

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
if (ncol(meta)>2) {
  confounders=meta[,3:ncol(meta),drop=F]
  # scale continuous confounders
  for (i in seq_len(ncol(confounders)))
    if (is.numeric(confounders[,i]))
      confounders[,i]=scale(confounders[,i])
  # convert factors to one-of-K encoding
  confounders=model.matrix( ~., data=confounders )
  confounders=confounders[,2:ncol(confounders),drop=F] # remove intercept
}
confounders
```

```
##      donorsubject2 donorsubject3 donorsubject4 donorsubject5 donorsubject6
## 1              0              1              0              0              0
## 2              0              1              0              0              0
## 3              0              0              1              0              0
## 4              1              0              0              0              0
## 5              1              0              0              0              0
## 6              0              0              0              0              0
## 7              0              0              0              0              0
## 8              0              0              0              0              1
## 9              0              0              0              1              0
## 10             0              0              0              1              0
## 11             0              1              0              0              0
## 12             0              1              0              0              0
## 13             0              0              1              0              0
## 14             1              0              0              0              0
## 15             1              0              0              0              0
## 16             0              0              0              0              0
## 17             0              0              0              0              0
## 18             0              0              0              0              1
## 19             0              0              0              1              0
## 20             0              0              0              1              0
## 21             0              1              0              0              0
## 22             0              1              0              0              0
## 23             0              0              1              0              0
## 24             1              0              0              0              0
## 25             1              0              0              0              0
## 26             0              0              0              0              0
## 27             0              0              0              0              0
## 28             0              0              0              0              1
## 29             0              0              0              1              0
```

```
## 30      0      0      0      1      0
## 31      0      0      1      0      0
## 32      0      0      1      0      0
## 33      0      0      1      0      0
## 34      0      0      0      0      1
## 35      0      0      0      0      1
## 36      0      0      0      0      1
```

```
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 322.4896 27 1.216097e-118 3.889929e-114  
## 6463  chr11:clu_1700_- Success 268.8594 17 2.100632e-103 3.359645e-99  
## 40089 chrX:clu_15162_+ Success 262.6567 14 3.968708e-103 4.231569e-99
```

```
## 27027 chr3:clu_18227_+ Success 231.0444 10 5.505743e-93 4.402805e-89
## 27798 chr3:clu_18998_+ Success 230.2288 15 1.279969e-88 8.188475e-85
## 40824 chrX:clu_291_- Success 216.3523 12 4.427017e-85 2.360117e-81
##
##          genes      chr cluster_id
## 34962          COA1   chr7 clu_35616_-
## 6463           PC chr11 clu_1700_-
## 40089      CA5BP1,CA5B chrX clu_15162_+
## 27027          PPARG   chr3 clu_18227_+
## 27798          MME   chr3 clu_18998_+
## 40824 CITED1,AL133500.1,HDAC8 chrX clu_291_-
```

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

```
##          cluster status      loglr df          p      p.adjust genes
## 16237 chr17:clu_19605_- Success 179.099707 7 2.163800e-73 6.921348e-70 PEMT
## 16236 chr17:clu_19604_- Success  5.928540 6 6.523417e-02 1.945227e-01 PEMT
## 16235 chr17:clu_19603_- Success  2.683392 2 6.833098e-02 2.008364e-01 PEMT
##          chr cluster_id
## 16237 chr17 clu_19605_-
## 16236 chr17 clu_19604_-
## 16235 chr17 clu_19603_-
```

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_-
## chrX:136197134:136206407:clu_15610_+     chrX:136197134:136206407:clu_15610_+
## chr18:5489194:5543913:clu_21093_-        chr18:5489194:5543913:clu_21093_-
## chr10:102449750:102450187:clu_29976_+    chr10:102449750:102450187:clu_29976_+
## chr11:111844723:111853380:clu_2011_-     chr11:111844723:111853380:clu_2011_-
##                                           logef      white      beige      deltapsi
## chr3:157428488:157436883:clu_4326_-     -2.347828  0.7407624  0.1781880 -0.5625744
## chrX:72330076:72351733:clu_291_-         -1.281808  0.7810052  0.2949279 -0.4860773
## chrX:136197134:136206407:clu_15610_+    -1.491351  0.7096092  0.2908517 -0.4187574
## chr18:5489194:5543913:clu_21093_-       -1.558692  0.8652771  0.4470775 -0.4181996
## chr10:102449750:102450187:clu_29976_+   -1.054931  0.5404988  0.1248251 -0.4156738
## chr11:111844723:111853380:clu_2011_-    -1.595982  0.6637130  0.2556499 -0.4080631
```

```
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_-
## chr17:17577027:17591967:clu_19605_-     chr17:17577027:17591967:clu_19605_-
## chr17:17577027:17577201:clu_19605_-     chr17:17577027:17577201:clu_19605_-
## chr17:17577027:17577122:clu_19605_-     chr17:17577027:17577122: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_-    -1.727506635  0.8075632767  0.56399898
## chr17:17577027:17591597:clu_19605_-    -1.727506332  0.0869669441  0.06073739
## chr17:17577027:17591967:clu_19605_-    -1.727505980  0.0462264505  0.03228439
## chr17:17577027:17577201:clu_19605_-    -0.000454917  0.0038486891  0.01511693
## chr17:17577027:17577122:clu_19605_-    -0.147398212  0.0053622523  0.01818366
## chr17:17577027:17577414:clu_19605_-     1.633629752  0.0023880001  0.04806834
## chr17:17577027:17582267:clu_19605_-     3.556160713  0.0003464926  0.04769396
## chr17:17577027:17577107:clu_19605_-     0.140581612  0.0472978946  0.21391635
##                                           deltapsi
## chr17:17577027:17591531:clu_19605_-    -0.24356429
## chr17:17577027:17591597:clu_19605_-    -0.02622956
## chr17:17577027:17591967:clu_19605_-    -0.01394206
## chr17:17577027:17577201:clu_19605_-     0.01126824
## chr17:17577027:17577122:clu_19605_-     0.01282141
## chr17:17577027:17577414:clu_19605_-     0.04568034
## chr17:17577027:17582267:clu_19605_-     0.04734746
## chr17:17577027:17577107:clu_19605_-     0.16661846
```

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

```
## [1] -1.214306e-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_+
## chr3:12301849:12312380:clu_18227_+ chr3:12301849:12312380:clu_18227_+
## chr3:12287675:12312380:clu_18227_+ chr3:12287675:12312380: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_+ -1.09350541 0.220931619 0.077888771
## chr3:12312453:12379704:clu_18227_+ -0.33506556 0.433072333 0.325959509
## chr3:12289134:12379704:clu_18227_+ -0.85539700 0.034992383 0.015653075
## chr3:12312453:12344807:clu_18227_+ -0.08279275 0.006436835 0.006234998
## chr3:12344894:12379704:clu_18227_+ -0.04871710 0.006730821 0.006745759
## chr3:12301849:12312380:clu_18227_+ 0.29219678 0.005117662 0.007212592
## chr3:12287675:12312380:clu_18227_+ 0.42415013 0.003578710 0.005755101
## chr3:12287675:12379704:clu_18227_+ 0.37551072 0.005951616 0.009116696
## chr3:12288020:12379704:clu_18227_+ 0.08661448 0.067274177 0.077194245
## chr3:12288020:12312380:clu_18227_+ 0.15408874 0.107834896 0.132373042
## chr3:12351674:12379704:clu_18227_+ 1.08291698 0.108078948 0.335866213
##                                     deltapsi
## chr3:12289134:12312380:clu_18227_+ -1.430428e-01
## chr3:12312453:12379704:clu_18227_+ -1.071128e-01
## chr3:12289134:12379704:clu_18227_+ -1.933931e-02
## chr3:12312453:12344807:clu_18227_+ -2.018375e-04
## chr3:12344894:12379704:clu_18227_+ 1.493785e-05
## chr3:12301849:12312380:clu_18227_+ 2.094930e-03
## chr3:12287675:12312380:clu_18227_+ 2.176391e-03
## chr3:12287675:12379704:clu_18227_+ 3.165080e-03
## chr3:12288020:12379704:clu_18227_+ 9.920068e-03
## chr3:12288020:12312380:clu_18227_+ 2.453815e-02
## chr3:12351674:12379704:clu_18227_+ 2.277873e-01
```

```
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_-
## chrX:136197134:136206407:clu_15610_+ chrX:136197134:136206407:clu_15610_+
## chr18:5489194:5543913:clu_21093_- chr18:5489194:5543913:clu_21093_-
## chr10:102449750:102450187:clu_29976_+ chr10:102449750:102450187:clu_29976_+
## chr11:111844723:111853380:clu_2011_- chr11:111844723:111853380:clu_2011_-
```

```
##               logef      white      beige      deltapsi
## chr3:157428488:157436883:clu_4326_- -2.347828 0.7407624 0.1781880 -0.5625744
## chrX:72330076:72351733:clu_291_- -1.281808 0.7810052 0.2949279 -0.4860773
## chrX:136197134:136206407:clu_15610_+ -1.491351 0.7096092 0.2908517 -0.4187574
## chr18:5489194:5543913:clu_21093_- -1.558692 0.8652771 0.4470775 -0.4181996
## chr10:102449750:102450187:clu_29976_+ -1.054931 0.5404988 0.1248251 -0.4156738
## chr11:111844723:111853380:clu_2011_- -1.595982 0.6637130 0.2556499 -0.4080631
##               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
## chrX:136197134:136206407:clu_15610_+ chrX clu_15610_+ 136197134:136206407
## chr18:5489194:5543913:clu_21093_- chr18 clu_21093_- 5489194:5543913
## chr10:102449750:102450187:clu_29976_+ chr10 clu_29976_+ 102449750:102450187
## 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 0.00033 0.00227 0.03479 0.01143 39.19099
```

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

```
##      chr cluster_id      intron      logef
## 131205 chrX clu_291_- chrX:72330076:72351733:clu_291_- -1.2818076
## 131210 chrX clu_291_- chrX:72302934:72307025:clu_291_- 1.6297925
## 22208 chr11 clu_2011_- chr11:111844723:111853380:clu_2011_- -1.5959820
## 130331 chrX clu_15610_+ chrX:136197134:136206407:clu_15610_+ -1.4913514
## 87511 chr3 clu_18227_+ chr3:12351674:12379704:clu_18227_+ 1.0829170
## 128926 chrX clu_15162_+ chrX:15675778:15688661:clu_15162_+ -0.4496297
##      white      beige      deltapsi      intron_coords      cluster
## 131205 0.78100518 0.2949279 -0.4860773 72330076:72351733 chrX:clu_291_-
## 131210 0.06279804 0.4360134 0.3732153 72302934:72307025 chrX:clu_291_-
## 22208 0.66371298 0.2556499 -0.4080631 111844723:111853380 chr11:clu_2011_-
## 130331 0.70960915 0.2908517 -0.4187574 136197134:136206407 chrX:clu_15610_+
## 87511 0.10807895 0.3358662 0.2277873 12351674:12379704 chr3:clu_18227_+
## 128926 0.46914400 0.2669770 -0.2021670 15675778:15688661 chrX:clu_15162_+
##      status  loglr df      p      p.adjust      genes
## 131205 Success 216.3523 12 4.427017e-85 2.360117e-81 CITED1,AL133500.1,HDAC8
## 131210 Success 216.3523 12 4.427017e-85 2.360117e-81 CITED1,AL133500.1,HDAC8
## 22208 Success 157.0918 5 8.925273e-66 1.679369e-62 ALG9,AP001781.2
## 130331 Success 144.6837 6 1.550467e-59 2.361656e-56 FHL1
## 87511 Success 231.0444 10 5.505743e-93 4.402805e-89 PPARG
```



```
## 128926 Success 262.6567 14 3.968708e-103 4.231569e-99 CA5BP1,CA5B
##          score
## 131205 39.19099
## 131210 30.09126
## 22208 25.20804
## 130331 23.29413
## 87511 20.12643
## 128926 19.88787
```

Save knitted file (human-useful)

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

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

```
##      Mode  FALSE    TRUE
## logical  95491   37096
```

```
summary(combo$p.adjust < 0.01)
```

```
##      Mode  FALSE    TRUE
## logical 107608   24979
```

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

```
##      Mode  FALSE    TRUE
## logical 131894    693
```

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

```
##      Mode  FALSE    TRUE
## logical 131976    611
```

```
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
## 131205 chrX  clu_291_- chrX:72330076:72351733:clu_291_- -1.2818076
## 131210 chrX  clu_291_- chrX:72302934:72307025:clu_291_-  1.6297925
## 22208  chr11 clu_2011_- chr11:111844723:111853380:clu_2011_- -1.5959820
## 130331 chrX  clu_15610_+ chrX:136197134:136206407:clu_15610_+ -1.4913514
## 87511  chr3  clu_18227_+ chr3:12351674:12379704:clu_18227_+  1.0829170
## 128926 chrX  clu_15162_+ chrX:15675778:15688661:clu_15162_+ -0.4496297
##      white      beige      deltappsi      intron_coords      cluster
## 131205 0.78100518 0.2949279 -0.4860773 72330076:72351733 chrX:clu_291_-
## 131210 0.06279804 0.4360134 0.3732153 72302934:72307025 chrX:clu_291_-
## 22208 0.66371298 0.2556499 -0.4080631 111844723:111853380 chr11:clu_2011_-
## 130331 0.70960915 0.2908517 -0.4187574 136197134:136206407 chrX:clu_15610_+
```

```
## 87511 0.10807895 0.3358662 0.2277873 12351674:12379704 chr3:clu_18227_+
## 128926 0.46914400 0.2669770 -0.2021670 15675778:15688661 chrX:clu_15162_+
##      status      loglr df      p      p.adjust      genes
## 131205 Success 216.3523 12 4.427017e-85 2.360117e-81 CITED1,AL133500.1,HDAC8
## 131210 Success 216.3523 12 4.427017e-85 2.360117e-81 CITED1,AL133500.1,HDAC8
## 22208 Success 157.0918 5 8.925273e-66 1.679369e-62 ALG9,AP001781.2
## 130331 Success 144.6837 6 1.550467e-59 2.361656e-56 FHL1
## 87511 Success 231.0444 10 5.505743e-93 4.402805e-89 PPARG
## 128926 Success 262.6567 14 3.968708e-103 4.231569e-99 CA5BP1,CA5B
##      score
## 131205 39.19099
## 131210 30.09126
## 22208 25.20804
## 130331 23.29413
## 87511 20.12643
## 128926 19.88787
```

```
## [1] 693
```

```
##      chr cluster_id      intron      logef
## 65824 chr19 clu_7219_- chr19:8304088:8305031:clu_7219_- -0.3345042
## 23655 chr11 clu_38492_+ chr11:36446440:36451209:clu_38492_+ -0.1610268
## 75483 chr2 clu_31501_+ chr2:96015325:96019201:clu_31501_+ -0.3036990
## 30256 chr12 clu_24959_- chr12:3713317:3732942:clu_24959_- -0.2907672
## 115660 chr7 clu_40416_+ chr7:30227745:30323642:clu_40416_+ 0.4163937
## 115661 chr7 clu_40416_+ chr7:30285826:30323642:clu_40416_+ -0.4163937
## 103689 chr5 clu_5751_- chr5:43015150:43067087:clu_5751_- -0.5312610
## 22845 chr11 clu_38251_+ chr11:819905:821628:clu_38251_+ 0.2646699
## 30221 chr12 clu_24942_- chr12:925525:927145:clu_24942_- 0.3236105
## 30257 chr12 clu_24959_- chr12:3733210:3753015:clu_24959_- 0.2049634
##      white      beige      deltapsi      intron_coords      cluster
## 65824 0.6439680 0.5400651 -0.1039029 8304088:8305031 chr19:clu_7219_-
## 23655 0.4203571 0.3108948 -0.1094623 36446440:36451209 chr11:clu_38492_+
## 75483 0.6009490 0.4906412 -0.1103079 96015325:96019201 chr2:clu_31501_+
## 30256 0.4732403 0.3563892 -0.1168511 3713317:3732942 chr12:clu_24959_-
## 115660 0.1087930 0.2191986 0.1104056 30227745:30323642 chr7:clu_40416_+
## 115661 0.8912070 0.7808014 -0.1104056 30285826:30323642 chr7:clu_40416_+
## 103689 0.7819260 0.6700056 -0.1119204 43015150:43067087 chr5:clu_5751_-
## 22845 0.5648130 0.6678739 0.1030609 819905:821628 chr11:clu_38251_+
## 30221 0.6154741 0.7251624 0.1096883 925525:927145 chr12:clu_24942_-
## 30257 0.4717471 0.5832371 0.1114900 3733210:3753015 chr12:clu_24959_-
##      status      loglr df      p      p.adjust      genes      score
## 65824 Success 5.368848 2 0.004659494 0.02917789 CD320 0.1594854
## 23655 Success 5.059982 2 0.006345673 0.03675824 PRR5L 0.1570396
## 75483 Success 7.951199 5 0.007128327 0.04008682 FAHD2CP 0.1540999
## 30256 Success 4.667336 2 0.009397266 0.04884471 CRACR2A 0.1532131
## 115660 Success 3.579025 1 0.007462830 0.04146553 ZNRF2 0.1526151
## 115661 Success 3.579025 1 0.007462830 0.04146553 ZNRF2 0.1526151
## 103689 Success 7.781685 5 0.008207802 0.04428115 AC025171.1,ANXA2R 0.1515157
## 22845 Success 5.074068 2 0.006256913 0.03639569 PNPLA2 0.1482995
## 30221 Success 4.796174 2 0.008261297 0.04451720 RAD52 0.1482407
## 30257 Success 4.667336 2 0.009397266 0.04884471 CRACR2A 0.1461837
```

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

```
##          genes  deltapsi    p.adjust          intron
## 70550      AAK1 -0.1709020 7.284469e-27 chr2:69466551:69505569:clu_11994_-
## 99636     ABHD18 0.1617273 3.597588e-04 chr4:128011700:128017363:clu_40017_+
## 114170 AC002467.1 0.1446007 3.508544e-05 chr7:107743403:107743933:clu_36052_-
## 118117 AC004889.1 -0.1424899 1.459308e-10 chr7:144195701:144207693:clu_41123_+
## 113070 AC006001.3 0.1074440 1.157585e-03 chr7:66531399:66531791:clu_35714_-
## 113071 AC006001.3 -0.1074440 1.157585e-03 chr7:66530137:66531791:clu_35714_-
```

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

```
##      Mode  FALSE  TRUE
## logical    687    6
```

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

```
##      genes  deltapsi    p.adjust          intron
## 52846  PEMT -0.2435643 6.921348e-70 chr17:17577027:17591531:clu_19605_-
## 52850  PEMT 0.1666185 6.921348e-70 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
## 202 200   14      8      1      1
```

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

```
##
##      ALG9,AP001781.2  CITED1,AL133500.1,HDAC8      CPT1B,CHKB-CPT1B
##              4              3              3
##      DIXDC1              EHBP1      FCRLB,DUSP12
##              4              4              3
##      FHL1              FMN2              KANK1
##              5              3              4
##      LRBA              LYPLAL1-DT      MKNK2
##              4              3              4
##      NAV1      PEX19,AL139011.2      PHTF1
##              3              3              3
##      PLD1              PPARG      SNHG10
##              3              3              4
##      SPTBN1      STON1,STON1-GTF2A1L      SVIL
##              3              3              6
##      TNS3              TTC7B      VEPH1
##              3              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.s4  logFC.s3  logFC.s2  logFC.s1  logFC.s6
## 1 protein_coding  0.37343465  0.57278654  0.8603088  0.49422154  0.6439646
## 2 protein_coding  1.81772689  2.58660597 -0.2720670  2.38787369  1.9785258
## 3 protein_coding  0.06689829 -0.06185359  0.1194288 -0.11790050  0.1710766
## 4 protein_coding  0.32546768  0.49896973  0.5484816  0.39439050  0.3469994
## 5 protein_coding -0.17866146  0.13017667  0.3071896 -0.05290602 -0.2082301
## 6 protein_coding  3.24138536  2.88724528  2.3497110  1.91496687  2.7442577
##      logFC.s5  AveExpr      F all.donors.P.Value all.donors.adj.P.Val
## 1  0.56429720  5.212308  31.7140824      2.969288e-11      7.806158e-10
## 2  1.94504039  1.961126  20.9163666      3.809066e-09      5.434087e-08
## 3 -0.01815705  5.314282  0.4712428      8.237264e-01      8.380174e-01
## 4  0.35682721  3.626998  10.0207970      6.300732e-06      3.633461e-05
## 5 -0.36021700  1.367959  1.3456734      2.705375e-01      3.177382e-01
## 6  2.88754674  2.951839  95.9233119      2.476214e-17      5.025045e-15
##      all.donors.AveLogFC  P.Value.s3  adj.P.Val.s3  AveExpr.s3  P.Value.s4
## 1      0.5848355  1.180377e-05  2.253571e-04  5.212308  1.164889e-02
## 2      1.7406176  1.956120e-06  5.218110e-05  1.961126  1.864017e-03
## 3      0.0265821  6.755915e-01  8.196868e-01  5.314282  7.245401e-01
## 4      0.4118560  8.348144e-04  6.877477e-03  3.626998  6.456487e-02
## 5     -0.0604414  5.128644e-01  7.062247e-01  1.367959  4.826271e-01
## 6      2.6708521  3.418267e-11  6.015586e-09  2.951839  1.506080e-10
##      adj.P.Val.s4  AveExpr.s4  P.Value.s2  adj.P.Val.s2  AveExpr.s2  P.Value.s1
## 1  4.171991e-02  5.212308  1.031683e-08  4.254162e-07  5.212308  7.423497e-05
## 2  1.155716e-02  1.961126  5.374467e-01  6.575098e-01  1.961126  4.606466e-06
## 3  8.149446e-01  5.314282  4.196868e-01  5.530398e-01  5.314282  4.216597e-01
## 4  1.441980e-01  3.626998  3.317889e-04  2.112990e-03  3.626998  5.556707e-03
## 5  6.174632e-01  1.367959  1.314679e-01  2.355830e-01  1.367959  7.868527e-01
## 6  5.681113e-08  2.951839  3.346207e-09  1.798686e-07  2.951839  8.493511e-08
##      adj.P.Val.s1  AveExpr.s1  P.Value.s6  adj.P.Val.s6  AveExpr.s6  P.Value.s5
## 1  2.280200e-03  5.212308  5.072093e-06  1.086679e-04  5.212308  1.989400e-05
## 2  2.307037e-04  1.961126  1.584704e-04  1.496932e-03  1.961126  1.460798e-04
## 3  7.813318e-01  5.314282  2.804584e-01  4.204665e-01  5.314282  9.043584e-01
## 4  6.606388e-02  3.626998  2.076997e-02  6.094661e-02  3.626998  1.359659e-02
```

```
## 5 9.473852e-01 1.367959 3.288567e-01 4.715745e-01 1.367959 8.133352e-02
## 6 8.063249e-06 2.951839 2.754188e-10 5.292092e-08 2.951839 6.097364e-11
## adj.P.Val.s5 AveExpr.s5
## 1 3.539955e-04 5.212308
## 2 1.739187e-03 1.961126
## 3 9.430495e-01 5.314282
## 4 5.250544e-02 3.626998
## 5 1.835518e-01 1.367959
## 6 1.596007e-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 AL049749.1
##
##                                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.s4 logFC.s3 logFC.s2 logFC.s1 logFC.s6 logFC.s5
## 7739  protein_coding 5.027284 5.161372 5.411840 4.437363 4.016895 4.246215
## 11938 protein_coding 5.182334 5.137902 4.667298 4.253555 4.210080 4.960927
## 5735  protein_coding 2.936617 3.280520 3.427829 2.815387 2.820063 3.307616
## 4442  protein_coding 3.933886 3.894205 2.892675 4.011275 3.982569 4.074751
## 8659  protein_coding 6.099764 5.985825 6.027110 4.912675 6.362162 5.615838
## 14686 lncRNA 4.312277 5.400274 5.348771 3.795672 4.078279 4.633162
##
## 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.s3 adj.P.Val.s3 AveExpr.s3 P.Value.s4
## 7739      4.716828 8.443781e-20 1.525031e-15 1.9579092 4.235916e-18
## 11938      4.735349 1.790497e-19 1.616908e-15 3.7244987 1.133118e-17
## 5735      3.098005 1.178200e-17 7.093159e-14 5.1643525 7.703710e-14
## 4442      3.798227 3.147594e-17 1.421217e-13 5.1094596 8.329436e-15
## 8659      5.833896 8.121645e-17 2.004566e-13 2.7473056 9.668098e-16
## 14686      4.594739 7.558276e-17 2.004566e-13 -0.7401688 4.500483e-13
```

```
##      adj.P.Val.s4 AveExpr.s4      P.Value.s2 adj.P.Val.s2 AveExpr.s2      P.Value.s1
## 7739 7.650489e-14 1.9579092 3.544941e-20 6.402518e-16 1.9579092 1.409533e-18
## 11938 1.023262e-13 3.7244987 2.789538e-18 2.191318e-14 3.7244987 1.029357e-17
## 5735 1.739209e-10 5.1643525 3.639862e-18 2.191318e-14 5.1643525 4.005202e-16
## 4442 3.008759e-11 5.1094596 5.699717e-14 2.859516e-11 5.1094596 9.064438e-18
## 8659 5.820518e-12 2.7473056 1.050525e-16 3.023505e-13 2.7473056 3.777763e-15
## 14686 4.332568e-10 -0.7401688 1.339242e-16 3.023505e-13 -0.7401688 1.728359e-13
##      adj.P.Val.s1 AveExpr.s1      P.Value.s6 adj.P.Val.s6 AveExpr.s6      P.Value.s5
## 7739 2.545758e-14 1.9579092 2.095342e-16 9.460994e-13 1.9579092 3.771704e-17
## 11938 6.197073e-14 3.7244987 1.002399e-16 6.034778e-13 3.7244987 9.551628e-19
## 5735 1.808449e-12 5.1643525 2.431191e-15 5.488718e-12 5.1643525 1.536313e-17
## 4442 6.197073e-14 5.1094596 7.220198e-17 6.034778e-13 5.1094596 1.578802e-17
## 8659 1.364604e-11 2.7473056 6.184633e-17 6.034778e-13 2.7473056 1.058378e-15
## 14686 2.401223e-10 -0.7401688 2.413864e-13 2.004799e-10 -0.7401688 9.631028e-15
##      adj.P.Val.s5 AveExpr.s5
## 7739 1.703019e-13 1.9579092
## 11938 1.725120e-14 3.7244987
## 5735 9.504912e-14 5.1643525
## 4442 9.504912e-14 5.1094596
## 8659 3.185893e-12 2.7473056
## 14686 1.932733e-11 -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 PC          0.176 3.36e-99 chr11:66872159:66907821:clu_1700_-
## 2 CA5BP1     -0.202 4.23e-99 chrX:15675778:15688661:clu_15162_+
## 3 CA5B       -0.202 4.23e-99 chrX:15675778:15688661:clu_15162_+
## 4 CA5BP1      0.148 4.23e-99 chrX:15738352:15749971:clu_15162_+
## 5 CA5B        0.148 4.23e-99 chrX:15738352:15749971:clu_15162_+
## 6 PPARG       0.228 4.40e-89 chr3:12351674:12379704:clu_18227_+
```

```
nrow(lc_genes)
```

```
## [1] 833
```

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

```
## [1] 519
```

```
summary(unique(lc_genes$genes) %in% all_sig$gene_name) #41 genes are also sig
```

```
##      Mode  FALSE    TRUE
## logical    464     55
```

```
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 PC       0.176 3.36e-99 chr11:66872159:66907821:clu_1700_- FALSE
## 2 CA5BP1   -0.202 4.23e-99 chrX:15675778:15688661:clu_15162_+ FALSE
## 3 CA5B     -0.202 4.23e-99 chrX:15675778:15688661:clu_15162_+ TRUE
## 4 CA5BP1   0.148 4.23e-99 chrX:15738352:15749971:clu_15162_+ FALSE
## 5 CA5B     0.148 4.23e-99 chrX:15738352:15749971:clu_15162_+ TRUE
## 6 PPARG    0.228 4.40e-89 chr3:12351674:12379704:clu_18227_+ FALSE
```

```
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    519
```

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

```
##      Mode   FALSE   TRUE
## logical    518     1
```

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

```
##      Mode   FALSE   TRUE
## logical    518     1
```

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

```
##      Mode   FALSE   TRUE
## logical    518     1
```