splicing_factors_heatmap_day15

2023-10-20

```
library(tidyr)
library(dplyr)
library(ComplexHeatmap)
library(clusterProfiler)
library(here)
here::i_am("R/03_splicing_factors_heatmap_day15.Rmd")
```

```
rpkm = read.delim(here("03limma", "beige_day15_rpkm_tmm_means.tab"))
head(rpkm)
```

```
##            Geneid Length gene_name
## 1 ENSG00000000003   4536    TSPAN6
## 2 ENSG00000000005   1476      TNMD
## 3 ENSG00000000419   1207      DPM1
## 4 ENSG00000000457   6883     SCYL3
## 5 ENSG00000000460   5970  C1orf112
## 6 ENSG00000000938   3382       FGR
##                                                    description   gene_biotype
## 1                                              tetraspanin 6  protein_coding
## 2                                               tenomodulin  protein_coding
## 3 dolichyl-phosphate mannosyltransferase subunit 1, catalytic  protein_coding
## 4                                 SCY1 like pseudokinase 3  protein_coding
## 5                         chromosome 1 open reading frame 112  protein_coding
## 6             FGR proto-oncogene, Src family tyrosine kinase  protein_coding
##   ensembl_gene_id_version subject1.beige subject1.white subject2.beige
## 1      ENSG00000000003.15       7.015370      5.0752183     12.5300941
## 2       ENSG00000000005.6       1.627583      0.3200113     30.6612441
## 3      ENSG00000000419.12      30.635853     33.4089794     32.3055705
## 4      ENSG00000000457.14       1.862880      1.4176901      2.3720554
## 5      ENSG00000000460.17       0.393713      0.4077405      0.4788014
## 6      ENSG00000000938.13       2.977751      0.7902128      4.6470874
##   subject2.white subject3.beige subject3.white subject4.beige subject4.white
## 1      6.4851444      9.7247952      6.2324136      9.3056202      9.0232992
## 2     30.2448972      2.7906896      0.4912616      2.3021246      0.6401144
## 3     29.1252400     34.1100563     34.7548040     32.7208903     34.6618668
## 4      1.5836657      2.1136964      1.5022350      2.1423301      1.7105704
## 5      0.3762281      0.4816479      0.4403470      0.4698593      0.5608791
## 6      0.8865716      7.1807572      0.9509210      5.8448505      0.6362375
##   subject5.beige subject5.white subject6.beige subject6.white
## 1     10.0681020      6.3019890     11.7900221      8.3815053
## 2      6.2119587      1.6856768      8.8600561      2.1400068
## 3     31.6549387     30.9930433     38.2724535     35.8339890
## 4      1.9104027      1.5130770      2.0849212      1.6548458
## 5      0.3557019      0.4699746      0.3751493      0.4361585
```

```
## 6      5.6321927      0.7380337     12.7832922      1.9632742
```

set up heatmap matrices

```
rpkm$gene_name[duplicated(rpkm$gene_name)]
```

```
##  [1] "CD99"     "SLC25A6"  "GTPBP6"   "Y_RNA"    "Y_RNA"    "Y_RNA"
##  [7] "Y_RNA"    "Y_RNA"    "Y_RNA"    "Y_RNA"    "BMS1P4"   "POLR2J4"
## [13] "MATR3"    "HSPA14"   "TBCE"     "POLR2J3"  "AHRR"     "C2orf27A"
```

```
rpkm = rpkm[!duplicated(rpkm$gene_name),]
rownames(rpkm) = rpkm$gene_name
rmat = rpkm[,grep("beige|white", colnames(rpkm))]
tail (rmat)
```

```
##            subject1.beige subject1.white subject2.beige subject2.white
## AC005618.4     0.19327565     0.26156522     0.27244232     0.21231920
## AL022318.5     0.05008334     0.06833633     0.08021512     0.15143889
## FAM106C        0.38080708     0.37914769     0.03539279     0.19205603
## AC010332.3     0.03458649     0.05052052     0.04345704     0.06817003
## CRIPAK         1.15433350     1.45605812     1.04181365     1.38704983
## AL109627.1     0.98793519     1.39978120     0.29364086     0.82971890
##            subject3.beige subject3.white subject4.beige subject4.white
## AC005618.4    0.193890828    0.209541955    0.187030458     0.19207329
## AL022318.5    0.027763170    0.054401627    0.007548142     0.02258869
## FAM106C       0.002384258    0.004406799    0.105184357     0.16605825
## AC010332.3    0.034930087    0.050085109    0.038787029     0.07133976
## CRIPAK        0.950599560    0.971724626    1.094710630     1.02932258
## AL109627.1    0.691190239    1.135051838    0.312630332     0.71429497
##            subject5.beige subject5.white subject6.beige subject6.white
## AC005618.4     0.15569489     0.13545189     0.17030631     0.27480056
## AL022318.5     0.07566827     0.08769766     0.00000000     0.01049002
## FAM106C        0.03577993     0.04848370     0.16900469     0.22934149
## AC010332.3     0.03473264     0.05364728     0.02968104     0.02685822
## CRIPAK         0.87594789     1.08185436     0.85676642     1.04604229
## AL109627.1     0.51394167     0.62222821     0.55453845     1.27463757
```

```
norm = t(scale(t(rmat)))
tail(norm)
```

```
##            subject1.beige subject1.white subject2.beige subject2.white
## AC005618.4    -0.25966141      1.2702403     1.51392165      0.1669741
## AL022318.5    -0.06778105      0.3536206     0.62786267      2.2721855
## FAM106C        1.76277678      1.7503367    -0.82673319      0.3477435
## AC010332.3    -0.70461973      0.4019181    -0.08860458      1.6275876
## CRIPAK         0.41349700      2.0663809    -0.20290031      1.6883452
## AL109627.1     0.58699210      1.7356168    -1.34937148      0.1457322
##            subject3.beige subject3.white subject4.beige subject4.white
## AC005618.4     -0.2458795     0.10475510    -0.39957348     -0.2865981
## AL022318.5     -0.5830804     0.03191402    -1.04977894     -0.7025421
## FAM106C        -1.0741924    -1.05902974    -0.30351815      0.1528427
## AC010332.3     -0.6807584     0.37168138    -0.41291324      1.8477095
```

```
## CRIPAK         -0.7025821    -0.58685644     0.08687573     -0.2713279
## AL109627.1     -0.2406196     0.99729549    -1.29641047     -0.1761813
##            subject5.beige subject5.white subject6.beige subject6.white
## AC005618.4     -1.1015891    -1.55509619     -0.7742471      1.5667537
## AL022318.5      0.5228910     0.80061006     -1.2240407     -0.9818606
## FAM106C        -0.8238309    -0.72859298      0.1749316      0.6272660
## AC010332.3     -0.6944700     0.61905647     -1.0452781     -1.2413090
## CRIPAK         -1.1115329     0.01644752     -1.2166113     -0.1797353
## AL109627.1     -0.7349599    -0.43295240     -0.6217368      1.3865954
```

Get splicing factors

```
molsig <- clusterProfiler::read.gmt(here("annotations/msigdb.v2023.1.Hs.symbols.gmt"))
head(molsig); nrow(molsig)
```

```
##      term       gene
## 1 chr1p11  LINC02798
## 2 chr1p11    MTIF2P1
## 3 chr1p11    SRGAP2C
## 4 chr1p11 SRGAP2-AS1
## 5 chr1p11  LINC01691
## 6 chr1p11     NBPF26
```

```
## [1] 3961711
```

```
prefixes = c("HALLMARK", "KEGG","REACTOME", "WP", "GOBP", "GOCC","GOMF")
colnames(molsig) = c("term","gene")
some.molsig = molsig[gsub("_.*","", molsig$term) %in% prefixes,]
some.molsig$term = factor(some.molsig$term)
table(gsub("_.*","", some.molsig$term))
```

```
##
##     GOBP     GOCC     GOMF HALLMARK     KEGG REACTOME       WP
##   642656    98915   108833     7322    12796    92769    31635
```

```
reg_rnasplice = some.molsig$gene[grep("GOBP_REGULATION_OF_RNA_SPLICING", some.molsig$term)]
head(reg_rnasplice, n=50)
```

```
##  [1] "PQBP1"   "RBM12"   "MBNL2"   "RBM7"    "RBM5"    "SRRM1"   "SF3B4"
##  [8] "SAP18"   "PRMT5"   "TADA3"   "DDX17"   "TAF6L"   "KHDRBS3" "KHDRBS1"
## [15] "CELF1"   "CELF2"   "SRSF10"  "RNPS1"   "FASTK"   "HNRNPA0" "CELF3"
## [22] "SGF29"   "AHNAK2"  "U2AF2"   "CIRBP"   "TADA1"   "CLK1"    "CLK2"
## [29] "CLK3"    "CLNS1A"  "SRSF12"  "RBFOX3"  "RBMY1F"  "DDX5"    "DYRK1A"
## [36] "KHDRBS2" "ERN1"    "RBM24"   "PUF60"   "HABP4"   "SNW1"    "SETX"
## [43] "RRP1B"   "JMJD6"   "FMR1"    "USP22"   "AFF2"    "SF3B3"   "RBFOX2"
## [50] "STH"
```

```
summary(reg_rnasplice %in% rpkm$gene_name) #24 unexpressed
```

```
##    Mode   FALSE    TRUE
## logical      24     157
```

regulation of RNA splicing looks like a decent list; but perhaps the splicing paper have some more specific/curated checking ones ... Castella shows reactome mRNA splicing; 212 genes And then checks 47 "representative" spliceaid genes; though they list 71 proteins in the paper abstract. the site is incredibly slow, http://www.intron i.it/splicing.html Tissue specific search tool also exists: http://193.206.120.249/splicing_tissue.html

```
splicing_proteins = read.delim(here("annotations", "SpliceAidF_Table1.csv"), sep = ";")
head(splicing_proteins)
```

```
##   Splicing.factor Binding.sites Conditional.binding.sites No.binding.sites
## 1             9G8            70                         1               29
## 2         CUG-BP1            42                         3               32
## 3          DAZAP1            12                         0                3
## 4           ESRP1             1                         0                1
## 5           ESRP2             1                         0                1
## 6           ETR-3            31                         4               36
```

```
splicing_proteins$no_punct = gsub("[ /]", "", splicing_proteins$Splicing.factor)
splicing_proteins$no_punct = gsub("Nova-", "NOVA", splicing_proteins$no_punct)
splicing_proteins
```

```
##    Splicing.factor Binding.sites Conditional.binding.sites No.binding.sites
## 1              9G8            70                         1               29
## 2          CUG-BP1            42                         3               32
## 3           DAZAP1            12                         0                3
## 4            ESRP1             1                         0                1
## 5            ESRP2             1                         0                1
## 6            ETR-3            31                         4               36
## 7             FMRP            43                         4                4
## 8            Fox-1            12                         0                2
## 9            Fox-2            13                         0                3
## 10        hnRNP A0             1                         0                0
## 11        hnRNP A1           143                        17               39
## 12     hnRNP A2/B1            42                         1                8
## 13        hnRNP A3             2                         0                1
## 14         hnRNP C            21                         7               13
## 15        hnRNP C1            11                         2               19
## 16        hnRNP C2            10                         0               16
## 17         hnRNP D            29                         2               14
## 18        hnRNP D0             1                         0                0
## 19        hnRNP DL            34                         0                0
## 20        hnRNP E1            43                        17               14
## 21        hnRNP E2            39                        13               23
## 22         hnRNP F            67                         8               26
## 23         hnRNP G             1                         0                0
## 24        hnRNP H1            85                         8               45
## 25        hnRNP H2           101                         8               42
## 26        hnRNP H3            60                         8               44
## 27   hnRNP I (PTB)           129                        13               53
```

```
## 28         hnRNP J            1                  0                12
## 29         hnRNP K           58                 15                13
## 30         hnRNP L          172                  4                 9
## 31        hnRNP LL           13                  0                 2
## 32         hnRNP M            1                  0                 3
## 33   hnRNP P (TLS)           16                  4                 8
## 34         hnRNP Q           10                  0                 7
## 35         hnRNP U           19                  3                 0
## 36          HTra2?            7                  0                 3
## 37         HTra2?1           20                  1                14
## 38             HuB           44                  0                 1
## 39             HuC            2                  0                 2
## 40             HuD           51                  6                 5
## 41             HuR           72                 25                26
## 42            KSRP           22                  2                 7
## 43           MBNL1           92                 11                34
## 44          Nova-1           25                  4                18
## 45          Nova-2           12                  4                 9
## 46            nPTB            3                  0                 1
## 47             PSF           32                  0                 7
## 48             QKI            1                  0                 0
## 49           RBM25            1                  0                 1
## 50            RBM4            8                  0                 2
## 51            RBM5            7                  0                 2
## 52           Sam68           16                  0                 5
## 53          SAP155            1                  0                 0
## 54            SC35          172                  5                47
## 55             SF1           24                  1                 5
## 56         SF2/ASF          248                 15                52
## 57           SLM-1            1                  0                 0
## 58           SLM-2            6                  0                 0
## 59          SRm160            1                  0                 0
## 60           SRp20           74                  0                23
## 61          SRp30c           25                  9                 6
## 62           SRp38           10                  0                 0
## 63           SRp40           68                  7                27
## 64           SRp54            1                  0                 0
## 65           SRp55           64                  7                24
## 66           SRp75            8                  0                18
## 67           TDP43           22                  1                 8
## 68           TIA-1           39                  2                 7
## 69           TIAL1           37                  2                 2
## 70             YB-1          21                  1                14
## 71          ZRANB2           19                  0                 4
## 72           Total         2590                245               896
##       no_punct
## 1          9G8
## 2      CUG-BP1
## 3       DAZAP1
## 4        ESRP1
## 5        ESRP2
## 6        ETR-3
## 7         FMRP
## 8        Fox-1
```

```
## 9        Fox-2
## 10     hnRNPA0
## 11     hnRNPA1
## 12   hnRNPA2B1
## 13     hnRNPA3
## 14      hnRNPC
## 15     hnRNPC1
## 16     hnRNPC2
## 17      hnRNPD
## 18     hnRNPD0
## 19     hnRNPDL
## 20     hnRNPE1
## 21     hnRNPE2
## 22      hnRNPF
## 23      hnRNPG
## 24     hnRNPH1
## 25     hnRNPH2
## 26     hnRNPH3
## 27 hnRNPI(PTB)
## 28      hnRNPJ
## 29      hnRNPK
## 30      hnRNPL
## 31     hnRNPLL
## 32      hnRNPM
## 33 hnRNPP(TLS)
## 34      hnRNPQ
## 35      hnRNPU
## 36      HTra2?
## 37     HTra2?1
## 38         HuB
## 39         HuC
## 40         HuD
## 41         HuR
## 42        KSRP
## 43       MBNL1
## 44       NOVA1
## 45       NOVA2
## 46        nPTB
## 47         PSF
## 48         QKI
## 49       RBM25
## 50        RBM4
## 51        RBM5
## 52       Sam68
## 53      SAP155
## 54        SC35
## 55         SF1
## 56      SF2ASF
## 57       SLM-1
## 58       SLM-2
## 59      SRm160
## 60       SRp20
## 61      SRp30c
## 62       SRp38
```

```
## 63       SRp40
## 64       SRp54
## 65       SRp55
## 66       SRp75
## 67       TDP43
## 68       TIA-1
## 69       TIAL1
## 70        YB-1
## 71      ZRANB2
## 72       Total
```

```
splicing_genes = HGNChelper::checkGeneSymbols(splicing_proteins$no_punct)
```

```
## Maps last updated on: Thu Oct 24 12:31:05 2019
```

```
## Warning in HGNChelper::checkGeneSymbols(splicing_proteins$no_punct): Human gene
## symbols should be all upper-case except for the 'orf' in open reading frames.
## The case of some letters was corrected.
```

```
## Warning in HGNChelper::checkGeneSymbols(splicing_proteins$no_punct): x contains
## non-approved gene symbols
```

```
splicing_genes$gene_name = splicing_genes$Suggested.Symbol
splicing_genes$gene_name[splicing_genes$x == "CUG-BP1"] = "CELF1" #genecards
splicing_genes$gene_name[splicing_genes$x == "hnRNPE1"] = "PCBP1" #genecards
splicing_genes$gene_name[splicing_genes$x == "hnRNPI(PTB)"] = "PTBP1" #genecards
splicing_genes$gene_name[splicing_genes$x == "hnRNPP(TLS)"] = "FUS" #genecards
splicing_genes$gene_name[splicing_genes$x == "HTra2?"] = "TRA2A" #genecards
splicing_genes$gene_name[splicing_genes$x == "HTra2?1"] = "TRA2B" #genecards
splicing_genes$gene_name[splicing_genes$x == "PSF"] = "SFPQ" #genecards to check which
splicing_genes$gene_name[splicing_genes$x == "SF2ASF"] ="SRSF1"
splicing_genes$gene_name[splicing_genes$x == "TDP43"] = "TARDBP" #genecards
splicing_genes = separate_rows(splicing_genes, gene_name, sep=" /// ")
splicing_genes
```

```
## # A tibble: 73 x 4
##    x       Approved Suggested.Symbol gene_name
##    <chr>   <lgl>    <chr>            <chr>
##  1 9G8     FALSE    SLU7 /// SRSF7   SLU7
##  2 9G8     FALSE    SLU7 /// SRSF7   SRSF7
##  3 CUG-BP1 FALSE    <NA>             CELF1
##  4 DAZAP1  TRUE     DAZAP1           DAZAP1
##  5 ESRP1   TRUE     ESRP1            ESRP1
##  6 ESRP2   TRUE     ESRP2            ESRP2
##  7 ETR-3   FALSE    CELF2            CELF2
##  8 FMRP    FALSE    FMR1             FMR1
##  9 Fox-1   FALSE    RBFOX1           RBFOX1
## 10 Fox-2   FALSE    RBFOX2           RBFOX2
## # i 63 more rows
```

9G8, both are possible, separate rows

C1/c2 here are just one gene by the looks, which is already included, same with hnrnpd0

couldn't find any mention of hnrnpJ, all that came up was k.

hnRNPP - aka TLS, aka FUS, aka hnrnpP-P2 the text format doesn't like the alpha and beta

SF2/asf= SRSF1

```
spliceaid = splicing_genes$gene_name[!is.na(splicing_genes$gene_name)]
length(spliceaid)
```

```
## [1] 68
```

```
summary(spliceaid %in% rownames(norm)) #not all are expressed
```

```
##    Mode   FALSE    TRUE
## logical       8      60
```

```
spliceaid[!spliceaid %in% rownames(norm)]
```

```
## [1] "ESRP1"   "ESRP2"   "RBFOX1"  "ELAVL2"  "ELAVL3"  "ELAVL4"  "NOVA2"
## [8] "KHDRBS2"
```

```
#elavl2 for e.g. has ensemblid ENSG00000107105
summary("ENSG00000107105" == rpkm$Geneid)
```

```
##    Mode   FALSE
## logical   18043
```

```
summary("ENSG00000104967" == rpkm$Geneid) #NOVA2 not expressed
```

```
##    Mode   FALSE
## logical   18043
```

```
summary("ENSG00000139910" == rpkm$Geneid) #NOVA1 is expressed.
```

```
##    Mode   FALSE    TRUE
## logical   18042       1
```

Check DE

```
sig = read.delim(here("03limma","any_and_all_donor_DGE.tsv"))
colnames(sig)
```

```
##  [1] "Geneid"              "gene_name"           "description"
##  [4] "Length"              "gene_biotype"        "logFC.s1"
##  [7] "logFC.s2"            "logFC.s3"            "logFC.s4"
## [10] "logFC.s5"            "logFC.s6"            "AveExpr"
## [13] "F"                   "all.donors.P.Value"  "all.donors.adj.P.Val"
```

```
## [16] "all.donors.AvelogFC"  "P.Value.s1"           "adj.P.Val.s1"
## [19] "AveExpr.s1"           "P.Value.s2"           "adj.P.Val.s2"
## [22] "AveExpr.s2"           "P.Value.s3"           "adj.P.Val.s3"
## [25] "AveExpr.s3"           "P.Value.s4"           "adj.P.Val.s4"
## [28] "AveExpr.s4"           "P.Value.s5"           "adj.P.Val.s5"
## [31] "AveExpr.s5"           "P.Value.s6"           "adj.P.Val.s6"
## [34] "AveExpr.s6"
```

```
all_sig = sig$gene_name[rowSums(sig[startsWith(colnames(sig), "adj.P.Val.s")] < 0.05) == 6]
length(all_sig)
```

```
## [1] 853
```

```
any_sig = sig$gene_name[sig$all.donors.adj.P.Val < 0.01]
length(any_sig)
```

```
## [1] 7554
```

```
summary(reg_rnasplice %in% all_sig) #NONE are sig in all donors
```

```
##    Mode   FALSE
## logical     181
```

```
summary(reg_rnasplice %in% any_sig) #only 52 are significant in any donor
```

```
##    Mode   FALSE    TRUE
## logical     129      52
```

```
52/157 #only 33% of expressed factors are DE in any donor
```

```
## [1] 0.3312102
```

```
summary(spliceaid %in% all_sig) #NONE are sig in all donors
```

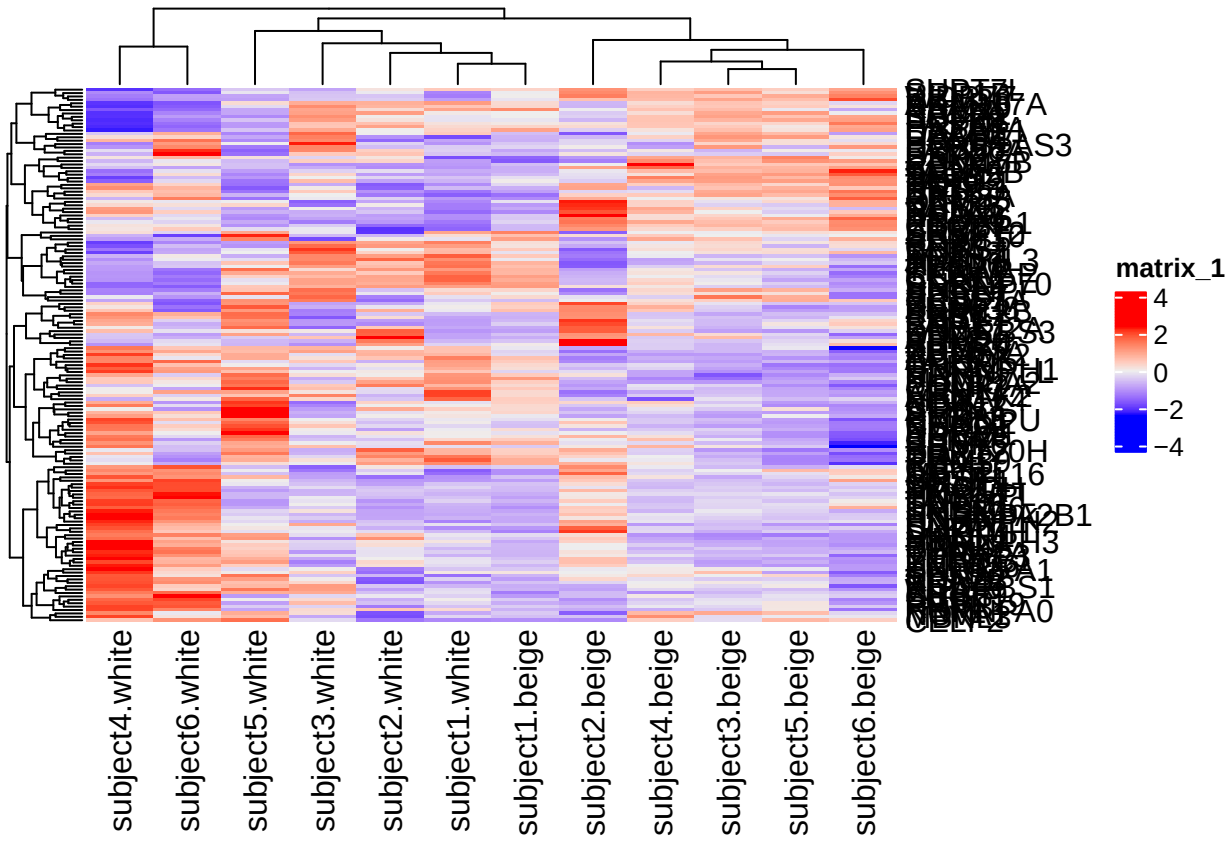
```
##    Mode   FALSE
## logical      68
```

```
summary(spliceaid %in% any_sig) #only 17 are significant in any donor
```

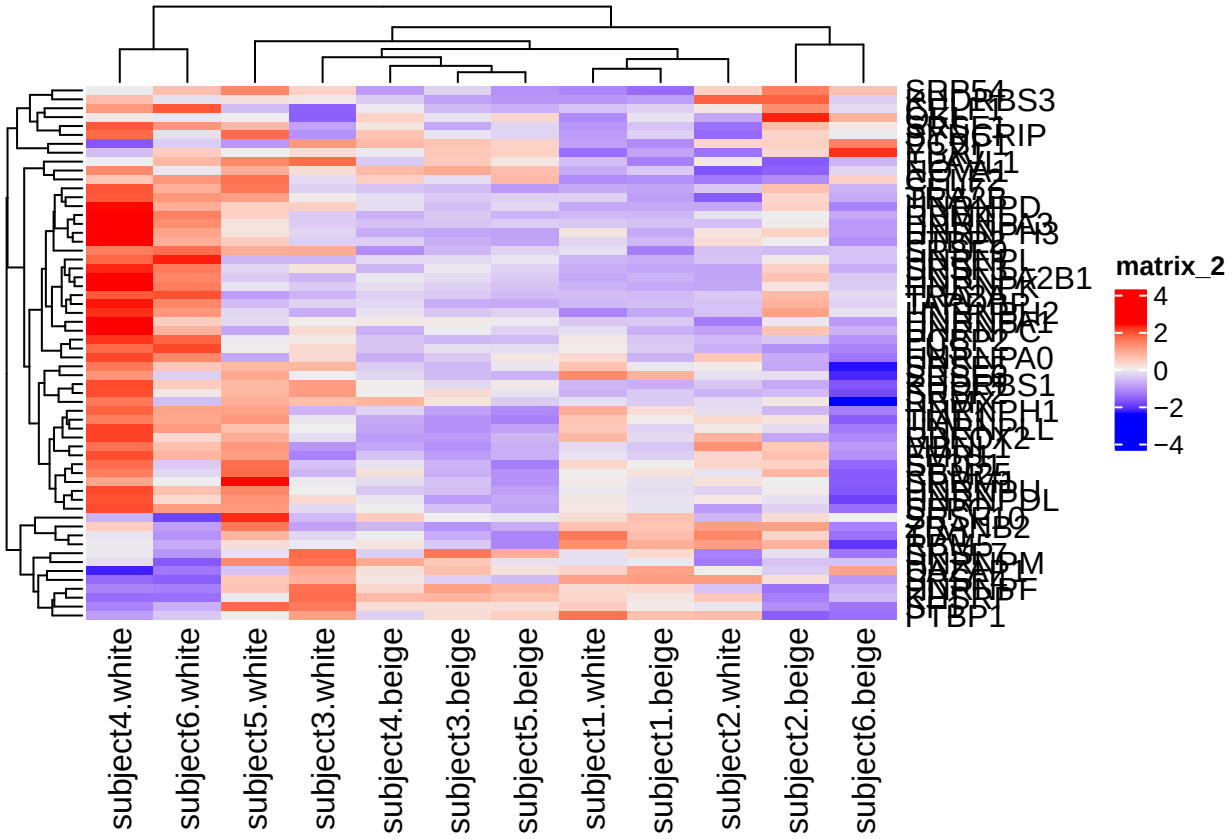
```
##    Mode   FALSE    TRUE
## logical      51      17
```

Heatmaps

```
Heatmap(norm[rownames(norm) %in% reg_rnasplice,])
```




```
Heatmap(norm[rownames(norm) %in% spliceaid,])
```

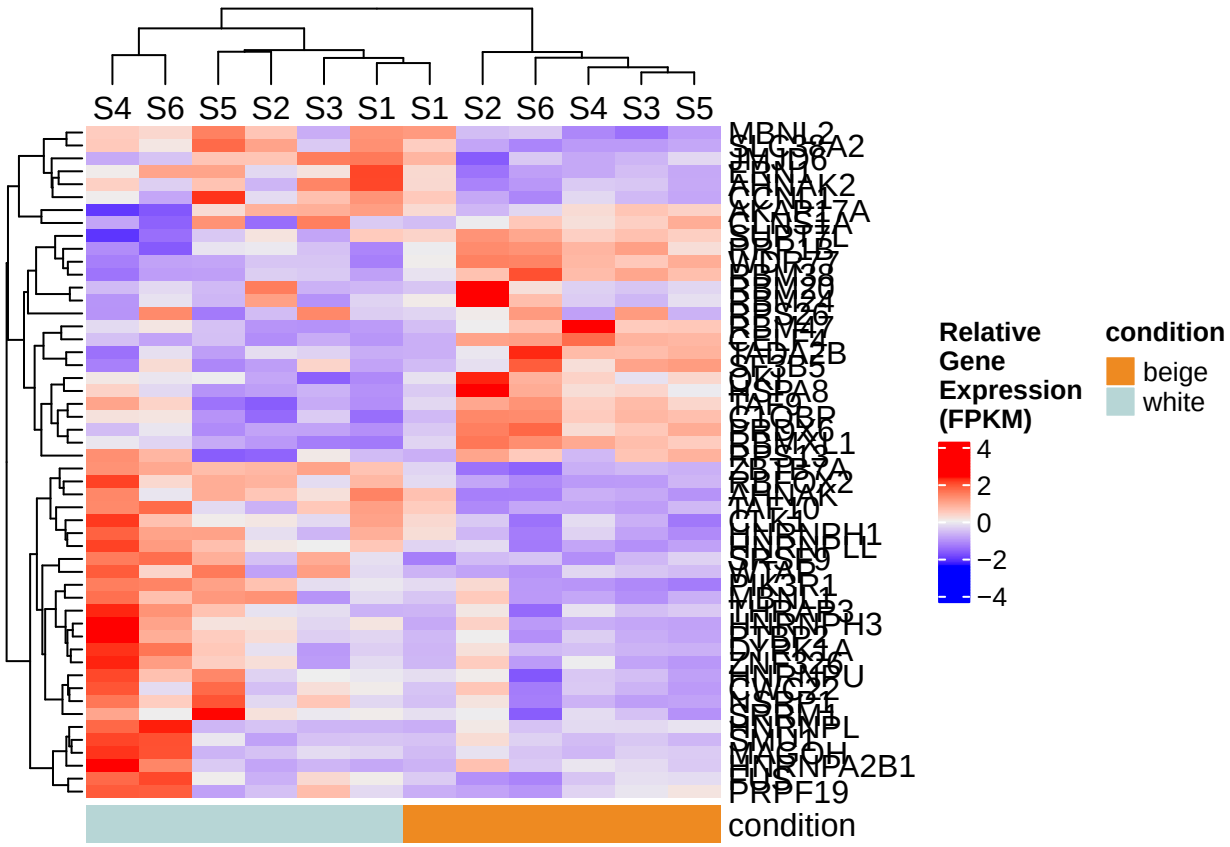



```
treat = gsub(".*\\.","",colnames(norm))
treat
```

```
##  [1] "beige" "white" "beige" "white" "beige" "white" "beige" "white" "beige"
## [10] "white" "beige" "white"
```

```
treatann = HeatmapAnnotation(condition=treat,
                             col=list(condition =c(white="#B7D6D6",beige="#EE8A21")))

donors = gsub("\\..*","",colnames(norm))
hm = Heatmap(norm[rownames(norm) %in% reg_rnasplice &
               rownames(norm) %in% any_sig,], name="Relative \nGene \nExpression \n(FPKM)",
             bottom_annotation = treatann, column_title_side ="top", ,
        column_labels = gsub("subject","S",donors), column_names_side = "top",
        column_names_rot=0, column_names_centered = T)
hm
```




```
pdf(here("R/plots", "reg_rnasplice_factors_any_sig.pdf"), width=7, height=8.5)
hm
dev.off
```

```
## function (which = dev.cur())
## {
##     if (which == 1)
##         stop("cannot shut down device 1 (the null device)")
##     .External(C_devoff, as.integer(which))
##     dev.cur()
## }
## <bytecode: 0x4ee5c58>
## <environment: namespace:grDevices>
```

```
Heatmap(norm[rownames(norm) %in% c(spliceaid, "UCP1") &
               rownames(norm) %in% any_sig,])
```

```
sig[sig$gene_name == "QKI",]
```

```
##               Geneid gene_name                              description Length
## 3380 ENSG00000112531       QKI QKI, KH domain containing RNA binding   17364
##        gene_biotype  logFC.s1 logFC.s2  logFC.s3    logFC.s4  logFC.s5
## 3380 protein_coding 0.4747053 1.092032 0.8504834 -0.05491998 0.2000905
##       logFC.s6  AveExpr        F all.donors.P.Value all.donors.adj.P.Val
```

```
## 3380 0.2602345 8.189493 11.50957       1.794058e-06         1.202765e-05
##      all.donors.AvelogFC P.Value.s1 adj.P.Val.s1 AveExpr.s1   P.Value.s2
## 3380           0.4704376 0.01385392    0.1248653   8.189493 1.549648e-06
##      adj.P.Val.s2 AveExpr.s2   P.Value.s3 adj.P.Val.s3 AveExpr.s3 P.Value.s4
## 3380 2.533872e-05   8.189493 6.981838e-05 0.0009419605   8.189493  0.8176884
##      adj.P.Val.s4 AveExpr.s4 P.Value.s5 adj.P.Val.s5 AveExpr.s5 P.Value.s6
## 3380    0.8836397   8.189493   0.297592    0.4526537   8.189493  0.1906393
##      adj.P.Val.s6 AveExpr.s6
## 3380    0.3173286   8.189493
```