

## TRIFID\_using\_extra\_introns

Makes Supplementary Table 5. trifid volcano plot - Figure2F

Required for: Trifid\_stats.Rmd - Figure 2E Go\_networks\_plus\_trifid.Rmd - Figure 2G,H

Theres a refseq TRIFID table, so for introns with only a refseq id I use the TRIFID score for refseq transcripts.

```
library(bioRxiv)
library(tidyverse)
library(dplyr)
library(ggplot2)
library(ggrepel)

library(clusterProfiler)
library(here)
i_am("R/15_TRIFID_using_extra_introns.Rmd")

lc = read.delim(here("31_leafcutter", "three_database_info_all_junctions.tsv"))
lc = unite(lc, "intron_coords", chr, start, end, strand, sep = ":")
lc[grep("PEMT", lc$gene),]

##      annotation      intron_coords cluster_id     deltapsi
## 12      gencode chr17:17577027:17577414:- clu_19605_- 7.101235e-02
## 13      gencode chr17:17577027:17582267:- clu_19605_- 7.985252e-02
## 14      gencode chr17:17577027:17591531:- clu_19605_- -4.296507e-01
## 15      gencode chr17:17577027:17591597:- clu_19605_- 2.523690e-03
## 38726    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 38727    gencode chr17:17506301:17509434:- clu_19603_- -1.104112e-03
## 39268    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 39269    gencode chr17:17509545:17576920:- clu_19604_- 3.879182e-03
## 39270    gencode chr17:17512654:17519027:- clu_19604_- 2.816277e-03
## 39271    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 39272    gencode chr17:17512654:17576920:- clu_19604_- 3.072487e-05
## 39273    gencode chr17:17522395:17576920:- clu_19604_- 1.226553e-02
## 94116      refseq chr17:17577027:17577107:- clu_19605_- 1.918093e-01
## 94117      refseq chr17:17577027:17591967:- clu_19605_- 4.230864e-02
##          p.adjust
## 12  4.360252e-104
## 13  4.360252e-104
## 14  4.360252e-104
## 15  4.360252e-104
## 38726  1.846905e-01
## 38727  1.846905e-01
## 39268  1.914447e-01
## 39269  1.914447e-01
## 39270  1.914447e-01
## 39271  1.914447e-01
```



```

## 38726      TRUE
## 38727     FALSE
## 39268     FALSE
## 39269     FALSE
## 39270      TRUE
## 39271     FALSE
## 39272     FALSE
## 39273     FALSE
## 94116      TRUE
## 94117      TRUE

head(lc)

##   annotation      intron_coords cluster_id    deltapsi      p.adjust
## 1  gencode chr7:43648652:43650493:- clu_35616_- -0.0141028170 2.192287e-123
## 2  gencode chr7:43648652:43650612:- clu_35616_- -0.0408035987 2.192287e-123
## 3  gencode chr7:43648652:43665658:- clu_35616_- -0.0009734716 2.192287e-123
## 4  gencode chr7:43648652:43711400:- clu_35616_- -0.0003668545 2.192287e-123
## 5  gencode chr7:43648652:43729429:- clu_35616_- -0.0681581659 2.192287e-123
## 6  gencode chr7:43650712:43656033:- clu_35616_- -0.0034620531 2.192287e-123
##
##                                     trans
## 1
## 2 ENST00000446564.5,ENST00000448704.5,ENST00000451651.5,ENST00000418140.5,ENST00000431651.5,ENST00000
## 3
## 4
## 5           ENST00000457939.1,ENST00000420441.1,ENST00000415076.6,ENST00000223336.10,ENST00000
## 6
##   min_intron_number mode_intron_number gene
## 1                  2                 2 COA1
## 2                  2                 2 COA1
## 3                  2                 2 COA1
## 4                  2                 2 COA1
## 5                  1                 1 COA1
## 6                  4                 4 COA1
##
##                   biotype genes_in_cluster
## 1
## 2      protein_coding          COA1
## 3      nonsense-mediated_decay COA1
## 4      protein_coding          COA1
## 5      nonsense-mediated_decay COA1
## 6      protein_coding          COA1
##
##   is_first_intron
## 1      FALSE
## 2      FALSE
## 3      FALSE
## 4      FALSE
## 5      TRUE
## 6      FALSE

dim(lc) # some duplications based on gene names/ antisense transcripts.

```

```
## [1] 132587 12
```

```
trifid = read.delim(here("annotations/trifid/gencode37_trifid_predictions.tsv"))
head(trifid)
```

```
##          gene_id gene_name transcript_id translation_id      flags
## 1 ENSG00000187010        RHD ENST00000342055 ENSP00000339577 protein_coding
## 2 ENSG00000187010        RHD ENST00000328664 ENSP00000331871 protein_coding
## 3 ENSG00000187010        RHD ENST00000417538 ENSP00000396420 protein_coding
## 4 ENSG00000187010        RHD ENST00000423810 ENSP00000399640 protein_coding
## 5 ENSG00000187010        RHD ENST00000622561 ENSP00000478087 protein_coding
## 6 ENSG00000187010        RHD ENST00000454452 ENSP00000413849 protein_coding
##          ccdsid      appris           ann_type length trifid_score
## 1 CCDS60028.1      MINOR Alternative     493    0.2105
## 2 CCDS262.1 PRINCIPAL:1 Principal       417    0.4021
## 3 CCDS60031.1      MINOR Alternative     378    0.0572
## 4 CCDS60027.1      MINOR Alternative Duplication     431    0.0415
## 5 CCDS60027.1      MINOR Alternative     431    0.0223
## 6 CCDS53285.1      MINOR Alternative     321    0.0652
##      norm_trifid_score
## 1            0.4210
## 2            0.8043
## 3            0.1145
## 4            0.0830
## 5            0.0447
## 6            0.1303
```

```
length(unique(trifid$transcript_id))#104 688
```

```
## [1] 104688
```

```
trefseq = read.delim(here("annotations/trifid/refseq110_trifid_predictions.tsv"))
nrow(trefseq)
```

```
## [1] 129456
```

```
head(trefseq)
```

```
##      gene_id gene_name transcript_id translation_id flags      ccdsid      appris
## 1      9997      SC02   NM_001169111   NP_001162582 mRNA CCDS14095.1 PRINCIPAL:1
## 2      9997      SC02   NM_001169110   NP_001162581 mRNA CCDS14095.1 PRINCIPAL:1
## 3      9997      SC02   NM_001169109   NP_001162580 mRNA CCDS14095.1 PRINCIPAL:1
## 4      9997      SC02   NM_005138      NP_005129 mRNA CCDS14095.1 PRINCIPAL:1
## 5      9994  CASP8AP2   NM_001137667   NP_001131139 mRNA          - PRINCIPAL:1
## 6      9994  CASP8AP2   NM_012115      NP_036247 mRNA          - PRINCIPAL:1
##      ann_type length trifid_score norm_trifid_score
## 1      -      266    0.5779      1.0000
## 2      -      266    0.5779      1.0000
## 3      -      266    0.5779      1.0000
## 4      -      266    0.5779      1.0000
## 5      -     1982    0.2909      0.5818
## 6      -     1982    0.2909      0.5818
```

```
trefseq$gene_id = as.character(trefseq$gene_id)
trifid = bind_rows(trifid, trefseq)
nrow(trifid)
```

```
## [1] 234144
```

The reason trifid has so few transcripts is only those with translated sequences are included.

## Convert lc introns to transcript table

```
transcripts = separate_longer_delim(lc, transcript_ids, delim=",")
transcripts$transcript_ids = gsub("\\.\\[0-9]*", "", gsub("rna-", "", transcripts$transcript_ids))
head(transcripts)

##   annotation      intron_coords cluster_id    deltapsi      p.adjust
## 1  gencode chr7:43648652:43650493:- clu_35616_- -0.01410282 2.192287e-123
## 2  gencode chr7:43648652:43650612:- clu_35616_- -0.04080360 2.192287e-123
## 3  gencode chr7:43648652:43650612:- clu_35616_- -0.04080360 2.192287e-123
## 4  gencode chr7:43648652:43650612:- clu_35616_- -0.04080360 2.192287e-123
## 5  gencode chr7:43648652:43650612:- clu_35616_- -0.04080360 2.192287e-123
## 6  gencode chr7:43648652:43650612:- clu_35616_- -0.04080360 2.192287e-123
##   transcript_ids min_intron_number mode_intron_number gene
## 1 ENST00000310564                  2                 2 COA1
## 2 ENST00000446564                  2                 2 COA1
## 3 ENST00000448704                  2                 2 COA1
## 4 ENST00000451651                  2                 2 COA1
## 5 ENST00000418140                  2                 2 COA1
## 6 ENST00000431651                  2                 2 COA1
##                                     biotype genes_in_cluster
## 1                               protein_coding          COA1
## 2 nonsense-mediated_decay,protein_coding,lncRNA COA1
## 3 nonsense-mediated_decay,protein_coding,lncRNA COA1
## 4 nonsense-mediated_decay,protein_coding,lncRNA COA1
## 5 nonsense-mediated_decay,protein_coding,lncRNA COA1
## 6 nonsense-mediated_decay,protein_coding,lncRNA COA1
##   is_first_intron
## 1        FALSE
## 2        FALSE
## 3        FALSE
## 4        FALSE
## 5        FALSE
## 6        FALSE

dim(transcripts)

## [1] 379523      12

transcripts[grep("PEMT", transcripts$gene),]
```

```

##      annotation      intron_coords cluster_id      deltapsi
## 27      gencode chr17:17577027:17577414:- clu_19605_- 7.101235e-02
## 28      gencode chr17:17577027:17582267:- clu_19605_- 7.985252e-02
## 29      gencode chr17:17577027:17591531:- clu_19605_- -4.296507e-01
## 30      gencode chr17:17577027:17591531:- clu_19605_- -4.296507e-01
## 31      gencode chr17:17577027:17591531:- clu_19605_- -4.296507e-01
## 32      gencode chr17:17577027:17591531:- clu_19605_- -4.296507e-01
## 33      gencode chr17:17577027:17591531:- clu_19605_- -4.296507e-01
## 34      gencode chr17:17577027:17591597:- clu_19605_- 2.523690e-03
## 35      gencode chr17:17577027:17591597:- clu_19605_- 2.523690e-03
## 130114    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130115    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130116    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130117    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130118    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130119    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130120    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130121    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130122    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130123    gencode chr17:17505848:17506227:- clu_19603_- 3.798946e-03
## 130124    gencode chr17:17506301:17509434:- clu_19603_- -1.104112e-03
## 130125    gencode chr17:17506301:17509434:- clu_19603_- -1.104112e-03
## 130126    gencode chr17:17506301:17509434:- clu_19603_- -1.104112e-03
## 130127    gencode chr17:17506301:17509434:- clu_19603_- -1.104112e-03
## 130128    gencode chr17:17506301:17509434:- clu_19603_- -1.104112e-03
## 130129    gencode chr17:17506301:17509434:- clu_19603_- -1.104112e-03
## 132056    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132057    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132058    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132059    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132060    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132061    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132062    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132063    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132064    gencode chr17:17509545:17512509:- clu_19604_- -2.027199e-02
## 132065    gencode chr17:17509545:17576920:- clu_19604_- 3.879182e-03
## 132066    gencode chr17:17512654:17519027:- clu_19604_- 2.816277e-03
## 132067    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 132068    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 132069    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 132070    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 132071    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 132072    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 132073    gencode chr17:17512654:17522280:- clu_19604_- -9.071787e-04
## 132074    gencode chr17:17512654:17576920:- clu_19604_- 3.072487e-05
## 132075    gencode chr17:17522395:17576920:- clu_19604_- 1.226553e-02
## 132076    gencode chr17:17522395:17576920:- clu_19604_- 1.226553e-02
## 132077    gencode chr17:17522395:17576920:- clu_19604_- 1.226553e-02
## 132078    gencode chr17:17522395:17576920:- clu_19604_- 1.226553e-02
## 132079    gencode chr17:17522395:17576920:- clu_19604_- 1.226553e-02
## 132080    gencode chr17:17522395:17576920:- clu_19604_- 1.226553e-02
## 320814    refseq chr17:17577027:17577107:- clu_19605_- 1.918093e-01
## 320815    refseq chr17:17577027:17591967:- clu_19605_- 4.230864e-02
##          p.adjust transcript_ids min_intron_number mode_intron_number gene

```

			biotype	genes_in_cluster	
			protein_coding		PEMT
##	27	4.360252e-104	ENST00000395782	1	1 PEMT
##	28	4.360252e-104	ENST00000395783	1	1 PEMT
##	29	4.360252e-104	ENST00000421096	1	1 PEMT
##	30	4.360252e-104	ENST00000580147	1	1 PEMT
##	31	4.360252e-104	ENST00000461404	1	1 PEMT
##	32	4.360252e-104	ENST00000255389	1	1 PEMT
##	33	4.360252e-104	ENST00000395781	1	1 PEMT
##	34	4.360252e-104	ENST00000435340	1	1 PEMT
##	35	4.360252e-104	ENST00000472446	1	1 PEMT
##	130114	1.846905e-01	ENST00000255389	1	4 PEMT
##	130115	1.846905e-01	ENST00000435340	1	4 PEMT
##	130116	1.846905e-01	ENST00000582268	1	4 PEMT
##	130117	1.846905e-01	ENST00000490392	1	4 PEMT
##	130118	1.846905e-01	ENST00000395783	1	4 PEMT
##	130119	1.846905e-01	ENST00000395781	1	4 PEMT
##	130120	1.846905e-01	ENST00000580147	1	4 PEMT
##	130121	1.846905e-01	ENST00000395782	1	4 PEMT
##	130122	1.846905e-01	ENST00000477595	1	4 PEMT
##	130123	1.846905e-01	ENST00000484838	1	4 PEMT
##	130124	1.846905e-01	ENST00000255389	3	3 PEMT
##	130125	1.846905e-01	ENST00000395783	3	3 PEMT
##	130126	1.846905e-01	ENST00000580147	3	3 PEMT
##	130127	1.846905e-01	ENST00000484838	3	3 PEMT
##	130128	1.846905e-01	ENST00000490392	3	3 PEMT
##	130129	1.846905e-01	ENST00000395782	3	3 PEMT
##	132056	1.914447e-01	ENST00000395783	2	4 PEMT
##	132057	1.914447e-01	ENST00000421096	2	4 PEMT
##	132058	1.914447e-01	ENST00000472446	2	4 PEMT
##	132059	1.914447e-01	ENST00000395781	2	4 PEMT
##	132060	1.914447e-01	ENST00000490392	2	4 PEMT
##	132061	1.914447e-01	ENST00000484838	2	4 PEMT
##	132062	1.914447e-01	ENST00000435340	2	4 PEMT
##	132063	1.914447e-01	ENST00000255389	2	4 PEMT
##	132064	1.914447e-01	ENST00000395782	2	4 PEMT
##	132065	1.914447e-01	ENST00000580147	2	2 PEMT
##	132066	1.914447e-01	ENST00000490392	1	1 PEMT
##	132067	1.914447e-01	ENST00000395783	3	3 PEMT
##	132068	1.914447e-01	ENST00000421096	3	3 PEMT
##	132069	1.914447e-01	ENST00000395782	3	3 PEMT
##	132070	1.914447e-01	ENST00000395781	3	3 PEMT
##	132071	1.914447e-01	ENST00000435340	3	3 PEMT
##	132072	1.914447e-01	ENST00000255389	3	3 PEMT
##	132073	1.914447e-01	ENST00000461404	3	3 PEMT
##	132074	1.914447e-01	ENST00000472446	2	2 PEMT
##	132075	1.914447e-01	ENST00000395783	2	2 PEMT
##	132076	1.914447e-01	ENST00000395782	2	2 PEMT
##	132077	1.914447e-01	ENST00000395781	2	2 PEMT
##	132078	1.914447e-01	ENST00000421096	2	2 PEMT
##	132079	1.914447e-01	ENST00000435340	2	2 PEMT
##	132080	1.914447e-01	ENST00000255389	2	2 PEMT
##	320814	4.360252e-104	XM_006721418	1	1 PEMT
##	320815	4.360252e-104	XM_024450532	1	1 PEMT

biotype genes\_in\_cluster  
protein\_coding PEMT

## 28	protein_coding	PEMT
## 29	lncRNA,nonsense-mediated_decay,protein_coding	PEMT
## 30	lncRNA,nonsense-mediated_decay,protein_coding	PEMT
## 31	lncRNA,nonsense-mediated_decay,protein_coding	PEMT
## 32	lncRNA,nonsense-mediated_decay,protein_coding	PEMT
## 33	lncRNA,nonsense-mediated_decay,protein_coding	PEMT
## 34	protein_coding,lncRNA	PEMT
## 35	protein_coding,lncRNA	PEMT
## 130114	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130115	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130116	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130117	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130118	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130119	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130120	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130121	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130122	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130123	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 130124	protein_coding,nonsense-mediated_decay,lncRNA	PEMT
## 130125	protein_coding,nonsense-mediated_decay,lncRNA	PEMT
## 130126	protein_coding,nonsense-mediated_decay,lncRNA	PEMT
## 130127	protein_coding,nonsense-mediated_decay,lncRNA	PEMT
## 130128	protein_coding,nonsense-mediated_decay,lncRNA	PEMT
## 130129	protein_coding,nonsense-mediated_decay,lncRNA	PEMT
## 132056	protein_coding,lncRNA	PEMT
## 132057	protein_coding,lncRNA	PEMT
## 132058	protein_coding,lncRNA	PEMT
## 132059	protein_coding,lncRNA	PEMT
## 132060	protein_coding,lncRNA	PEMT
## 132061	protein_coding,lncRNA	PEMT
## 132062	protein_coding,lncRNA	PEMT
## 132063	protein_coding,lncRNA	PEMT
## 132064	protein_coding,lncRNA	PEMT
## 132065	nonsense-mediated_decay	PEMT
## 132066	lncRNA	PEMT
## 132067	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 132068	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 132069	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 132070	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 132071	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 132072	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 132073	protein_coding,lncRNA,nonsense-mediated_decay	PEMT
## 132074	lncRNA	PEMT
## 132075	protein_coding,lncRNA	PEMT
## 132076	protein_coding,lncRNA	PEMT
## 132077	protein_coding,lncRNA	PEMT
## 132078	protein_coding,lncRNA	PEMT
## 132079	protein_coding,lncRNA	PEMT
## 132080	protein_coding,lncRNA	PEMT
## 320814	<NA>	PEMT
## 320815	<NA>	PEMT
## is_first_intron		
## 27	TRUE	
## 28	TRUE	

```
## 29      TRUE
## 30      TRUE
## 31      TRUE
## 32      TRUE
## 33      TRUE
## 34      TRUE
## 35      TRUE
## 130114   TRUE
## 130115   TRUE
## 130116   TRUE
## 130117   TRUE
## 130118   TRUE
## 130119   TRUE
## 130120   TRUE
## 130121   TRUE
## 130122   TRUE
## 130123   TRUE
## 130124   FALSE
## 130125   FALSE
## 130126   FALSE
## 130127   FALSE
## 130128   FALSE
## 130129   FALSE
## 132056   FALSE
## 132057   FALSE
## 132058   FALSE
## 132059   FALSE
## 132060   FALSE
## 132061   FALSE
## 132062   FALSE
## 132063   FALSE
## 132064   FALSE
## 132065   FALSE
## 132066   TRUE
## 132067   FALSE
## 132068   FALSE
## 132069   FALSE
## 132070   FALSE
## 132071   FALSE
## 132072   FALSE
## 132073   FALSE
## 132074   FALSE
## 132075   FALSE
## 132076   FALSE
## 132077   FALSE
## 132078   FALSE
## 132079   FALSE
## 132080   FALSE
## 320814   TRUE
## 320815   TRUE
```

some numbers on this

```

sig_trans = filter(transcripts, p.adjust < 0.05 & abs(deltapsi) >= 0.1)
sprintf("Leafcutter significant introns (%d) correspond to %d unique transcripts (represented in %d rows)
       length(unique(sig_trans$intron_coords)),
       length(unique(sig_trans$transcript_ids)),
       length(sig_trans$transcript_ids))

## [1] "Leafcutter significant introns (777) correspond to 2094 unique transcripts (represented in 2233 rows)

sprintf("%d of these transcripts are represented in the trifid database.\n",
        sum(unique(sig_trans$transcript_ids) %in% trifid$transcript_id))

## [1] "1484 of these transcripts are represented in the trifid database.\n"

transcripts$transcript_ids[grep("PEMT", transcripts$gene,)] %in% trifid$transcript_id

## [1] TRUE TRUE FALSE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE FALSE
## [13] FALSE TRUE TRUE TRUE TRUE FALSE FALSE TRUE TRUE TRUE FALSE FALSE
## [25] TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE TRUE TRUE TRUE FALSE
## [37] TRUE FALSE TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE FALSE
## [49] TRUE TRUE TRUE TRUE

```

## Merge trifid and leafcutter

```

mtrans = merge(transcripts, trifid, by.x ="transcript_ids",by.y="transcript_id")
mtrans = mutate(mtrans, condition = if_else(deltapsi > 0, "beige", if_else(deltapsi < 0, "white", "none"))
               sig = p.adjust < 0.05 & abs(deltapsi) >= 0.1)

length(unique(mtrans$transcript_ids)) #82 326 transcripts in total

## [1] 82326

length(unique(filter(mtrans, sig) %>% pull(transcript_ids)))

## [1] 1484

mtrans = arrange(mtrans, desc(trifid_score), desc(norm_trifid_score), p.adjust, desc(abs(deltapsi)))
head(mtrans[c("gene_name", "trifid_score", "norm_trifid_score", "deltapsi", "p.adjust")], n=20)

##   gene_name trifid_score norm_trifid_score      deltapsi      p.adjust
## 1    LRRC28          1             1 -0.1277105151 2.653936e-14
## 2     COPS5          1             1  0.0595450863 1.754366e-08
## 3     STK38          1             1  0.0312487644 1.250549e-07
## 4      C1D          1             1  0.0099164702 4.253925e-07
## 5     AAMDC          1             1 -0.0008870103 9.403490e-07
## 6     AAMDC          1             1 -0.0006973601 9.403490e-07
## 7     ACTR6          1             1  0.0581605204 8.495921e-06
## 8    ARFIP1          1             1  0.0068791997 1.189622e-05

```

```

## 9      ARFIP1          1          1  0.0064822954 1.189622e-05
## 10     GNAI2          1          1 -0.0279736041 6.917548e-05
## 11     GNB4          1          1  0.0282956006 2.413151e-04
## 12     DR1           1          1  0.0045084324 1.478700e-03
## 13     GDI2          1          1 -0.0327184175 2.460331e-03
## 14     GDI2          1          1  0.0319582980 2.460331e-03
## 15     ACTR10         1          1  0.0275113044 2.660836e-03
## 16     ACTR10         1          1 -0.0265122782 2.660836e-03
## 17     ATP6V1C1        1          1  0.0274511700 4.226843e-03
## 18     ATP6V1C1        1          1 -0.0153237824 4.226843e-03
## 19     ATP6V1C1        1          1 -0.0153237824 4.226843e-03
## 20     ATP6V1C1        1          1 -0.0036512432 4.226843e-03

```

```
head(mtrans[mtrans$gene == "PEMT",])
```

```

##      transcript_ids annotation      intron_coords cluster_id
## 39508 XM_024450532    refseq chr17:17577027:17591967:- clu_19605_-
## 49409 ENST00000395783   gencode chr17:17577027:17582267:- clu_19605_-
## 49410 ENST00000395782   gencode chr17:17577027:17577414:- clu_19605_-
## 49425 ENST00000395782   gencode chr17:17505848:17506227:- clu_19603_-
## 49426 ENST00000395783   gencode chr17:17505848:17506227:- clu_19603_-
## 49427 ENST00000395782   gencode chr17:17506301:17509434:- clu_19603_-
##      deltapsi      p.adjust min_intron_number mode_intron_number gene
## 39508 0.042308642 4.360252e-104          1                  1 PEMT
## 49409 0.079852522 4.360252e-104          1                  1 PEMT
## 49410 0.071012350 4.360252e-104          1                  1 PEMT
## 49425 0.003798946 1.846905e-01          1                  4 PEMT
## 49426 0.003798946 1.846905e-01          1                  4 PEMT
## 49427 -0.001104112 1.846905e-01          3                  3 PEMT
##                                biotype genes_in_cluster
## 39508 <NA>                  PEMT
## 49409 protein_coding          PEMT
## 49410 protein_coding          PEMT
## 49425 protein_coding,lncRNA,nonsense-mediated_decay PEMT
## 49426 protein_coding,lncRNA,nonsense-mediated_decay PEMT
## 49427 protein_coding,nonsense-mediated_decay,lncRNA PEMT
##      is_first_intron      gene_id gene_name translation_id flags
## 39508      TRUE       10400    PEMT  XP_024306300 mRNA
## 49409      TRUE ENSG00000133027    PEMT ENSP00000379129 protein_coding
## 49410      TRUE ENSG00000133027    PEMT ENSP00000379128 protein_coding
## 49425      TRUE ENSG00000133027    PEMT ENSP00000379128 protein_coding
## 49426      TRUE ENSG00000133027    PEMT ENSP00000379129 protein_coding
## 49427      FALSE ENSG00000133027   PEMT ENSP00000379128 protein_coding
##      ccdsid      appris      ann_type length trifid_score
## 39508 - PRINCIPAL:1          -      199  0.8739
## 49409 CCDS11187.1 PRINCIPAL:1 Principal Duplication 199  0.8038
## 49410 CCDS11187.1 PRINCIPAL:1          Principal 199  0.8038
## 49425 CCDS11187.1 PRINCIPAL:1          Principal 199  0.8038
## 49426 CCDS11187.1 PRINCIPAL:1 Principal Duplication 199  0.8038
## 49427 CCDS11187.1 PRINCIPAL:1          Principal 199  0.8038
##      norm_trifid_score condition sig
## 39508      0.8914    beige FALSE
## 49409      1.0000    beige FALSE
## 49410      1.0000    beige FALSE

```

```

## 49425      1.0000    beige FALSE
## 49426      1.0000    beige FALSE
## 49427      1.0000    white FALSE

## Log missing introns,

missing = transcripts[!transcripts$transcript_ids %in% trifid$transcript_id &
                      !transcripts$intron_coords %in% mtrans$intron_coords,] #annotate later

cat("Unrepresented transcripts include those from genes like: \n")

## Unrepresented transcripts include those from genes like:

cat(head(unique(missing$gene[missing$p.adjust < 0.05 & abs(missing$deltapsi)>0.1])))

## CA5BP1 LYRM4 ALG9 NAV1 LINC01140 SPON2

table(missing$annotation)

##  

##      cryptic fantom_cat      gencode      refseq  

##      21831       18742       26204       2694

missing[grep( "PEMT", missing$gene),] #Missing PEMT transcripts from non-significant cluster

##      annotation      intron_coords cluster_id     deltapsi p.adjust  

## 132066   gencode chr17:17512654:17519027:- clu_19604_- 2.816277e-03 0.1914447  

## 132074   gencode chr17:17512654:17576920:- clu_19604_- 3.072487e-05 0.1914447  

##      transcript_ids min_intron_number mode_intron_number gene biotype  

## 132066 ENST00000490392           1                  1 PEMT lncRNA  

## 132074 ENST00000472446           2                  2 PEMT lncRNA  

##      genes_in_cluster is_first_intron  

## 132066          PEMT        TRUE  

## 132074          PEMT       FALSE

```

Of course fantom and cryptic transcripts cannot be annotated by this trifid (which is based on the ensembl + refseq annotation); but also many transcripts from gencode cannot be found there either - perhaps they're new transcripts, or not protein coding versions. The gencode annotation is always changing.

### summarise transcripts with introns in multiple directions

```
table(mtrans$condition)
```

```
##  

##  beige  white  

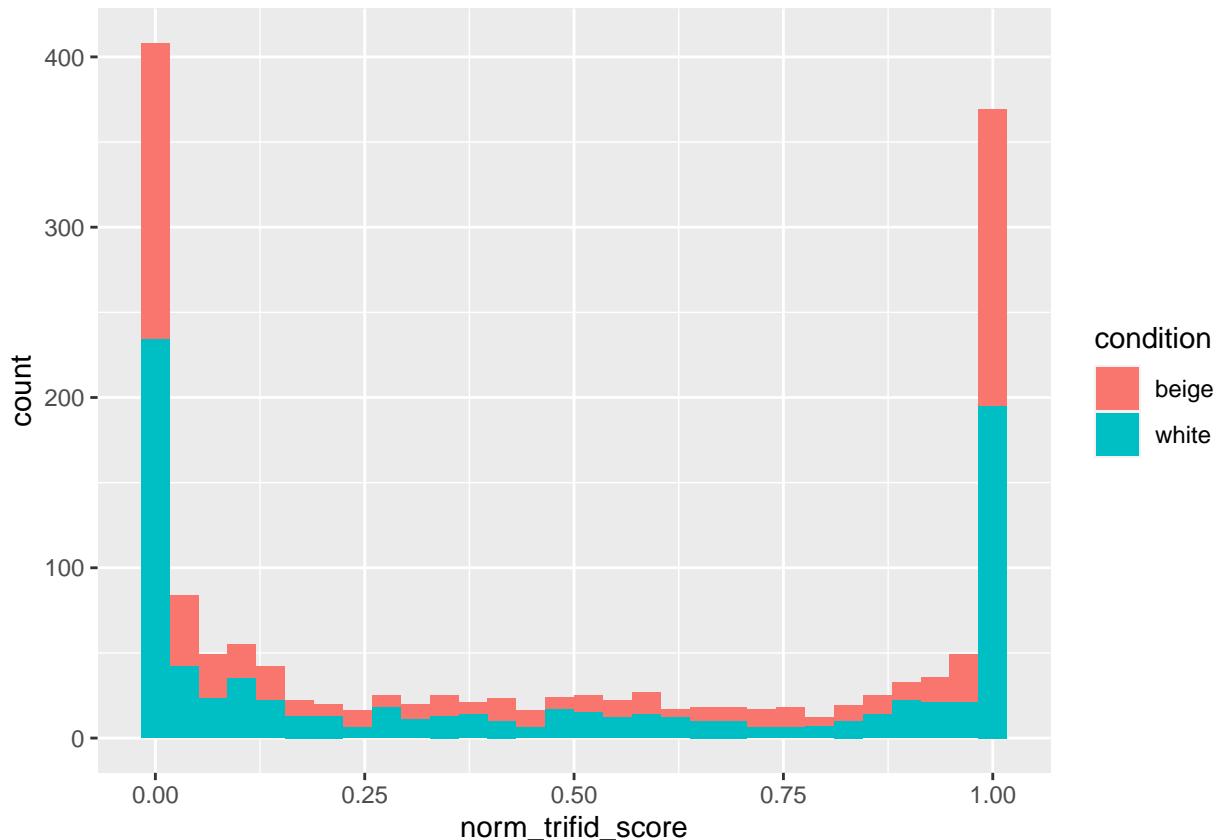
## 134112 125512
```

```
table(paste(mtrans$sig, mtrans$condition))
```

```
##  
## FALSE beige FALSE white TRUE beige TRUE white  
##      133409      124660       703       852
```

## Histogram

```
ggplot(filter(mtrans, sig)) + geom_histogram(aes(x=norm_trifid_score, fill=condition), bins=30)
```



```
## Flags
```

```
table(mtrans$flags)
```

```
##  
##          mRNA           non_stop_decay  
##          25475            146  
## nonsense-mediated_decay nonsense-mediated_decay,RT  
##          41896            1836  
## protein_coding           protein_coding,RT  
##          189018            1251  
## TR_C_gene                  2
```

```
table(filter(mtrans, sig) %>% pull(flags))
```

```
##          mRNA      non_stop_decay
##             79                  2
##    nonsense_mediated_decay nonsense_mediated_decay,RT
##            197                 13
##      protein_coding      protein_coding,RT
##            1256                  8
```

## Correlation between dPSI and TRIFID score?

There is actually, if you look within the same cluster if we have a intron with a higher PSI it tends to have higher functionality - or maybe it just has more transcripts?

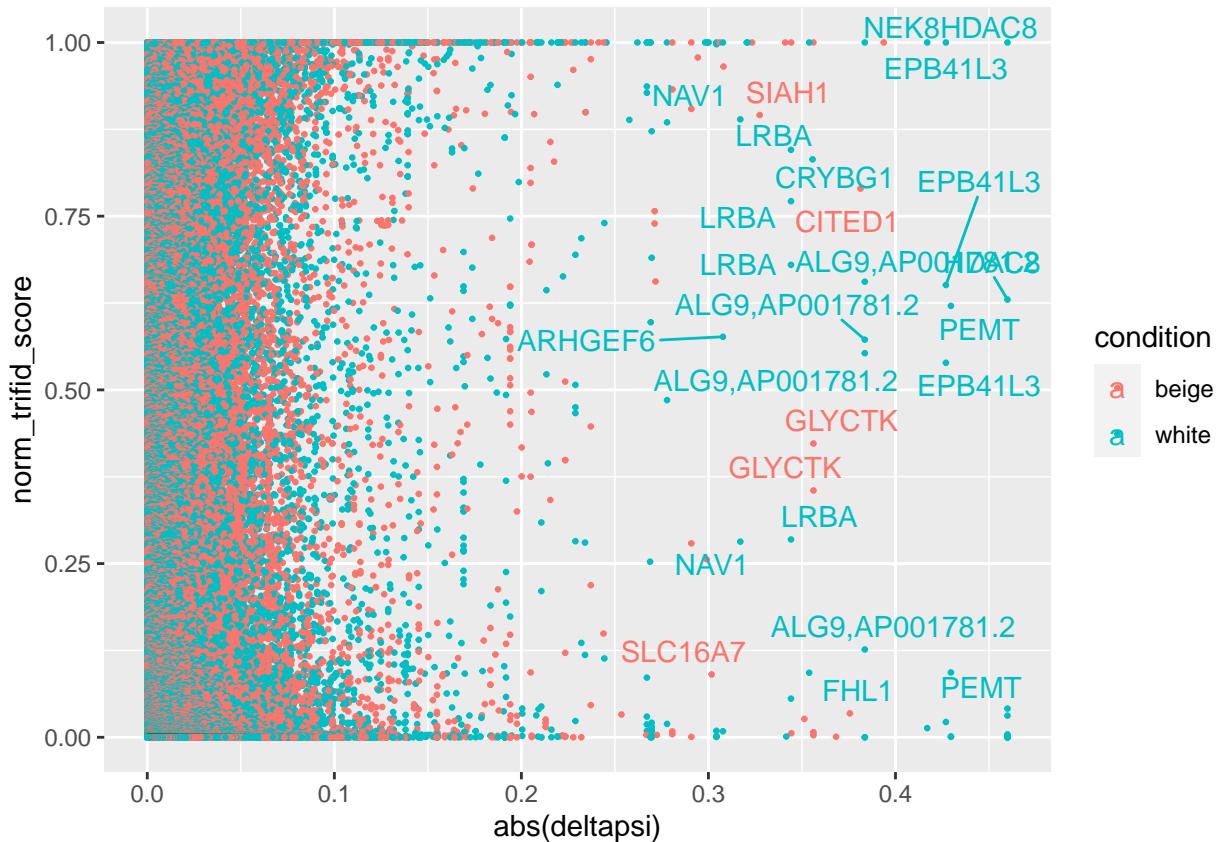
```
ggplot(mtrans, aes(x=deltapsi, y=norm_trifid_score, colour=condition)) + geom_point(size=0.5) +
  geom_text_repel(data= filter(mtrans, abs(deltapsi) >0.3), aes(label=gene))
```

```
## Warning: ggrepel: 45 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
ggplot(mtrans, aes(x=abs(deltapsi), y=norm_trifid_score, colour=condition)) + geom_point(size=0.5) +
  geom_text_repel(data= filter(mtrans, abs(deltapsi) >0.3), aes(label=gene))
```

```
## Warning: ggrepel: 45 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



## Annotate transcript names

```
mart <- useMart(biomart = "ensembl",
  dataset = "hsapiens_gene_ensembl",
  host = "https://sep2019.archive.ensembl.org")

biomaRt::searchAttributes(mart = mart, pattern = "transcript.*name")
```

	name	description	page
## 24	external_transcript_name	Transcript name	feature_page
## 25	external_transcript_source_name	Source of transcript name	feature_page
## 58	entrezgene_trans_name	EntrezGene transcript name	ID feature_page
## 77	mirbase_trans_name	miRBase transcript name	ID feature_page
## 92	rfam_trans_name	RFAM transcript name	ID feature_page

```

annot = getBM(c("external_transcript_name", "ensembl_gene_id", "ensembl_transcript_id"),
              filters = "ensembl_transcript_id",
              values = mtrans$transcript_ids,
              mart = mart, useCache = F)

head(annot, n=2); dim(annot)

##   external_transcript_name ensembl_gene_id ensembl_transcript_id
## 1                      VTI1B-204 ENSG00000100568      ENST00000554659
## 2                      VPS36-201 ENSG00000136100      ENST00000378060

## [1] 69071      3

#Add gene names to filt_series
mtrans = merge(mtrans, annot, by.x= "transcript_ids",
                by.y = "ensembl_transcript_id", sort=FALSE,
                all.x = T)
#head(mtrans)
remove(annot)

```

## Averaging score across isoforms at each intron

```

summary(mtrans$trifid_score)

##      Min. 1st Qu. Median    Mean 3rd Qu.    Max.
##  0.0000  0.0009  0.0785  0.2993  0.6374  1.0000

head(mtrans)

##   transcript_ids annotation           intron_coords cluster_id     deltapsi
## 1 ENST00000301981  gencode chr15:99352471:99361336:+ clu_30900_+  0.007505714
## 2 ENST00000301981  gencode chr15:99276616:99287257:+ clu_30898_+  0.011562718
## 3 ENST00000301981  gencode chr15:99251541:99255898:+ clu_30897_- -0.127710515
## 4 ENST00000301981  gencode chr15:99256125:99276576:+ clu_30898_+  0.012183427
## 5 ENST00000301981  gencode chr15:99287951:99333923:+ clu_30899_+  0.042272131
## 6 ENST00000301981  gencode chr15:99361511:99363106:+ clu_30901_+  0.004034933
##          p.adjust min_intron_number mode_intron_number    gene
## 1 8.537649e-01            2                 4 LRRC28
## 2 1.657108e-01            3                 3 LRRC28
## 3 2.653936e-14            1                 1 LRRC28
## 4 1.657108e-01            2                 2 LRRC28
## 5 3.285558e-02            4                 5 LRRC28
## 6 5.303164e-01            1                 8 LRRC28
##                                     biotype
## 1 protein_coding,nonsense-mediated_decay,retained_intron
## 2                                         lncRNA,protein_coding
## 3 nonsense-mediated_decay,lncRNA,protein_coding,retained_intron
## 4 nonsense-mediated_decay,lncRNA,protein_coding
## 5 protein_coding,nonsense-mediated_decay

```

```

## 6      protein_coding,nonsense-mediated_decay,retained_intron
## genes_in_cluster is_first_intron      gene_id gene_name translation_id
## 1      LRRC28      FALSE ENSG00000168904      LRRC28 ENSP00000304923
## 2      LRRC28      FALSE ENSG00000168904      LRRC28 ENSP00000304923
## 3      LRRC28      TRUE  ENSG00000168904      LRRC28 ENSP00000304923
## 4      LRRC28      FALSE ENSG00000168904      LRRC28 ENSP00000304923
## 5      LRRC28      FALSE ENSG00000168904      LRRC28 ENSP00000304923
## 6      LRRC28      TRUE  ENSG00000168904      LRRC28 ENSP00000304923
##      flags      ccdsid      appris ann_type length trifid_score
## 1 protein_coding CCDS10380.1 PRINCIPAL:1 Principal    367      1
## 2 protein_coding CCDS10380.1 PRINCIPAL:1 Principal    367      1
## 3 protein_coding CCDS10380.1 PRINCIPAL:1 Principal    367      1
## 4 protein_coding CCDS10380.1 PRINCIPAL:1 Principal    367      1
## 5 protein_coding CCDS10380.1 PRINCIPAL:1 Principal    367      1
## 6 protein_coding CCDS10380.1 PRINCIPAL:1 Principal    367      1
##      norm_trifid_score condition sig external_transcript_name ensembl_gene_id
## 1           1     beige FALSE      LRRC28-201 ENSG00000168904
## 2           1     beige FALSE      LRRC28-201 ENSG00000168904
## 3           1     white  TRUE      LRRC28-201 ENSG00000168904
## 4           1     beige FALSE      LRRC28-201 ENSG00000168904
## 5           1     beige FALSE      LRRC28-201 ENSG00000168904
## 6           1     beige FALSE      LRRC28-201 ENSG00000168904

dim(mtrans)

## [1] 288536      26

#mtrans = filter(mtrans, sig)

#unknown gene names get given a ".", change to NA
mtrans = mutate(mtrans, gene = if_else(gene==".", cluster_id, gene))

#this step averages the score if multiple transcripts are implicated at an intron
introns = group_by(mtrans, intron_coords, condition, deltapsi, p.adjust, gene, cluster_id, annotation) %
  mean_norm_score = mean(norm_trifid_score)
  median_norm_score = median(norm_trifid_score)
  transcripts = paste(transcript_ids, sep = ", ")
  transcript_names = paste(external_transcript_name, sep = ", ")

## `summarise()` has grouped output by 'intron_coords', 'condition', 'deltapsi',
## 'p.adjust', 'gene', 'cluster_id'. You can override using the '.groups'
## argument.

introns = arrange(introns, desc(abs(deltapsi)), desc(mean_trifid_score), p.adjust, )
dim(introns) #introns with annotations

## [1] 87024      12

head(introns)

## # A tibble: 6 x 12

```

```

## # Groups:  intron_coords, condition, deltapsi, p.adjust, gene, cluster_id [6]
##   intron_coords      condition deltapsi p.adjust gene  cluster_id annotation
##   <chr>            <chr>     <dbl>    <dbl> <chr> <chr>      <chr>
## 1 chrX:72330076:723517~ white      -0.460 6.88e- 82 HDAC8 clu_291_- gencode
## 2 chr17:17577027:17591~ white      -0.430 4.36e-104 PEMT clu_19605~ gencode
## 3 chr18:5489194:554391~ white      -0.427 4.23e- 13 EPB4~ clu_21093~ gencode
## 4 chr17:28728860:28733~ white      -0.417 5.30e- 40 NEK8  clu_34526~ gencode
## 5 chr2:48569086:485805~ beige       0.394 3.17e- 40 STON~ clu_31216~ gencode
## 6 chr11:111844723:1118~ white      -0.384 2.04e- 62 ALG9~ clu_2011_- gencode
## # i 5 more variables: mean_trifid_score <dbl>, mean_norm_score <dbl>,
## # median_norm_score <dbl>, transcripts <chr>, transcript_names <chr>

length(unique(introns$intron_coords))

```

```
## [1] 87024
```

If an intron is not in trifid; report as -0.1

```

sig_clusters = unique(lc$cluster_id[lc$p.adjust < 0.05 & abs(lc$deltapsi) > 0.1])
summary(sig_clusters %in% introns$cluster_id)#20 clusters have at least 1 trifid annotation; 78 do not

##      Mode   FALSE    TRUE
## logical      78     420

#must have another annotated intron in the cluster, but not have a score already for that intron
to_add = missing#[missing$cluster_id %in% sig_clusters &
                  #      !missing$intron_coords %in% introns$intron_coords,]
nrow(to_add) #unannotated trifid transcripts, have an annotated cluster and are for an unscored intron

## [1] 69471

to_add = mutate(to_add, gene = if_else(gene==".", cluster_id, gene))
to_add = group_by(to_add, intron_coords, deltapsi, p.adjust, gene, cluster_id, annotation) %>%
  summarise(transcripts = paste(transcript_ids,collapse=","),
            )

## `summarise()` has grouped output by 'intron_coords', 'deltapsi', 'p.adjust',
## 'gene', 'cluster_id'. You can override using the '.groups' argument.

to_add = mutate(to_add, mean_trifid_score = -0.1,
               mean_norm_score = -0.1,
               median_norm_score = -0.1,
               condition= if_else(deltapsi > 0, "beige", "white"))
head(to_add)

## # A tibble: 6 x 11
## # Groups:  intron_coords, deltapsi, p.adjust, gene, cluster_id [6]
##   intron_coords      deltapsi p.adjust gene  cluster_id annotation transcripts
##   <chr>            <dbl>    <dbl> <chr> <chr>      <chr>
```

```

## 1 chr10:100006342:100~ 4.80e-4 0.989 ENSG~ clu_37954~ fantom_cat FTMT237000~
## 2 chr10:1001013:10070~ -2.57e-4 0.0811 <NA> clu_29373~ cryptic Unknown
## 3 chr10:100190968:100~ 1.20e-2 0.147 CHUK clu_37959~ gencode ENST000005~
## 4 chr10:100200780:100~ -8.79e-4 0.920 <NA> clu_37960~ cryptic Unknown
## 5 chr10:100246935:100~ -1.12e-2 0.0387 CWF1~ clu_37962~ gencode ENST000004~
## 6 chr10:100246935:100~ 2.18e-4 0.0387 <NA> clu_37962~ cryptic Unknown
## # i 4 more variables: mean_trifid_score <dbl>, mean_norm_score <dbl>,
## # median_norm_score <dbl>, condition <chr>

nrow(to_add)

## [1] 45563

length(unique(to_add$intron_coords))

## [1] 45563

to_add= distinct(to_add) #managed to duplicate
nrow(to_add)

## [1] 45563

all_introns = bind_rows(in_trifid=introns, not_in_trifid=to_add[to_add$cluster_id %in% introns$cluster_id,
                                                               cluster_not_in_trifid= to_add[!to_add$cluster_id %in% introns$cluster_id,],
                                                               .id = "in_trifid" ])
nrow(all_introns)

## [1] 132587

length(unique(all_introns$intron_coords))

## [1] 132587

table(all_introns$in_trifid)

## 
##   cluster_not_in_trifid      in_trifid      not_in_trifid
##               11293          87024          34270

table(filter(all_introns, p.adjust < 0.05 & abs(deltapsi) > 0.1 & annotation %in% c("gencode", "refseq")))

## 
##   cluster_not_in_trifid      in_trifid      not_in_trifid
##               74                  577                  29

```

Gencode + Refseq = 647 + 33 = 680 junctions; of which 577 are in trifid; 29 are not in trifid; and 74 of which none of the junctions in the cluster are in trifid, e.g.

```

filter(all_introns, in_trifid == "cluster_not_in_trifid" & p.adjust < 0.05 & abs(deltapsi) > 0.1) %>%
  arrange(gene) %>% pull(gene) %>% unique()

## [1] "AC002074.1"          "AC002467.1"          "AC004889.1"
## [4] "AC006001.3"          "AC008771.1"          "AC016924.1"
## [7] "AC021739.2"          "AC022167.2"          "AC025171.1"
## [10] "AC078883.1"          "AC093724.3"          "AC138207.8"
## [13] "AC244154.1"          "AC244669.2"          "AF165147.1"
## [16] "AL451165.2"          "C2orf27A"            "CATG00000105473.1"
## [19] "ENSG00000124003.9"    "ENSG00000174804.3"  "ENSG00000188185.7"
## [22] "ENSG00000188681.7"    "ENSG00000228063.1"  "ENSG00000228782.3"
## [25] "ENSG00000229043.2"    "ENSG00000229180.4"  "ENSG00000246090.2"
## [28] "ENSG00000249042.1"    "ENSG00000272622.1"  "FAHD2CP"
## [31] "FAM66B"               "FRG1HP"              "GABPB1-AS1"
## [34] "GCC2-AS1"              "GPAT2P1"              "HAND2-AS1"
## [37] "HCG18"                 "HSD11B1-AS1"          "ID2-AS1"
## [40] "KCNK15-AS1"            "KCNK15-AS1, AL139352.1" "KTN1-AS1"
## [43] "LINC-PINT"             "LINC00847"            "LINC00886"
## [46] "LINC01119"              "LINC01239"            "LINC01347"
## [49] "LINC01547"              "LINC02202"            "LINC02607"
## [52] "LINC02749"              "LOC100130027"          "LOC105375587"
## [55] "LOC105379814"           "LOC128966744"          "LYPLAL1-DT"
## [58] "MEG8"                  "MIR31HG"              "MIR99AHG"
## [61] "MRPL45P2"              "NDUFS2"                "PCOTH"
## [64] "PI4KAP2"                "PKD1P5"                "RPARP-AS1"
## [67] "SDCBP2-AS1"              "SNHG10"                "TP73-AS1"
## [70] "WASH8P"                 "ZEB1-AS1"              "ZNF583, ZNF582-AS1"
## [73] NA

filter(all_introns, p.adjust < 0.05 & abs(deltapsi) > 0.1 & annotation %in% c("gencode", "refseq")) %>%
  pull(cluster_id) %>% unique() %>% length()

## [1] 470

```

### Violin plots on average TRIFID score

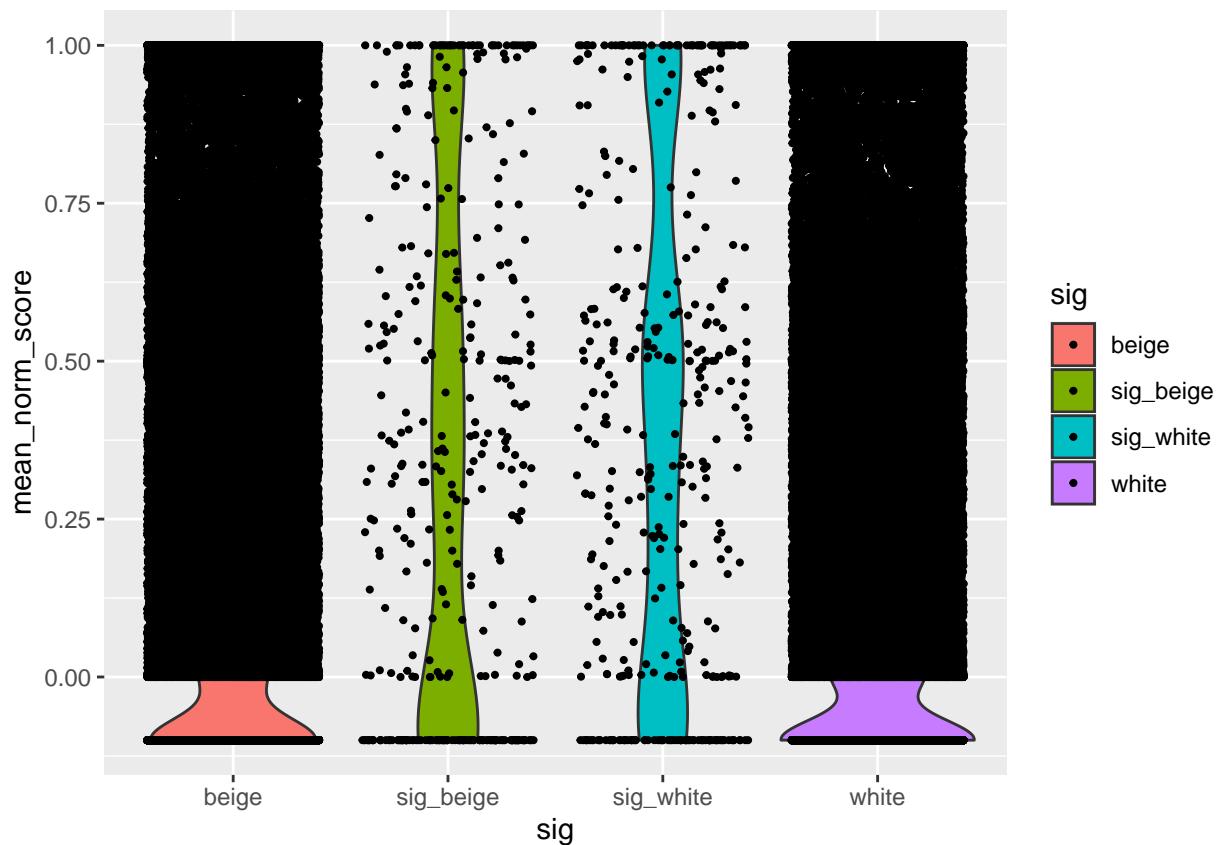
```

all_introns = mutate(all_introns, sig = if_else(p.adjust < 0.05 & deltapsi > 0.1, "sig_beige",
                                                if_else(p.adjust < 0.05 & deltapsi < -0.1, "sig_white",
                                                       if_else(deltapsi > 0, "beige",
                                                               if_else(deltapsi < -0, "white", "neither"))
table(all_introns$sig)

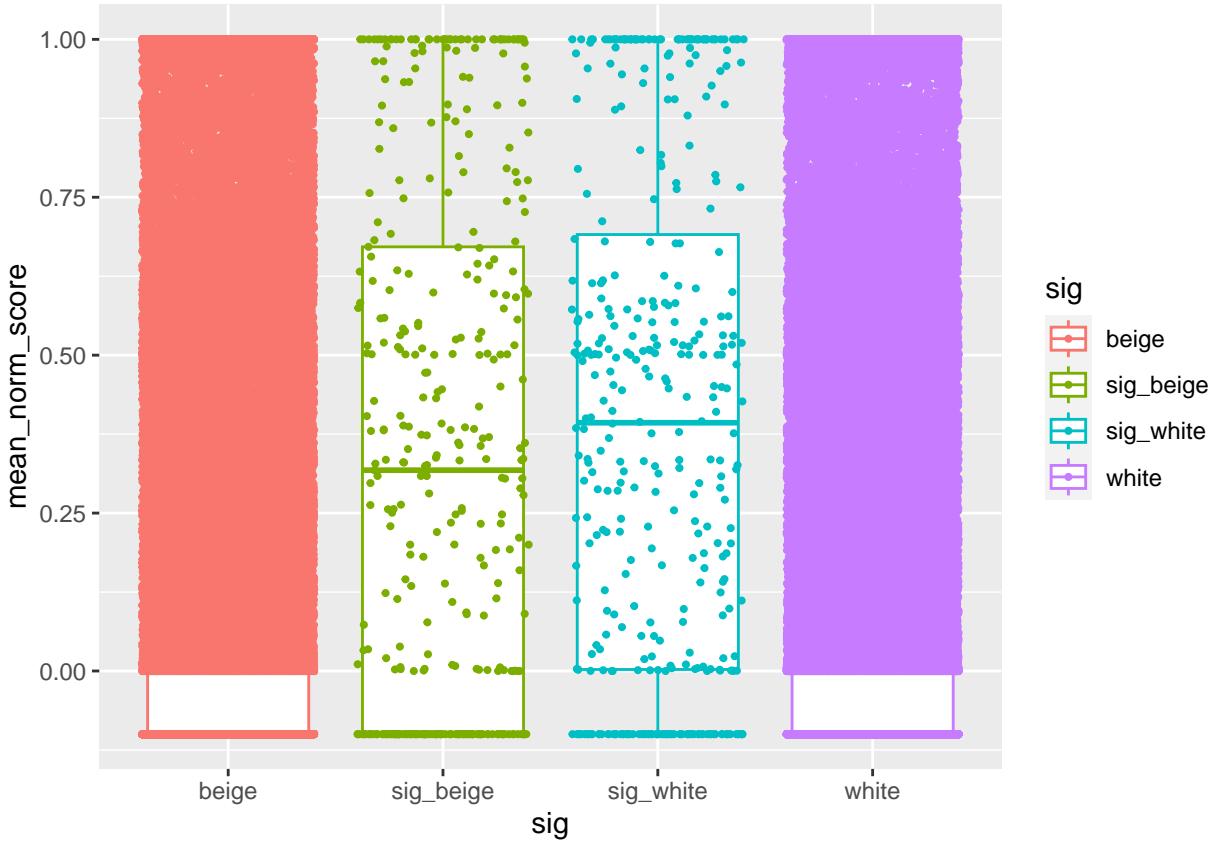
##
##      beige sig_beige sig_white     white
##      63426       385      392    68384

ggplot(all_introns, aes(x=sig, y=mean_norm_score, fill=sig)) + geom_violin() +
  geom_jitter(size=0.75)

```



```
ggplot(all_introns, aes(x= sig, y=mean_norm_score, colour=sig)) + geom_boxplot() +  
  geom_jitter(size=0.75)
```



The non-averaged score would be appropriate if the number of transcripts per intron was similar between the groups. Trifid scores (more appropriate if we're comparing across gene instead of pairwise between within each gene); has the average score HIGHER in white than beige. also a nice validation of the psi threshold.

```
all_introns[grep("NDUF", all_introns$gene) & all_introns$p.adjust < 0.05 & abs(all_introns$deltapsi)>0

## # A tibble: 4 x 14
## # Groups:   intron_coords, condition, deltapsi, p.adjust, gene, cluster_id [4]
##   in_trifid      intron_coords condition deltapsi p.adjust gene cluster_id
##   <chr>          <chr>       <chr>     <dbl>    <dbl>   <chr> <chr>
## 1 in_trifid    chr11:475657~ white     -0.103  8.65e-5 NDUFA~ clu_38578~
## 2 in_trifid    chr20:138089~ beige      0.103  1.01e-4 NDUFB~ clu_6504_+
## 3 cluster_not_in_trifid chr1:1611971~ beige      0.136  5.55e-8 NDUFC~ clu_14507~
## 4 cluster_not_in_trifid chr1:1611974~ white     -0.136  5.55e-8 NDUFD~ clu_14507~

## # i 7 more variables: annotation <chr>, mean_trifid_score <dbl>,
## #   mean_norm_score <dbl>, median_norm_score <dbl>, transcripts <chr>,
## #   transcript_names <chr>, sig <chr>

write.table(all_introns, here("31_leafcutter/trifid_all_introns.tsv"), sep="\t", quote=F, row.names = F)
```

Select significant introns + alt introns

```

alt_introns = read.delim(here("31_leafcutter/alt_introns_195.tsv"))
alt_introns = unite(alt_introns, "intron_coords", chr, start, end, strand, sep = ":")
sig_junctions = filter(all_introns, (p.adjust < 0.05 & abs(deltapsi) > 0.1) |
                           intron_coords %in% alt_introns$intron_coords) %>% arrange(p.adjust)
nrow(sig_junctions)

## [1] 1009

table(sig_junctions$sig)

##
##      beige sig_beige sig_white     white
##      118       385       392       114

write.table(sig_junctions, here("31_leafcutter/trifid_with_alt_introns.tsv"), sep="\t", quote=F, row.names=F)

```

## Plot the TRIFID difference

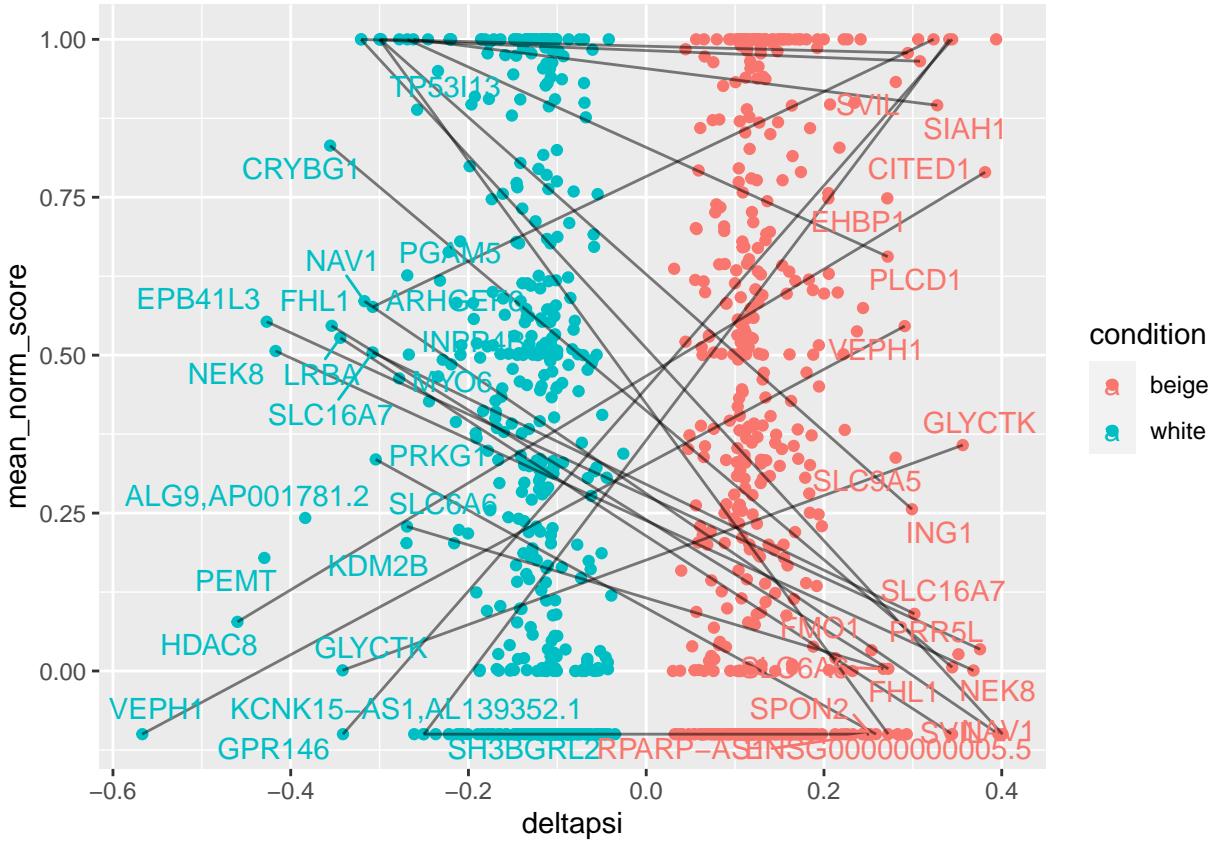
```

ggplot(sig_junctions, aes(colour=condition, y=mean_norm_score, x=deltapsi)) + geom_jitter() +
  geom_line(data=filter(sig_junctions, abs(deltapsi) > 0.25), aes(group=cluster_id), colour="black", size=1) +
  geom_text_repel(data=filter(sig_junctions, abs(deltapsi) > 0.25), aes(label=gene))

## Warning: Removed 5 rows containing missing values ('geom_text_repel()').

## Warning: ggrepel: 18 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



## Calculate the TRIFID difference

```
#if three introns are significant for a cluster
#average TRIFID score between the two introns in the same direction

paste_uq = function(x){
  return = paste(unique(x), collapse=",")
}

mean_of_3s = filter(sig_junctions, in_trifid != "cluster_not_in_trifid") %>%
  group_by(cluster_id, condition, p.adjust) %>% summarise(mean_norm_score = mean(mean_norm_score),
  deltapsi = mean(deltapsi),
  gene = paste(unique(gene), collapse=","),
  transcript_names= paste(unique(transcript_names)))

## `summarise()` has grouped output by 'cluster_id', 'condition'. You can override
## using the '.groups' argument.

head(mean_of_3s)

## # A tibble: 6 x 7
## # Groups:   cluster_id, condition [6]
##   cluster_id  condition p.adjust mean_norm_score deltapsi gene transcript_names
##   <chr>        <chr>      <dbl>          <dbl>      <dbl> <chr> <chr>
## 1 1             1          0.001       -0.001     -0.001 1       1
## 2 1             2          0.001       -0.001     -0.001 1       2
## 3 2             1          0.001       -0.001     -0.001 2       1
## 4 2             2          0.001       -0.001     -0.001 2       2
## 5 3             1          0.001       -0.001     -0.001 3       1
## 6 3             2          0.001       -0.001     -0.001 3       2
```

```

## 1 clu_10104_- beige      1.44e- 2      1      0.0562 DENN~ DENND1A-203
## 2 clu_10104_- white     1.44e- 2      0.316 -0.108 DENN~ DENND1A-202,DEN-
## 3 clu_10181_- beige     8.76e-15     0.513  0.0625 GOLG~ GOLGA2-214,GOLG-
## 4 clu_10181_- white     8.76e-15     0      -0.113 GOLG~ GOLGA2-203
## 5 clu_10209_- beige     2.10e- 8      0.502  0.107 CRAT  CRAT-201,CRAT-2~
## 6 clu_10209_- white     2.10e- 8      0.0049 -0.0449 CRAT  CRAT-207

filter(mean_of_3s, is.na(deltapsi))

## # A tibble: 0 x 7
## # Groups:   cluster_id, condition [0]
## # i 7 variables: cluster_id <chr>, condition <chr>, p.adjust <dbl>,
## #   mean_norm_score <dbl>, deltapsi <dbl>, gene <chr>, transcript_names <chr>

trifid_diff = pivot_wider(mean_of_3s, names_from = condition, values_from = c(deltapsi, mean_norm_score
                           id_cols=c("cluster_id", "p.adjust"))
head(trifid_diff); nrow(trifid_diff)#420 clusters -


## # A tibble: 6 x 8
## # Groups:   cluster_id [6]
##   cluster_id  p.adjust deltapsi_beige deltapsi_white mean_norm_score_beige
##   <chr>        <dbl>       <dbl>        <dbl>           <dbl>
## 1 clu_10104_- 1.44e- 2      0.0562     -0.108          1
## 2 clu_10181_- 8.76e-15     0.0625     -0.113          0.513
## 3 clu_10209_- 2.10e- 8      0.107      -0.0449         0.502
## 4 clu_10638_+ 1.87e- 2      0.106      -0.122          0.551
## 5 clu_10654_+ 1.08e- 7      0.192      -0.192          -0.1
## 6 clu_10672_+ 1.39e-62     0.241      -0.246          1
## # i 3 more variables: mean_norm_score_white <dbl>, gene_beige <chr>,
## #   gene_white <chr>

## [1] 420

#should be gencode + refseq - cluster_not_in_trifid
#470 - 74


## If your other intron is not in trifid speak now
## deltapsi we should have from the other table, just the trifid score we could set to -1
filter(trifid_diff, is.na(deltapsi_beige) | is.na(deltapsi_white)) # just 2 with an intron pair missi


## # A tibble: 2 x 8
## # Groups:   cluster_id [2]
##   cluster_id  p.adjust deltapsi_beige deltapsi_white mean_norm_score_beige
##   <chr>        <dbl>       <dbl>        <dbl>           <dbl>
## 1 clu_30508_+ 2.04e-34      0.149       NA            -0.1
## 2 clu_3320_-  1.58e-16      NA          -0.104         NA
## # i 3 more variables: mean_norm_score_white <dbl>, gene_beige <chr>,
## #   gene_white <chr>
```

```

sig_junctions[sig_junctions$cluster_id %in% c("clu_30508_+","clu_3320_-"),] #because the deltapsi goes i

## # A tibble: 4 x 14
## # Groups:   intron_coords, condition, deltapsi, p.adjust, gene, cluster_id [4]
##   in_trifid     intron_coords     condition deltapsi p.adjust gene  cluster_id
##   <chr>          <chr>           <chr>        <dbl>    <dbl> <chr> <chr>
## 1 not_in_trifid chr15:62570849:625~ beige      0.0540  2.04e-34 <NA> clu_30508~
## 2 not_in_trifid chr15:62570852:625~ beige      0.243   2.04e-34 <NA> clu_30508~
## 3 in_trifid      chr3:12941865:1296~ white     -0.147   1.58e-16 IQSE~ clu_3320_-
## 4 in_trifid      chr3:12941865:1302~ white     -0.0618  1.58e-16 IQSE~ clu_3320_-
## # i 7 more variables: annotation <chr>, mean_trifid_score <dbl>,
## #   mean_norm_score <dbl>, median_norm_score <dbl>, transcripts <chr>,
## #   transcript_names <chr>, sig <chr>

trifid_diff = filter(trifid_diff,!is.na(deltapsi_beige ) & !is.na(deltapsi_white ))

trifid_diff[grep("NDUF",trifid_diff$gene_to_plot),]

## Warning: Unknown or uninitialized column: 'gene_to_plot'.

## # A tibble: 0 x 8
## # Groups:   cluster_id [0]
## # i 8 variables: cluster_id <chr>, p.adjust <dbl>, deltapsi_beige <dbl>,
## #   deltapsi_white <dbl>, mean_norm_score_beige <dbl>,
## #   mean_norm_score_white <dbl>, gene_beige <chr>, gene_white <chr>

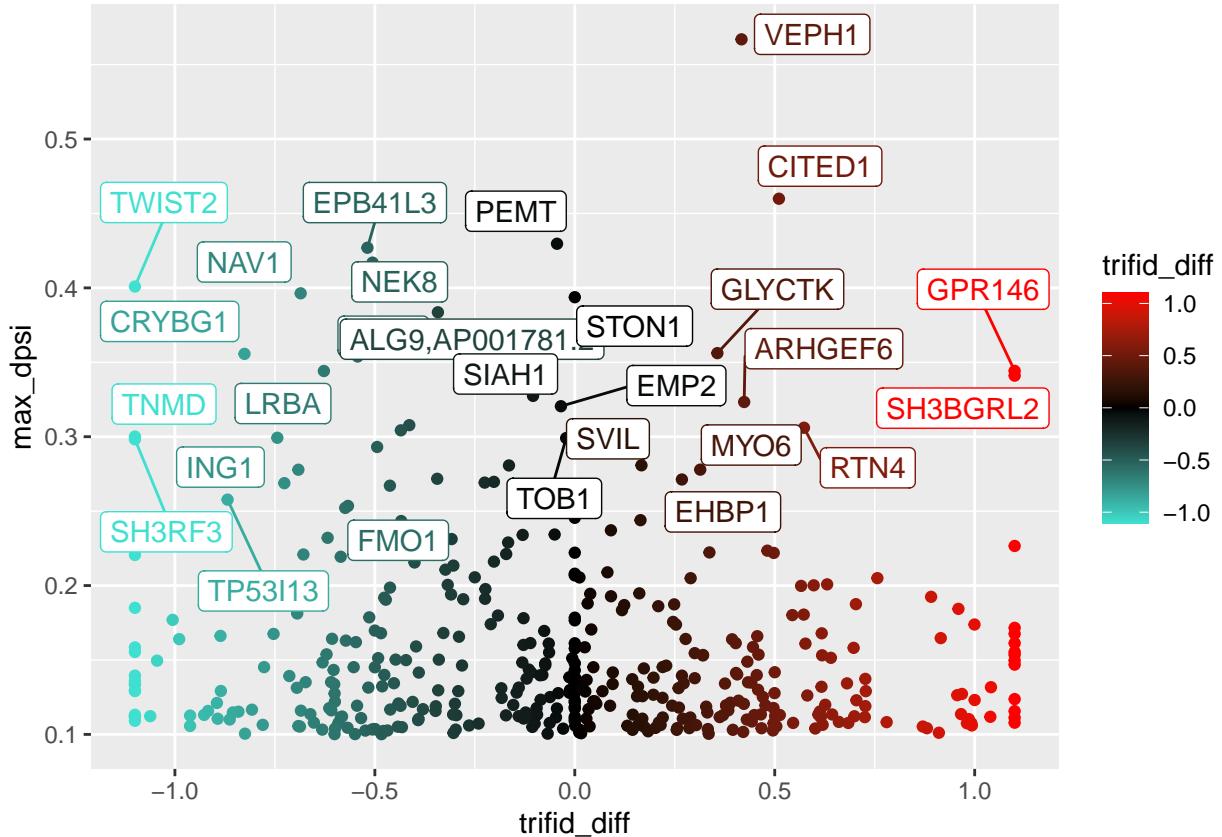
trifid_diff = mutate(trifid_diff, trifid_diff = mean_norm_score_beige-mean_norm_score_white, max_dpsi =
  gene_to_plot = if_else(trifid_diff > 0, gene_beige, gene_white))
head(trifid_diff)

## # A tibble: 6 x 11
## # Groups:   cluster_id [6]
##   cluster_id  p.adjust deltapsi_beige deltapsi_white mean_norm_score_beige
##   <chr>        <dbl>       <dbl>        <dbl>            <dbl>
## 1 clu_10104_- 1.44e- 2      0.0562      -0.108          1
## 2 clu_10181_- 8.76e-15     0.0625      -0.113          0.513
## 3 clu_10209_- 2.10e- 8      0.107       -0.0449         0.502
## 4 clu_10638_+ 1.87e- 2      0.106       -0.122          0.551
## 5 clu_10654_+ 1.08e- 7      0.192       -0.192          -0.1
## 6 clu_10672_+ 1.39e-62     0.241       -0.246          1
## # i 6 more variables: mean_norm_score_white <dbl>, gene_beige <chr>,
## #   gene_white <chr>, trifid_diff <dbl>, max_dpsi <dbl>, gene_to_plot <chr>

figs = here("R/plots")
ggplot(trifid_diff, aes(y=max_dpsi, x=trifid_diff, colour=trifid_diff)) + geom_point() +
  geom_label_repel(data=filter(trifid_diff, max_dpsi > 0.25), aes(label=gene_to_plot)) +
  scale_color_gradient2(low="turquoise", mid="black", high="red")

## Warning: ggrepel: 11 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



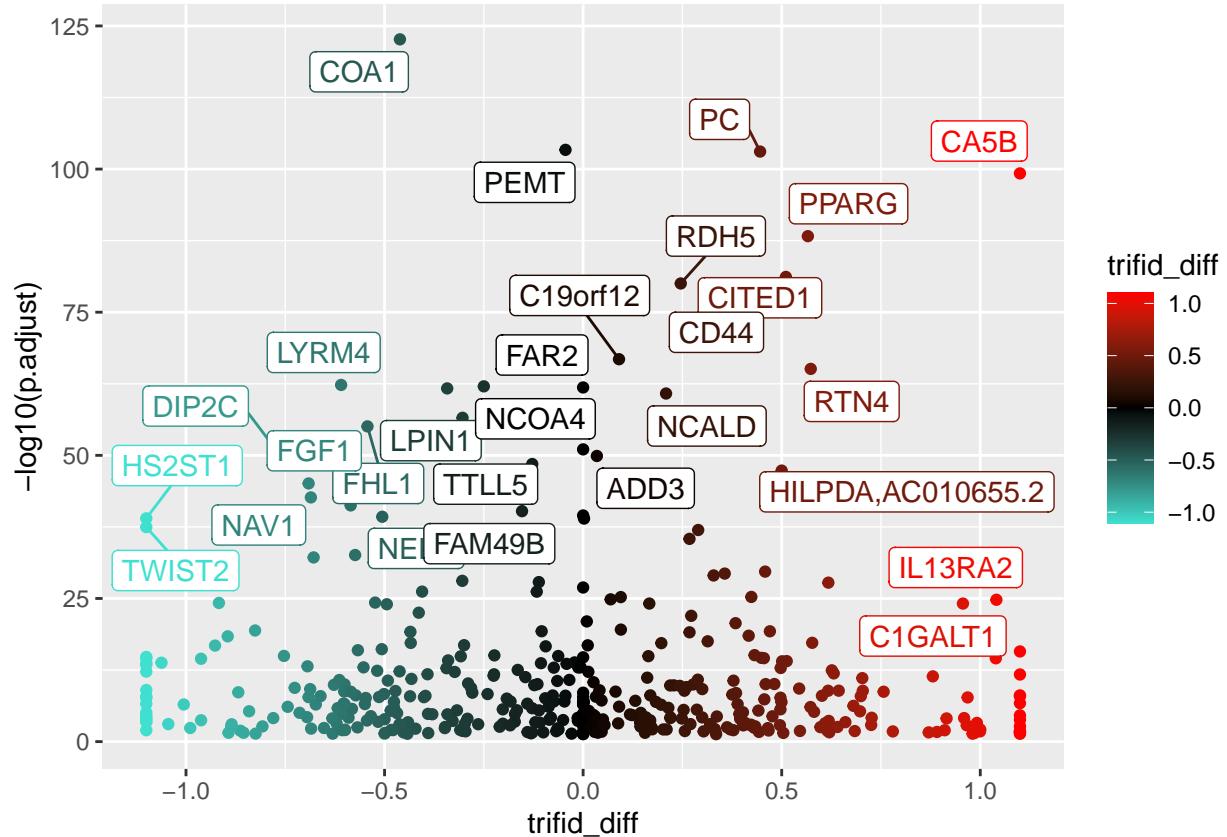
```
ggsave(file.path(figs, "trifid_difference_v_dpsi.pdf"))
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: ggrepel: 11 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
ggplot(trifid_diff, aes(y=-log10(p.adjust), x=trifid_diff, colour=trifid_diff)) + geom_point() +
  geom_label_repel(data=filter(trifid_diff, p.adjust < 0.01), aes(label=gene_to_plot)) +
  scale_color_gradient2(low="turquoise", mid="black", high="red")
```

```
## Warning: ggrepel: 336 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
ggsave(file.path(figs, "trifid_difference_v_pvalue.pdf"))
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: ggrepel: 336 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

Hmm.... CITED1 really shouldn't be here because we're comparing across genes which doesn't necessarily make sense.

```
write.table(trifid_diff, here("31_leafcutter/trifid_DIFFERENCE_with_Alt_introns.tsv"), sep="\t", quote=F)
```

## GSEA

### Setup

```
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   LINCO1691
## 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

rm(molsig)

shorten = function(ont) {
  abbreviate(gsub("_", " ", tolower(ont)), minlength=40, dot=T, named = F)
}

genelist = trifid_diff$trifid_diff
names(genelist) = trifid_diff$gene_to_plot
genelist = genelist[order(genelist, decreasing = T)]
length(genelist)

## [1] 418

summary(genelist)

##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## -1.10000 -0.46190  0.00000 -0.02781  0.37770  1.10000

gse = GSEA(genelist, TERM2GENE= some.molsig, pvalueCutoff=1)

## preparing geneSet collections...

## GSEA analysis...

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize,
## gseaParam, : There are duplicate gene names, fgsea may produce unexpected
## results.

## leading edge analysis...

## done...

```

```
head(gse)
```

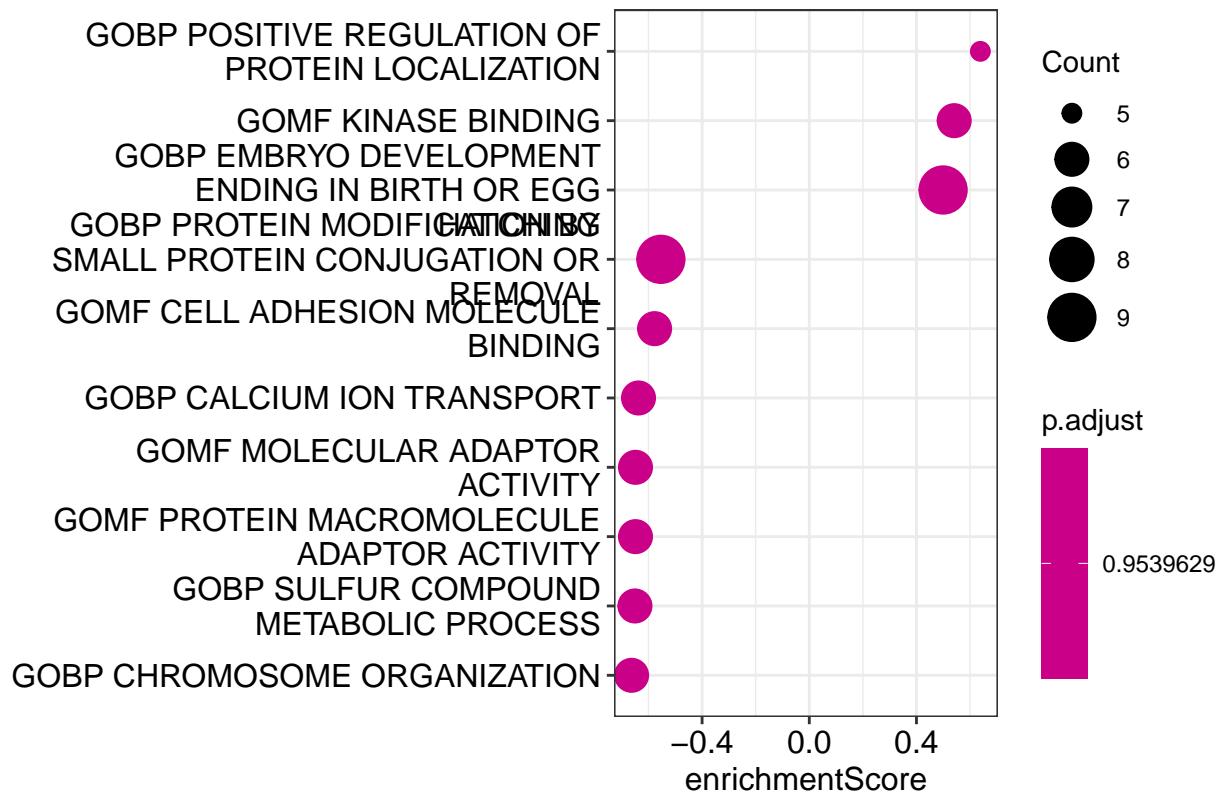
```
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL GOBP_PROTEIN_MODIFICATION_BY_SMALL_ GOBP_PROTEIN_MODIFICATION_BY_SMALL_  
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING GOBP_EMBRYO_DEVELOPMENT_1  
## GOBP_CHROMOSOME_ORGANIZATION  
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS  
## GOMF_KINASE_BINDING GOMF_KINASE_BINDING  
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY GOMF_MOLECULAR_ADAPTOR_ACTIVITY  
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL GOBP_PROTEIN_MODIFICATION_BY_SMALL_ GOBP_PROTEIN_MODIFICATION_BY_SMALL_  
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING GOBP_EMBRYO_DEVELOPMENT_1  
## GOBP_CHROMOSOME_ORGANIZATION  
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS  
## GOMF_KINASE_BINDING GOMF_KINASE_BINDING  
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY GOMF_MOLECULAR_ADAPTOR_ACTIVITY  
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL setSize 21  
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING 24  
## GOBP_CHROMOSOME_ORGANIZATION 10  
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS 11  
## GOMF_KINASE_BINDING 17  
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY 11  
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL enrichmentScore -0.5537734  
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING 0.4994379  
## GOBP_CHROMOSOME_ORGANIZATION -0.6631608  
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS -0.6505769  
## GOMF_KINASE_BINDING 0.5408111  
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY -0.6485722  
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL NES -1.721993  
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING 1.684443  
## GOBP_CHROMOSOME_ORGANIZATION -1.659121  
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS -1.655164  
## GOMF_KINASE_BINDING 1.653228  
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY -1.650064  
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL pvalue 0.009541377  
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING 0.014414858  
## GOBP_CHROMOSOME_ORGANIZATION 0.029342914  
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS 0.013004751  
## GOMF_KINASE_BINDING 0.020089297  
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY 0.013866936  
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL p.adjust 0.9539629  
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING 0.9539629  
## GOBP_CHROMOSOME_ORGANIZATION 0.9539629  
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS 0.9539629  
## GOMF_KINASE_BINDING 0.9539629  
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY 0.9539629  
##  
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL qvalue 0.9539629
```

```

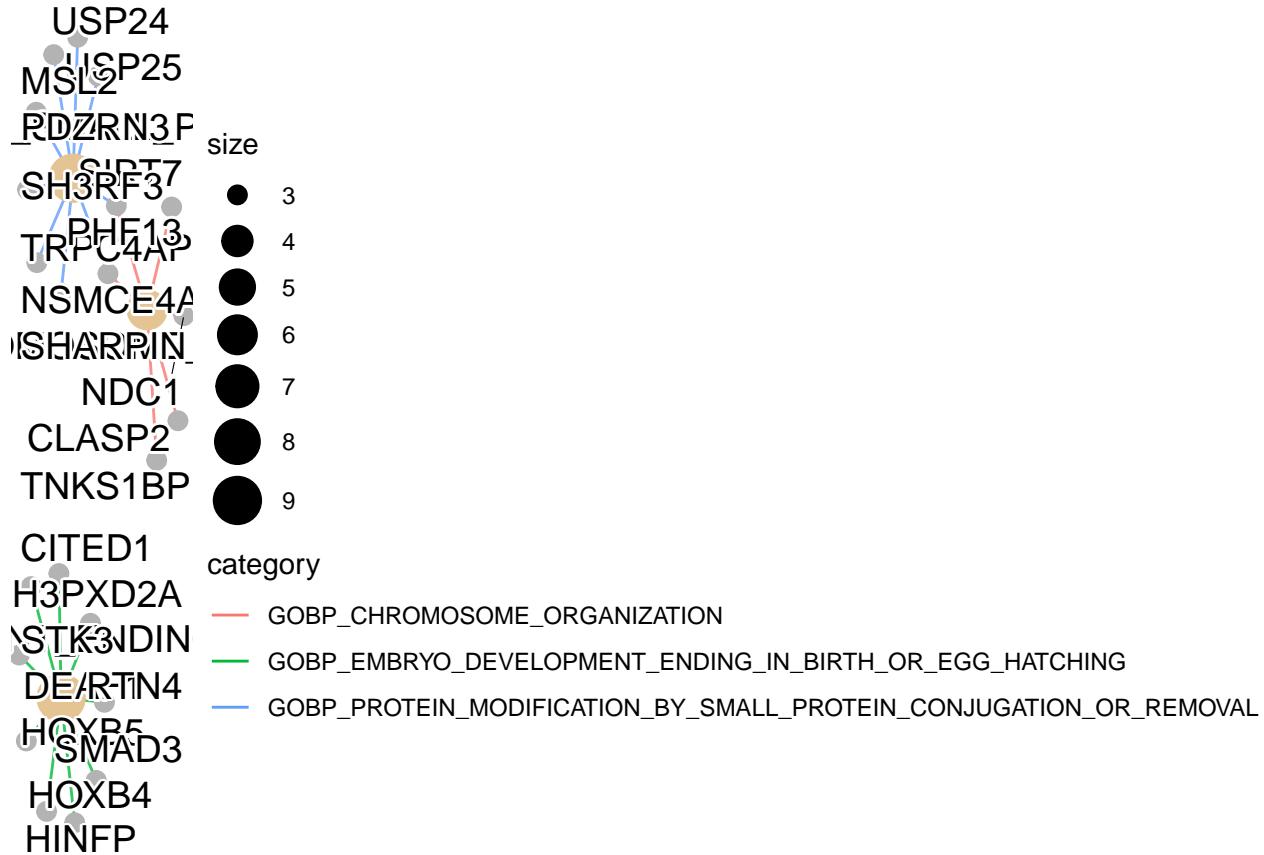
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING      0.9539629
## GOBP_CHROMOSOME_ORGANIZATION                                0.9539629
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS                      0.9539629
## GOMF_KINASE_BINDING                                         0.9539629
## GOMF_MOLECULAR_ADAPTER_ACTIVITY                           0.9539629
##                                                               rank
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL 73
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING       66
## GOBP_CHROMOSOME_ORGANIZATION                                63
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS                      85
## GOMF_KINASE_BINDING                                         65
## GOMF_MOLECULAR_ADAPTER_ACTIVITY                           87
##                                                               leading_edge
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL tags=43%, list=17%, signal=37%
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING       tags=38%, list=16%, signal=34%
## GOBP_CHROMOSOME_ORGANIZATION                                tags=60%, list=15%, signal=52%
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS                      tags=55%, list=20%, signal=45%
## GOMF_KINASE_BINDING                                         tags=33%, list=16%, signal=29%
## GOMF_MOLECULAR_ADAPTER_ACTIVITY                           tags=55%, list=21%, signal=44%
##                                                              
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL USP25/SHARPIN/NSMCE4A/SIRT7/TRPC4A/
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING        DEAF1/HOXB4/HOXB5/SMAD3/STK3/NSMCE4A/TNKS1B/
## GOBP_CHROMOSOME_ORGANIZATION                                ACSL1/FM01/
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS                      PRKRIP1/INKA/
## GOMF_KINASE_BINDING                                         VAMP8/EPB41/
## GOMF_MOLECULAR_ADAPTER_ACTIVITY

```

```
dotplot(gse, showCategory = 10, x="enrichmentScore")
```



```
cnetplot(gse, showCategory=3, categorySize="pvalue", color.params = list(edge=T))
```



```

to_print = gse[,c("ID", "setSize", "enrichmentScore", "NES",
                 "pvalue", "p.adjust", "qvalue", "core_enrichment", "rank")]
write.table(to_print, here("31_leafcutter", "TRIFID_GSEA.txt"))

custom_ora_to_df = function(res, annot=NULL, other_cols=NULL){
  res_df = res[,c(other_cols, "ID", "setSize", "enrichmentScore", "NES",
                  "pvalue", "p.adjust", "qvalue", "core_enrichment", "rank")]

  print(dim(res_df))
  if (length(annot)>1){
    res_df = merge(res_df, annot, by.x="ID", by.y="term", sort=F)
  }

  res_df = res_df[order(res_df$p.adjust),]
  return(res_df)
}

sep_go = list()
for (db in prefixes){
  print(db)
  t2g = some.molsig[grep(db, some.molsig$term),]
  ea = GSEA(geneList, TERM2GENE = t2g, pvalueCutoff = 1, minGSSize = 3)
  df = custom_ora_to_df(ea)
  print(head(df[2:8], n=10))
  sep_go[[db]] = ea
}

```

```

    #print(dotplot(ea, showCategory=20) +ggtitle(db))
    #print(cnetplot(ea, geneSetID = 1:5) + ggtitle(db))
}

## [1] "HALLMARK"

## preparing geneSet collections...

## GSEA analysis...

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize,
## gseaParam, : There are duplicate gene names, fgsea may produce unexpected
## results.

## leading edge analysis...

## done...

## [1] 29  9

##                                     setSize enrichmentScore      NES      pvalue
## HALLMARK_IL2_STAT5_SIGNALING          5     0.7267555  1.496353 0.07186858
## HALLMARK_COAGULATION                 3     0.8506621  1.427607 0.06144068
## HALLMARK_HEME_METABOLISM              8     0.5646877  1.367787 0.12244898
## HALLMARK_ANDROGEN_RESPONSE             6    -0.6259678 -1.338996 0.14044944
## HALLMARK_INTERFERON_GAMMA_RESPONSE      3    -0.7579490 -1.302771 0.13935970
## HALLMARK_MTORC1_SIGNALING                3    -0.7318584 -1.257926 0.16949153
## HALLMARK_TGF_BETA_SIGNALING               4     0.7008750  1.294078 0.20042194
## HALLMARK_UV_RESPONSE_DN                  7     0.4606469  1.084190 0.35394456
## HALLMARK_SPERMATOGENESIS                 4     0.5853886  1.080847 0.39451477
## HALLMARK_ESTROGEN_RESPONSE_LATE           9    -0.4464292 -1.075248 0.36678832
##                                     p.adjust      qvalue core_enrichment
## HALLMARK_IL2_STAT5_SIGNALING        0.8146067 0.8146067 SH3BGRL2/MY01C
## HALLMARK_COAGULATION                0.8146067 0.8146067 PECAM1
## HALLMARK_HEME_METABOLISM            0.8146067 0.8146067 PC/TNRC6B/SLC25A37
## HALLMARK_ANDROGEN_RESPONSE          0.8146067 0.8146067 TNFAIP8/INPP4B/PGM3
## HALLMARK_INTERFERON_GAMMA_RESPONSE   0.8146067 0.8146067 VAMP8/PTPN1
## HALLMARK_MTORC1_SIGNALING            0.8192090 0.8192090 SLC6A6/SLC1A4
## HALLMARK_TGF_BETA_SIGNALING          0.8303195 0.8303195 SMAD3
## HALLMARK_UV_RESPONSE_DN              0.9961240 0.9961240 SMAD3/PPARG/PMP22/DBP
## HALLMARK_SPERMATOGENESIS             0.9961240 0.9961240 IL13RA2
## HALLMARK_ESTROGEN_RESPONSE_LATE       0.9961240 0.9961240 AFF1/LAMC2/SLC1A4
## [1] "KEGG"

## preparing geneSet collections...

## GSEA analysis...

```

```

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are duplica

## leading edge analysis...

## done...

## [1] 19 9

## setSize enrichmentScore      NES
## KEGG_AXON_GUIDANCE          3     0.7769543  1.3379366
## KEGG_WNT_SIGNALING_PATHWAY   5     0.6065796  1.2483613
## KEGG_PANCREATIC_CANCER       3     0.6467982  1.1138042
## KEGG_INOSITOL_PHOSPHATE_METABOLISM 4     -0.5676856 -1.0728843
## KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM 4     -0.5676856 -1.0728843
## KEGG_FOCAL_ADHESION          3     -0.6205295 -1.0425137
## KEGG_CARDIAC_MUSCLE_CONTRACTION 3     -0.6154223 -1.0339334
## KEGG_PPAR_SIGNALING_PATHWAY   4     0.4922018  0.9307427
## KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS    3     -0.5381482 -0.9041099
## KEGG_PATHWAYS_IN_CANCER        9     0.3552171  0.8810921

## pvalue  p.adjust    qvalue
## KEGG_AXON_GUIDANCE          0.1149897 0.9425051 0.9425051
## KEGG_WNT_SIGNALING_PATHWAY   0.2274590 0.9425051 0.9425051
## KEGG_PANCREATIC_CANCER       0.3860370 0.9425051 0.9425051
## KEGG_INOSITOL_PHOSPHATE_METABOLISM 0.4034417 0.9425051 0.9425051
## KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM 0.4034417 0.9425051 0.9425051
## KEGG_FOCAL_ADHESION          0.4767442 0.9425051 0.9425051
## KEGG_CARDIAC_MUSCLE_CONTRACTION 0.4825581 0.9425051 0.9425051
## KEGG_PPAR_SIGNALING_PATHWAY   0.5530146 0.9425051 0.9425051
## KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS    0.6395349 0.9425051 0.9425051
## KEGG_PATHWAYS_IN_CANCER        0.6373166 0.9425051 0.9425051

## core_enrichment
## KEGG_AXON_GUIDANCE          RGS3
## KEGG_WNT_SIGNALING_PATHWAY   SMAD3/WNT5A
## KEGG_PANCREATIC_CANCER       SMAD3/ARHGEF6
## KEGG_INOSITOL_PHOSPHATE_METABOLISM  PLCD1/INPP4B/PI4KA
## KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM  PLCD1/INPP4B/PI4KA
## KEGG_FOCAL_ADHESION          MYLK/LAMC2
## KEGG_CARDIAC_MUSCLE_CONTRACTION  CACNA2D1/ATP1A2
## KEGG_PPAR_SIGNALING_PATHWAY   PPARG/ACADL
## KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS    PPP3CC/TOMM40L
## KEGG_PATHWAYS_IN_CANCER        SMAD3/STK36/PPARG/WNT5A/BID
## [1] "REACTOME"

## preparing geneSet collections...

## GSEA analysis...

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are duplica

```

```

## leading edge analysis...

## done...

## [1] 139    9

## setSize
## REACTOME_RAB_GEFS_EXCHANGE_GTP_FOR_GDP_ON_RABS      5
## REACTOME_RAB_REGULATION_OF_TRAFFICKING             5
## REACTOME_LEISHMANIA_INFECTATION                   4
## REACTOME_SUMOYLATION_OF_DNA_DAMAGE_RESPONSE_AND_REPAIR_PROTEINS 3
## REACTOME_INNATE_IMMUNE_SYSTEM                     12
## REACTOME_ESR_MEDIANED_SIGNALING                  3
## REACTOME_ESTROGEN_DEPENDENT_GENE_EXPRESSION       3
## REACTOME_G_ALPHA_I_SIGNALLING_EVENTS              3
## REACTOME_RHO_GTPASE_EFFECTORS                    5
## REACTOME_SARS_COV_2_INFECTION                   4
## enrichmentScore
## REACTOME_RAB_GEFS_EXCHANGE_GTP_FOR_GDP_ON_RABS     0.7217667
## REACTOME_RAB_REGULATION_OF_TRAFFICKING           0.7217667
## REACTOME_LEISHMANIA_INFECTATION                 0.7589379
## REACTOME_SUMOYLATION_OF_DNA_DAMAGE_RESPONSE_AND_REPAIR_PROTEINS -0.8265060
## REACTOME_INNATE_IMMUNE_SYSTEM                   0.5025444
## REACTOME_ESR_MEDIANED_SIGNALING                0.7952007
## REACTOME_ESTROGEN_DEPENDENT_GENE_EXPRESSION     0.7952007
## REACTOME_G_ALPHA_I_SIGNALLING_EVENTS            0.7800926
## REACTOME_RHO_GTPASE_EFFECTORS                  -0.6680017
## REACTOME_SARS_COV_2_INFECTION                  -0.7125604
## NES
## REACTOME_RAB_GEFS_EXCHANGE_GTP_FOR_GDP_ON_RABS   1.492899
## REACTOME_RAB_REGULATION_OF_TRAFFICKING          1.492899
## REACTOME_LEISHMANIA_INFECTATION                 1.412674
## REACTOME_SUMOYLATION_OF_DNA_DAMAGE_RESPONSE_AND_REPAIR_PROTEINS -1.390513
## REACTOME_INNATE_IMMUNE_SYSTEM                   1.366008
## REACTOME_ESR_MEDIANED_SIGNALING                1.354703
## REACTOME_ESTROGEN_DEPENDENT_GENE_EXPRESSION     1.354703
## REACTOME_G_ALPHA_I_SIGNALLING_EVENTS            1.328965
## REACTOME_RHO_GTPASE_EFFECTORS                  -1.323465
## REACTOME_SARS_COV_2_INFECTION                  -1.316028
## pvalue
## REACTOME_RAB_GEFS_EXCHANGE_GTP_FOR_GDP_ON_RABS   0.07024793
## REACTOME_RAB_REGULATION_OF_TRAFFICKING          0.07024793
## REACTOME_LEISHMANIA_INFECTATION                 0.09766454
## REACTOME_SUMOYLATION_OF_DNA_DAMAGE_RESPONSE_AND_REPAIR_PROTEINS 0.08187135
## REACTOME_INNATE_IMMUNE_SYSTEM                   0.12087912
## REACTOME_ESR_MEDIANED_SIGNALING                0.13008130
## REACTOME_ESTROGEN_DEPENDENT_GENE_EXPRESSION     0.13008130
## REACTOME_G_ALPHA_I_SIGNALLING_EVENTS            0.15040650
## REACTOME_RHO_GTPASE_EFFECTORS                  0.16570328
## REACTOME_SARS_COV_2_INFECTION                  0.15601504
## p.adjust
## REACTOME_RAB_GEFS_EXCHANGE_GTP_FOR_GDP_ON_RABS   0.9962406
## REACTOME_RAB_REGULATION_OF_TRAFFICKING          0.9962406
## REACTOME_LEISHMANIA_INFECTATION                 0.9962406
## REACTOME_SUMOYLATION_OF_DNA_DAMAGE_RESPONSE_AND_REPAIR_PROTEINS 0.9962406

```

```

## REACTOME_INNATE_IMMUNE_SYSTEM          0.9962406
## REACTOME_ESR_MEDIATED_SIGNALING       0.9962406
## REACTOME_ESTROGEN_DEPENDENT_GENE_EXPRESSION 0.9962406
## REACTOME_G_ALPHA_I_SIGNALLING_EVENTS   0.9962406
## REACTOME_RHO_GTPASE_EFFECTORS         0.9962406
## REACTOME_SARS_COV_2_INFECTATION      0.9962406
##                                         qvalue
## REACTOME_RAB_GEFS_EXCHANGE_GTP_FOR_GDP_ON_RABS 0.9962406
## REACTOME_RAB_REGULATION_OF_TRAFFICKING    0.9962406
## REACTOME_LEISHMANIA_INFECTION           0.9962406
## REACTOME_SUMOYLATION_OF_DNA_DAMAGE_RESPONSE_AND_REPAIR_PROTEINS 0.9962406
## REACTOME_INNATE_IMMUNE_SYSTEM           0.9962406
## REACTOME_ESR_MEDIATED_SIGNALING        0.9962406
## REACTOME_ESTROGEN_DEPENDENT_GENE_EXPRESSION 0.9962406
## REACTOME_G_ALPHA_I_SIGNALLING_EVENTS   0.9962406
## REACTOME_RHO_GTPASE_EFFECTORS         0.9962406
## REACTOME_SARS_COV_2_INFECTATION      0.9962406
##                                         core_enrichment
## REACTOME_RAB_GEFS_EXCHANGE_GTP_FOR_GDP_ON_RABS SBF2/SBF2/DENND1A
## REACTOME_RAB_REGULATION_OF_TRAFFICKING    SBF2/SBF2/DENND1A
## REACTOME_LEISHMANIA_INFECTION           MY01C/WNT5A
## REACTOME_SUMOYLATION_OF_DNA_DAMAGE_RESPONSE_AND_REPAIR_PROTEINS NSMCE4A/NDC1
## REACTOME_INNATE_IMMUNE_SYSTEM           PECAM1/MY01C/ATAD3B/ATP8B4
## REACTOME_ESR_MEDIATED_SIGNALING        CITED1/TNRC6B
## REACTOME_ESTROGEN_DEPENDENT_GENE_EXPRESSION  CITED1/TNRC6B
## REACTOME_G_ALPHA_I_SIGNALLING_EVENTS   RGS3
## REACTOME_RHO_GTPASE_EFFECTORS         CLASP2/FMNL2
## REACTOME_SARS_COV_2_INFECTATION      GPC3/MGAT1/NDC1
## [1] "WP"

## preparing geneSet collections...

## GSEA analysis...

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are dupl...
## leading edge analysis...

## done...

## [1] 58 9
##                                         setSize
## WP_THYROID_HORMONES_PRODUCTION_AND_PERIPHERAL_DOWNSTREAM_SIGNALING_EFFECTS 3
## WP_ANDROGEN_RECECTOR_SIGNALING_PATHWAY           3
## WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY 3
## WP_TGFBETA_SIGNALING_PATHWAY                    3
## WP_ALZHEIMERS_DISEASE                         4
## WP_ALZHEIMERS_DISEASE_AND_MIRNA_EFFECTS       4
## WP,GLYCEROLIPIDS_AND,GLYCERO,PHOSPHOLIPIDS    3

```

## WP_MARKERS_OF_KIDNEY_CELL_LINEAGE	3
## WP_EXERCISEINDUCED_CIRCADIAN_REGULATION	3
## WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II	4
##	enrichmentScore
## WP_THYROID_HORMONES_PRODUCTION_AND_PERIPHERAL_DOWNSTREAM_SIGNALING_EFFECTS	0.8578313
## WP_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY	0.8416832
## WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY	-0.8462425
## WP_TGFBETA_SIGNALING_PATHWAY	0.8131015
## WP_ALZHEIMERS_DISEASE	0.7173236
## WP_ALZHEIMERS_DISEASE_AND_MIRNA_EFFECTS	0.7173236
## WP_GLYCEROLIPIDS_AND,GLYCEROPHOSPHOLIPIDS	0.7880114
## WP_MARKERS_OF_KIDNEY_CELL_LINEAGE	0.7734940
## WP_EXERCISEINDUCED_CIRCADIAN_REGULATION	0.7571049
## WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II	-0.6974097
##	NES
## WP_THYROID_HORMONES_PRODUCTION_AND_PERIPHERAL_DOWNSTREAM_SIGNALING_EFFECTS	1.445693
## WP_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY	1.418479
## WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY	-1.389019
## WP_TGFBETA_SIGNALING_PATHWAY	1.370310
## WP_ALZHEIMERS_DISEASE	1.350661
## WP_ALZHEIMERS_DISEASE_AND_MIRNA_EFFECTS	1.350661
## WP_GLYCEROLIPIDS_AND,GLYCEROPHOSPHOLIPIDS	1.328026
## WP_MARKERS_OF_KIDNEY_CELL_LINEAGE	1.303560
## WP_EXERCISEINDUCED_CIRCADIAN_REGULATION	1.275940
## WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II	-1.267417
##	pvalue
## WP_THYROID_HORMONES_PRODUCTION_AND_PERIPHERAL_DOWNSTREAM_SIGNALING_EFFECTS	0.05895197
## WP_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY	0.07641921
## WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY	0.06776557
## WP_TGFBETA_SIGNALING_PATHWAY	0.10480349
## WP_ALZHEIMERS_DISEASE	0.14035088
## WP_ALZHEIMERS_DISEASE_AND_MIRNA_EFFECTS	0.14035088
## WP_GLYCEROLIPIDS_AND,GLYCEROPHOSPHOLIPIDS	0.12445415
## WP_MARKERS_OF_KIDNEY_CELL_LINEAGE	0.15502183
## WP_EXERCISEINDUCED_CIRCADIAN_REGULATION	0.18122271
## WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II	0.19195612
##	p.adjust
## WP_THYROID_HORMONES_PRODUCTION_AND_PERIPHERAL_DOWNSTREAM_SIGNALING_EFFECTS	0.9679638
## WP_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY	0.9679638
## WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY	0.9679638
## WP_TGFBETA_SIGNALING_PATHWAY	0.9679638
## WP_ALZHEIMERS_DISEASE	0.9679638
## WP_ALZHEIMERS_DISEASE_AND_MIRNA_EFFECTS	0.9679638
## WP_GLYCEROLIPIDS_AND,GLYCEROPHOSPHOLIPIDS	0.9679638
## WP_MARKERS_OF_KIDNEY_CELL_LINEAGE	0.9679638
## WP_EXERCISEINDUCED_CIRCADIAN_REGULATION	0.9679638
## WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II	0.9679638
##	qvalue
## WP_THYROID_HORMONES_PRODUCTION_AND_PERIPHERAL_DOWNSTREAM_SIGNALING_EFFECTS	0.9679638
## WP_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY	0.9679638
## WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY	0.9679638
## WP_TGFBETA_SIGNALING_PATHWAY	0.9679638
## WP_ALZHEIMERS_DISEASE	0.9679638
## WP_ALZHEIMERS_DISEASE_AND_MIRNA_EFFECTS	0.9679638

```

## WP_GLYCEROLIPIDS_AND_GLYCEROPOHOSPHOLIPIDS          0.9679638
## WP_MARKERS_OF_KIDNEY_CELL_LINEAGE                   0.9679638
## WP_EXERCISEINDUCED_CIRCADIAN_REGULATION           0.9679638
## WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II    0.9679638
## core_enrichment
## WP_THYROID_HORMONES_PRODUCTION_AND_PERIPHERAL_DOWNSTREAM_SIGNALING_EFFECTS PNPLA2/THRA/PPARG
## WP_ANDROGEN_RECECTOR_SIGNALING_PATHWAY             SMAD3
## WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY     FGF1/LAMC2
## WP_TGFBETA_SIGNALING_PATHWAY                      SMAD3/CITED1
## WP_ALZHEIMERS_DISEASE                            RTN4/WNT5A/BID
## WP_ALZHEIMERS_DISEASE_AND_MIRNA_EFFECTS          RTN4/WNT5A/BID
## WP_GLYCEROLIPIDS_AND_GLYCEROPOHOSPHOLIPIDS        PNPLA2
## WP_MARKERS_OF_KIDNEY_CELL_LINEAGE                  PECAM1/CITED1/WNT5A
## WP_EXERCISEINDUCED_CIRCADIAN_REGULATION           QKI
## WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II    FM01/CHST11/HS2ST1
## [1] "GOBP"

## preparing geneSet collections...

## GSEA analysis...

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are duplica

## leading edge analysis...

## done...

## [1] 1245      9

## setSize
## GOBP_REGULATION_OF_BINDING                         8
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL 21
## GOBP_EMBRYONIC_SKELETAL_SYSTEM_DEVELOPMENT       8
## GOBP_MICROTUBULE_BASED_MOVEMENT                  6
## GOBP_POLYSACCHARIDE_METABOLIC_PROCESS            5
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS           11
## GOBP_CHROMOSOME_ORGANIZATION                     10
## GOBP_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION 10
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING 24
## GOBP_CARBOHYDRATE BIOSYNTHETIC_PROCESS          8
## enrichmentScore
## GOBP_REGULATION_OF_BINDING                        0.7549030
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL -0.5537734
## GOBP_EMBRYONIC_SKELETAL_SYSTEM_DEVELOPMENT       0.7156521
## GOBP_MICROTUBULE_BASED_MOVEMENT                 0.7797329
## GOBP_POLYSACCHARIDE_METABOLIC_PROCESS            -0.8425691
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS           -0.6505769
## GOBP_CHROMOSOME_ORGANIZATION                    -0.6631608
## GOBP_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION 0.6385118
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING 0.4994379

```

		-0.6997039
## GOBP_CARBOHYDRATE BIOSYNTHETIC PROCESS	NES	
## GOBP_REGULATION_OF_BINDING		1.850986
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL		-1.765383
## GOBP_EMBRYONIC_SKELETAL_SYSTEM_DEVELOPMENT		1.754745
## GOBP_MICROTUBULE_BASED_MOVEMENT		1.717462
## GOBP_POLYSACCHARIDE_METABOLIC_PROCESS		-1.698379
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS		-1.694532
## GOBP_CHROMOSOME_ORGANIZATION		-1.679926
## GOBP_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION		1.667273
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING		1.660378
## GOBP_CARBOHYDRATE BIOSYNTHETIC PROCESS		-1.656912
##	pvalue	
## GOBP_REGULATION_OF_BINDING		0.005453285
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL		0.008935050
## GOBP_EMBRYONIC_SKELETAL_SYSTEM_DEVELOPMENT		0.013483654
## GOBP_MICROTUBULE_BASED_MOVEMENT		0.008827424
## GOBP_POLYSACCHARIDE_METABOLIC_PROCESS		0.013723238
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS		0.012924795
## GOBP_CHROMOSOME_ORGANIZATION		0.015859052
## GOBP_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION		0.026316861
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING		0.009405051
## GOBP_CARBOHYDRATE BIOSYNTHETIC PROCESS		0.019863170
##	p.adjust	
## GOBP_REGULATION_OF_BINDING		0.9986399
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL		0.9986399
## GOBP_EMBRYONIC_SKELETAL_SYSTEM_DEVELOPMENT		0.9986399
## GOBP_MICROTUBULE_BASED_MOVEMENT		0.9986399
## GOBP_POLYSACCHARIDE_METABOLIC_PROCESS		0.9986399
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS		0.9986399
## GOBP_CHROMOSOME_ORGANIZATION		0.9986399
## GOBP_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION		0.9986399
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING		0.9986399
## GOBP_CARBOHYDRATE BIOSYNTHETIC PROCESS		0.9986399
##	qvalue	
## GOBP_REGULATION_OF_BINDING		0.9986399
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL		0.9986399
## GOBP_EMBRYONIC_SKELETAL_SYSTEM_DEVELOPMENT		0.9986399
## GOBP_MICROTUBULE_BASED_MOVEMENT		0.9986399
## GOBP_POLYSACCHARIDE_METABOLIC_PROCESS		0.9986399
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS		0.9986399
## GOBP_CHROMOSOME_ORGANIZATION		0.9986399
## GOBP_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION		0.9986399
## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING		0.9986399
## GOBP_CARBOHYDRATE BIOSYNTHETIC PROCESS		0.9986399
##		SMAD3/STK31
## GOBP_REGULATION_OF_BINDING		
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL	USP25/SHARPIN/NSMCE4A/SIRT7/TRPC4A	
## GOBP_EMBRYONIC_SKELETAL_SYSTEM_DEVELOPMENT		
## GOBP_MICROTUBULE_BASED_MOVEMENT		
## GOBP_POLYSACCHARIDE_METABOLIC_PROCESS		
## GOBP_SULFUR_COMPOUND_METABOLIC_PROCESS		
## GOBP_CHROMOSOME_ORGANIZATION		
## GOBP_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION		
##		ACSL1/FM01,
## GOBP_REGULATION_OF_BINDING		NSMCE4A/TNKS1B
## GOBP_PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION_OR_REMOVAL		PEC

```

## GOBP_EMBRYO_DEVELOPMENT_ENDING_IN_BIRTH_OR_EGG_HATCHING DEAF1/HOXB4/HOXB5/SMAD3/STK3/1
## GOBP_CARBOHYDRATE BIOSYNTHETIC PROCESS
## [1] "GOCC"

## preparing geneSet collections...

## GSEA analysis...

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are duplica

## leading edge analysis...

## done...

## [1] 175 9

##                                     setSize enrichmentScore      NES      pvalue
## GOCC_TIGHT_JUNCTION             5       0.8135593  1.722004 0.01530244
## GOCC_NUCLEAR_CHROMOSOME          5      -0.7814929 -1.569556 0.02431119
## GOCC_DISTAL_AXON                5      -0.7558595 -1.518074 0.04807692
## GOCC_CONTRACTILE_FIBER           7      -0.6570526 -1.495993 0.06921676
## GOCC_RECECTOR_COMPLEX            4       0.7630071  1.469504 0.06118143
## GOCC_CATALYTIC_COMPLEX           30      -0.4300823 -1.459151 0.04804270
## GOCC_VESICLE_MEMBRANE            31       0.4040671  1.450137 0.06018519
## GOCC_APICAL_JUNCTION_COMPLEX     6       0.6591580  1.439126 0.07708779
## GOCC_MAIN_AXON                  4      -0.7764133 -1.424091 0.06792453
## GOCC_SPINDLE_POLE                3       0.8192771  1.399689 0.07838983

##                                     p.adjust      qvalue
## GOCC_TIGHT_JUNCTION            0.9978355 0.9978355
## GOCC_NUCLEAR_CHROMOSOME         0.9978355 0.9978355
## GOCC_DISTAL_AXON               0.9978355 0.9978355
## GOCC_CONTRACTILE_FIBER          0.9978355 0.9978355
## GOCC_RECECTOR_COMPLEX           0.9978355 0.9978355
## GOCC_CATALYTIC_COMPLEX          0.9978355 0.9978355
## GOCC_VESICLE_MEMBRANE           0.9978355 0.9978355
## GOCC_APICAL_JUNCTION_COMPLEX    0.9978355 0.9978355
## GOCC_MAIN_AXON                 0.9978355 0.9978355
## GOCC_SPINDLE_POLE               0.9978355 0.9978355

##                                     GOCC_TIGHT_JUNCTION
##                                     GOCC_NUCLEAR_CHROMOSOME
##                                     GOCC_DISTAL_AXON
##                                     GOCC_CONTRACTILE_FIBER
##                                     GOCC_RECECTOR_COMPLEX
##                                     GOCC_CATALYTIC_COMPLEX
##                                     GOCC_VESICLE_MEMBRANE SBF2/PECAM1/FMN2/SBF2/ARHGAP32/MY01C/DENND1A/ATAD3B/ATP8B4/VPS4A/WNT5A/PHC2/SHARPIN
##                                     GOCC_APICAL_JUNCTION_COMPLEX
##                                     GOCC_MAIN_AXON
##                                     GOCC_SPINDLE_POLE
## [1] "GOMF"

```

```

## preparing geneSet collections...

## GSEA analysis...

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are ties in
## The order of those tied genes will be arbitrary, which may produce unexpected results.

## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are duplica

## leading edge analysis...

## done...

## [1] 196   9

##                                     setSize enrichmentScore      NES
## GOMF_MOLECULAR_TRANSDUCER_ACTIVITY          9       0.7041565 1.732447
## GOMF_HISTONE_BINDING                         9      -0.7038299 -1.729354
## GOMF_STRUCTURAL_MOLECULE_ACTIVITY           8       -0.7180670 -1.710835
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY             11      -0.6485722 -1.699577
## GOMF_PROTEIN_MACROMOLECULE_ADAPTOR_ACTIVITY 11      -0.6485722 -1.699577
## GOMF ubiquitin_like_protein_ligase_ACTIVITY    8      -0.7088733 -1.688930
## GOMF_modification_dependent_protein_binding  7      -0.7387111 -1.666998
## GOMF_cell_adhesion_molecule_binding          15      -0.5772856 -1.659491
## GOMF_kinase_binding                          17       0.5408111 1.612947
## GOMF_carbon_oxygen_lyase_ACTIVITY            3       0.8936254 1.512432
##                                     pvalue  p.adjust  qvalue
## GOMF_MOLECULAR_TRANSDUCER_ACTIVITY          0.01193079 0.4072646 0.395891
## GOMF_HISTONE_BINDING                         0.02054889 0.4072646 0.395891
## GOMF_STRUCTURAL_MOLECULE_ACTIVITY           0.01884710 0.4072646 0.395891
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY             0.01440160 0.4072646 0.395891
## GOMF_PROTEIN_MACROMOLECULE_ADAPTOR_ACTIVITY 0.01440160 0.4072646 0.395891
## GOMF ubiquitin_like_protein_ligase_ACTIVITY 0.02206127 0.4072646 0.395891
## GOMF_modification_dependent_protein_binding 0.01945766 0.4072646 0.395891
## GOMF_cell_adhesion_molecule_binding          0.01172809 0.4072646 0.395891
## GOMF_kinase_binding                          0.01930952 0.4072646 0.395891
## GOMF_carbon_oxygen_lyase_ACTIVITY            0.02488660 0.4072646 0.395891
##                                     core_enrichment
## GOMF_MOLECULAR_TRANSDUCER_ACTIVITY          PECAM1/GPR146/IL13RA2/MCC/MCC/THRA/PPARG
## GOMF_HISTONE_BINDING                         FAM156B/SBN02/BRD4/ING1/DEK/PHF13
## GOMF_STRUCTURAL_MOLECULE_ACTIVITY           EPB41L3/LAMC2/CMTM8/NDC1
## GOMF_MOLECULAR_ADAPTOR_ACTIVITY             VAMP8/EPB41L3/SYNE3/BRD4/TRPC4AP/NDC1
## GOMF_PROTEIN_MACROMOLECULE_ADAPTOR_ACTIVITY VAMP8/EPB41L3/SYNE3/BRD4/TRPC4AP/NDC1
## GOMF ubiquitin_like_protein_ligase_ACTIVITY SH3RF3/PDZRN3/MSL2
## GOMF_modification_dependent_protein_binding SHARPIN/FAM156B/BRD4/ING1/PHF13
## GOMF_cell_adhesion_molecule_binding          COBLL1/PI4KA/PTPN1/TNKS1BP1/FGF1/FMNL2
## GOMF_kinase_binding                          PRKRIP1/INKA1/SMAD3/STAU2/TTC28/GOLGA2
## GOMF_carbon_oxygen_lyase_ACTIVITY            CA5B/AUH

```