

leafcutter_custom_diffsplice

activate leafcutter module before loading.

```
here::i_am("R/10_leafcutter_custom_diffsplice.Rmd")
```

```
## here() starts at /projects/imb-pkbphil/sp/rnaseq/six_donor_trans/splicing_paper
```

```
library(here)
library(tidyr, lib.loc=.libPaths()[1])
library(leafcutter, lib.loc=.libPaths()[1])
```

```
## Loading required package: Rcpp
```

```
dir = "/home/sarahhp/home/rnaseq/six_donor_trans/output/31_leafcutter"
exon_file = here("annotations/hg38_ensembl98.exons.txt.gz")
```

```
counts=read.table(here("31_leafcutter/beige_splicing_perind_numers.counts.gz"), header=T, check.names =
meta=read.table(here("sample_info/beige_splicing_groups.txt"), header=F, stringsAsFactors = F)
colnames(meta)=c("sample","group","donor")
str(meta)
```

```
## 'data.frame': 36 obs. of 3 variables:
## $ sample: chr "1-22589_S146_Aligned.out" "2-22590_S149_Aligned.out" "3-22591_S154_Aligned.out" "4-
## $ group : chr "white" "beige" "beige" "white" ...
## $ donor : chr "subject3" "subject3" "subject4" "subject2" ...
```

Set up options

```
opt = list(
  min_samples_per_intron = 14,  #(5 donors 2-3 reps)
  min_coverage = 20,
  min_samples_per_group = 12,  #(4 donors x 3reps),
  num_threads = 12,
  timeout=30,
  max_cluster_size=10000,
  output_prefix = here("31_leafcutter/with_donor_info")
)
```

```
counts=counts[,meta$sample]
```

```
group_names=unique(meta$group) # keep order from groups_file unless numeric
if (is.numeric(meta$group)) group_names=sort(group_names)
```

```
meta$group=factor(meta$group, group_names)

stopifnot(length(group_names)==2)

cat("Encoding as",group_names[1],"=0,",group_names[2],"=1\n")
```

```
## Encoding as white =0, beige =1
```

```
numeric_x=as.numeric(meta$group)-1

confounders=meta$donor
confounders=model.matrix( ~0+confounders )
confounders
```

```
##      confounderssubject1 confounderssubject2 confounderssubject3
## 1              0              0              1
## 2              0              0              1
## 3              0              0              0
## 4              0              1              0
## 5              0              1              0
## 6              1              0              0
## 7              1              0              0
## 8              0              0              0
## 9              0              0              0
## 10             0              0              0
## 11             0              0              1
## 12             0              0              1
## 13             0              0              0
## 14             0              1              0
## 15             0              1              0
## 16             1              0              0
## 17             1              0              0
## 18             0              0              0
## 19             0              0              0
## 20             0              0              0
## 21             0              0              1
## 22             0              0              1
## 23             0              0              0
## 24             0              1              0
## 25             0              1              0
## 26             1              0              0
## 27             1              0              0
## 28             0              0              0
## 29             0              0              0
## 30             0              0              0
## 31             0              0              0
## 32             0              0              0
## 33             0              0              0
## 34             0              0              0
## 35             0              0              0
## 36             0              0              0
##      confounderssubject4 confounderssubject5 confounderssubject6
## 1              0              0              0
```

```

## 2          0          0          0
## 3          1          0          0
## 4          0          0          0
## 5          0          0          0
## 6          0          0          0
## 7          0          0          0
## 8          0          0          1
## 9          0          1          0
## 10         0          1          0
## 11         0          0          0
## 12         0          0          0
## 13         1          0          0
## 14         0          0          0
## 15         0          0          0
## 16         0          0          0
## 17         0          0          0
## 18         0          0          1
## 19         0          1          0
## 20         0          1          0
## 21         0          0          0
## 22         0          0          0
## 23         1          0          0
## 24         0          0          0
## 25         0          0          0
## 26         0          0          0
## 27         0          0          0
## 28         0          0          1
## 29         0          1          0
## 30         0          1          0
## 31         1          0          0
## 32         1          0          0
## 33         1          0          0
## 34         0          0          1
## 35         0          0          1
## 36         0          0          1
## attr("assign")
## [1] 1 1 1 1 1 1
## attr("contrasts")
## attr("contrasts")$confounders
## [1] "contr.treatment"

```

```

minimum_group_size=min(sum(numeric_x==0),sum(numeric_x==1))
if (minimum_group_size < opt$min_samples_per_intron)
  stop("The number of samples in the smallest group is less than min_samples_per_intron, which means no")
if (minimum_group_size < opt$min_samples_per_group)
  stop("The number of samples in the smallest group is less than min_samples_per_group, which means no")

```

```

cat("Settings:\n")

```

```

## Settings:

```

```

print(opt)

## $min_samples_per_intron
## [1] 14
##
## $min_coverage
## [1] 20
##
## $min_samples_per_group
## [1] 12
##
## $num_threads
## [1] 12
##
## $timeout
## [1] 30
##
## $max_cluster_size
## [1] 10000
##
## $output_prefix
## [1] "/projects/imb-pkbphil/sp/rnaseq/six_donor_trans/splicing_paper/31_leafcutter/with_donor_info"

cat("Running differential splicing analysis...\n")

## Running differential splicing analysis...

require(doMC)

## Loading required package: doMC

## Loading required package: foreach

## Loading required package: iterators

## Loading required package: parallel

registerDoMC(opt$num_threads)

results <- differential_splicing(counts, numeric_x, confounders=confounders,
                                max_cluster_size=opt$max_cluster_size,
                                min_samples_per_intron=opt$min_samples_per_intron,
                                min_samples_per_group=opt$min_samples_per_group,
                                min_coverage=opt$min_coverage,
                                timeout=opt$timeout,)

## Differential splicing summary:
##
##          statuses  Freq
## 1      <=1 sample with coverage>min_coverage 2197
## 2 <2 introns used in >=min_samples_per_intron samples 1597
## 3              Not enough valid samples 5456
## 4              Success 31987
## 5              timeout      1

```

```

# Make cluster table
clusters          = cluster_results_table(results)
clusters$cluster  = add_chr(clusters$cluster)

# Add gene names to clusters
exons_table       = read.table(exon_file, header=T, stringsAsFactors = F)
intron_meta       = get_intron_meta(rownames(counts))
exons_table$chr    = add_chr(exons_table$chr)
intron_meta$chr    = add_chr(intron_meta$chr)
clu_gene_map      = map_clusters_to_genes(intron_meta, exons_table)
clusters          = merge(clusters, clu_gene_map, by.x="cluster", by.y="clu", all.x=TRUE)

```

Save raw files (machine useful)

```
write.table(clusters, paste0(opt$output_prefix, "_cluster_significance.txt"), quote=F, sep="\t", row.names=FALSE)
```

Format clusters

```

clusters = clusters[order(clusters$p.adjust),]
clusters$chr = gsub(".*:", "", clusters$cluster)
clusters$cluster_id = gsub(".*:", "", clusters$cluster)
head(clusters)

```

```

##              cluster status   loglr df          p      p.adjust
## 34962 chr7:clu_35616_- Success 344.6131 27 6.853680e-128 2.192287e-123
## 16237 chr17:clu_19605_- Success 260.3903  7 2.726265e-108 4.360252e-104
## 6463  chr11:clu_1700_- Success 279.2982 17 8.173667e-108 8.715036e-104
## 40089 chrX:clu_15162_+ Success 264.4180 14 7.096651e-104 5.675014e-100
## 27027 chr3:clu_18227_+ Success 230.6475 10 8.132307e-93  5.202562e-89
## 27798 chr3:clu_18998_+ Success 230.5377 15 9.479469e-89  5.053663e-85
##              genes  chr cluster_id
## 34962      COA1  chr7  clu_35616_-
## 16237      PEMT chr17  clu_19605_-
## 6463        PC chr11  clu_1700_-
## 40089 CA5BP1,CA5B chrX  clu_15162_+
## 27027      PPARG chr3  clu_18227_+
## 27798        MME chr3  clu_18998_+

```

```
table(table(clusters$cluster_id))
```

```

##
##      1
## 41238

```

PEMT cluster chr17:clu_19605_- is highly DE PEMT contains three expressed clusters. The top cluster contains 8 different introns; probably this AF event. Not a bad way of looking at it, as a event type.

```
clusters[grep("PEMT", clusters$genes),]
```

```
##               cluster status      loglr df                p      p.adjust genes
## 16237 chr17:clu_19605_- Success 260.390348 7 2.726265e-108 4.360252e-104 PEMT
## 16235 chr17:clu_19603_- Success  2.764428 2 6.301215e-02  1.846905e-01 PEMT
## 16236 chr17:clu_19604_- Success  5.901093 6 6.653012e-02  1.914447e-01 PEMT
##           chr cluster_id
## 16237 chr17 clu_19605_-
## 16235 chr17 clu_19603_-
## 16236 chr17 clu_19604_-
```

Get effect sizes

```
registerDoMC(opt$num_threads)
effect_size = leaf_cutter_effect_sizes(results)
colnames(effect_size)[3:4] = group_names
effect_size$intron = add_chr(effect_size$intron)
```

```
write.table(effect_size, paste0(opt$output_prefix, "_effect_sizes.txt"), quote=F, col.names = T, row.names = F)
```

Only 2 of these PEMT introns has a large delta PSI (>0.1); chr17:17577027:17591531 which is ensembl 201/202. And our novel intron chr17:17577027:17577107:clu_19605_- has a positive 0.20 PSI.

Do the delta PSIs add to 1?

PPARG has many introns in this Diffspliced cluster; but chr3:12289134:12312380:clu_18227_+ decreases in beige by 16% and chr3:12351674:12379704:clu_18227_+ increases by 18% in beige

```
effect_size = effect_size[order(effect_size$deltapsi),]
effect_size = as.data.frame(effect_size)
head(effect_size, n=6)
```

```
##                                     intron
## chr3:157428488:157436883:clu_4326_- chr3:157428488:157436883:clu_4326_-
## chrX:72330076:72351733:clu_291_- chrX:72330076:72351733:clu_291_-
## chr17:17577027:17591531:clu_19605_- chr17:17577027:17591531:clu_19605_-
## chr18:5489194:5543913:clu_21093_- chr18:5489194:5543913:clu_21093_-
## chr17:28728860:28733983:clu_34526_+ chr17:28728860:28733983:clu_34526_+
## chr11:111844723:111853380:clu_2011_- chr11:111844723:111853380:clu_2011_-
##           logef      white      beige      deltappsi
## chr3:157428488:157436883:clu_4326_- -2.348086 0.7218727 0.1548895 -0.5669832
## chrX:72330076:72351733:clu_291_- -1.428589 0.6534584 0.1935577 -0.4599006
## chr17:17577027:17591531:clu_19605_- -2.220196 0.8095034 0.3798527 -0.4296507
## chr18:5489194:5543913:clu_21093_- -1.560274 0.8645888 0.4376253 -0.4269635
## chr17:28728860:28733983:clu_34526_+ -1.615703 0.7648865 0.3479208 -0.4169657
## chr11:111844723:111853380:clu_2011_- -1.598078 0.5727442 0.1891304 -0.3836138
```

```
effect_size[grep("clu_19605_-", effect_size$intron),]#PEMT
```

```
##                                     intron
```

```
## chr17:17577027:17591531:clu_19605_- chr17:17577027:17591531:clu_19605_-
## chr17:17577027:17591597:clu_19605_- chr17:17577027:17591597:clu_19605_-
## chr17:17577027:17577201:clu_19605_- chr17:17577027:17577201:clu_19605_-
## chr17:17577027:17577122:clu_19605_- chr17:17577027:17577122:clu_19605_-
## chr17:17577027:17591967:clu_19605_- chr17:17577027:17591967:clu_19605_-
## chr17:17577027:17577414:clu_19605_- chr17:17577027:17577414:clu_19605_-
## chr17:17577027:17582267:clu_19605_- chr17:17577027:17582267:clu_19605_-
## chr17:17577027:17577107:clu_19605_- chr17:17577027:17577107:clu_19605_-
##
## logef white beige
## chr17:17577027:17591531:clu_19605_- -2.2201955 0.8095034006 0.37985273
## chr17:17577027:17591597:clu_19605_- -1.4313947 0.0772095282 0.07973322
## chr17:17577027:17577201:clu_19605_- -0.0524112 0.0054240446 0.02224214
## chr17:17577027:17577122:clu_19605_- -0.2190489 0.0102483794 0.03557449
## chr17:17577027:17591967:clu_19605_- -0.3871993 0.0218764959 0.06418514
## chr17:17577027:17577414:clu_19605_- 1.3170798 0.0046936948 0.07570605
## chr17:17577027:17582267:clu_19605_- 3.1400782 0.0008078418 0.08066036
## chr17:17577027:17577107:clu_19605_- -0.1469083 0.0702366148 0.26204587
##
## deltapsi
## chr17:17577027:17591531:clu_19605_- -0.42965067
## chr17:17577027:17591597:clu_19605_- 0.00252369
## chr17:17577027:17577201:clu_19605_- 0.01681810
## chr17:17577027:17577122:clu_19605_- 0.02532611
## chr17:17577027:17591967:clu_19605_- 0.04230864
## chr17:17577027:17577414:clu_19605_- 0.07101235
## chr17:17577027:17582267:clu_19605_- 0.07985252
## chr17:17577027:17577107:clu_19605_- 0.19180926
```

```
pemt = effect_size[grep("clu_19605_-", effect_size$intron),]
sum(pemt["deltapsi"]) #yep almost zero
```

```
## [1] -1.040834e-16
```

```
effect_size[grep("clu_18227_+", effect_size$intron),] #PPARG
```

```
##
## intron
## chr3:12289134:12312380:clu_18227_+ chr3:12289134:12312380:clu_18227_+
## chr3:12312453:12379704:clu_18227_+ chr3:12312453:12379704:clu_18227_+
## chr3:12289134:12379704:clu_18227_+ chr3:12289134:12379704:clu_18227_+
## chr3:12312453:12344807:clu_18227_+ chr3:12312453:12344807:clu_18227_+
## chr3:12344894:12379704:clu_18227_+ chr3:12344894:12379704:clu_18227_+
## chr3:12287675:12312380:clu_18227_+ chr3:12287675:12312380:clu_18227_+
## chr3:12301849:12312380:clu_18227_+ chr3:12301849:12312380:clu_18227_+
## chr3:12287675:12379704:clu_18227_+ chr3:12287675:12379704:clu_18227_+
## chr3:12288020:12379704:clu_18227_+ chr3:12288020:12379704:clu_18227_+
## chr3:12288020:12312380:clu_18227_+ chr3:12288020:12312380:clu_18227_+
## chr3:12351674:12379704:clu_18227_+ chr3:12351674:12379704:clu_18227_+
##
## logef white beige
## chr3:12289134:12312380:clu_18227_+ -1.08362348 0.266987890 0.104052904
## chr3:12312453:12379704:clu_18227_+ -0.32988833 0.430996417 0.356926934
## chr3:12289134:12379704:clu_18227_+ -0.84274297 0.040588221 0.020126846
## chr3:12312453:12344807:clu_18227_+ -0.08899966 0.008405269 0.008856740
## chr3:12344894:12379704:clu_18227_+ -0.05300910 0.007945375 0.008678950
## chr3:12287675:12312380:clu_18227_+ 0.39603949 0.004039287 0.006913163
```

```
## chr3:12301849:12312380:clu_18227_+ 0.27977262 0.006049529 0.009217212
## chr3:12287675:12379704:clu_18227_+ 0.37948947 0.006436726 0.010835515
## chr3:12288020:12379704:clu_18227_+ 0.09113406 0.060627193 0.076492846
## chr3:12288020:12312380:clu_18227_+ 0.16046767 0.085725666 0.115924553
## chr3:12351674:12379704:clu_18227_+ 1.09136024 0.082198428 0.281974336
##
##                                deltapsi
## chr3:12289134:12312380:clu_18227_+ -0.162934986
## chr3:12312453:12379704:clu_18227_+ -0.074069484
## chr3:12289134:12379704:clu_18227_+ -0.020461374
## chr3:12312453:12344807:clu_18227_+ 0.000451471
## chr3:12344894:12379704:clu_18227_+ 0.000733575
## chr3:12287675:12312380:clu_18227_+ 0.002873876
## chr3:12301849:12312380:clu_18227_+ 0.003167684
## chr3:12287675:12379704:clu_18227_+ 0.004398789
## chr3:12288020:12379704:clu_18227_+ 0.015865653
## chr3:12288020:12312380:clu_18227_+ 0.030198888
## chr3:12351674:12379704:clu_18227_+ 0.199775908
```

```
effect_size$chr = gsub(".*", "", effect_size$intron)
effect_size$cluster_id = gsub(".*:clu", "clu", effect_size$intron)
effect_size$intron_coords = gsub("chr.?:", "", gsub(":clu_.*", "", effect_size$intron))
head(effect_size)
```

```
##
##                                intron
## chr3:157428488:157436883:clu_4326_- chr3:157428488:157436883:clu_4326_-
## chrX:72330076:72351733:clu_291_- chrX:72330076:72351733:clu_291_-
## chr17:17577027:17591531:clu_19605_- chr17:17577027:17591531:clu_19605_-
## chr18:5489194:5543913:clu_21093_- chr18:5489194:5543913:clu_21093_-
## chr17:28728860:28733983:clu_34526_+ chr17:28728860:28733983:clu_34526_+
## chr11:111844723:111853380:clu_2011_- chr11:111844723:111853380:clu_2011_-
##
##          logef      white      beige      deltapsi
## chr3:157428488:157436883:clu_4326_- -2.348086 0.7218727 0.1548895 -0.5669832
## chrX:72330076:72351733:clu_291_- -1.428589 0.6534584 0.1935577 -0.4599006
## chr17:17577027:17591531:clu_19605_- -2.220196 0.8095034 0.3798527 -0.4296507
## chr18:5489194:5543913:clu_21093_- -1.560274 0.8645888 0.4376253 -0.4269635
## chr17:28728860:28733983:clu_34526_+ -1.615703 0.7648865 0.3479208 -0.4169657
## chr11:111844723:111853380:clu_2011_- -1.598078 0.5727442 0.1891304 -0.3836138
##
##          chr cluster_id      intron_coords
## chr3:157428488:157436883:clu_4326_- chr3 clu_4326_- 157428488:157436883
## chrX:72330076:72351733:clu_291_- chrX clu_291_- 72330076:72351733
## chr17:17577027:17591531:clu_19605_- chr17 clu_19605_- 17577027:17591531
## chr18:5489194:5543913:clu_21093_- chr18 clu_21093_- 5489194:5543913
## chr17:28728860:28733983:clu_34526_+ chr17 clu_34526_+ 28728860:28733983
## chr11:111844723:111853380:clu_2011_- chr11 clu_2011_- 111844723:111853380
```

```
combo = merge(effect_size, clusters, by=c("chr", "cluster_id"))
nrow(combo);nrow(effect_size)
```

```
## [1] 132587
```

```
## [1] 132587
```



```

combo$score = -log10(combo$p.adjust) * abs(combo$deltapsi)
summary(combo$score)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00053 0.00306 0.03863 0.01372 44.40890

```

```

combo = combo[order(combo$score, decreasing = T),]
head(combo)

```

```

##      chr cluster_id      intron      logef
## 52850 chr17 clu_19605_- chr17:17577027:17591531:clu_19605_- -2.2201955
## 131212 chrX  clu_291_-   chrX:72330076:72351733:clu_291_- -1.4285887
## 131210 chrX  clu_291_-   chrX:72302934:72307025:clu_291_-  1.4361095
## 22207 chr11 clu_2011_- chr11:111844723:111853380:clu_2011_- -1.5980779
## 70171  chr2 clu_11888_- chr2:54987698:55010087:clu_11888_-  1.0163245
## 52844 chr17 clu_19605_- chr17:17577027:17577107:clu_19605_- -0.1469083
##      white      beige      deltappsi      intron_coords      cluster
## 52850 0.80950340 0.3798527 -0.4296507 17577027:17591531 chr17:clu_19605_-
## 131212 0.65345837 0.1935577 -0.4599006 72330076:72351733 chrX:clu_291_-
## 131210 0.09086171 0.4721675 0.3813058 72302934:72307025 chrX:clu_291_-
## 22207 0.57274421 0.1891304 -0.3836138 111844723:111853380 chr11:clu_2011_-
## 70171 0.24681456 0.5528444 0.3060298 54987698:55010087 chr2:clu_11888_-
## 52844 0.07023661 0.2620459 0.1918093 17577027:17577107 chr17:clu_19605_-
##      status      loglr df      p      p.adjust      genes
## 52850 Success 260.3903 7 2.726265e-108 4.360252e-104 PEMT
## 131212 Success 217.4563 12 1.505375e-85 6.878918e-82 CITED1,AL133500.1,HDAC8
## 131210 Success 217.4563 12 1.505375e-85 6.878918e-82 CITED1,AL133500.1,HDAC8
## 22207 Success 156.8379 5 1.147804e-65 2.039712e-62 ALG9,AP001781.2
## 70171 Success 165.0909 5 3.227202e-69 7.940654e-66 RTN4
## 52844 Success 260.3903 7 2.726265e-108 4.360252e-104 PEMT
##      score
## 52850 44.40890
## 131212 37.32668
## 131210 30.94772
## 22207 23.66530
## 70171 19.92259
## 52844 19.82550

```

Save knitted file (human-useful)

```

write.table(combo, paste0(opt$output_prefix, "_leafcutter.txt"), quote=F, col.names = T, row.names = F,

```

```

summary(combo$p.adjust < 0.05) #36 thousand exon excision events?!

```

```

##      Mode  FALSE  TRUE
## logical  93573 39014

```

```

summary(combo$p.adjust < 0.01)

```

```
##      Mode    FALSE    TRUE
## logical 106239  26348
```

```
summary(combo$p.adjust < 0.05 & abs(combo$deltapsi) > 0.1) #with a delta psi threshold though, we have on
```

```
##      Mode    FALSE    TRUE
## logical 131810    777
```

```
summary(combo$p.adjust < 0.01 & abs(combo$deltapsi) > 0.1)
```

```
##      Mode    FALSE    TRUE
## logical 131913    674
```

```
robust = combo[combo$p.adjust < 0.05 & abs(combo$deltapsi) > 0.1,]
head(robust); nrow(robust); tail(robust, n=10)
```

```
##      chr cluster_id      intron      logef
## 52850 chr17 clu_19605_- chr17:17577027:17591531:clu_19605_- -2.2201955
## 131212 chrX  clu_291_-   chrX:72330076:72351733:clu_291_- -1.4285887
## 131210 chrX  clu_291_-   chrX:72302934:72307025:clu_291_-  1.4361095
## 22207  chr11 clu_2011_- chr11:111844723:111853380:clu_2011_- -1.5980779
## 70171  chr2  clu_11888_- chr2:54987698:55010087:clu_11888_-  1.0163245
## 52844  chr17 clu_19605_- chr17:17577027:17577107:clu_19605_- -0.1469083
##
##      white      beige      deltappsi      intron_coords      cluster
## 52850 0.80950340 0.3798527 -0.4296507 17577027:17591531 chr17:clu_19605_-
## 131212 0.65345837 0.1935577 -0.4599006 72330076:72351733 chrX:clu_291_-
## 131210 0.09086171 0.4721675  0.3813058 72302934:72307025 chrX:clu_291_-
## 22207  0.57274421 0.1891304 -0.3836138 111844723:111853380 chr11:clu_2011_-
## 70171  0.24681456 0.5528444  0.3060298 54987698:55010087 chr2:clu_11888_-
## 52844  0.07023661 0.2620459  0.1918093 17577027:17577107 chr17:clu_19605_-
##
##      status      loglr df      p      p.adjust      genes
## 52850 Success 260.3903 7 2.726265e-108 4.360252e-104 PEMT
## 131212 Success 217.4563 12 1.505375e-85 6.878918e-82 CITED1,AL133500.1,HDAC8
## 131210 Success 217.4563 12 1.505375e-85 6.878918e-82 CITED1,AL133500.1,HDAC8
## 22207  Success 156.8379 5 1.147804e-65 2.039712e-62 ALG9,AP001781.2
## 70171  Success 165.0909 5 3.227202e-69 7.940654e-66 RTN4
## 52844  Success 260.3903 7 2.726265e-108 4.360252e-104 PEMT
##
##      score
## 52850 44.40890
## 131212 37.32668
## 131210 30.94772
## 22207 23.66530
## 70171 19.92259
## 52844 19.82550
```

```
## [1] 777
```

```
##      chr cluster_id      intron      logef
## 30256 chr12 clu_24959_- chr12:3713317:3732942:clu_24959_- -0.2910272
## 22843 chr11 clu_38251_+ chr11:819905:821628:clu_38251_+ 0.2739194
## 75478 chr2  clu_31501_+ chr2:96015325:96019201:clu_31501_+ -0.3037609
```

```
## 30255 chr12 clu_24959_- chr12:3733210:3753015:clu_24959_- 0.2051720
## 1064 chr1 clu_13371_+ chr1:27963693:27966503:clu_13371_+ 0.4575920
## 104902 chr5 clu_6087_- chr5:132826182:132830594:clu_6087_- 0.5036644
## 20424 chr11 clu_1475_- chr11:61333008:61333145:clu_1475_- 0.4968729
## 20423 chr11 clu_1475_- chr11:61333008:61335311:clu_1475_- -0.4968729
## 104556 chr5 clu_5967_- chr5:96763228:96776404:clu_5967_- -0.3271111
## 82050 chr20 clu_6946_+ chr20:64255870:64259942:clu_6946_+ 0.3218514
##          white      beige      deltapsi      intron_coords      cluster
## 30256 0.4913872 0.37286368 -0.1185235 3713317:3732942 chr12:clu_24959_-
## 22843 0.4927449 0.60393986 0.1111950 819905:821628 chr11:clu_38251_+
## 75478 0.6489860 0.54099681 -0.1079892 96015325:96019201 chr2:clu_31501_+
## 30255 0.4604723 0.57388698 0.1134146 3733210:3753015 chr12:clu_24959_-
## 1064 0.5526962 0.66019953 0.1075033 27963693:27966503 chr1:clu_13371_+
## 104902 0.2063283 0.31147687 0.1051485 132826182:132830594 chr5:clu_6087_-
## 20424 0.8218989 0.92573936 0.1038405 61333008:61333145 chr11:clu_1475_-
## 20423 0.1781011 0.07426064 -0.1038405 61333008:61335311 chr11:clu_1475_-
## 104556 0.7643356 0.66383467 -0.1005009 96763228:96776404 chr5:clu_5967_-
## 82050 0.5907431 0.69115714 0.1004140 64255870:64259942 chr20:clu_6946_+
##          status      loglr df      p      p.adjust      genes      score
## 30256 Success 4.667763 2 0.009393262 0.04731689 CRACR2A 0.1570418
## 22843 Success 4.841984 2 0.007891378 0.04157824 PNPLA2 0.1535752
## 75478 Success 7.953900 5 0.007112298 0.03862497 FAHD2CP 0.1526029
## 30255 Success 4.667763 2 0.009393262 0.04731689 CRACR2A 0.1502726
## 1064 Success 9.584257 7 0.007675524 0.04079711 XKR8 0.1493619
## 104902 Success 11.143438 9 0.008012917 0.04202478 SHROOM1 0.1447364
## 20424 Success 3.544716 1 0.007753979 0.04106030 DDB1 0.1439829
## 20423 Success 3.544716 1 0.007753979 0.04106030 DDB1 0.1439829
## 104556 Success 5.993976 3 0.007424549 0.03984491 ERAP1 0.1406638
## 82050 Success 6.625853 4 0.010109202 0.04991711 PCMTD2 0.1307140
```

```
robust = robust[order(robust$genes),c("genes","deltapsi","p.adjust","intron")]
head(robust)
```

```
##          genes      deltapsi      p.adjust      intron
## 113696 AC002074.1 -0.1231060 3.610163e-03 chr7:94391533:94392118:clu_35879_-
## 114171 AC002467.1 0.1294696 3.478625e-05 chr7:107743403:107743933:clu_36052_-
## 118113 AC004889.1 -0.1414560 2.368093e-11 chr7:144195701:144207693:clu_41123_+
## 113071 AC006001.3 0.1023289 1.137120e-03 chr7:66531399:66531791:clu_35714_-
## 113070 AC006001.3 -0.1023289 1.137120e-03 chr7:66530137:66531791:clu_35714_-
## 104258 AC008771.1 0.1882644 2.703407e-04 chr5:80485561:80487778:clu_5896_+
```

```
summary(is.na(robust$genes)) #6 exons unable to be assigned to genes
```

```
##      Mode  FALSE    TRUE
## logical    764     13
```

```
robust[grepl("PEMT", robust$genes),]
```

```
##          genes      deltapsi      p.adjust      intron
## 52850 PEMT -0.4296507 4.360252e-104 chr17:17577027:17591531:clu_19605_-
## 52844 PEMT 0.1918093 4.360252e-104 chr17:17577027:17577107:clu_19605_-
```

Many of these events only involve one intron; Many involve 2 introns

```
table(table(robust$genes))
```

```
##
##      1      2      3      4      5      6
## 219 229   13      8      2      1
```

```
table(robust$genes)[table(robust$genes) > 2]
```

```
##
##      ALG9,AP001781.2  CITED1,AL133500.1,HDAC8      DIP2C
##              3              3              3
##      DIXDC1              DST      EHBP1
##              4              4              4
##      ENPP2      FCRLB,DUSP12      FHL1
##              3              3              4
##      FMN2      KANK1      KAZN
##              5              4              3
##      MKNK2      MYO6      PEX19,AL139011.2
##              4              3              3
##      PHTF1      PRR5L      RGS3
##              3              6              3
##      SLC25A45      SNHG10      STON1,STON1-GTF2A1L
##              3              4              3
##      SVIL      TTC7B      VEPH1
##              5              3              4
```

Differential expression

```
de_file= here("03limma/any_and_all_donor_DGE.tsv")
sig = read.delim(de_file)
head(sig)
```

```
##      Geneid gene_name
## 1 ENSG00000000003      TSPAN6
## 2 ENSG00000000005      TNMD
## 3 ENSG00000000419      DPM1
## 4 ENSG00000000457      SCYL3
## 5 ENSG00000000460  C1orf112
## 6 ENSG00000000938      FGR
##
##      description Length
## 1      tetraspanin 6    4536
## 2      tenomodulin     1476
## 3 dolichyl-phosphate mannosyltransferase subunit 1, catalytic 1207
## 4      SCY1 like pseudokinase 3    6883
## 5      chromosome 1 open reading frame 112    5970
## 6      FGR proto-oncogene, Src family tyrosine kinase    3382
##      gene_biotype  logFC.s1 logFC.s2 logFC.s3 logFC.s4 logFC.s5
## 1 protein_coding  0.49422154 0.8603088 0.57278654 0.37343465 0.56429720
```

```
## 2 protein_coding 2.38787369 -0.2720670 2.58660597 1.81772689 1.94504039
## 3 protein_coding -0.11790050 0.1194288 -0.06185359 0.06689829 -0.01815705
## 4 protein_coding 0.39439050 0.5484816 0.49896973 0.32546768 0.35682721
## 5 protein_coding -0.05290602 0.3071896 0.13017667 -0.17866146 -0.36021700
## 6 protein_coding 1.91496687 2.3497110 2.88724528 3.24138536 2.88754674
##      logFC.s6 AveExpr      F all.donors.P.Value all.donors.adj.P.Val
## 1  0.6439646 5.212308 31.7140824      2.969288e-11      7.806158e-10
## 2  1.9785258 1.961126 20.9163666      3.809066e-09      5.434087e-08
## 3  0.1710766 5.314282  0.4712428      8.237264e-01      8.380174e-01
## 4  0.3469994 3.626998 10.0207970      6.300732e-06      3.633461e-05
## 5 -0.2082301 1.367959  1.3456734      2.705375e-01      3.177382e-01
## 6  2.7442577 2.951839 95.9233119      2.476214e-17      5.025045e-15
##      all.donors.AveLogFC P.Value.s1 adj.P.Val.s1 AveExpr.s1 P.Value.s2
## 1      0.5848355 7.423497e-05 2.280200e-03  5.212308 1.031683e-08
## 2      1.7406176 4.606466e-06 2.307037e-04  1.961126 5.374467e-01
## 3      0.0265821 4.216597e-01 7.813318e-01  5.314282 4.196868e-01
## 4      0.4118560 5.556707e-03 6.606388e-02  3.626998 3.317889e-04
## 5     -0.0604414 7.868527e-01 9.473852e-01  1.367959 1.314679e-01
## 6      2.6708521 8.493511e-08 8.063249e-06  2.951839 3.346207e-09
##      adj.P.Val.s2 AveExpr.s2 P.Value.s3 adj.P.Val.s3 AveExpr.s3 P.Value.s4
## 1 4.254162e-07  5.212308 1.180377e-05 2.253571e-04  5.212308 1.164889e-02
## 2 6.575098e-01  1.961126 1.956120e-06 5.218110e-05  1.961126 1.864017e-03
## 3 5.530398e-01  5.314282 6.755915e-01 8.196868e-01  5.314282 7.245401e-01
## 4 2.112990e-03  3.626998 8.348144e-04 6.877477e-03  3.626998 6.456487e-02
## 5 2.355830e-01  1.367959 5.128644e-01 7.062247e-01  1.367959 4.826271e-01
## 6 1.798686e-07  2.951839 3.418267e-11 6.015586e-09  2.951839 1.506080e-10
##      adj.P.Val.s4 AveExpr.s4 P.Value.s5 adj.P.Val.s5 AveExpr.s5 P.Value.s6
## 1 4.171991e-02  5.212308 1.989400e-05 3.539955e-04  5.212308 5.072093e-06
## 2 1.155716e-02  1.961126 1.460798e-04 1.739187e-03  1.961126 1.584704e-04
## 3 8.149446e-01  5.314282 9.043584e-01 9.430495e-01  5.314282 2.804584e-01
## 4 1.441980e-01  3.626998 1.359659e-02 5.250544e-02  3.626998 2.076997e-02
## 5 6.174632e-01  1.367959 8.133352e-02 1.835518e-01  1.367959 3.288567e-01
## 6 5.681113e-08  2.951839 6.097364e-11 1.596007e-08  2.951839 2.754188e-10
##      adj.P.Val.s6 AveExpr.s6
## 1 1.086679e-04  5.212308
## 2 1.496932e-03  1.961126
## 3 4.204665e-01  5.314282
## 4 6.094661e-02  3.626998
## 5 4.715745e-01  1.367959
## 6 5.292092e-08  2.951839
```

```
all_sig = sig[rowSums(sig[grep("adj.P.Val.s[1-6]", colnames(sig))] < 0.05 ) == 6,]
nrow(all_sig)
```

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

```
all_sig = all_sig[order(all_sig$all.donors.adj.P.Val),]
head(all_sig)
```

```
##      Geneid  gene_name
## 7739 ENSG00000153283      CD96
## 11938 ENSG00000185640      KRT79
## 5735  ENSG00000135447      PPP1R1A
```

```

## 4442 ENSG00000123360      PDE1B
## 8659 ENSG00000162877      PM20D1
## 14686 ENSG00000234688 ALO49749.1
##
##                                description Length
## 7739                                CD96 molecule      5621
## 11938                               keratin 79        2560
## 5735 protein phosphatase 1 regulatory inhibitor subunit 1A 3984
## 4442                                phosphodiesterase 1B 7367
## 8659                                peptidase M20 domain containing 1 2611
## 14686                               novel transcript    632
##
##      gene_biotype logFC.s1 logFC.s2 logFC.s3 logFC.s4 logFC.s5 logFC.s6
## 7739 protein_coding 4.437363 5.411840 5.161372 5.027284 4.246215 4.016895
## 11938 protein_coding 4.253555 4.667298 5.137902 5.182334 4.960927 4.210080
## 5735 protein_coding 2.815387 3.427829 3.280520 2.936617 3.307616 2.820063
## 4442 protein_coding 4.011275 2.892675 3.894205 3.933886 4.074751 3.982569
## 8659 protein_coding 4.912675 6.027110 5.985825 6.099764 5.615838 6.362162
## 14686 lncRNA 3.795672 5.348771 5.400274 4.312277 4.633162 4.078279
##
##      AveExpr      F all.donors.P.Value all.donors.adj.P.Val
## 7739 1.9579092 466.0315      1.144536e-26      2.067147e-22
## 11938 3.7244987 441.5264      2.413563e-26      2.179568e-22
## 5735 5.1643525 333.1993      1.144380e-24      6.889551e-21
## 4442 5.1094596 326.3615      1.554287e-24      7.017996e-21
## 8659 2.7473056 308.8168      3.321825e-24      1.199910e-20
## 14686 -0.7401688 243.1292      8.491714e-23      2.556148e-19
##
##      all.donors.AveLogFC P.Value.s1 adj.P.Val.s1 AveExpr.s1 P.Value.s2
## 7739      4.716828 1.409533e-18 2.545758e-14 1.9579092 3.544941e-20
## 11938      4.735349 1.029357e-17 6.197073e-14 3.7244987 2.789538e-18
## 5735      3.098005 4.005202e-16 1.808449e-12 5.1643525 3.639862e-18
## 4442      3.798227 9.064438e-18 6.197073e-14 5.1094596 5.699717e-14
## 8659      5.833896 3.777763e-15 1.364604e-11 2.7473056 1.050525e-16
## 14686      4.594739 1.728359e-13 2.401223e-10 -0.7401688 1.339242e-16
##
##      adj.P.Val.s2 AveExpr.s2 P.Value.s3 adj.P.Val.s3 AveExpr.s3 P.Value.s4
## 7739 6.402518e-16 1.9579092 8.443781e-20 1.525031e-15 1.9579092 4.235916e-18
## 11938 2.191318e-14 3.7244987 1.790497e-19 1.616908e-15 3.7244987 1.133118e-17
## 5735 2.191318e-14 5.1643525 1.178200e-17 7.093159e-14 5.1643525 7.703710e-14
## 4442 2.859516e-11 5.1094596 3.147594e-17 1.421217e-13 5.1094596 8.329436e-15
## 8659 3.023505e-13 2.7473056 8.121645e-17 2.004566e-13 2.7473056 9.668098e-16
## 14686 3.023505e-13 -0.7401688 7.558276e-17 2.004566e-13 -0.7401688 4.500483e-13
##
##      adj.P.Val.s4 AveExpr.s4 P.Value.s5 adj.P.Val.s5 AveExpr.s5 P.Value.s6
## 7739 7.650489e-14 1.9579092 3.771704e-17 1.703019e-13 1.9579092 2.095342e-16
## 11938 1.023262e-13 3.7244987 9.551628e-19 1.725120e-14 3.7244987 1.002399e-16
## 5735 1.739209e-10 5.1643525 1.536313e-17 9.504912e-14 5.1643525 2.431191e-15
## 4442 3.008759e-11 5.1094596 1.578802e-17 9.504912e-14 5.1094596 7.220198e-17
## 8659 5.820518e-12 2.7473056 1.058378e-15 3.185893e-12 2.7473056 6.184633e-17
## 14686 4.332568e-10 -0.7401688 9.631028e-15 1.932733e-11 -0.7401688 2.413864e-13
##
##      adj.P.Val.s6 AveExpr.s6
## 7739 9.460994e-13 1.9579092
## 11938 6.034778e-13 3.7244987
## 5735 5.488718e-12 5.1643525
## 4442 6.034778e-13 5.1094596
## 8659 6.034778e-13 2.7473056
## 14686 2.004799e-10 -0.7401688

```

```
robust = robust[order(robust$p.adjust),]
lc_genes = separate_rows(robust, genes, sep=",")
head(lc_genes)
```

```
## # A tibble: 6 x 4
##   genes   deltapsi p.adjust intron
##   <chr>     <dbl>     <dbl> <chr>
## 1 COA1      0.117 2.19e-123 chr7:43749288:43750147:clu_35616_-
## 2 PEMT     -0.430 4.36e-104 chr17:17577027:17591531:clu_19605_-
## 3 PEMT      0.192 4.36e-104 chr17:17577027:17577107:clu_19605_-
## 4 PC        0.159 8.72e-104 chr11:66872159:66907821:clu_1700_-
## 5 CA5BP1   -0.147 5.68e-100 chrX:15675778:15688661:clu_15162_+
## 6 CA5B     -0.147 5.68e-100 chrX:15675778:15688661:clu_15162_+
```

```
nrow(lc_genes)
```

```
## [1] 920
```

```
length(unique(lc_genes$genes))
```

```
## [1] 562
```

```
summary(unique(lc_genes$genes) %in% all_sig$gene_name)
```

```
##      Mode   FALSE    TRUE
## logical    504     58
```

```
lc_genes$is_DEG = lc_genes$genes %in% all_sig$gene_name
head(lc_genes)
```

```
## # A tibble: 6 x 5
##   genes   deltapsi p.adjust intron      is_DEG
##   <chr>     <dbl>     <dbl> <chr>     <lgl>
## 1 COA1      0.117 2.19e-123 chr7:43749288:43750147:clu_35616_- FALSE
## 2 PEMT     -0.430 4.36e-104 chr17:17577027:17591531:clu_19605_- TRUE
## 3 PEMT      0.192 4.36e-104 chr17:17577027:17577107:clu_19605_- TRUE
## 4 PC        0.159 8.72e-104 chr11:66872159:66907821:clu_1700_- FALSE
## 5 CA5BP1   -0.147 5.68e-100 chrX:15675778:15688661:clu_15162_+ FALSE
## 6 CA5B     -0.147 5.68e-100 chrX:15675778:15688661:clu_15162_+ TRUE
```

```
any_sig = sig[sig$adj.P.Val < 0.01,]
nrow(any_sig)
```

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

```
summary(unique(lc_genes$genes) %in% any_sig$gene_name)
```

```
##      Mode   FALSE
## logical    562
```

```
summary(unique(lc_genes$genes) %in% any_sig$gene_name[abs(sig$all.donors.AvelogFC) > 1])
```

```
##      Mode   FALSE    TRUE  
## logical    561      1
```

```
summary(unique(lc_genes$genes) %in% any_sig$gene_name[abs(sig$all.donors.AvelogFC) > 0.5])
```

```
##      Mode   FALSE    TRUE  
## logical    561      1
```

```
summary(unique(lc_genes$genes) %in% any_sig$gene_name[abs(sig$all.donors.AvelogFC) > 0.1])
```

```
##      Mode   FALSE    TRUE  
## logical    561      1
```