312 26 ...
##  $ s95_t0 : int  0 0 0 0 223 49 0 0 457 33 ...
##  $ s2_t28 : int  0 19 0 8 115 30 26 19 364 6 ...
##  $ s36_t28: int  0 0 0 4 128 0 10 9 254 10 ...
```

```
##  $ s47_t28: int  0 0 0 4 123 42 0 0 126 1161 ...
##  $ s51_t28: int  0 0 0 0 123 0 0 0 153 11 ...
##  $ s58_t28: int  0 0 0 0 164 11 0 0 403 0 ...
##  $ s81_t28: int  0 0 0 0 165 19 0 0 401 22 ...
##  $ s85_t28: int  0 0 0 3 164 9 5 11 244 39 ...
##  $ s90_t28: int  0 0 0 3 244 31 12 0 424 10 ...
##  $ s91_t28: int  0 0 0 9 144 31 0 0 351 14 ...
##  $ s94_t28: int  0 17 0 0 167 7 21 0 338 8 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
## zoom in at small piece
rc.dt[,1:4]
```

```
##                                    gene_id s2_t0 s3_t0 s47_t0
##     1: ENSG00000272438.1|ENSG00000272438     0     0      0
##     2: ENSG00000230699.2|ENSG00000230699     0     0      0
##     3: ENSG00000241180.1|ENSG00000241180     0     0      0
##     4:        ENSG00000223764.2|LINC02593     7     0      5
##     5:         ENSG00000187961.15|KLHL17   142    78    101
##    ---
## 61902: ENSG00000278510.1|ENSG00000278510     0     0      0
## 61903: ENSG00000277836.1|ENSG00000277836     0     0      0
## 61904: ENSG00000276017.1|ENSG00000276017     0     0      0
## 61905: ENSG00000277196.4|ENSG00000277196     0     0      0
## 61906:             ENSG00000277374.1|U1     0     0      0
```

Filter raw data

```
## input data
rc.dt[,1:5]
```

```
##                                    gene_id s2_t0 s3_t0 s47_t0 s51_t0
##     1: ENSG00000272438.1|ENSG00000272438     0     0      0      0
##     2: ENSG00000230699.2|ENSG00000230699     0     0      0     10
##     3: ENSG00000241180.1|ENSG00000241180     0     0      0      0
##     4:        ENSG00000223764.2|LINC02593     7     0      5     11
##     5:         ENSG00000187961.15|KLHL17   142    78    101    197
##    ---
## 61902: ENSG00000278510.1|ENSG00000278510     0     0      0      0
## 61903: ENSG00000277836.1|ENSG00000277836     0     0      0      0
## 61904: ENSG00000276017.1|ENSG00000276017     0     0      0      0
## 61905: ENSG00000277196.4|ENSG00000277196     0     0      0      2
## 61906:             ENSG00000277374.1|U1     0     0      0      0
```

```
## summarize raw data
#?.SD

rc.ss.dt <- rc.dt[,.(rowsum=sum(.SD),
                 rowmax=max(.SD),
                 rowmin=min(.SD)),
              by=gene_id]
rc.ss.dt
```

```
##                                    gene_id rowsum rowmax rowmin
##     1: ENSG00000272438.1|ENSG00000272438      0      0      0
##     2: ENSG00000230699.2|ENSG00000230699     71     25      0
##     3: ENSG00000241180.1|ENSG00000241180      0      0      0
```

```
##     4:          ENSG00000223764.2|LINC02593    74     11      0
##     5:          ENSG00000187961.15|KLHL17    2931    244     78
##    ---
## 61902: ENSG00000278510.1|ENSG00000278510     0      0      0
## 61903: ENSG00000277836.1|ENSG00000277836     5      2      0
## 61904: ENSG00000276017.1|ENSG00000276017    13      6      0
## 61905: ENSG00000277196.4|ENSG00000277196    32     14      0
## 61906:            ENSG00000277374.1|U1       1      1      0
```

filter the table
```r
rc.ss.dt[rowsum > 0] ## always required
```

```
##                                 gene_id rowsum rowmax rowmin
##     1: ENSG00000230699.2|ENSG00000230699     71     25      0
##     2:          ENSG00000223764.2|LINC02593    74     11      0
##     3:          ENSG00000187961.15|KLHL17    2931    244     78
##     4:          ENSG00000187583.11|PLEKHN1    407     49      0
##     5:          ENSG00000187642.10|PERM1     134     30      0
##    ---
## 55888: ENSG00000274792.1|ENSG00000274792     9      2      0
## 55889: ENSG00000277836.1|ENSG00000277836     5      2      0
## 55890: ENSG00000276017.1|ENSG00000276017    13      6      0
## 55891: ENSG00000277196.4|ENSG00000277196    32     14      0
## 55892:            ENSG00000277374.1|U1       1      1      0
```

```r
rc.ss.dt[rowsum > 50] ## always DESIRED
```

```
##                                 gene_id rowsum rowmax rowmin
##     1: ENSG00000230699.2|ENSG00000230699     71     25      0
##     2:          ENSG00000223764.2|LINC02593    74     11      0
##     3:          ENSG00000187961.15|KLHL17    2931    244     78
##     4:          ENSG00000187583.11|PLEKHN1    407     49      0
##     5:          ENSG00000187642.10|PERM1     134     30      0
##    ---
## 30265:          ENSG00000012817.16|KDM5D   23036   4058      0
## 30266: ENSG00000288049.1|ENSG00000288049    465     76      0
## 30267:          ENSG00000198692.10|EIF1AY   4212    843      0
## 30268: ENSG00000254488.1|ENSG00000254488    161     37      0
## 30269: ENSG00000289707.1|ENSG00000289707   18983   2684      0
```

extract gene ids that are expressed
```r
usefull_gid <- rc.ss.dt[rowsum > 50][rowmax > 50][["gene_id"]]
usefull_gid[1:10]
```

```
##  [1] "ENSG00000187961.15|KLHL17"         "ENSG00000188976.11|NOC2L"
##  [3] "ENSG00000272512.1|ENSG00000272512" "ENSG00000188290.11|HES4"
##  [5] "ENSG00000187608.10|ISG15"          "ENSG00000188157.15|AGRN"
##  [7] "ENSG00000131591.18|C1orf159"       "ENSG00000186827.12|TNFRSF4"
##  [9] "ENSG00000176022.7|B3GALT6"         "ENSG00000078808.20|SDF4"
```

##
Remove not useful genes
##

```r
c(rc.dt[["gene_id"]] %in% usefull_gid)[1:10]
```

```
##  [1] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE
```

4

```r
rc.filt.dt <- rc.dt[gene_id %in% usefull_gid]
rc.filt.dt[,1:5]
```

```
##                                  gene_id s2_t0 s3_t0 s47_t0 s51_t0
##     1:          ENSG00000187961.15|KLHL17   142    78    101    197
##     2:           ENSG00000188976.11|NOC2L   324   208    242    451
##     3: ENSG00000272512.1|ENSG00000272512   265    25     20    157
##     4:           ENSG00000188290.11|HES4   940   237      0    239
##     5:           ENSG00000187608.10|ISG15  8337  3267    582   2688
##    ---
## 18713: ENSG00000260197.1|ENSG00000260197     0     0    135     98
## 18714:          ENSG00000012817.16|KDM5D     0     0   2250   2335
## 18715: ENSG00000288049.1|ENSG00000288049     0     0     76     76
## 18716:          ENSG00000198692.10|EIF1AY     0     0    508    295
## 18717: ENSG00000289707.1|ENSG00000289707     0     0   2066   2469
```

Group comparison

```r
## Input data
rc.filt.dt
```

```
##                                  gene_id s2_t0 s3_t0 s47_t0 s51_t0 s58_t0
##     1:          ENSG00000187961.15|KLHL17   142    78    101    197    152
##     2:           ENSG00000188976.11|NOC2L   324   208    242    451    350
##     3: ENSG00000272512.1|ENSG00000272512   265    25     20    157     14
##     4:           ENSG00000188290.11|HES4   940   237      0    239     12
##     5:           ENSG00000187608.10|ISG15  8337  3267    582   2688    387
##    ---
## 18713: ENSG00000260197.1|ENSG00000260197     0     0    135     98    195
## 18714:          ENSG00000012817.16|KDM5D     0     0   2250   2335   3647
## 18715: ENSG00000288049.1|ENSG00000288049     0     0     76     76     70
## 18716:          ENSG00000198692.10|EIF1AY     0     0    508    295    416
## 18717: ENSG00000289707.1|ENSG00000289707     0     0   2066   2469   2003
##        s78_t0 s81_t0 s94_t0 s95_t0 s2_t28 s36_t28 s47_t28 s51_t28 s58_t28
##     1:    201    187    113    223    115     128     123     123     164
##     2:    332    344    312    457    364     254     126     153     403
##     3:     56     34     26     33      6      10    1161      11       0
##     4:    315    138     34     68     99      22    1748      23       6
##     5:   1362   2905    221    590   1331      84   25883     145      81
##    ---
## 18713:      0      0      0    164      0       0      58     165     250
## 18714:      0      0      0   3819      0       0    1136    1630    4058
## 18715:      0      0      0     63      0       0       0      39      48
## 18716:      0      0      0    232      0       0     384     843     658
## 18717:      0      0      0   2515      0       0     933    2568    2684
##        s81_t28 s85_t28 s90_t28 s91_t28 s94_t28
##     1:     165     164     244     144     167
##     2:     401     244     424     351     338
##     3:      22      39      10      14       8
##     4:      70      63      54      38      47
##     5:    1038    1127     301     281     368
##    ---
## 18713:       0     220       0     151       0
## 18714:       0    1847       0    2314       0
## 18715:       0      25       0      68       0
```

```
## 18716:         0      326       0      550        0
## 18717:         0     2123       0     1622        0
```

```r
## Melt
rc.filt.dtm <- melt.data.table(data = rc.filt.dt,
                               id.vars = "gene_id",
                               variable.name = "sample",
                               value.name = "raw_count")
rc.filt.dtm
```

```
##                                 gene_id  sample raw_count
##      1:         ENSG00000187961.15|KLHL17   s2_t0       142
##      2:         ENSG00000188976.11|NOC2L   s2_t0       324
##      3: ENSG00000272512.1|ENSG00000272512   s2_t0       265
##      4:         ENSG00000188290.11|HES4   s2_t0       940
##      5:         ENSG00000187608.10|ISG15   s2_t0      8337
##     ---
## 355619: ENSG00000260197.1|ENSG00000260197 s94_t28         0
## 355620:         ENSG00000012817.16|KDM5D s94_t28         0
## 355621: ENSG00000288049.1|ENSG00000288049 s94_t28         0
## 355622:         ENSG00000198692.10|EIF1AY s94_t28         0
## 355623: ENSG00000289707.1|ENSG00000289707 s94_t28         0
```

```r
## extract the time variable from name
##?tstrsplit

lapply(tstrsplit(rc.filt.dtm[["sample"]], split="_"),head)
```

```
## [[1]]
## [1] "s2" "s2" "s2" "s2" "s2" "s2"
##
## [[2]]
## [1] "t0" "t0" "t0" "t0" "t0" "t0"
```

```r
str(tstrsplit("My_name_is_Pavol", split="_", keep = 3))
```

```
## List of 1
##  $ : chr "is"
```

```r
str(unlist(tstrsplit("My_name_is_Pavol", split="_", keep = 3)))
```

```
##  chr "is"
```

```r
rc.filt.dtm[,"time" := tstrsplit(sample, split="_", keep = 2)]
rc.filt.dtm
```

```
##                                 gene_id  sample raw_count time
##      1:         ENSG00000187961.15|KLHL17   s2_t0       142   t0
##      2:         ENSG00000188976.11|NOC2L   s2_t0       324   t0
##      3: ENSG00000272512.1|ENSG00000272512   s2_t0       265   t0
##      4:         ENSG00000188290.11|HES4   s2_t0       940   t0
##      5:         ENSG00000187608.10|ISG15   s2_t0      8337   t0
##     ---
## 355619: ENSG00000260197.1|ENSG00000260197 s94_t28         0  t28
## 355620:         ENSG00000012817.16|KDM5D s94_t28         0  t28
## 355621: ENSG00000288049.1|ENSG00000288049 s94_t28         0  t28
## 355622:         ENSG00000198692.10|EIF1AY s94_t28         0  t28
## 355623: ENSG00000289707.1|ENSG00000289707 s94_t28         0  t28
```

```r
## Lets compare gene expression by group
mc.dtm <- rc.filt.dtm[,.(mean_rc=mean(raw_count),
                         sd_rc=sd(raw_count),
                         N=length(raw_count)),
                      by=list(gene_id,time)]
mc.dtm
```

```
##                                   gene_id time   mean_rc       sd_rc  N
##     1:        ENSG00000187961.15|KLHL17    t0   154.8889    50.39207  9
##     2:         ENSG00000188976.11|NOC2L    t0   335.5556    82.21939  9
##     3: ENSG00000272512.1|ENSG00000272512    t0    70.0000    85.18509  9
##     4:          ENSG00000188290.11|HES4    t0   220.3333   292.43846  9
##     5:          ENSG00000187608.10|ISG15    t0  2259.8889  2564.65945  9
##    ---
## 37430: ENSG00000260197.1|ENSG00000260197  t28    84.4000   101.68382 10
## 37431:        ENSG00000012817.16|KDM5D    t28  1098.5000  1378.95864 10
## 37432: ENSG00000288049.1|ENSG00000288049  t28    18.0000    25.46021 10
## 37433:         ENSG00000198692.10|EIF1AY  t28   276.1000   322.75428 10
## 37434: ENSG00000289707.1|ENSG00000289707  t28   993.0000  1152.12326 10
```

```r
## How do you search for patterns/geen names
## search for a particular gene to check
##?grep

grep("IL10",mc.dtm[["gene_id"]]) # location of IL10 term - index
```

```
## [1]   1621 10488 17419 17420 20338 29205 36136 36137
```

```r
mc.dtm[grep("IL10",gene_id)] # table subset to show IL10 containing terms
```

```
##                           gene_id time     mean_rc       sd_rc  N
## 1:       ENSG00000136634.7|IL10     t0    24.66667    19.33908  9
## 2:    ENSG00000110324.12|IL10RA    t0 10120.77778  2651.75549  9
## 3:    ENSG00000243646.11|IL10RB    t0  3161.00000   650.10230  9
## 4: ENSG00000223799.3|IL10RB-DT    t0   166.55556    93.11433  9
## 5:       ENSG00000136634.7|IL10    t28    17.50000    18.07546 10
## 6:    ENSG00000110324.12|IL10RA   t28  8862.80000  2214.11923 10
## 7:    ENSG00000243646.11|IL10RB   t28  2849.80000   728.84274 10
## 8: ENSG00000223799.3|IL10RB-DT   t28   187.20000    99.14725 10
```

```r
mc.dtm[grep("IL10$",gene_id)] # table showing only "IL10" word itself
```

```
##                   gene_id time mean_rc    sd_rc  N
## 1: ENSG00000136634.7|IL10    t0 24.66667 19.33908  9
## 2: ENSG00000136634.7|IL10   t28 17.50000 18.07546 10
```

```r
mc.dtm[grep("CXCL10",gene_id)] # another example
```

```
##                   gene_id time   mean_rc     sd_rc  N
## 1: ENSG00000169245.6|CXCL10    t0  494.1111  1028.457  9
## 2: ENSG00000169245.6|CXCL10   t28 2090.8000  6352.551 10
```

```r
#
# Calculate the degree of change - fold change
#


mc.dtm
```

```
##                                gene_id time   mean_rc       sd_rc  N
##     1:          ENSG00000187961.15|KLHL17   t0  154.8889    50.39207  9
##     2:          ENSG00000188976.11|NOC2L   t0  335.5556    82.21939  9
##     3: ENSG00000272512.1|ENSG00000272512   t0   70.0000    85.18509  9
##     4:          ENSG00000188290.11|HES4   t0  220.3333   292.43846  9
##     5:          ENSG00000187608.10|ISG15   t0 2259.8889  2564.65945  9
##    ---
## 37430: ENSG00000260197.1|ENSG00000260197  t28   84.4000   101.68382 10
## 37431:          ENSG00000012817.16|KDM5D  t28 1098.5000  1378.95864 10
## 37432: ENSG00000288049.1|ENSG00000288049  t28   18.0000    25.46021 10
## 37433:          ENSG00000198692.10|EIF1AY  t28  276.1000   322.75428 10
## 37434: ENSG00000289707.1|ENSG00000289707  t28  993.0000  1152.12326 10
```

```r
mc.dtw <- dcast.data.table(data = mc.dtm,
                           formula = gene_id~time,
                           value.var = "mean_rc")


## calculate the fold change
## log transformation simplifies plotting

log2(10/5)
```

```
## [1] 1
```

```r
log2(5/10)
```

```
## [1] -1
```

```r
log2(0) # <- zeros are possible and therefore need to be accounted for
```

```
## [1] -Inf
```

```r
min(mc.dtw[["t0"]]) # can't divide by zero
```
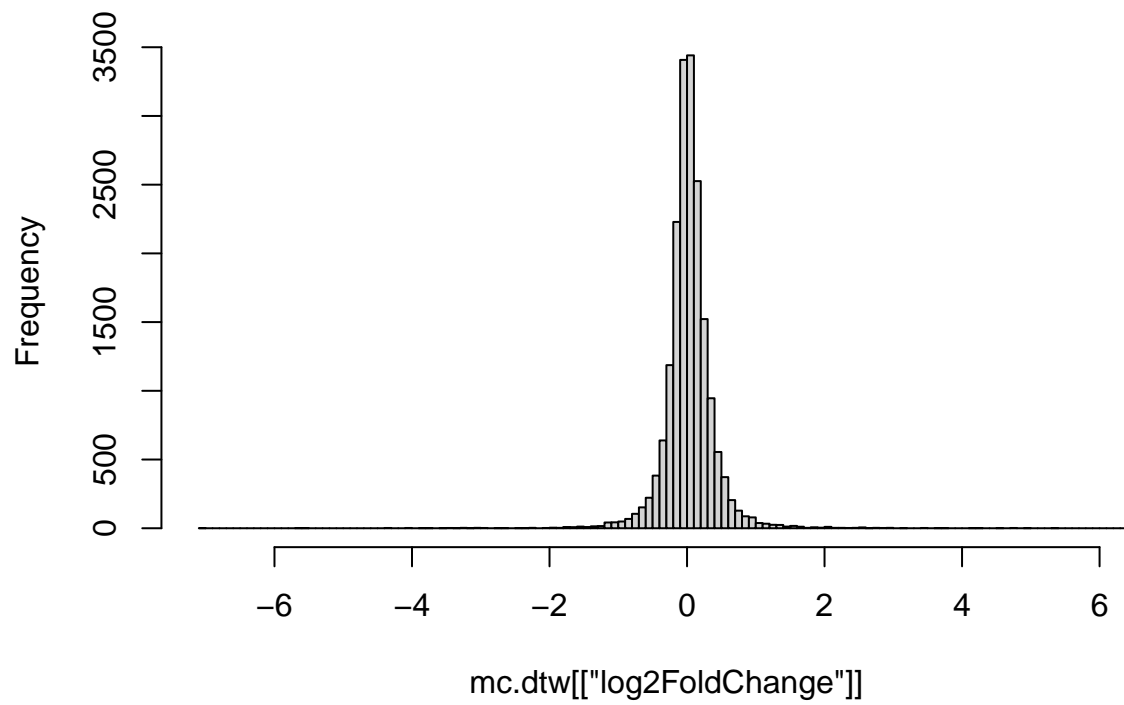
```
## [1] 0
```

```r
min(mc.dtw[["t28"]]) #  can't divide by zero
```

```
## [1] 0
```

```r
## calculate L2FC with constant to avoid zeros
mc.dtw[,"log2FoldChange" := log2(t28+1) - log2(t0+1)]
mc.dtw
```

```
##                                gene_id         t0      t28 log2FoldChange
##     1:          ENSG00000000419.14|DPM1  982.22222 1132.9     0.20570398
##     2:          ENSG00000000457.14|SCYL3 1275.11111 1463.1     0.19826014
##     3:        ENSG00000000460.17|C1orf112  528.77778  469.5    -0.17119261
##     4:          ENSG00000000938.13|FGR  9417.88889 8255.4    -0.19004401
##     5:          ENSG00000000971.17|CFH   157.11111  192.5     0.29139481
##    ---
## 18713: ENSG00000290114.1|ENSG00000290114   23.77778   21.0    -0.17154328
## 18714: ENSG00000290117.1|ENSG00000290117  102.66667  144.0     0.48410082
## 18715: ENSG00000290121.1|ENSG00000290121   47.66667   58.2     0.28266321
## 18716: ENSG00000290126.1|ENSG00000290126   54.88889   57.8     0.07325466
## 18717: ENSG00000290146.1|ENSG00000290146   34.66667   40.4     0.21505438
```
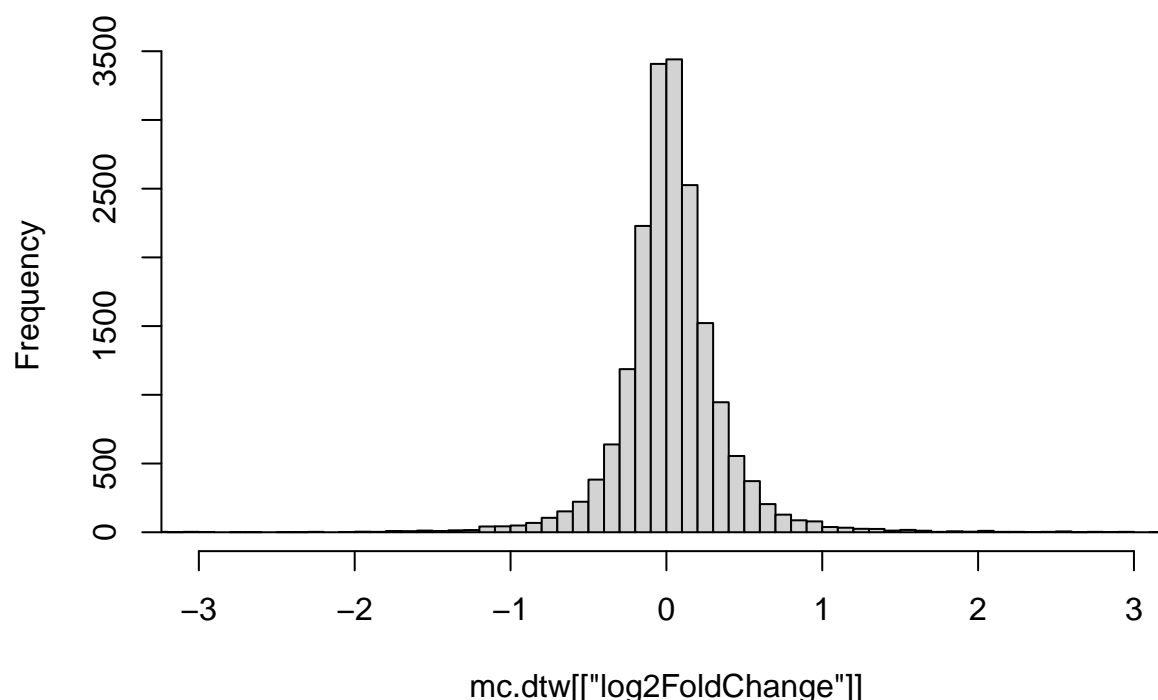
```r
## sanity check - data should be normally distributed
hist(mc.dtw[["log2FoldChange"]], breaks = 100)
```

Histogram of mc.dtw[["log2FoldChange"]]


```r
hist(mc.dtw[["log2FoldChange"]], breaks = 100, xlim = c(-3,3))
```

Histogram of mc.dtw[["log2FoldChange"]]


```
log2(1.1534050)
```

```
## [1] 0.2058992
## order your data for simpler reading
mc.dtw <- mc.dtw[order(log2FoldChange)]
mc.dtw[1:50]
```

```
##                               gene_id           t0     t28 log2FoldChange
##  1:    ENSG00000124208.16|PEDS1-UBE2V1   268.444444     1.0      -7.073844
##  2:  ENSG00000260836.2|ENSG00000260836    50.777778     0.0      -5.694261
##  3:  ENSG00000255730.5|ENSG00000255730    60.000000     0.3      -5.552226
##  4:  ENSG00000248710.1|ENSG00000248710    19.555556     0.0      -4.361456
##  5:  ENSG00000284057.1|ENSG00000284057    19.000000     0.0      -4.321928
##  6:         ENSG00000165949.13|IFI27   2789.444444   165.5      -4.066901
##  7:      ENSG00000269955.2|FMC1-LUC7L2    15.000000     0.0      -4.000000
##  8:         ENSG00000145029.14|NICN1     14.777778     0.1      -3.842319
##  9:       ENSG00000224041.3|IGKV3D-15   655.888889    46.9      -3.777552
## 10:  ENSG00000234484.1|ENSG00000234484    10.888889     0.0      -3.571542
## 11:      ENSG00000262500.1|MAPK8IP1P1    23.111111     1.2      -3.454123
## 12:  ENSG00000285668.1|ENSG00000285668     9.888889     0.0      -3.444785
## 13:  ENSG00000272410.5|ENSG00000272410     9.555556     0.0      -3.399931
## 14:  ENSG00000276087.2|ENSG00000276087    24.333333     1.6      -3.284453
## 15:      ENSG00000263503.1|MAPK8IP1P2    28.222222     2.1      -3.236726
## 16:  ENSG00000254559.1|ENSG00000254559     8.222222     0.0      -3.205114
## 17:  ENSG00000228318.3|ENSG00000228318     8.000000     0.0      -3.169925
## 18:       ENSG00000273003.1|ARL2-SNX15    81.555556     8.9      -3.059865
```

```
## 19: ENSG00000260618.1|ENSG00000260618     7.222222    0.0    -3.039528
## 20: ENSG00000268107.6|ENSG00000268107    13.444444    0.8    -3.004446
## 21:          ENSG00000115155.19|OTOF     794.555556  104.1    -2.920200
## 22: ENSG00000254732.2|ENSG00000254732    20.111111    1.8    -2.914504
## 23: ENSG00000289750.1|ENSG00000289750    16.111111    1.6    -2.718350
## 24: ENSG00000266923.1|ENSG00000266923    12.555556    1.2    -2.623309
## 25: ENSG00000285238.2|ENSG00000285238   580.666667  103.4    -2.478071
## 26: ENSG00000236540.7|ENSG00000236540    16.777778    2.5    -2.344648
## 27:          ENSG00000269900.3|RMRP   20517.777778 4261.9    -2.267038
## 28:        ENSG00000244437.1|IGKV3-15   1186.555556  250.3    -2.240512
## 29:         ENSG00000115474.7|KCNJ13     47.666667    9.4    -2.226350
## 30: ENSG00000271793.2|ENSG00000271793    78.666667   18.5    -2.030502
## 31:          ENSG00000204936.10|CD177    759.888889  187.9    -2.010063
## 32: ENSG00000268659.2|ENSG00000268659    25.111111    5.6    -1.984126
## 33: ENSG00000269481.1|ENSG00000269481    39.333333    9.2    -1.983403
## 34: ENSG00000224738.2|ENSG00000224738    30.888889    7.3    -1.941871
## 35: ENSG00000267952.1|ENSG00000267952    22.222222    5.2    -1.905166
## 36: ENSG00000279762.3|ENSG00000279762    13.888889    3.0    -1.896164
## 37:        ENSG00000241244.1|IGKV1D-16    83.666667   21.8    -1.892760
## 38:   ENSG00000273294.1|C1QTNF3-AMACR    14.555556    3.4    -1.821854
## 39:          ENSG00000159189.13|C1QC     51.000000   14.0    -1.793549
## 40:         ENSG00000112309.11|B3GAT2     39.444444   10.7    -1.789433
## 41:       ENSG00000248672.5|LY75-CD302   346.555556  100.1    -1.781461
## 42:         ENSG00000239855.1|IGKV1-6    192.555556   55.8    -1.768785
## 43:         ENSG00000117399.14|CDC20     91.333333   26.5    -1.747420
## 44:          ENSG00000173369.18|C1QB    182.777778   54.5    -1.727403
## 45:          ENSG00000284194.3|SCO2     167.777778   50.5    -1.712481
## 46: ENSG00000258035.2|ENSG00000258035    11.444444    2.8    -1.711431
## 47:         ENSG00000211905.1|IGHJ1      12.333333    3.1    -1.701342
## 48:        ENSG00000111057.11|KRT18      16.666667    4.5    -1.683526
## 49:          ENSG00000206172.8|HBA1   12249.777778 3814.2    -1.683043
## 50:         ENSG00000188536.13|HBA2   15047.888889 4705.2    -1.677022
##                          gene_id        t0    t28 log2FoldChange
```

```
##
## Calculate the p.value
##


## extract names of genes you want to test
gid_list <- mc.dtw[["gene_id"]]
str(gid_list)
```

```
##  chr [1:18717] "ENSG00000124208.16|PEDS1-UBE2V1" ...
```

```
## use the sample specific table to have all the observations
rc.filt.dtm
```

```
##                              gene_id  sample raw_count time
##      1:       ENSG00000187961.15|KLHL17  s2_t0       142   t0
##      2:        ENSG00000188976.11|NOC2L  s2_t0       324   t0
##      3: ENSG00000272512.1|ENSG00000272512  s2_t0       265   t0
##      4:        ENSG00000188290.11|HES4   s2_t0       940   t0
##      5:        ENSG00000187608.10|ISG15  s2_t0      8337   t0
##     ---
## 355619: ENSG00000260197.1|ENSG00000260197 s94_t28        0  t28
## 355620:        ENSG00000012817.16|KDM5D s94_t28        0  t28
```

```
## 355621: ENSG00000288049.1|ENSG00000288049 s94_t28          0  t28
## 355622:            ENSG00000198692.10|EIF1AY s94_t28          0  t28
## 355623: ENSG00000289707.1|ENSG00000289707 s94_t28          0  t28
```

```r
## Run Students t-test
##?t.test


## example calculation
a_gene <- "ENSG00000124208.16|PEDS1-UBE2V1"
t.test(x = rc.filt.dtm[gene_id %in% a_gene][time %in% "t28"][["raw_count"]],
       y = rc.filt.dtm[gene_id %in% a_gene][time %in% "t0"][["raw_count"]])
```

```
##
##  Welch Two Sample t-test
##
## data:  rc.filt.dtm[gene_id %in% a_gene][time %in% "t28"][["raw_count"]] and rc.filt.dtm[gene_id %in%
## t = -1.0383, df = 8.0002, p-value = 0.3295
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -861.3942  326.5053
## sample estimates:
## mean of x mean of y
##     1.0000  268.4444
```

```r
## to extract p-value
t.test(x = rc.filt.dtm[gene_id %in% a_gene][time %in% "t28"][["raw_count"]],
       y = rc.filt.dtm[gene_id %in% a_gene][time %in% "t0"][["raw_count"]])$p.value
```

```
## [1] 0.3294707
```

```r
## Run a loop through all gene names to calculate the p.values
## -> results in new table and takes a little time

pval.dt <- rbindlist(lapply(X = gid_list, function(g){
  # make gene sub table
  sub.dt <- rc.filt.dtm[gene_id %in% g]
  # calculate p.value
  pval <- t.test(x = sub.dt[time %in% "t28"][["raw_count"]],
                 y = sub.dt[time %in% "t0"][["raw_count"]])$p.value
  # make new table
  res.dt <- data.table("gene_id"=g,
                       "p.value"=pval)
  # return the last created variable
  return(res.dt)}))

## adjust p.value to account for false positives
##?p.adjust

pval.dt[,"p.adjust" := p.adjust(p = p.value,method = "BH")]
pval.dt
```

```
##                                 gene_id    p.value  p.adjust
##     1:    ENSG00000124208.16|PEDS1-UBE2V1 0.32947073 0.9999167
##     2: ENSG00000260836.2|ENSG00000260836 0.31622159 0.9999167
##     3: ENSG00000255730.5|ENSG00000255730 0.11617525 0.9999167
##     4: ENSG00000248710.1|ENSG00000248710 0.33189885 0.9999167
```

```
##     5:    ENSG00000284057.1|ENSG00000284057 0.34659351 0.9999167
##      ---
## 18713:    ENSG00000171570.11|RAB4B-EGLN2 0.22166860 0.9999167
## 18714:          ENSG00000205790.1|DPP9-AS1 0.09960765 0.9999167
## 18715: ENSG00000288534.1|ENSG00000288534 0.15234476 0.9999167
## 18716: ENSG00000282885.3|ENSG00000282885 0.34651682 0.9999167
## 18717: ENSG00000256500.6|ENSG00000256500 0.16901650 0.9999167
```

```
##
## Combine two tables
##
```

```
## tables to combine
mc.dtw
```

```
##                                    gene_id        t0    t28 log2FoldChange
##     1:    ENSG00000124208.16|PEDS1-UBE2V1 268.44444    1.0      -7.073844
##     2: ENSG00000260836.2|ENSG00000260836  50.77778    0.0      -5.694261
##     3: ENSG00000255730.5|ENSG00000255730  60.00000    0.3      -5.552226
##     4: ENSG00000248710.1|ENSG00000248710  19.55556    0.0      -4.361456
##     5: ENSG00000284057.1|ENSG00000284057  19.00000    0.0      -4.321928
##      ---
## 18713:    ENSG00000171570.11|RAB4B-EGLN2   0.00000   26.0       4.754888
## 18714:          ENSG00000205790.1|DPP9-AS1   0.00000   26.7       4.791814
## 18715: ENSG00000288534.1|ENSG00000288534   0.00000   29.3       4.921246
## 18716: ENSG00000282885.3|ENSG00000282885  83.00000 3483.4       5.374377
## 18717: ENSG00000256500.6|ENSG00000256500   0.00000   84.5       6.417853
```

```
pval.dt
```

```
##                                    gene_id   p.value  p.adjust
##     1:    ENSG00000124208.16|PEDS1-UBE2V1 0.32947073 0.9999167
##     2: ENSG00000260836.2|ENSG00000260836 0.31622159 0.9999167
##     3: ENSG00000255730.5|ENSG00000255730 0.11617525 0.9999167
##     4: ENSG00000248710.1|ENSG00000248710 0.33189885 0.9999167
##     5: ENSG00000284057.1|ENSG00000284057 0.34659351 0.9999167
##      ---
## 18713:    ENSG00000171570.11|RAB4B-EGLN2 0.22166860 0.9999167
## 18714:          ENSG00000205790.1|DPP9-AS1 0.09960765 0.9999167
## 18715: ENSG00000288534.1|ENSG00000288534 0.15234476 0.9999167
## 18716: ENSG00000282885.3|ENSG00000282885 0.34651682 0.9999167
## 18717: ENSG00000256500.6|ENSG00000256500 0.16901650 0.9999167
```

```
## add p.val table to mean count table
res.dt <- mc.dtw[pval.dt,on=.(gene_id=gene_id)]
res.dt
```

```
##                                    gene_id        t0    t28 log2FoldChange
##     1:    ENSG00000124208.16|PEDS1-UBE2V1 268.44444    1.0      -7.073844
##     2: ENSG00000260836.2|ENSG00000260836  50.77778    0.0      -5.694261
##     3: ENSG00000255730.5|ENSG00000255730  60.00000    0.3      -5.552226
##     4: ENSG00000248710.1|ENSG00000248710  19.55556    0.0      -4.361456
##     5: ENSG00000284057.1|ENSG00000284057  19.00000    0.0      -4.321928
##      ---
## 18713:    ENSG00000171570.11|RAB4B-EGLN2   0.00000   26.0       4.754888
## 18714:          ENSG00000205790.1|DPP9-AS1   0.00000   26.7       4.791814
## 18715: ENSG00000288534.1|ENSG00000288534   0.00000   29.3       4.921246
```

```
## 18716:  ENSG00000282885.3|ENSG00000282885   83.00000 3483.4        5.374377
## 18717:  ENSG00000256500.6|ENSG00000256500    0.00000   84.5        6.417853
##            p.value  p.adjust
##     1: 0.32947073 0.9999167
##     2: 0.31622159 0.9999167
##     3: 0.11617525 0.9999167
##     4: 0.33189885 0.9999167
##     5: 0.34659351 0.9999167
##    ---
## 18713: 0.22166860 0.9999167
## 18714: 0.09960765 0.9999167
## 18715: 0.15234476 0.9999167
## 18716: 0.34651682 0.9999167
## 18717: 0.16901650 0.9999167
```

```r
## add count table to p.val table
## both ways work as long as there "on" columns are properly defined
pval.dt[mc.dtw,on=.(gene_id=gene_id)]
```

```
##                                  gene_id    p.value  p.adjust        t0    t28
##     1:    ENSG00000124208.16|PEDS1-UBE2V1 0.32947073 0.9999167 268.44444    1.0
##     2: ENSG00000260836.2|ENSG00000260836 0.31622159 0.9999167  50.77778    0.0
##     3: ENSG00000255730.5|ENSG00000255730 0.11617525 0.9999167  60.00000    0.3
##     4: ENSG00000248710.1|ENSG00000248710 0.33189885 0.9999167  19.55556    0.0
##     5: ENSG00000284057.1|ENSG00000284057 0.34659351 0.9999167  19.00000    0.0
##    ---
## 18713:    ENSG00000171570.11|RAB4B-EGLN2 0.22166860 0.9999167   0.00000   26.0
## 18714:          ENSG00000205790.1|DPP9-AS1 0.09960765 0.9999167   0.00000   26.7
## 18715: ENSG00000288534.1|ENSG00000288534 0.15234476 0.9999167   0.00000   29.3
## 18716: ENSG00000282885.3|ENSG00000282885 0.34651682 0.9999167  83.00000 3483.4
## 18717: ENSG00000256500.6|ENSG00000256500 0.16901650 0.9999167   0.00000   84.5
##        log2FoldChange
##     1:      -7.073844
##     2:      -5.694261
##     3:      -5.552226
##     4:      -4.361456
##     5:      -4.321928
##    ---
## 18713:       4.754888
## 18714:       4.791814
## 18715:       4.921246
## 18716:       5.374377
## 18717:       6.417853
```

Plot the results

```r
## Input data
res.dt
```

```
##                                  gene_id        t0    t28 log2FoldChange
##     1:    ENSG00000124208.16|PEDS1-UBE2V1 268.44444    1.0      -7.073844
##     2: ENSG00000260836.2|ENSG00000260836  50.77778    0.0      -5.694261
##     3: ENSG00000255730.5|ENSG00000255730  60.00000    0.3      -5.552226
##     4: ENSG00000248710.1|ENSG00000248710  19.55556    0.0      -4.361456
##     5: ENSG00000284057.1|ENSG00000284057  19.00000    0.0      -4.321928
##    ---
```

```
## 18713:       ENSG00000171570.11|RAB4B-EGLN2   0.00000   26.0      4.754888
## 18714:           ENSG00000205790.1|DPP9-AS1   0.00000   26.7      4.791814
## 18715: ENSG00000288534.1|ENSG00000288534   0.00000   29.3      4.921246
## 18716: ENSG00000282885.3|ENSG00000282885  83.00000 3483.4      5.374377
## 18717: ENSG00000256500.6|ENSG00000256500   0.00000   84.5      6.417853
##           p.value  p.adjust
##     1: 0.32947073 0.9999167
##     2: 0.31622159 0.9999167
##     3: 0.11617525 0.9999167
##     4: 0.33189885 0.9999167
##     5: 0.34659351 0.9999167
##    ---
## 18713: 0.22166860 0.9999167
## 18714: 0.09960765 0.9999167
## 18715: 0.15234476 0.9999167
## 18716: 0.34651682 0.9999167
## 18717: 0.16901650 0.9999167
```
Sort your data: p.adjust
```
res.dt[order(p.adjust)]
```

```
##                                 gene_id        t0     t28 log2FoldChange
##     1:          ENSG00000122729.19|ACO1  229.44444   348.1    5.992214e-01
##     2:    ENSG00000124208.16|PEDS1-UBE2V1  268.44444     1.0   -7.073844e+00
##     3: ENSG00000260836.2|ENSG00000260836   50.77778     0.0   -5.694261e+00
##     4: ENSG00000255730.5|ENSG00000255730   60.00000     0.3   -5.552226e+00
##     5: ENSG00000248710.1|ENSG00000248710   19.55556     0.0   -4.361456e+00
##    ---
## 18713:          ENSG00000122644.13|ARL4A  364.77778   364.8    8.764588e-05
## 18714:        ENSG00000156475.19|PPP2R2B  479.55556   479.6    1.334223e-04
## 18715:            ENSG00000229314.6|ORM1   49.88889    49.9    3.149645e-04
## 18716:         ENSG00000018280.17|SLC11A1 7878.77778  7878.6   -3.254940e-05
## 18717: ENSG00000269951.1|ENSG00000269951   33.00000    33.0    0.000000e+00
##           p.value  p.adjust
##     1: 2.226443e-05 0.4167233
##     2: 3.294707e-01 0.9999167
##     3: 3.162216e-01 0.9999167
##     4: 1.161752e-01 0.9999167
##     5: 3.318989e-01 0.9999167
##    ---
## 18713: 9.998218e-01 0.9999544
## 18714: 9.995339e-01 0.9999544
## 18715: 9.997315e-01 0.9999544
## 18716: 9.999137e-01 0.9999671
## 18717: 1.000000e+00 1.0000000
```

```
res.dt[order(-p.adjust)]
```

```
##                                 gene_id        t0     t28 log2FoldChange
##     1: ENSG00000269951.1|ENSG00000269951   33.0000    33.0    0.0000000000
##     2:         ENSG00000018280.17|SLC11A1 7878.7778  7878.6   -0.0000325494
##     3:          ENSG00000186318.18|BACE1  132.1111   132.1   -0.0001204303
##     4:         ENSG00000128609.16|NDUFA5  979.7778   979.7   -0.0001144133
##     5:         ENSG00000108588.15|CCDC47  616.3333   616.3   -0.0000779014
##    ---
```

```
## 18713:          ENSG00000205790.1|DPP9-AS1     0.0000   26.7   4.7918140712
## 18714: ENSG00000288534.1|ENSG00000288534     0.0000   29.3   4.9212458886
## 18715: ENSG00000282885.3|ENSG00000282885    83.0000 3483.4   5.3743771129
## 18716: ENSG00000256500.6|ENSG00000256500     0.0000   84.5   6.4178525149
## 18717:          ENSG00000122729.19|ACO1   229.4444  348.1   0.5992213693
##           p.value  p.adjust
##     1: 1.000000e+00 1.0000000
##     2: 9.999137e-01 0.9999671
##     3: 9.996315e-01 0.9999544
##     4: 9.996049e-01 0.9999544
##     5: 9.996428e-01 0.9999544
##    ---
## 18713: 9.960765e-02 0.9999167
## 18714: 1.523448e-01 0.9999167
## 18715: 3.465168e-01 0.9999167
## 18716: 1.690165e-01 0.9999167
## 18717: 2.226443e-05 0.4167233
```

Sort your data: p.value
```r
res.dt[order(p.value)]
```

```
##                                 gene_id       t0    t28 log2FoldChange
##     1:          ENSG00000122729.19|ACO1 229.44444  348.1   5.992214e-01
##     2: ENSG00000289935.1|ENSG00000289935  18.00000   37.0   1.000000e+00
##     3: ENSG00000262380.1|ENSG00000262380  52.77778   95.9   8.494865e-01
##     4: ENSG00000224738.2|ENSG00000224738  30.88889    7.3  -1.941871e+00
##     5:          ENSG00000115474.7|KCNJ13  47.66667    9.4  -2.226350e+00
##    ---
## 18713:          ENSG00000122644.13|ARL4A 364.77778  364.8   8.764588e-05
## 18714:          ENSG00000168395.16|ING5  432.88889  432.9   3.694434e-05
## 18715:          ENSG00000124459.12|ZNF45 366.88889  366.9   4.357213e-05
## 18716:         ENSG00000018280.17|SLC11A1 7878.77778 7878.6  -3.254940e-05
## 18717: ENSG00000269951.1|ENSG00000269951  33.00000   33.0   0.000000e+00
##           p.value  p.adjust
##     1: 2.226443e-05 0.4167233
##     2: 2.486792e-04 0.9999167
##     3: 3.400907e-04 0.9999167
##     4: 7.311745e-04 0.9999167
##     5: 7.749546e-04 0.9999167
##    ---
## 18713: 9.998218e-01 0.9999544
## 18714: 9.998310e-01 0.9999544
## 18715: 9.998475e-01 0.9999544
## 18716: 9.999137e-01 0.9999671
## 18717: 1.000000e+00 1.0000000
```

How many genes have significant p.value
```r
res.dt[p.value <= 0.05]
```

```
##                                 gene_id        t0    t28 log2FoldChange
##   1:        ENSG00000165949.13|IFI27 2789.444444  165.5      -4.066901
##   2:         ENSG00000115155.19|OTOF  794.555556  104.1      -2.920200
##   3:        ENSG00000115474.7|KCNJ13   47.666667    9.4      -2.226350
##   4: ENSG00000268659.2|ENSG00000268659   25.111111    5.6      -1.984126
##   5: ENSG00000224738.2|ENSG00000224738   30.888889    7.3      -1.941871
```

```
##    ---
## 397: ENSG00000290111.1|ENSG00000290111    7.000000  31.2         2.008989
## 398: ENSG00000236514.1|ENSG00000236514    4.666667  22.3         2.039758
## 399:           ENSG00000172232.10|AZU1    8.111111  52.4         2.551141
## 400:           ENSG00000122733.12|PHF24   6.111111  41.1         2.565673
## 401: ENSG00000288997.2|ENSG00000288997    3.666667  30.3         2.745698
##           p.value  p.adjust
##   1: 0.0329881832 0.9999167
##   2: 0.0317079264 0.9999167
##   3: 0.0007749546 0.9999167
##   4: 0.0266826745 0.9999167
##   5: 0.0007311745 0.9999167
##   ---
## 397: 0.0128887858 0.9999167
## 398: 0.0233251232 0.9999167
## 399: 0.0095128030 0.9999167
## 400: 0.0431931589 0.9999167
## 401: 0.0137273086 0.9999167
```

```
##
## GGPLOT - most versatile R plotting package
## -> please refer to many online manuals
## ?geom_point

## examine table
res.dt[abs(log2FoldChange) > 4] # There are a lot of low values - background
```

```
##                              gene_id           t0    t28 log2FoldChange
##  1:    ENSG00000124208.16|PEDS1-UBE2V1  268.4444444    1.0       -7.073844
##  2: ENSG00000260836.2|ENSG00000260836   50.7777778    0.0       -5.694261
##  3: ENSG00000255730.5|ENSG00000255730   60.0000000    0.3       -5.552226
##  4: ENSG00000248710.1|ENSG00000248710   19.5555556    0.0       -4.361456
##  5: ENSG00000284057.1|ENSG00000284057   19.0000000    0.0       -4.321928
##  6:          ENSG00000165949.13|IFI27 2789.4444444  165.5       -4.066901
##  7: ENSG00000287916.1|ENSG00000287916    0.6666667   27.6        4.100978
##  8: ENSG00000264058.2|ENSG00000264058    4.5555556  108.1        4.295576
##  9:           ENSG00000138135.7|CH25H    6.5555556  181.7        4.595795
## 10:      ENSG00000171570.11|RAB4B-EGLN2   0.0000000   26.0        4.754888
## 11:         ENSG00000205790.1|DPP9-AS1    0.0000000   26.7        4.791814
## 12: ENSG00000288534.1|ENSG00000288534    0.0000000   29.3        4.921246
## 13: ENSG00000282885.3|ENSG00000282885   83.0000000 3483.4        5.374377
## 14: ENSG00000256500.6|ENSG00000256500    0.0000000   84.5        6.417853
##         p.value  p.adjust
##  1: 0.32947073 0.9999167
##  2: 0.31622159 0.9999167
##  3: 0.11617525 0.9999167
##  4: 0.33189885 0.9999167
##  5: 0.34659351 0.9999167
##  6: 0.03298818 0.9999167
##  7: 0.30875788 0.9999167
##  8: 0.32504672 0.9999167
##  9: 0.35777310 0.9999167
## 10: 0.22166860 0.9999167
## 11: 0.09960765 0.9999167
## 12: 0.15234476 0.9999167
```

```
## 13: 0.34651682 0.9999167
## 14: 0.16901650 0.9999167
```

```
res.dt[,"baseMean" := (t0+t28)/2] # baseMean helps filtering these
res.dt[abs(log2FoldChange) > 4]
```

```
##                                gene_id          t0    t28 log2FoldChange
##  1:    ENSG00000124208.16|PEDS1-UBE2V1  268.4444444    1.0      -7.073844
##  2: ENSG00000260836.2|ENSG00000260836   50.7777778    0.0      -5.694261
##  3: ENSG00000255730.5|ENSG00000255730   60.0000000    0.3      -5.552226
##  4: ENSG00000248710.1|ENSG00000248710   19.5555556    0.0      -4.361456
##  5: ENSG00000284057.1|ENSG00000284057   19.0000000    0.0      -4.321928
##  6:           ENSG00000165949.13|IFI27 2789.4444444  165.5      -4.066901
##  7: ENSG00000287916.1|ENSG00000287916    0.6666667   27.6       4.100978
##  8: ENSG00000264058.2|ENSG00000264058    4.5555556  108.1       4.295576
##  9:           ENSG00000138135.7|CH25H    6.5555556  181.7       4.595795
## 10:     ENSG00000171570.11|RAB4B-EGLN2    0.0000000   26.0       4.754888
## 11:          ENSG00000205790.1|DPP9-AS1    0.0000000   26.7       4.791814
## 12: ENSG00000288534.1|ENSG00000288534    0.0000000   29.3       4.921246
## 13: ENSG00000282885.3|ENSG00000282885   83.0000000 3483.4       5.374377
## 14: ENSG00000256500.6|ENSG00000256500    0.0000000   84.5       6.417853
##         p.value  p.adjust    baseMean
##  1: 0.32947073 0.9999167  134.722222
##  2: 0.31622159 0.9999167   25.388889
##  3: 0.11617525 0.9999167   30.150000
##  4: 0.33189885 0.9999167    9.777778
##  5: 0.34659351 0.9999167    9.500000
##  6: 0.03298818 0.9999167 1477.472222
##  7: 0.30875788 0.9999167   14.133333
##  8: 0.32504672 0.9999167   56.327778
##  9: 0.35777310 0.9999167   94.127778
## 10: 0.22166860 0.9999167   13.000000
## 11: 0.09960765 0.9999167   13.350000
## 12: 0.15234476 0.9999167   14.650000
## 13: 0.34651682 0.9999167 1783.200000
## 14: 0.16901650 0.9999167   42.250000
```

```
res.dt
```

```
##                                 gene_id        t0    t28 log2FoldChange
##     1:    ENSG00000124208.16|PEDS1-UBE2V1 268.44444    1.0      -7.073844
##     2: ENSG00000260836.2|ENSG00000260836  50.77778    0.0      -5.694261
##     3: ENSG00000255730.5|ENSG00000255730  60.00000    0.3      -5.552226
##     4: ENSG00000248710.1|ENSG00000248710  19.55556    0.0      -4.361456
##     5: ENSG00000284057.1|ENSG00000284057  19.00000    0.0      -4.321928
##    ---
## 18713:     ENSG00000171570.11|RAB4B-EGLN2   0.00000   26.0       4.754888
## 18714:          ENSG00000205790.1|DPP9-AS1   0.00000   26.7       4.791814
## 18715: ENSG00000288534.1|ENSG00000288534   0.00000   29.3       4.921246
## 18716: ENSG00000282885.3|ENSG00000282885  83.00000 3483.4       5.374377
## 18717: ENSG00000256500.6|ENSG00000256500   0.00000   84.5       6.417853
##         p.value  p.adjust    baseMean
##     1: 0.32947073 0.9999167  134.722222
##     2: 0.31622159 0.9999167   25.388889
##     3: 0.11617525 0.9999167   30.150000
##     4: 0.33189885 0.9999167    9.777778
```

```
##     5: 0.34659351 0.9999167    9.500000
##      ---
## 18713: 0.22166860 0.9999167   13.000000
## 18714: 0.09960765 0.9999167   13.350000
## 18715: 0.15234476 0.9999167   14.650000
## 18716: 0.34651682 0.9999167 1783.200000
## 18717: 0.16901650 0.9999167   42.250000
```

```r
## remove lowly expressed genes <- at least one sample has to have 50 reads
res.dt <- res.dt[baseMean > 50]
res.dt
```

```
##                                    gene_id         t0    t28 log2FoldChange
##     1:    ENSG00000124208.16|PEDS1-UBE2V1  268.444444    1.0      -7.073844
##     2:          ENSG00000165949.13|IFI27 2789.444444  165.5      -4.066901
##     3:        ENSG00000224041.3|IGKV3D-15  655.888889   46.9      -3.777552
##     4:          ENSG00000115155.19|OTOF  794.555556  104.1      -2.920200
##     5: ENSG00000285238.2|ENSG00000285238  580.666667  103.4      -2.478071
##      ---
## 14955: ENSG00000272491.1|ENSG00000272491   40.666667  243.3       2.551688
## 14956:          ENSG00000108700.5|CCL8   58.777778  356.1       2.578647
## 14957: ENSG00000264058.2|ENSG00000264058    4.555556  108.1       4.295576
## 14958:          ENSG00000138135.7|CH25H    6.555556  181.7       4.595795
## 14959: ENSG00000282885.3|ENSG00000282885   83.000000 3483.4       5.374377
##          p.value  p.adjust    baseMean
##     1: 0.32947073 0.9999167   134.72222
##     2: 0.03298818 0.9999167  1477.47222
##     3: 0.33940932 0.9999167   351.39444
##     4: 0.03170793 0.9999167   449.32778
##     5: 0.25901085 0.9999167   342.03333
##      ---
## 14955: 0.40319539 0.9999167   141.98333
## 14956: 0.42062361 0.9999167   207.43889
## 14957: 0.32504672 0.9999167    56.32778
## 14958: 0.35777310 0.9999167    94.12778
## 14959: 0.34651682 0.9999167  1783.20000
```

```r
## significantly UP
res.up.dt <- res.dt[p.value <= 0.05][log2FoldChange > 0][order(-log2FoldChange)][1:10]
res.up.sig.dt <- res.dt[p.value <= 0.05][log2FoldChange > 0][order(p.value)][1:10]


res.dn.dt <- res.dt[p.value <= 0.05][log2FoldChange < 0][order(log2FoldChange)][1:10]

## view genes on extremes of x axis to set reasonable axis size
res.dt[abs(log2FoldChange)>4]
```

```
##                                    gene_id         t0    t28 log2FoldChange
## 1:    ENSG00000124208.16|PEDS1-UBE2V1  268.444444    1.0      -7.073844
## 2:          ENSG00000165949.13|IFI27 2789.444444  165.5      -4.066901
## 3: ENSG00000264058.2|ENSG00000264058    4.555556  108.1       4.295576
## 4:          ENSG00000138135.7|CH25H    6.555556  181.7       4.595795
## 5: ENSG00000282885.3|ENSG00000282885   83.000000 3483.4       5.374377
##       p.value  p.adjust   baseMean
## 1: 0.32947073 0.9999167  134.72222
## 2: 0.03298818 0.9999167 1477.47222
```
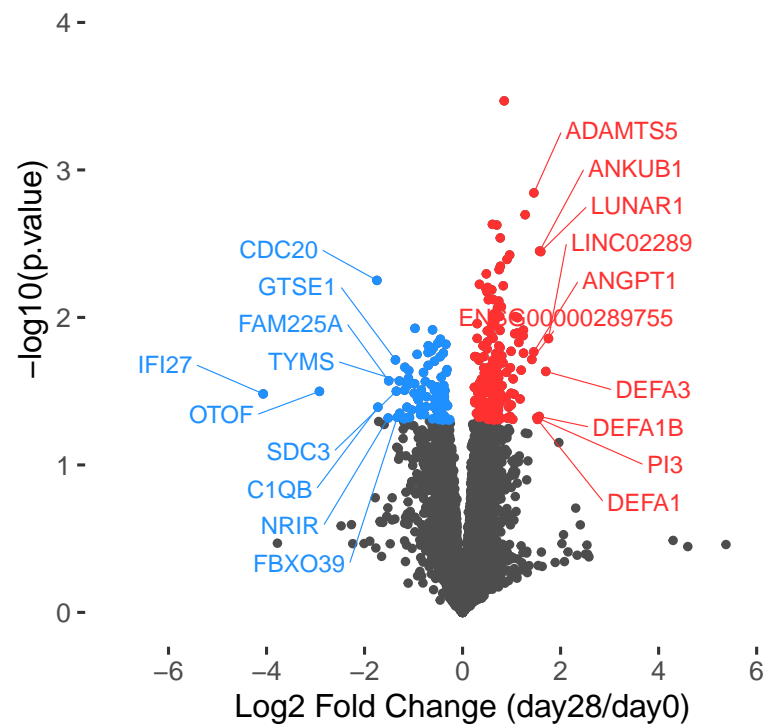
```
## 3: 0.32504672 0.9999167    56.32778
## 4: 0.35777310 0.9999167    94.12778
## 5: 0.34651682 0.9999167 1783.20000
```

```
res.dt[p.value < 0.001]
```

```
##                                gene_id        t0    t28 log2FoldChange
## 1:            ENSG00000122729.19|ACO1 229.44444 348.1      0.5992214
## 2: ENSG00000262380.1|ENSG00000262380  52.77778  95.9      0.8494865
##           p.value  p.adjust  baseMean
## 1: 2.226443e-05 0.4167233 288.77222
## 2: 3.400907e-04 0.9999167  74.33889
```

```r
## plot
ggp.volcano <- ggplot() + theme_pubclean() +
  geom_point(data = res.dt,
             aes(x = log2FoldChange, y = -log10(p.value)),
             size=1, colour = "grey30") +
  geom_point(data = res.dt[p.value <= 0.05][log2FoldChange > 0],
             aes(x = log2FoldChange, y = -log10(p.value)),
             size=1, colour = "firebrick1") +
  geom_point(data = res.dt[p.value <= 0.05][log2FoldChange < 0],
             aes(x = log2FoldChange, y = -log10(p.value)),
             size=1, colour = "dodgerblue") +
  geom_text_repel(data = res.up.dt,
             aes(x = log2FoldChange, y = -log10(p.value),
                 label = unlist(tstrsplit(gene_id,split="\\|",keep = 2))),
             size=3, colour = "firebrick1",
             nudge_x = 3,
             direction = "y",
             force = 2,
             force_pull = NA,
             vjust=1,
             hjust=1,
             segment.size = 0.2) +
  geom_text_repel(data = res.dn.dt,
             aes(x = log2FoldChange, y = -log10(p.value),
                 label = unlist(tstrsplit(gene_id,split="\\|",keep = 2))),
             size=3, colour = "dodgerblue",
             nudge_x = -2,
             force = 3,
             force_pull = NA,
             vjust=0,
             direction = "y",
             segment.size = 0.2) +
  scale_x_continuous(limits = c(-7,7), breaks = seq(-10,10,2)) +
  xlab("Log2 Fold Change (day28/day0)") +
  theme(aspect.ratio = 1,
        panel.grid.major.y = element_blank()); ggp.volcano
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```

```
## save to file
ggsave(filename = paste(FIG.DIR,"scope_batch1_volcano_time_",
                        version.date,".pdf"),
       plot = ggp.volcano, width = 6, height = 8)
```

Warning: Removed 1 rows containing missing values (`geom_point()`).

Runnging DGE using DE

```
## library
library(DESeq2)

## rc
```