leafcutter_post_analysis

Volcano plot - Figure 2A and gene ontology - saved overlaps for later

```
library(EnhancedVolcano)
library(tidyr)
library(clusterProfiler)
library(enrichplot)
here::i_am("R/11_leafcutter_post_analysis.Rmd")
library(here)
```

```
leaf = read.delim(here("31_leafcutter","with_donor_info_leafcutter.txt"))
dim(leaf)
```

```
## [1] 132587     16
```

```
head(leaf)
```

```
##     chr  cluster_id                               intron     logef      white
## 1 chr17 clu_19605_-  chr17:17577027:17591531:clu_19605_- -2.2201955 0.80950340
## 2  chrX   clu_291_-      chrX:72330076:72351733:clu_291_- -1.4285887 0.65345837
## 3  chrX   clu_291_-      chrX:72302934:72307025:clu_291_-  1.4361095 0.09086171
## 4 chr11  clu_2011_- chr11:111844723:111853380:clu_2011_- -1.5980779 0.57274421
## 5  chr2 clu_11888_-   chr2:54987698:55010087:clu_11888_-  1.0163245 0.24681456
## 6 chr17 clu_19605_-  chr17:17577027:17577107:clu_19605_- -0.1469083 0.07023661
##       beige   deltapsi        intron_coords            cluster  status    loglr
## 1 0.3798527 -0.4296507    17577027:17591531 chr17:clu_19605_- Success 260.3903
## 2 0.1935577 -0.4599006    72330076:72351733   chrX:clu_291_- Success 217.4563
## 3 0.4721675  0.3813058    72302934:72307025   chrX:clu_291_- Success 217.4563
## 4 0.1891304 -0.3836138 111844723:111853380  chr11:clu_2011_- Success 156.8379
## 5 0.5528444  0.3060298    54987698:55010087  chr2:clu_11888_- Success 165.0909
## 6 0.2620459  0.1918093    17577027:17577107 chr17:clu_19605_- Success 260.3903
##   df            p     p.adjust              genes    score
## 1  7 2.726265e-108 4.360252e-104               PEMT 44.40890
## 2 12   1.505375e-85  6.878918e-82 CITED1,AL133500.1,HDAC8 37.32668
## 3 12   1.505375e-85  6.878918e-82 CITED1,AL133500.1,HDAC8 30.94772
## 4  5   1.147804e-65  2.039712e-62       ALG9,AP001781.2 23.66530
## 5  5   3.227202e-69  7.940654e-66               RTN4 19.92259
## 6  7 2.726265e-108 4.360252e-104               PEMT 19.82550
```

Using raster for the points to make the file size smaller.
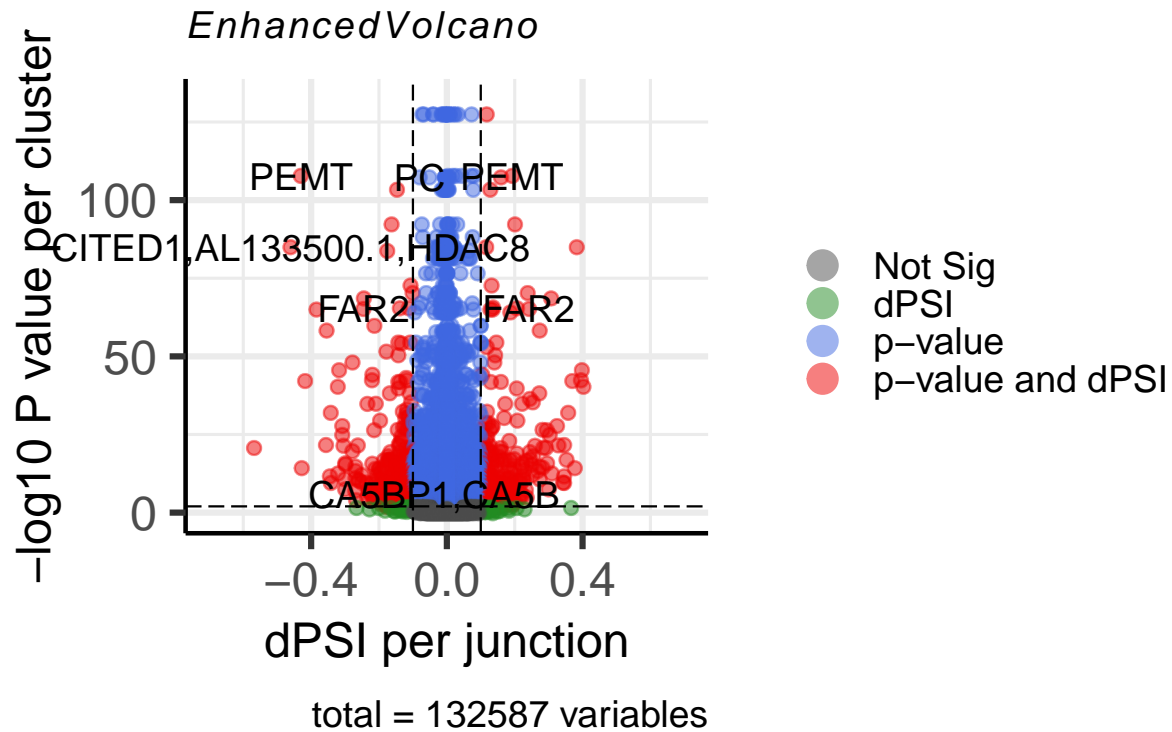
```
EnhancedVolcano(leaf, lab=leaf$genes, x="deltapsi", y="p",
                xlim=c(-0.7, 0.7), legendPosition = "right",
                pCutoffCol = "p.adjust", pCutoff = 0.05,
                FCcutoff = 0.1, ylab = "-log10 P value per cluster",
```

1

```
xlab="dPSI per junction", legendLabels = c("Not Sig", "dPSI", "p-value", "p-value and d
raster = T,
selectLab = grep("CITED1|PPARG|PEMT|FAR2|^PC$|CA5BP",leaf$genes, value=T)
)
```

Volcano plot


EnhancedVolcano

total = 132587 variables

```
ggsave(filename = here("R/plots", "volcanoplot.pdf"), width=8, height =7)

#random labelling
EnhancedVolcano(leaf, lab=leaf$genes, x="deltapsi", y="p",
                xlim=c(-0.7, 0.7), legendPosition = "right",
                pCutoffCol = "p.adjust", pCutoff = 0.05,
                FCcutoff = 0.1, ylab = "-log10 P value per cluster",
                raster=T,
                xlab="dPSI per junction",  legendLabels = c("Not Sig", "dPSI", "p-value", "p-value and
```

Volcano plot

EnhancedVolcano

total = 132587 variables
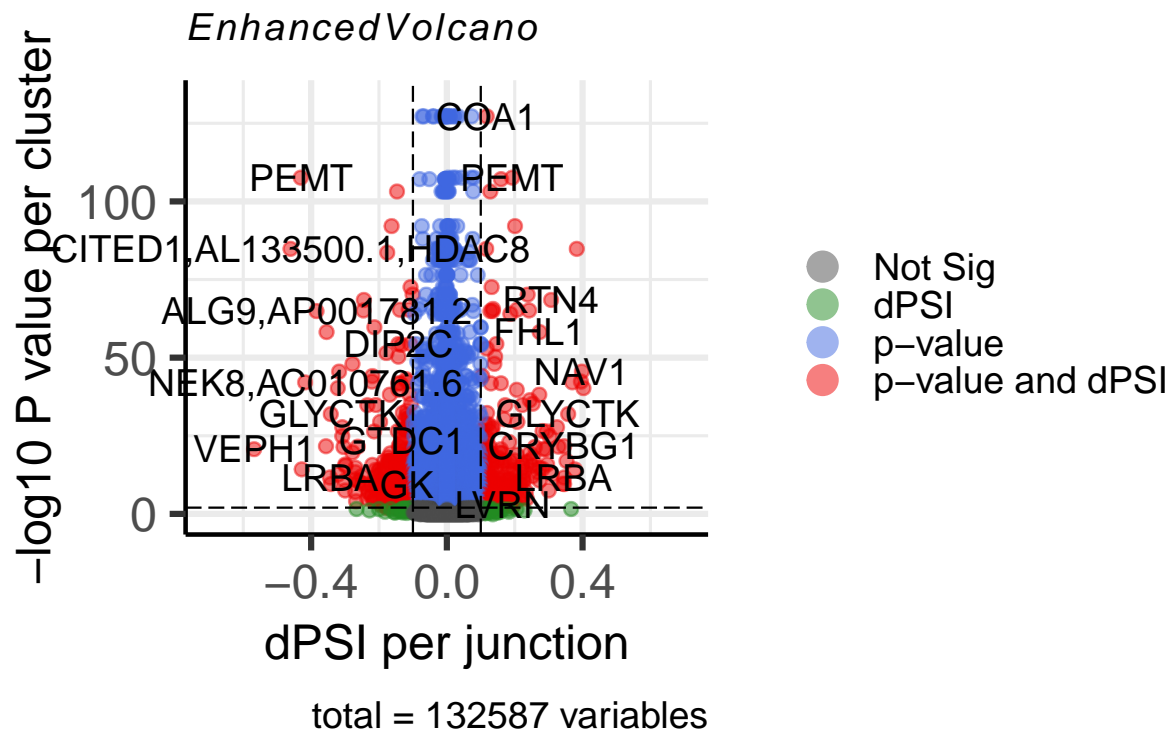
```
sig = filter(leaf, p.adjust < 0.05 & abs(deltapsi) > 0.1)
#tail(table(sig$genes))
```

Gene Ontology

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  LINC01691
## 6 chr1p11      NBPF26
```

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

```
prefixes = c("HALLMARK", "KEGG","REACTOME", "WP", "GOBP", "GOCC","GOMF")
colnames(molsig) = c("term","gene")
```

```r
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
```

```r
rm(molsig)

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

seperate gene clusters to gene names

```r
#robust = leaf[leaf$p.adjust < 0.05 & abs(leaf$deltapsi) > 0.1,]
lc_genes = separate_rows(leaf, genes, sep=",")
head(lc_genes)
```

```
## # A tibble: 6 x 16
##   chr   cluster_id   intron       logef  white beige deltapsi intron_coords cluster
##   <chr> <chr>        <chr>        <dbl>  <dbl> <dbl>    <dbl> <chr>          <chr>
## 1 chr17 clu_19605_-  chr17:175~  -2.22 0.810  0.380   -0.430 17577027:175~  chr17:~
## 2 chrX  clu_291_-    chrX:7233~  -1.43 0.653  0.194   -0.460 72330076:723~  chrX:c~
## 3 chrX  clu_291_-    chrX:7233~  -1.43 0.653  0.194   -0.460 72330076:723~  chrX:c~
## 4 chrX  clu_291_-    chrX:7233~  -1.43 0.653  0.194   -0.460 72330076:723~  chrX:c~
## 5 chrX  clu_291_-    chrX:7230~   1.44 0.0909 0.472    0.381 72302934:723~  chrX:c~
## 6 chrX  clu_291_-    chrX:7230~   1.44 0.0909 0.472    0.381 72302934:723~  chrX:c~
## # i 7 more variables: status <chr>, loglr <dbl>, df <int>, p <dbl>,
## #   p.adjust <dbl>, genes <chr>, score <dbl>
```

```r
nrow(lc_genes)
```

```
## [1] 150945
```

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

```
## [1] 12131
```

```r
lc_genes = lc_genes[order(lc_genes$p.adjust),]
```

Generic test

```r
go = enricher(unique(lc_genes$genes[lc_genes$p.adjust < 0.05 & abs(lc_genes$deltapsi) >= 0.1]),
              universe = unique(lc_genes$genes),
          TERM2GENE = some.molsig,
          pvalueCutoff = 1, qvalueCutoff = 0.9)
head(go)[3:9]
```

```
##                                             GeneRatio  BgRatio       pvalue
## GOBP_ACTIN_FILAMENT_BASED_PROCESS             37/410 484/9983 0.0001856485
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME       5/410  15/9983 0.0002433173
## GOMF_CALMODULIN_BINDING                       13/410 103/9983 0.0002895356
## GOCC_GOLGI_CIS_CISTERNA                        5/410  17/9983 0.0004683640
## GOBP_ACTIN_FILAMENT_ORGANIZATION              23/410 275/9983 0.0009433978
## WP_FATTY_ACID_BETAOXIDATION                    6/410  29/9983 0.0009827721
##                                             p.adjust    qvalue
## GOBP_ACTIN_FILAMENT_BASED_PROCESS          0.4529302 0.4529302
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME   0.4529302 0.4529302
## GOMF_CALMODULIN_BINDING                    0.4529302 0.4529302
## GOCC_GOLGI_CIS_CISTERNA                    0.5495081 0.5495081
## GOBP_ACTIN_FILAMENT_ORGANIZATION           0.6269841 0.6269841
## WP_FATTY_ACID_BETAOXIDATION                0.6269841 0.6269841
##
## GOBP_ACTIN_FILAMENT_BASED_PROCESS          ADD3/KANK1/SVIL/MYO1C/INPP5K/PRKG1/DIXDC1/MYO6/DPYSL3/IQSEC
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME
## GOMF_CALMODULIN_BINDING
## GOCC_GOLGI_CIS_CISTERNA
## GOBP_ACTIN_FILAMENT_ORGANIZATION
## WP_FATTY_ACID_BETAOXIDATION
##                                             Count
## GOBP_ACTIN_FILAMENT_BASED_PROCESS             37
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME       5
## GOMF_CALMODULIN_BINDING                       13
## GOCC_GOLGI_CIS_CISTERNA                        5
## GOBP_ACTIN_FILAMENT_ORGANIZATION              23
## WP_FATTY_ACID_BETAOXIDATION                    6
```

No significant terms but still a decent number of actin filament process genes have Diff splicing. Perhaps we need to run these separately.

Separate databases

```
colnames(go@result)
```

```
## [1] "ID"          "Description" "GeneRatio"   "BgRatio"      "pvalue"
## [6] "p.adjust"    "qvalue"      "geneID"      "Count"
```

```
custom_ora_to_df = function(res, annot=NULL, other_cols=NULL){
    res_df = res[,c(other_cols,"ID","GeneRatio", "BgRatio"  ,
                 "pvalue"    , "p.adjust" ,  "qvalue"    ,  "geneID"   ,   "Count" )]

    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)
}
```

```
summary(go[,c(NULL, "ID")])
```

```
##    Length     Class      Mode
##       162 character character
```

```
sep_go = list()
gl = unique(lc_genes$genes[lc_genes$p.adjust < 0.05 & abs(lc_genes$deltapsi) >= 0.1])
universe = unique(lc_genes$genes)
for (db in prefixes){
    print(db)
    t2g = some.molsig[grep(db, some.molsig$term),]
    ea = enricher(gl, universe= universe, TERM2GENE = t2g, pvalueCutoff = 1, qvalueCutoff = 0.5,minGSSi:
    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"
## [1] 9 8
##                                          GeneRatio  BgRatio      pvalue
## HALLMARK_MYOGENESIS                         13/118 126/2967 0.001212086
## HALLMARK_ESTROGEN_RESPONSE_EARLY            11/118 124/2967 0.009239323
## HALLMARK_MITOTIC_SPINDLE                    13/118 169/2967 0.015197367
## HALLMARK_ESTROGEN_RESPONSE_LATE              9/118 112/2967 0.032130256
## HALLMARK_ADIPOGENESIS                       12/118 181/2967 0.053146533
## HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY     4/118  37/2967 0.057154537
## HALLMARK_SPERMATOGENESIS                     5/118  56/2967 0.069607357
## HALLMARK_TGF_BETA_SIGNALING                  4/118  44/2967 0.095329227
## HALLMARK_ANDROGEN_RESPONSE                   6/118  80/2967 0.096300442
##                                             p.adjust     qvalue
## HALLMARK_MYOGENESIS                       0.05575597 0.05358697
## HALLMARK_ESTROGEN_RESPONSE_EARLY          0.21250443 0.20423767
## HALLMARK_MITOTIC_SPINDLE                  0.23302629 0.22396120
## HALLMARK_ESTROGEN_RESPONSE_LATE           0.36949795 0.35512388
## HALLMARK_ADIPOGENESIS                     0.43818478 0.42113869
## HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY  0.43818478 0.42113869
## HALLMARK_SPERMATOGENESIS                  0.45741978 0.43962542
## HALLMARK_TGF_BETA_SIGNALING               0.49220226 0.47305480
## HALLMARK_ANDROGEN_RESPONSE                0.49220226 0.47305480
##
## HALLMARK_MYOGENESIS                                             PC/LPIN1/FHL1/PFKM/FXYD1/ACSL1/SVIL/MYO1C/CRAT,
## HALLMARK_ESTROGEN_RESPONSE_EARLY                               CD44/ADD3/SVIL/TOB1/KAZN/SLC22A5/AFF1/TMEM164/:
## HALLMARK_MITOTIC_SPINDLE                   ARHGEF2/CLIP1/SPTBN1/PCGF5/AKAP13/BIN1/ARHGEF3/DST/NCK1/RAS/
## HALLMARK_ESTROGEN_RESPONSE_LATE                                          CD44/ADD3/TOB1/EMP2/SLC22A5/G/
## HALLMARK_ADIPOGENESIS                                          PEMT/PPARG/ENPP2/TOB1/BCL6/ACADL/CRAT/SLC19A1/NI
## HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY                                                                     I
## HALLMARK_SPERMATOGENESIS                                                                           LPIN1/IL13
## HALLMARK_TGF_BETA_SIGNALING                                                                             SMAL
## HALLMARK_ANDROGEN_RESPONSE                                                           NCOA4/TSC22D1/CAMKK:
##                                           Count
## HALLMARK_MYOGENESIS                          13
```

```
## HALLMARK_ESTROGEN_RESPONSE_EARLY                11
## HALLMARK_MITOTIC_SPINDLE                        13
## HALLMARK_ESTROGEN_RESPONSE_LATE                  9
## HALLMARK_ADIPOGENESIS                           12
## HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY         4
## HALLMARK_SPERMATOGENESIS                         5
## HALLMARK_TGF_BETA_SIGNALING                      4
## HALLMARK_ANDROGEN_RESPONSE                       6
## [1] "KEGG"
## [1] 1 8
##                                       GeneRatio BgRatio      pvalue  p.adjust
## KEGG_INOSITOL_PHOSPHATE_METABOLISM      6/106   40/2864 0.003070376 0.3316006
##                                         qvalue
## KEGG_INOSITOL_PHOSPHATE_METABOLISM 0.3316006
##                                                                         geneID
## KEGG_INOSITOL_PHOSPHATE_METABOLISM INPP5K/PLCD1/PLCB3/PI4KA/PIP5K1C/INPP4B
##                                    Count
## KEGG_INOSITOL_PHOSPHATE_METABOLISM     6
## [1] "REACTOME"
## [1] 0 8
## [1] GeneRatio BgRatio  pvalue   p.adjust  qvalue   geneID   Count
## <0 rows> (or 0-length row.names)
## [1] "WP"
## [1] 2 8
##                                         GeneRatio BgRatio      pvalue
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME   5/197   15/4496 0.000321810
## WP_FATTY_ACID_BETAOXIDATION                6/197   29/4496 0.001337412
##                                         p.adjust    qvalue
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME 0.1049101 0.1049101
## WP_FATTY_ACID_BETAOXIDATION              0.2179982 0.2179982
##                                                                         geneID
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME GOLGA8Q/GOLGA8H/ULK4P2/ULK4P1/GOLGA8O
## WP_FATTY_ACID_BETAOXIDATION                    ACSL1/ACADL/CPT1B/CRAT/GK/PNPLA2
##                                         Count
## WP_15Q133_COPY_NUMBER_VARIATION_SYNDROME     5
## WP_FATTY_ACID_BETAOXIDATION                  6
## [1] "GOBP"
## [1] 1 8
##                                    GeneRatio  BgRatio       pvalue p.adjust
## GOBP_ACTIN_FILAMENT_BASED_PROCESS    37/372   484/9243 0.0001187289 0.412108
##                                      qvalue
## GOBP_ACTIN_FILAMENT_BASED_PROCESS 0.4066777
##
## GOBP_ACTIN_FILAMENT_BASED_PROCESS ADD3/KANK1/SVIL/MYO1C/INPP5K/PRKG1/DIXDC1/MYO6/DPYSL3/IQSEC1/SH3KBI
##                                    Count
## GOBP_ACTIN_FILAMENT_BASED_PROCESS     37
## [1] "GOCC"
## [1] 52  8
##                                   GeneRatio  BgRatio       pvalue
## GOCC_GOLGI_CIS_CISTERNA              5/308    17/7873 0.0003722310
## GOCC_CELL_CORTEX                    17/308   184/7873 0.0008414744
## GOCC_CELL_PROJECTION_MEMBRANE       16/308   175/7873 0.0013281127
## GOCC_T_TUBULE                        5/308    25/7873 0.0024709528
## GOCC_MAIN_AXON                       5/308    26/7873 0.0029628746
```

```
## GOCC_TRANSPORTER_COMPLEX                     14/308 158/7873 0.0034904584
## GOCC_T_CELL_RECEPTOR_COMPLEX                   2/308   3/7873 0.0044584157
## GOCC_LEADING_EDGE_MEMBRANE                    10/308  99/7873 0.0051776565
## GOCC_CLATHRIN_COATED_ENDOCYTIC_VESICLE         6/308  43/7873 0.0061366814
## GOCC_BRUSH_BORDER_MEMBRANE                      4/308  20/7873 0.0067723230
##                                               p.adjust   qvalue
## GOCC_GOLGI_CIS_CISTERNA                       0.1421923 0.1324359
## GOCC_CELL_CORTEX                              0.1607216 0.1496939
## GOCC_CELL_PROJECTION_MEMBRANE                 0.1691130 0.1575095
## GOCC_T_TUBULE                                 0.2212155 0.2060370
## GOCC_MAIN_AXON                                0.2212155 0.2060370
## GOCC_TRANSPORTER_COMPLEX                      0.2212155 0.2060370
## GOCC_T_CELL_RECEPTOR_COMPLEX                  0.2212155 0.2060370
## GOCC_LEADING_EDGE_MEMBRANE                    0.2212155 0.2060370
## GOCC_CLATHRIN_COATED_ENDOCYTIC_VESICLE        0.2212155 0.2060370
## GOCC_BRUSH_BORDER_MEMBRANE                    0.2212155 0.2060370
##
## GOCC_GOLGI_CIS_CISTERNA
## GOCC_CELL_CORTEX                              ADD3/FGF1/MYO1C/MYO6/CLIP1/FMN2/NEDD4/MICAL3/EXOC6B/SPTB
## GOCC_CELL_PROJECTION_MEMBRANE        CD44/PEX19/KANK1/MYO1C/INPP5K/MYO6/ARHGEF2/EPB41L3/SLC6A6/SLC
## GOCC_T_TUBULE
## GOCC_MAIN_AXON
## GOCC_TRANSPORTER_COMPLEX                      FXYD1/LRRC8B/LRRC8C/ATP11C/PTPA/NDUFS2/CACNB3/
## GOCC_T_CELL_RECEPTOR_COMPLEX
## GOCC_LEADING_EDGE_MEMBRANE                                       CD44/KANK1/MYO1C/
## GOCC_CLATHRIN_COATED_ENDOCYTIC_VESICLE
## GOCC_BRUSH_BORDER_MEMBRANE
##                                               Count
## GOCC_GOLGI_CIS_CISTERNA                           5
## GOCC_CELL_CORTEX                                 17
## GOCC_CELL_PROJECTION_MEMBRANE                    16
## GOCC_T_TUBULE                                     5
## GOCC_MAIN_AXON                                    5
## GOCC_TRANSPORTER_COMPLEX                         14
## GOCC_T_CELL_RECEPTOR_COMPLEX                      2
## GOCC_LEADING_EDGE_MEMBRANE                       10
## GOCC_CLATHRIN_COATED_ENDOCYTIC_VESICLE            6
## GOCC_BRUSH_BORDER_MEMBRANE                        4
## [1] "GOMF"
## [1] 22  8
##                                                           GeneRatio  BgRatio
## GOMF_CALMODULIN_BINDING                                    13/350 103/8538
## GOMF_MONOATOMIC_ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY    7/350  44/8538
## GOMF_GALACTOSYLTRANSFERASE_ACTIVITY                         4/350  14/8538
## GOMF_ACTIN_BINDING                                         21/350 259/8538
## GOMF_PHOSPHATIDYLINOSITOL_PHOSPHATE_PHOSPHATASE_ACTIVITY    5/350  24/8538
## GOMF_CYCLIC_NUCLEOTIDE_BINDING                              3/350   8/8538
## GOMF_CARBONATE_DEHYDRATASE_ACTIVITY                         2/350   3/8538
## GOMF_COMPLEMENT_COMPONENT_C3B_BINDING                       2/350   3/8538
## GOMF_CATALYTIC_ACTIVITY_ACTING_ON_A_GLYCOPROTEIN            3/350  10/8538
## GOMF_PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY                   17/350 215/8538
##                                                               pvalue  p.adjust
## GOMF_CALMODULIN_BINDING                                    0.0002798409 0.1919709
## GOMF_MONOATOMIC_ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY   0.0018935943 0.3443594
```

```
## GOMF_GALACTOSYLTRANSFERASE_ACTIVITY                               0.0020034971 0.3443594
## GOMF_ACTIN_BINDING                                                0.0021674096 0.3443594
## GOMF_PHOSPHATIDYLINOSITOL_PHOSPHATE_PHOSPHATASE_ACTIVITY          0.0025099079 0.3443594
## GOMF_CYCLIC_NUCLEOTIDE_BINDING                                    0.0032792723 0.3749301
## GOMF_CARBONATE_DEHYDRATASE_ACTIVITY                               0.0048908712 0.4080324
## GOMF_COMPLEMENT_COMPONENT_C3B_BINDING                             0.0048908712 0.4080324
## GOMF_CATALYTIC_ACTIVITY_ACTING_ON_A_GLYCOPROTEIN                  0.0066096689 0.4080324
## GOMF_PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY                          0.0071150785 0.4080324
##                                                                        qvalue
## GOMF_CALMODULIN_BINDING                                            0.1852841
## GOMF_MONOATOMIC_ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY           0.3323646
## GOMF_GALACTOSYLTRANSFERASE_ACTIVITY                               0.3323646
## GOMF_ACTIN_BINDING                                                0.3323646
## GOMF_PHOSPHATIDYLINOSITOL_PHOSPHATE_PHOSPHATASE_ACTIVITY          0.3323646
## GOMF_CYCLIC_NUCLEOTIDE_BINDING                                    0.3618706
## GOMF_CARBONATE_DEHYDRATASE_ACTIVITY                               0.3938198
## GOMF_COMPLEMENT_COMPONENT_C3B_BINDING                             0.3938198
## GOMF_CATALYTIC_ACTIVITY_ACTING_ON_A_GLYCOPROTEIN                  0.3938198
## GOMF_PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY                          0.3938198
##
## GOMF_CALMODULIN_BINDING
## GOMF_MONOATOMIC_ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
## GOMF_GALACTOSYLTRANSFERASE_ACTIVITY
## GOMF_ACTIN_BINDING                                                NCALD/ADD3/TLN2/SVIL/MYO1C/DIXDC1/MYO6/LIMCI
## GOMF_PHOSPHATIDYLINOSITOL_PHOSPHATE_PHOSPHATASE_ACTIVITY
## GOMF_CYCLIC_NUCLEOTIDE_BINDING
## GOMF_CARBONATE_DEHYDRATASE_ACTIVITY
## GOMF_COMPLEMENT_COMPONENT_C3B_BINDING
## GOMF_CATALYTIC_ACTIVITY_ACTING_ON_A_GLYCOPROTEIN
## GOMF_PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY                          LPIN1/DUSP12/INPP5K/ENI
##                                                                  Count
## GOMF_CALMODULIN_BINDING                                             13
## GOMF_MONOATOMIC_ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY             7
## GOMF_GALACTOSYLTRANSFERASE_ACTIVITY                                 4
## GOMF_ACTIN_BINDING                                                 21
## GOMF_PHOSPHATIDYLINOSITOL_PHOSPHATE_PHOSPHATASE_ACTIVITY            5
## GOMF_CYCLIC_NUCLEOTIDE_BINDING                                      3
## GOMF_CARBONATE_DEHYDRATASE_ACTIVITY                                 2
## GOMF_COMPLEMENT_COMPONENT_C3B_BINDING                               2
## GOMF_CATALYTIC_ACTIVITY_ACTING_ON_A_GLYCOPROTEIN                    3
## GOMF_PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY                           17
```

Only two datasets contain significant hits

```
sig = dplyr::bind_rows(sep_go[["GOBP"]]@result, sep_go[["GOCC"]]@result)
sig = sig[order(sig$p.adjust),]
head(sig)
```

```
##                                                     ID
## GOCC_GOLGI_CIS_CISTERNA            GOCC_GOLGI_CIS_CISTERNA
## GOCC_CELL_CORTEX                      GOCC_CELL_CORTEX
## GOCC_CELL_PROJECTION_MEMBRANE GOCC_CELL_PROJECTION_MEMBRANE
## GOCC_T_TUBULE                            GOCC_T_TUBULE
```

```
## GOCC_MAIN_AXON                                      GOCC_MAIN_AXON
## GOCC_TRANSPORTER_COMPLEX              GOCC_TRANSPORTER_COMPLEX
##                                                  Description GeneRatio  BgRatio
## GOCC_GOLGI_CIS_CISTERNA            GOCC_GOLGI_CIS_CISTERNA      5/308   17/7873
## GOCC_CELL_CORTEX                          GOCC_CELL_CORTEX     17/308 184/7873
## GOCC_CELL_PROJECTION_MEMBRANE GOCC_CELL_PROJECTION_MEMBRANE   16/308 175/7873
## GOCC_T_TUBULE                                GOCC_T_TUBULE      5/308   25/7873
## GOCC_MAIN_AXON                              GOCC_MAIN_AXON      5/308   26/7873
## GOCC_TRANSPORTER_COMPLEX          GOCC_TRANSPORTER_COMPLEX     14/308 158/7873
##                                   pvalue  p.adjust    qvalue
## GOCC_GOLGI_CIS_CISTERNA        0.0003722310 0.1421923 0.1324359
## GOCC_CELL_CORTEX               0.0008414744 0.1607216 0.1496939
## GOCC_CELL_PROJECTION_MEMBRANE  0.0013281127 0.1691130 0.1575095
## GOCC_T_TUBULE                  0.0024709528 0.2212155 0.2060370
## GOCC_MAIN_AXON                 0.0029628746 0.2212155 0.2060370
## GOCC_TRANSPORTER_COMPLEX       0.0034904584 0.2212155 0.2060370
##
## GOCC_GOLGI_CIS_CISTERNA
## GOCC_CELL_CORTEX                         ADD3/FGF1/MYO1C/MYO6/CLIP1/FMN2/NEDD4/MICAL3/EXOC6B/SPTBN1/CLASP2/
## GOCC_CELL_PROJECTION_MEMBRANE CD44/PEX19/KANK1/MYO1C/INPP5K/MYO6/ARHGEF2/EPB41L3/SLC6A6/SLC7A8/SLC19/
## GOCC_T_TUBULE
## GOCC_MAIN_AXON
## GOCC_TRANSPORTER_COMPLEX                             FXYD1/LRRC8B/LRRC8C/ATP11C/PTPA/NDUFS2/CACNB3/CASQ2/ATP1
##                               Count
## GOCC_GOLGI_CIS_CISTERNA           5
## GOCC_CELL_CORTEX                 17
## GOCC_CELL_PROJECTION_MEMBRANE    16
## GOCC_T_TUBULE                     5
## GOCC_MAIN_AXON                    5
## GOCC_TRANSPORTER_COMPLEX         14
```
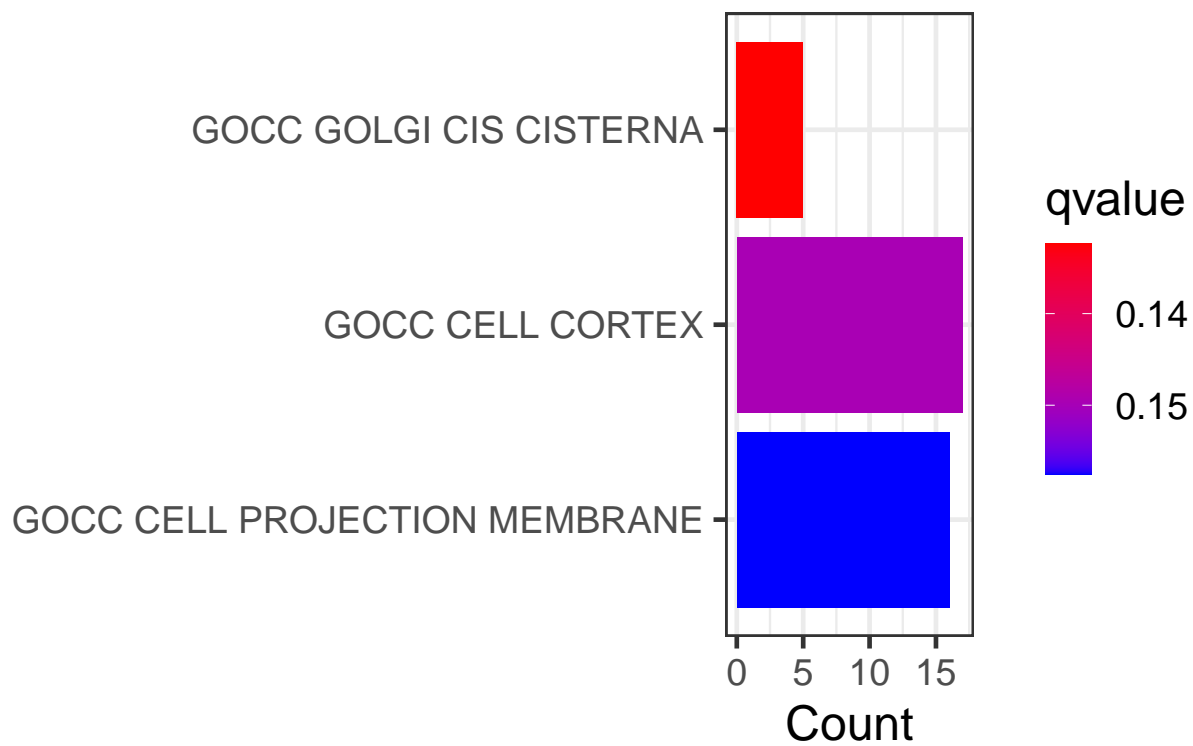
```r
sig_ob = multienrichjam::enrichDF2enrichResult(sig)
dotplot(sig_ob, color="qvalue", showCategory = 3)
```
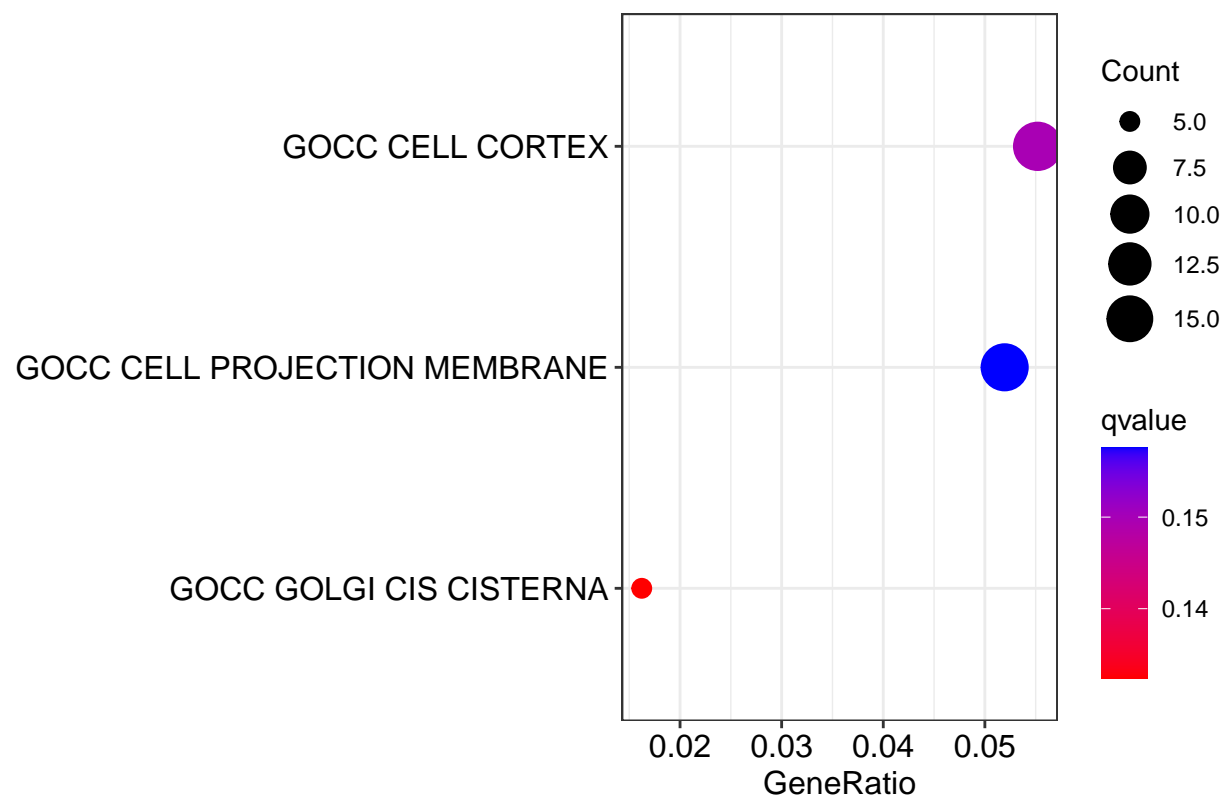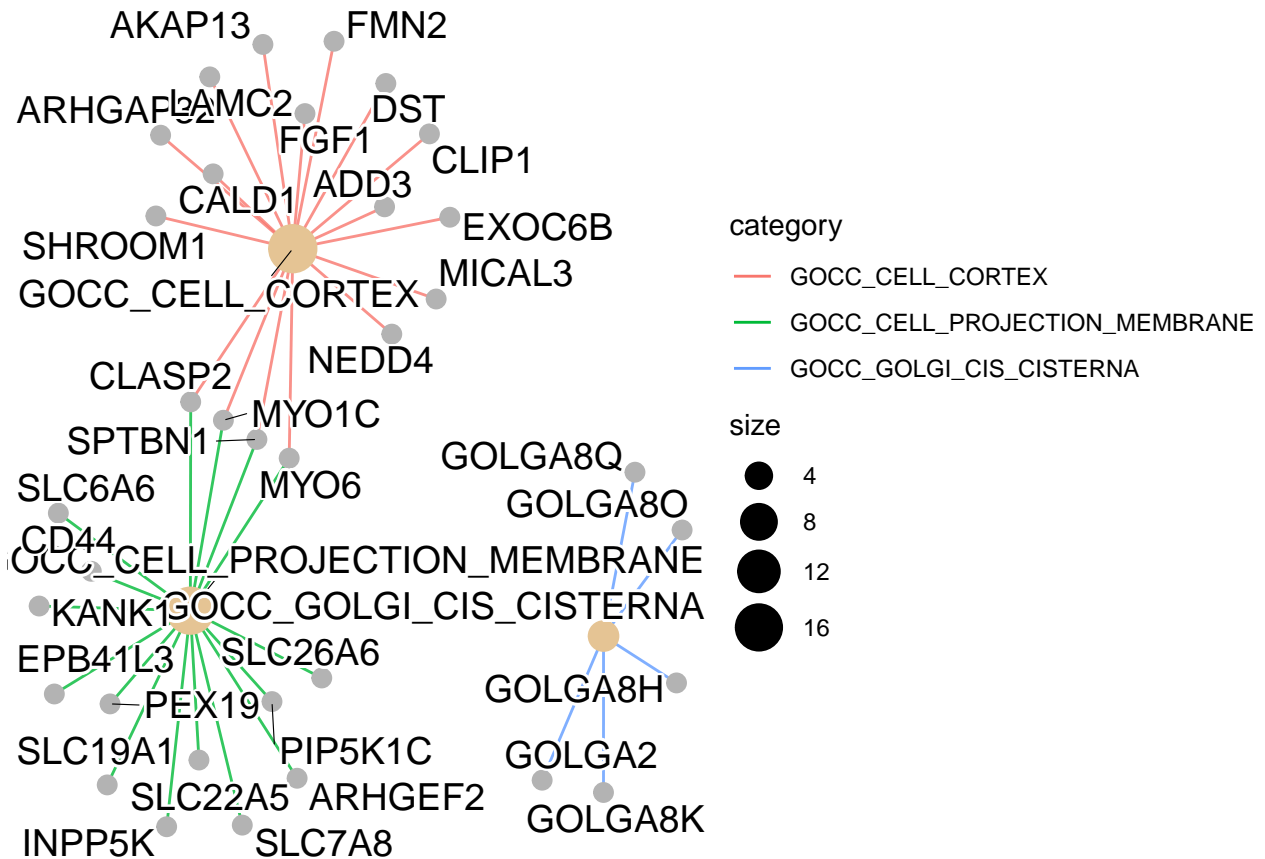
```
barplot(sig_ob, showCategory = 3, color = "qvalue") + theme_bw(base_size = 18)
```

```
ggsave(filename = here("R/plots", "geneontology.pdf"), width=8, height =3)
cnetplot(sig_ob, showCategory=3, categorySize="pvalue", color.params = list(edge=T))
```

The network diagram shows three categories connected to gene nodes:

category
- GOCC_CELL_CORTEX (red line)
- GOCC_CELL_PROJECTION_MEMBRANE (green line)
- GOCC_GOLGI_CIS_CISTERNA (blue line)

size
- 4
- 8
- 12
- 16

Gene labels: AKAP13, FMN2, ARHGAP, LAMC2, DST, FGF1, CLIP1, ADD3, CALD1, EXOC6B, SHROOM1, MICAL3, GOCC_CELL_CORTEX, NEDD4, CLASP2, MYO1C, SPTBN1, MYO6, SLC6A6, GOLGA8Q, GOLGA8O, CD44, GOCC_CELL_PROJECTION_MEMBRANE, GOCC_GOLGI_CIS_CISTERNA, KANK1, SLC26A6, EPB41L3, GOLGA8H, PEX19, SLC19A1, PIP5K1C, GOLGA2, SLC22A5, ARHGEF2, GOLGA8K, INPP5K, SLC7A8

```
save(sig_ob, file = here("R/plots", "GO_object.RData"))
```

Write table to get genes lists

```
to_print = sig_ob[,c("ID","GeneRatio", "BgRatio"  ,
                "pvalue"    , "p.adjust" ,  "qvalue"   , "geneID"   ,   "Count" )]
write.table(to_print, here("31_leafcutter","General_GO_ORA.txt"))
```

Thermogenic test

Limiting our scope to gene sets from DEGs in white and beige or thermogenic gene sets

```
grep_sets = some.molsig[grep("THERMO|BROWN|UNCOUPLING|COLD", some.molsig$term),]
from_rosi = some.molsig[grep("HALLMARK_FATTY_ACID_METABOLISM|HALLMARK_ADIPOGENESIS|GOBP_MONOCARBOXYLIC_

terms_to_test = rbind(grep_sets, from_rosi)
terms_to_test$term = factor(terms_to_test$term)

specific = enricher(unique(lc_genes$genes[lc_genes$p.adjust < 0.05 & abs(lc_genes$deltapsi) >= 0.1]),
            universe = unique(lc_genes$genes),
          TERM2GENE = terms_to_test,
          pvalueCutoff = 1, qvalueCutoff = 0.9)
specific = mutate(specific,Description = shorten(Description) )
head(specific, n=20)[3:9]
```

```
## 
## HALLMARK_ADIPOGENESIS
## GOBP_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOBP_ADAPTIVE_THERMOGENESIS
## WP_THERMOGENESIS
## HALLMARK_FATTY_ACID_METABOLISM
## REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY
## WP_DIFFERENTIATION_OF_WHITE_AND_BROWN_ADIPOCYTE
## GOBP_MONOCARBOXYLIC_ACID_METABOLIC_PROCESS
## GOBP_RESPONSE_TO_COLD
## GOBP_NEGATIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOCC_MITOCHONDRIAL_MATRIX
## 
## HALLMARK_ADIPOGENESIS
## GOBP_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOBP_ADAPTIVE_THERMOGENESIS
## WP_THERMOGENESIS
## HALLMARK_FATTY_ACID_METABOLISM
## REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY
## WP_DIFFERENTIATION_OF_WHITE_AND_BROWN_ADIPOCYTE
## GOBP_MONOCARBOXYLIC_ACID_METABOLIC_PROCESS
## GOBP_RESPONSE_TO_COLD
## GOBP_NEGATIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOCC_MITOCHONDRIAL_MATRIX
## 
## HALLMARK_ADIPOGENESIS
## GOBP_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOBP_ADAPTIVE_THERMOGENESIS
## WP_THERMOGENESIS
## HALLMARK_FATTY_ACID_METABOLISM
## REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY
## WP_DIFFERENTIATION_OF_WHITE_AND_BROWN_ADIPOCYTE
## GOBP_MONOCARBOXYLIC_ACID_METABOLIC_PROCESS
## GOBP_RESPONSE_TO_COLD
## GOBP_NEGATIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOCC_MITOCHONDRIAL_MATRIX
## 
## HALLMARK_ADIPOGENESIS
## GOBP_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOBP_ADAPTIVE_THERMOGENESIS
## WP_THERMOGENESIS
## HALLMARK_FATTY_ACID_METABOLISM
## REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY
## WP_DIFFERENTIATION_OF_WHITE_AND_BROWN_ADIPOCYTE
## GOBP_MONOCARBOXYLIC_ACID_METABOLIC_PROCESS
## GOBP_RESPONSE_TO_COLD
## GOBP_NEGATIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOCC_MITOCHONDRIAL_MATRIX
## 
## HALLMARK_ADIPOGENESIS
## GOBP_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOBP_ADAPTIVE_THERMOGENESIS
## WP_THERMOGENESIS
## HALLMARK_FATTY_ACID_METABOLISM
```

```
## REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_B
## WP_DIFFERENTIATION_OF_WHITE_AND_BROWN_ADIPOCYTE
## GOBP_MONOCARBOXYLIC_ACID_METABOLIC_PROCESS
## GOBP_RESPONSE_TO_COLD
## GOBP_NEGATIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOCC_MITOCHONDRIAL_MATRIX
##
## HALLMARK_ADIPOGENESIS
## GOBP_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOBP_ADAPTIVE_THERMOGENESIS
## WP_THERMOGENESIS
## HALLMARK_FATTY_ACID_METABOLISM
## REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_B
## WP_DIFFERENTIATION_OF_WHITE_AND_BROWN_ADIPOCYTE
## GOBP_MONOCARBOXYLIC_ACID_METABOLIC_PROCESS
## GOBP_RESPONSE_TO_COLD
## GOBP_NEGATIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOCC_MITOCHONDRIAL_MATRIX
##
## HALLMARK_ADIPOGENESIS
## GOBP_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOBP_ADAPTIVE_THERMOGENESIS
## WP_THERMOGENESIS
## HALLMARK_FATTY_ACID_METABOLISM
## REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_B
## WP_DIFFERENTIATION_OF_WHITE_AND_BROWN_ADIPOCYTE
## GOBP_MONOCARBOXYLIC_ACID_METABOLIC_PROCESS
## GOBP_RESPONSE_TO_COLD
## GOBP_NEGATIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS
## GOCC_MITOCHONDRIAL_MATRIX
```

```
#emapplot(pairwise_termsim(specific))
cnetplot(specific, showCategory=5)
```
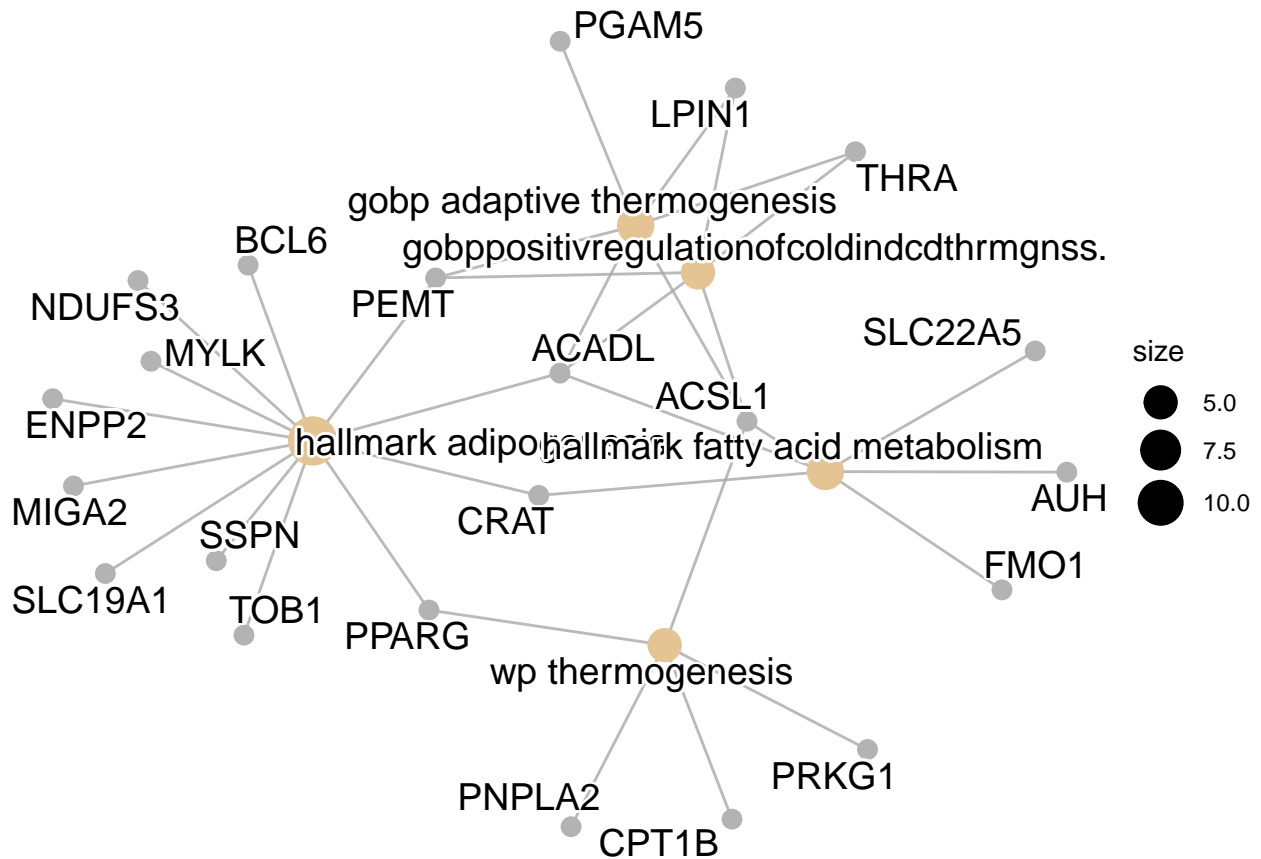
```
#cnetplot(specific, showCategory=10)

save(specific, file = here("R/plots", "Themogenesis_object.RData"))

terms = separate_longer_delim(as.data.frame(specific), cols = geneID, delim = "/") %>% select(ID, geneI
terms$ID = stringr::str_to_title(gsub("_"," ",terms$ID))

write.table(terms, here("R/plots", "thermogenesis_network.txt"), sep="-", quote = F, row.names = F)
```