revisions_adipog_cell_cycle_TFs_take2

Sarah Hp

2023-06-14

Relies on:

```r
library(biomaRt)
library(ComplexHeatmap)
library(circlize)
library(GSVA)
library(GSEABase)
library(ggplot2)
library(tidyr)
library(dplyr)
library(clusterProfiler)
```

```r
knitr::opts_chunk$set(echo = TRUE, dev = c("pdf"), fig.path = "heatmaps_adipogenesis_take2/")
```

```r
rpkm = read.delim("C:/Users/sarahhp/OneDrive - Universitetet i Oslo/Projects/dunia's_nucleolus_paper/nu
head(rpkm); dim(rpkm)
```

```
##         Geneid Length gene_name
## 1 ENSG00000000003   4535    TSPAN6
## 2 ENSG00000000005   1610     TNMD
## 3 ENSG00000000419   1207     DPM1
## 4 ENSG00000000457   6883     SCYL3
## 5 ENSG00000000460   5967   C1orf112
## 6 ENSG00000000938   3474      FGR
##                                              description day.2.D1G.bulk
## 1                                            tetraspanin 6   3.601999275
## 2                                            tenomodulin   0.007936845
## 3 dolichyl-phosphate mannosyltransferase subunit 1, catalytic  62.130005013
## 4                                      SCY1 like pseudokinase 3   1.426204636
## 5                           chromosome 1 open reading frame 112   1.906747475
## 6                  FGR proto-oncogene, Src family tyrosine kinase   0.005503088
##   day.2.D2A.bulk day0.D1G.bulk day0.D2A.bulk day1.D1G.bulk day1.D2A.bulk
## 1     6.76071437    2.69543757    6.28332533    2.21416364    6.40180026
## 2     0.00000000    0.00000000    0.02464032    0.03748234    0.25958703
## 3    41.18788108   46.57260499   38.24198388   62.18075805   44.14029328
## 4     1.33753816    1.58269525    1.83452213    1.67277113    1.78269269
## 5     1.65219243    0.36937923    0.56555920    0.35661924    0.73777550
## 6     0.01975517    0.06610962    0.03066300    0.02233808    0.04456323
##   day15.D1G.bulk day15.D1G.floating day15.D2A.bulk day15.D2A.floating
## 1      9.1422795         15.7180139      8.3310883         15.8591145
## 2      0.5985436          0.5635616      1.9493616          2.5943861
```

```
## 3      35.7545717          30.0118693        36.1959969             43.8454420
## 4       1.7310483           1.5638762         1.6389382              1.9744310
## 5       0.5774263           0.8213003         0.4421079              0.7067799
## 6       0.6295882           3.7540597         1.9196028              3.2901662
##   day3.D1G.bulk day3.D2A.bulk day9.D1G.bulk day9.D2A.bulk
## 1    2.76755264     7.6279970     2.3380573     8.8221868
## 2    0.06034418     0.1242015     0.3061671     2.7625195
## 3   56.03395361    46.3859709    49.9509805    45.7146885
## 4    1.98222151     1.9524608     2.0051649     2.0390762
## 5    0.39508980     0.5258706     0.3640591     0.6355628
## 6    0.16554146     0.1578544     0.3397660     0.3845655
```

```
## [1] 21174    18
```

```
combpval = read.delim("C:/Users/sarahhp/OneDrive - Universitetet i Oslo/Projects/dunia's_nucleolus_pape
head(combpval)
```

```
##          Geneid Length gene_name
## 1 ENSG00000151726   6284     ACSL1
## 2 ENSG00000099194   5362       SCD
## 3 ENSG00000042445   4005    RETSAT
## 4 ENSG00000056998   3655      GYG2
## 5 ENSG00000076555  14505     ACACB
## 6 ENSG00000101938   3920    CHRDL1
##                                    description      logFC    AveExpr        t
## 1 acyl-CoA synthetase long chain family member 1  12.916437   8.810891 50.67161
## 2                          stearoyl-CoA desaturase  16.490127  11.807874 43.63744
## 3                             retinol saturase   8.195969   7.100134 36.32129
## 4                                glycogenin 2  10.028582   4.980263 36.39842
## 5                    acetyl-CoA carboxylase beta  14.871196   7.351638 33.29351
## 6                               chordin like 1  14.207558   5.954537 32.38451
##        P.Value    adj.P.Val        B
## 1 8.291877e-37 1.755722e-32 68.92770
## 2 2.897205e-34 3.067271e-30 64.45829
## 3 2.309287e-31 9.779370e-28 59.33478
## 4 2.133268e-31 9.779370e-28 58.59981
## 5 7.038353e-30 2.474282e-26 55.41491
## 6 1.960845e-29 4.613214e-26 54.48242
```

```
sig = merge(rpkm, combpval[c("gene_name","adj.P.Val")])
head(sig); dim(sig)
```

```
##   gene_name          Geneid Length
## 1      A1BG ENSG00000121410   4006
## 2  A1BG-AS1 ENSG00000268895   2793
## 3       A2M ENSG00000175899   6384
## 4   A2M-AS1 ENSG00000245105   2816
## 5    A4GALT ENSG00000128274   3407
## 6      ABAT ENSG00000183044   9744
##                             description day.2.D1G.bulk
## 1                 alpha-1-B glycoprotein     0.01815337
## 2                     A1BG antisense RNA 1     0.90975838
```

```
## 3                         alpha-2-macroglobulin       0.21349869
## 4                            A2M antisense RNA 1       0.29287176
## 5 alpha 1,4-galactosyltransferase (P blood group)     4.67272390
## 6               4-aminobutyrate aminotransferase       0.81520901
##   day.2.D2A.bulk day0.D1G.bulk day0.D2A.bulk day1.D1G.bulk day1.D2A.bulk
## 1      0.1047458    0.05978501    0.06987478    0.04544531     0.0815610
## 2      1.3179450    0.90779680    1.23126472    1.87323332     2.0247272
## 3      0.8046160    7.33195064   14.48690784    6.17424097    12.9542154
## 4      0.2337510    0.49255301    0.29768876    0.40976779     0.4652485
## 5      2.0682521    8.79899844    2.55743385    4.76230682     1.3129724
## 6      0.7232442    1.13859664    0.99927577    1.94422327     1.6450301
##   day15.D1G.bulk day15.D1G.floating day15.D2A.bulk day15.D2A.floating
## 1     0.03912984          0.0185444     0.03159112         0.01945774
## 2     1.27934477          1.0516270     1.39386219         0.74776720
## 3    55.90851651         39.4762646    73.55229005        69.76889731
## 4     0.80400184          1.1234145     0.59104284         0.81515811
## 5     4.78652768          9.4369909     4.47173184         2.50902053
## 6     1.99799798          2.1548674     2.29141846         2.42183547
##   day3.D1G.bulk day3.D2A.bulk day9.D1G.bulk day9.D2A.bulk     adj.P.Val
## 1    0.04925457    0.05550053    0.02259233    0.05723545 4.206849e-01
## 2    1.72406789    1.63700087    1.06522748    1.13733587 1.941349e-01
## 3    7.65875953   10.30956590    9.76357919   14.88341115 2.466237e-18
## 4    0.46771549    0.33457401    0.48169174    0.48900749 2.002885e-06
## 5    6.63361018    2.40230131    4.70604627    0.93162324 1.130921e-01
## 6    2.28435365    1.54204016    0.79823994    0.69142124 1.171812e-08

## [1] 20590    19
```

formating...

```
sig = sig[!duplicated(sig$gene_name),]
dim(sig)
```

```
## [1] 20261    19
```

```
rownames(sig) = sig$gene_name
sig$gene_name = NULL
#select time points
sig = sig[!grepl("floating", colnames(sig))]
head(sig)
```

```
##                  Geneid Length
## A1BG     ENSG00000121410   4006
## A1BG-AS1 ENSG00000268895   2793
## A2M      ENSG00000175899   6384
## A2M-AS1  ENSG00000245105   2816
## A4GALT   ENSG00000128274   3407
## ABAT     ENSG00000183044   9744
##                                    description day.2.D1G.bulk
## A1BG                       alpha-1-B glycoprotein     0.01815337
## A1BG-AS1                      A1BG antisense RNA 1     0.90975838
## A2M                       alpha-2-macroglobulin     0.21349869
```

```
## A2M-AS1                                  A2M antisense RNA 1       0.29287176
## A4GALT     alpha 1,4-galactosyltransferase (P blood group)        4.67272390
## ABAT                  4-aminobutyrate aminotransferase            0.81520901
##           day.2.D2A.bulk day0.D1G.bulk day0.D2A.bulk day1.D1G.bulk day1.D2A.bulk
## A1BG           0.1047458    0.05978501    0.06987478    0.04544531     0.0815610
## A1BG-AS1       1.3179450    0.90779680    1.23126472    1.87323332     2.0247272
## A2M            0.8046160    7.33195064   14.48690784    6.17424097    12.9542154
## A2M-AS1        0.2337510    0.49255301    0.29768876    0.40976779     0.4652485
## A4GALT         2.0682521    8.79899844    2.55743385    4.76230682     1.3129724
## ABAT           0.7232442    1.13859664    0.99927577    1.94422327     1.6450301
##           day15.D1G.bulk day15.D2A.bulk day3.D1G.bulk day3.D2A.bulk
## A1BG          0.03912984     0.03159112    0.04925457    0.05550053
## A1BG-AS1      1.27934477     1.39386219    1.72406789    1.63700087
## A2M          55.90851651    73.55229005    7.65875953   10.30956590
## A2M-AS1       0.80400184     0.59104284    0.46771549    0.33457401
## A4GALT        4.78652768     4.47173184    6.63361018    2.40230131
## ABAT          1.99799798     2.29141846    2.28435365    1.54204016
##           day9.D1G.bulk day9.D2A.bulk    adj.P.Val
## A1BG         0.02259233    0.05723545 4.206849e-01
## A1BG-AS1     1.06522748    1.13733587 1.941349e-01
## A2M          9.76357919   14.88341115 2.466237e-18
## A2M-AS1      0.48169174    0.48900749 2.002885e-06
## A4GALT       4.70604627    0.93162324 1.130921e-01
## ABAT         0.79823994    0.69142124 1.171812e-08
```

```r
sig = sig[grepl("day(.2|0|3|15)", colnames(sig))]
# check for zeroes
summary(rowSums(sig) > 0)#there are none
```

```
##    Mode    TRUE
## logical   20261
```

```r
#separate by donor
d1 = sig[grep("D1G", colnames(sig), value=T)]
head(d1)
```

```
##           day.2.D1G.bulk day0.D1G.bulk day15.D1G.bulk day3.D1G.bulk
## A1BG          0.01815337    0.05978501     0.03912984    0.04925457
## A1BG-AS1      0.90975838    0.90779680     1.27934477    1.72406789
## A2M           0.21349869    7.33195064    55.90851651    7.65875953
## A2M-AS1       0.29287176    0.49255301     0.80400184    0.46771549
## A4GALT        4.67272390    8.79899844     4.78652768    6.63361018
## ABAT          0.81520901    1.13859664     1.99799798    2.28435365
```

```r
d2 = sig[grep("D2A", colnames(sig), value=T)]
head(d2)
```

```
##           day.2.D2A.bulk day0.D2A.bulk day15.D2A.bulk day3.D2A.bulk
## A1BG           0.1047458    0.06987478     0.03159112    0.05550053
## A1BG-AS1       1.3179450    1.23126472     1.39386219    1.63700087
## A2M            0.8046160   14.48690784    73.55229005   10.30956590
## A2M-AS1        0.2337510    0.29768876     0.59104284    0.33457401
## A4GALT         2.0682521    2.55743385     4.47173184    2.40230131
## ABAT           0.7232442    0.99927577     2.29141846    1.54204016
```

Get terms

```
molsig_list <- GSEABase::getGmt("C:/Users/sarahhp/OneDrive - Universitetet i Oslo/Projects/dunia's_nucl
molsig_list
```

```
## GeneSetCollection
##   names: chr1p12, chr1p13, ..., REACTOME_ION_CHANNEL_TRANSPORT (32284 total)
##   unique identifiers: VTCN1, LINC01525, ..., HMGB1P40 (39781 total)
##   types in collection:
##     geneIdType: NullIdentifier (1 total)
##     collectionType: NullCollection (1 total)
```

```
##Pick gene sets to look in bcos all in a lot
key_terms = c("HALLMARK_ADIPOGENESIS","GOBP_MITOTIC_CELL_CYCLE_ARREST", "GOBP_MITOTIC_CELL_CYCLE")
selected = molsig_list[names(molsig_list) %in% key_terms,]
selected
```

```
## GeneSetCollection
##   names: GOBP_MITOTIC_CELL_CYCLE, GOBP_MITOTIC_CELL_CYCLE_ARREST, HALLMARK_ADIPOGENESIS (3 total)
##   unique identifiers: SRA1, MIR892B, ..., UQCR11 (1225 total)
##   types in collection:
##     geneIdType: NullIdentifier (1 total)
##     collectionType: NullCollection (1 total)
```

```
rm(molsig_list)
```

Donor 1

heatmap formatting

```
colnames(d1) = gsub(".D1G.bulk","", colnames(d1))
d1 = d1[rowSums(d1) > 0,] #remove zeroes
d1 = d1[c("day.2","day0","day3","day15")]
zs = t(scale(t(log2(d1+1))))
summary(zs)
```

```
##      day.2               day0               day3               day15
## Min.   :-1.49964    Min.   :-1.49883    Min.   :-1.49931    Min.   :-1.49976
## 1st Qu.:-0.72768    1st Qu.:-0.60791    1st Qu.:-0.72592    1st Qu.:-0.85161
## Median :-0.05323    Median :-0.01355    Median :-0.09717    Median :-0.17937
## Mean   : 0.02035    Mean   : 0.03213    Mean   :-0.01549    Mean   :-0.03698
## 3rd Qu.: 0.78125    3rd Qu.: 0.66849    3rd Qu.: 0.72670    3rd Qu.: 0.88044
## Max.   : 1.50000    Max.   : 1.49998    Max.   : 1.49961    Max.   : 1.50000
```

Adipogenesis Heatmap

Showing DE genes for hallmark adipogenesis

```
Heatmap(as.matrix(zs[rownames(zs) %in% geneIds(selected)$HALLMARK_ADIPOGENESIS &
                     rownames(zs) %in% combpval$gene_name[combpval$adj.P.Val < 0.01],]),
  cluster_columns = F,
  name="Relative FPKM", column_title = "Hallmark Adipogenesis",
  col=circlize::colorRamp2(c(-2,0,2),c(rgb(0.2,0.3,0.6),rgb(1,1,1),rgb(0.9,0.1,0.1))),
  show_row_names = F)
```


Hallmark Adipogenesis
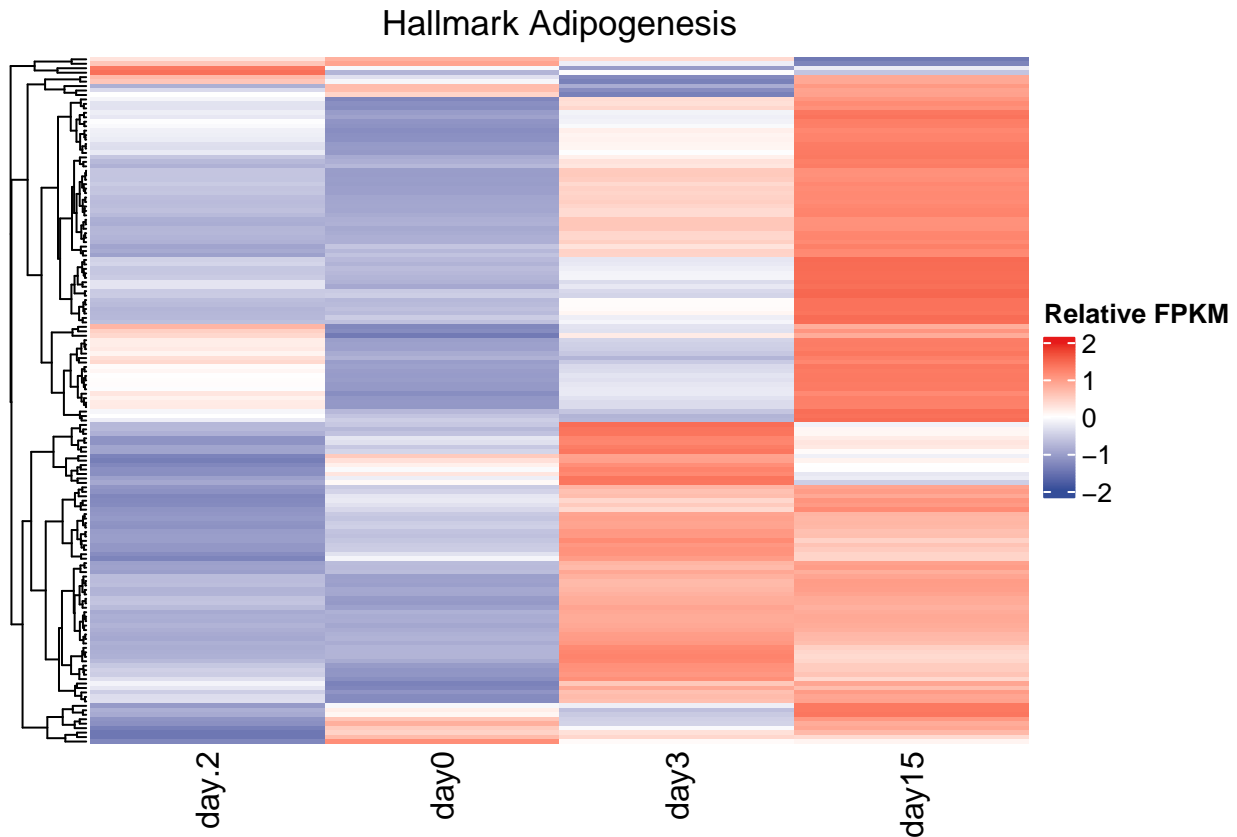
TF Heatmap And for Key Adipogenesis regulators

```
Heatmap(as.matrix(zs[rownames(zs) %in% c("PPARG","CEBPA","CEBPB","CEBPD"),]),
  cluster_columns = F,
  name="Relative FPKM", column_title = "Adipogenic TFs",
  col=circlize::colorRamp2(c(-2,0,2),c(rgb(0.2,0.3,0.6),rgb(1,1,1),rgb(0.9,0.1,0.1))))
```
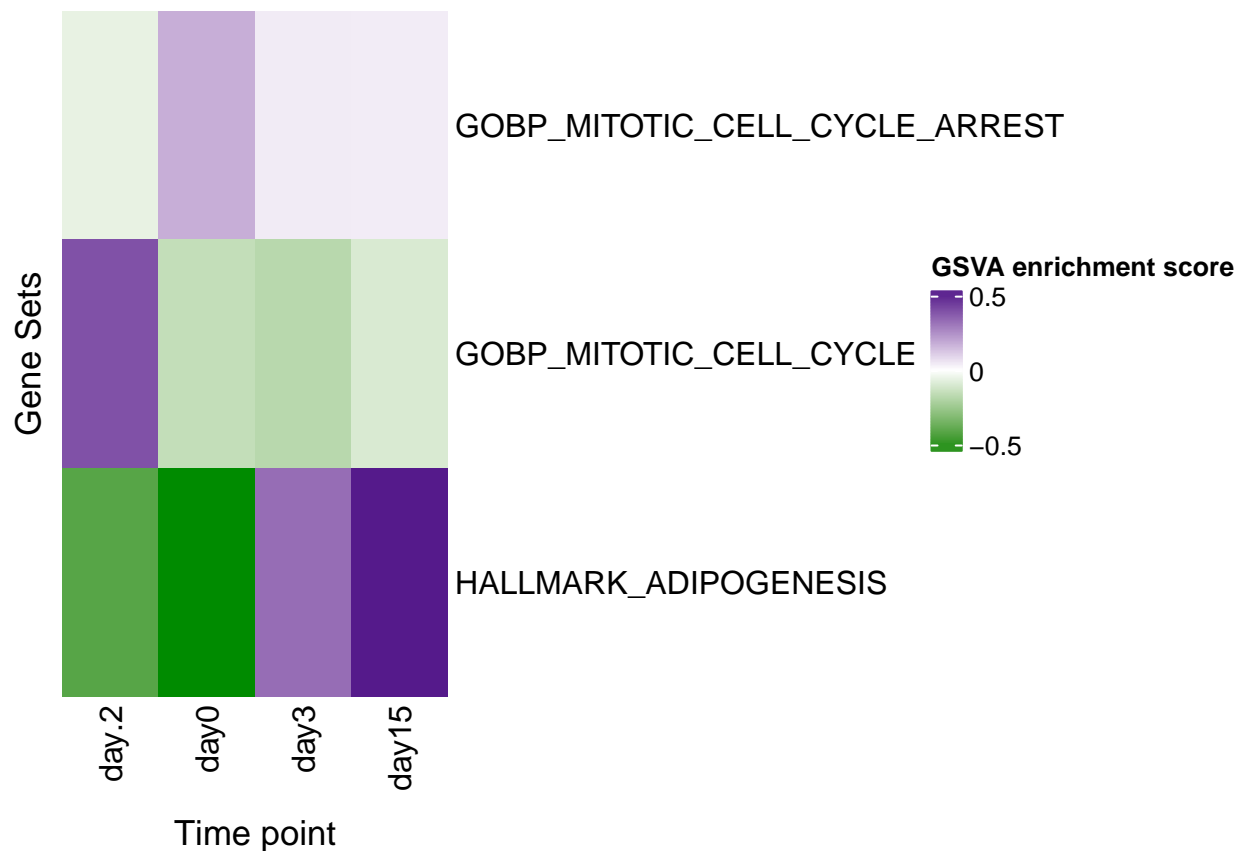
Adipogenic TFs
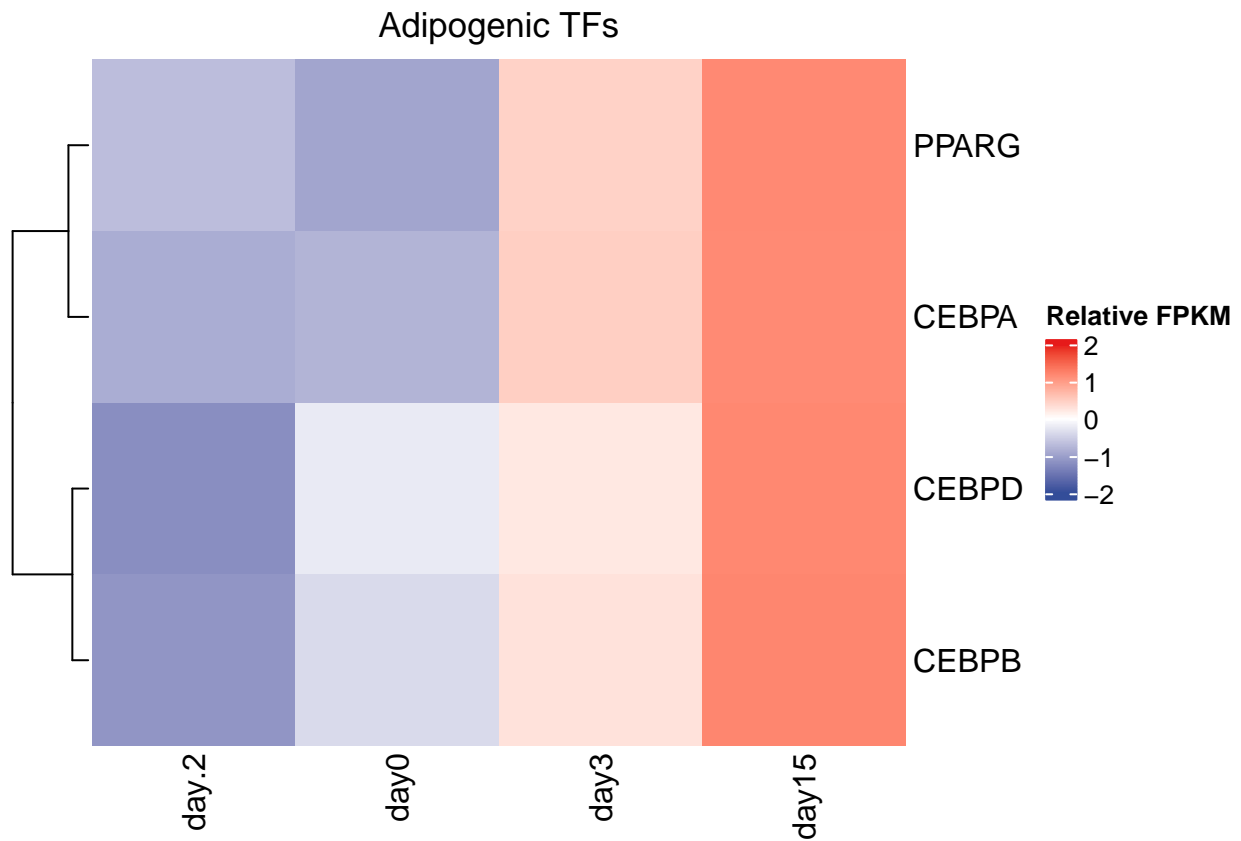
Average Gene set Expression

```
d1va = gsva(as.matrix(d1), selected)
```

```
## Estimating GSVA scores for 3 gene sets.
## Estimating ECDFs with Gaussian kernels
##   |                                                                      |
```

```
Heatmap(d1va, cluster_columns = F, show_row_dend = F, col = colorRamp2(c(min(d1va),0, max(d1va)), c("gr
        name="GSVA enrichment score", row_title = "Gene Sets",column_title = "Time point",column_title_s
```

D2

```r
colnames(d2) = gsub(".D2A.bulk","", colnames(d2))
d2 = d2[rowSums(d2) > 0,] #remove zeroes
d2 = d2[c("day.2","day0","day3","day15")]
zs2 = t(scale(t(log2(d2+1))))
summary(zs2)
```

```
##      day.2              day0              day3              day15
##  Min.   :-1.499194  Min.   :-1.499947  Min.   :-1.499994  Min.   :-1.49999
##  1st Qu.:-0.723028  1st Qu.:-0.597469  1st Qu.:-0.676075  1st Qu.:-0.98056
##  Median :-0.072619  Median :-0.043413  Median :-0.013964  Median :-0.12211
##  Mean   : 0.003168  Mean   : 0.003131  Mean   : 0.007173  Mean   :-0.01347
##  3rd Qu.: 0.719754  3rd Qu.: 0.584658  3rd Qu.: 0.688837  3rd Qu.: 1.05062
##  Max.   : 1.500000  Max.   : 1.500000  Max.   : 1.500000  Max.   : 1.50000
```

```r
Heatmap(as.matrix(zs2[rownames(zs2) %in% geneIds(selected)$HALLMARK_ADIPOGENESIS &
                      rownames(zs2) %in% combpval$gene_name[combpval$adj.P.Val < 0.01],]),
  cluster_columns = F,
  name="Relative FPKM", column_title = "Hallmark Adipogenesis",
  col=circlize::colorRamp2(c(-2,0,2),c(rgb(0.2,0.3,0.6),rgb(1,1,1),rgb(0.9,0.1,0.1))),
  show_row_names = F)
```

Hallmark Adipogenesis
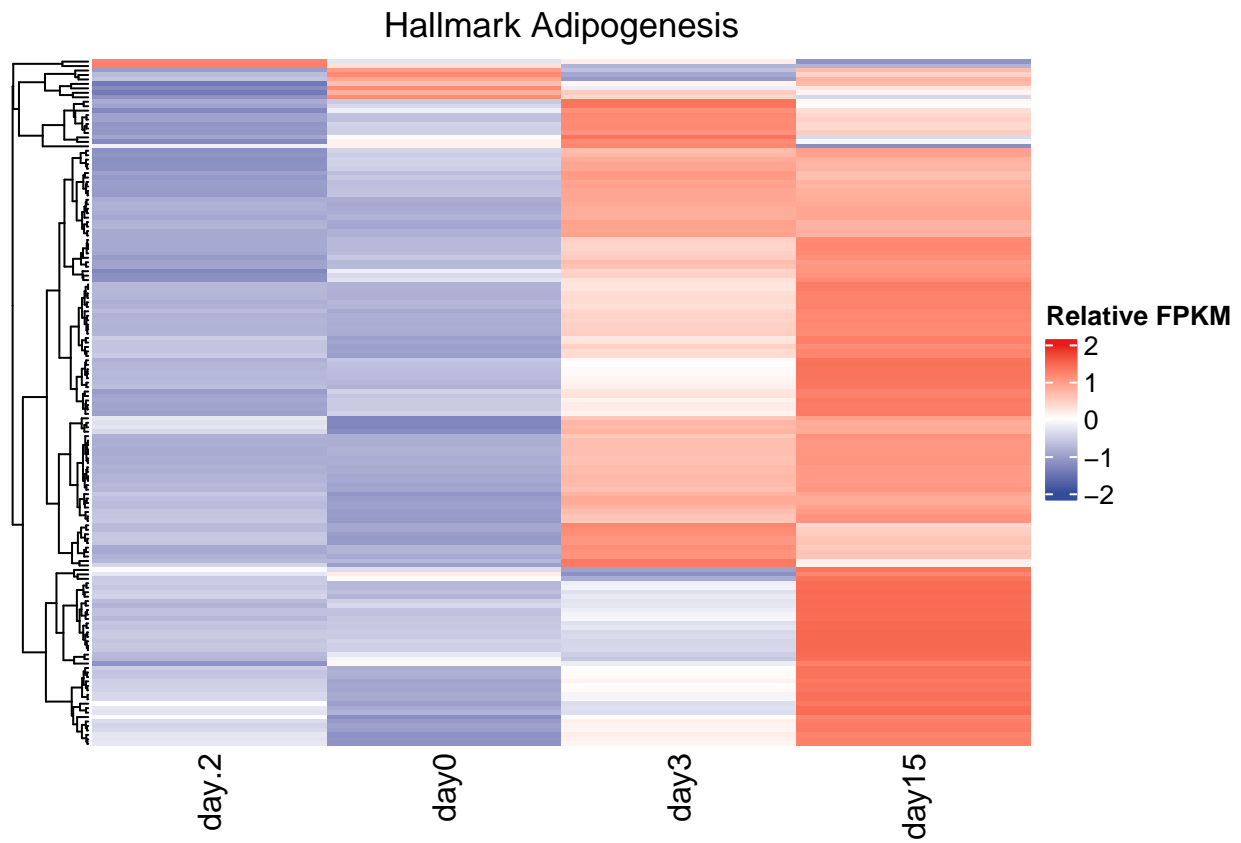

```
Heatmap(as.matrix(zs2[rownames(zs2) %in% c("PPARG","CEBPA","CEBPB","CEBPD"),]),
  cluster_columns = F,
  name="Relative FPKM", column_title = "Adipogenic TFs",
  col=circlize::colorRamp2(c(-2,0,2),c(rgb(0.2,0.3,0.6),rgb(1,1,1),rgb(0.9,0.1,0.1))))
```

Adipogenic TFs
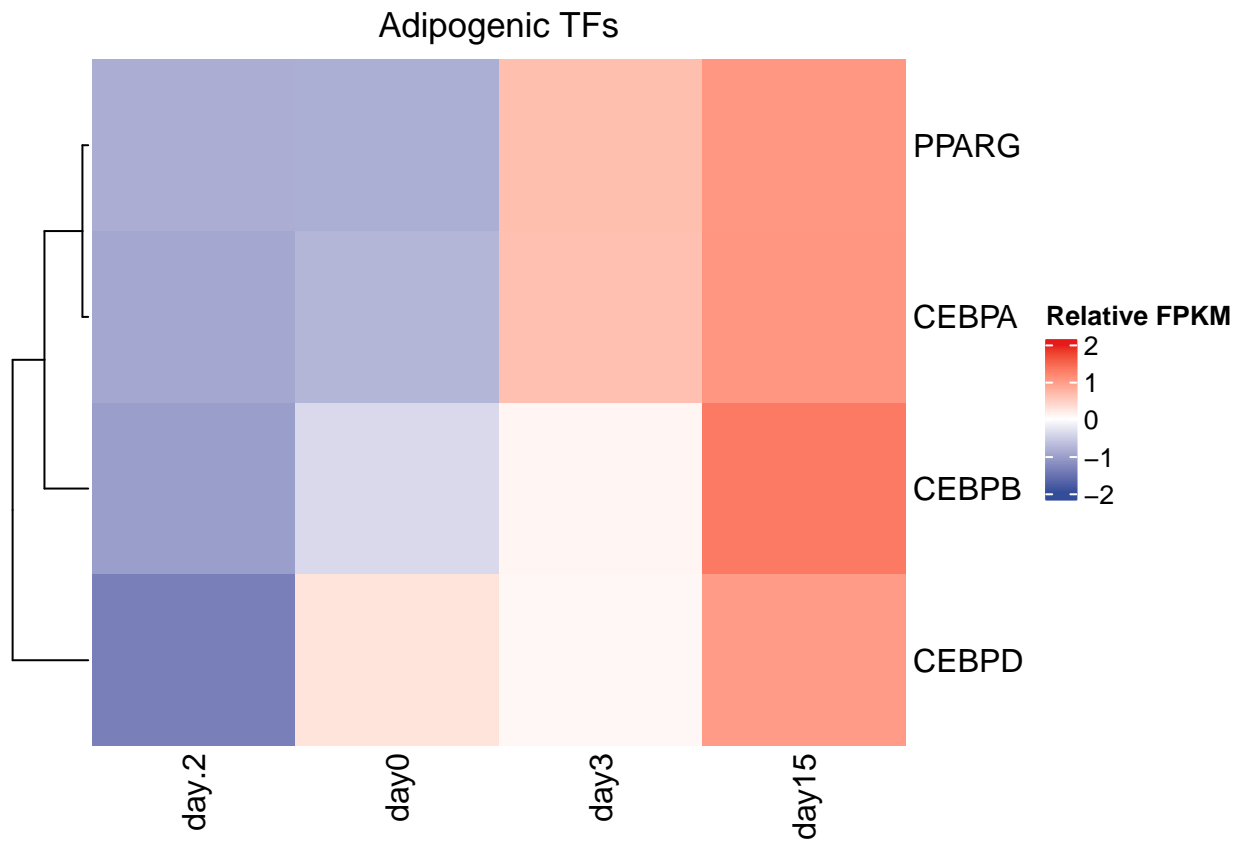
Average Gene set Expression

```
d2va = gsva(as.matrix(d2), selected)
```

```
## Estimating GSVA scores for 3 gene sets.
## Estimating ECDFs with Gaussian kernels
##   |                                                                      |
```

```
Heatmap(d2va, cluster_columns = F, show_row_dend = F, col = colorRamp2(c(min(d2va),0, max(d2va)), c("gr
        name="GSVA enrichment score", row_title = "Gene Sets",column_title = "Time point",column_title_s
```