Gene Ontology Term Analysis

Asad

2023-05-10

```
library(clusterProfiler)
```

```
## Warning: replacing previous import 'utils::findMatches' by
## 'S4Vectors::findMatches' when loading 'AnnotationDbi'
```

```
##
```

```
## Registered S3 method overwritten by 'ggtree':
##   method      from
##   identify.gg ggfun
```

```
## clusterProfiler v4.8.1  For help: https://yulab-smu.top/biomedical-knowledge-mining-book/
##
## If you use clusterProfiler in published research, please cite:
## T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo, an
d G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innova
tion. 2021, 2(3):100141
```

```
##
## Attaching package: 'clusterProfiler'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(DOSE)
```

```
## DOSE v3.26.1  For help: https://yulab-smu.top/biomedical-knowledge-mining-book/
##
## If you use DOSE in published research, please cite:
## Guangchuang Yu, Li-Gen Wang, Guang-Rong Yan, Qing-Yu He. DOSE: an R/Bioconductor package for
Disease Ontology Semantic and Enrichment analysis. Bioinformatics 2015, 31(4):608-609
```

```
library(R.utils)
```

```
## Loading required package: R.oo
```

```
## Loading required package: R.methodsS3
```

```
## R.methodsS3 v1.8.2 (2022-06-13 22:00:14 UTC) successfully loaded. See ?R.methodsS3 for help.
```

```
## R.oo v1.25.0 (2022-06-12 02:20:02 UTC) successfully loaded. See ?R.oo for help.
```

```
##
## Attaching package: 'R.oo'
```

```
## The following object is masked from 'package:R.methodsS3':
##
##     throw
```

```
## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods
```

```
## The following objects are masked from 'package:base':
##
##     attach, detach, load, save
```

```
## R.utils v2.12.2 (2022-11-11 22:00:03 UTC) successfully loaded. See ?R.utils for help.
```

```
##
## Attaching package: 'R.utils'
```

```
## The following object is masked from 'package:utils':
##
##     timestamp
```

```
## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, isOpen, nullfile, parse, warnings
```

```
library(gtools)
```

```
##
## Attaching package: 'gtools'
```

```
## The following object is masked from 'package:R.utils':
##
##     capture
```

```
library(org.Hs.eg.db)
```

```
## Loading required package: AnnotationDbi
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##     dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##     grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##     order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##     rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##     union, unique, unsplit, which.max, which.min
```

```
## Loading required package: Biobase
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
## Loading required package: IRanges
```

```
## Loading required package: S4Vectors
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:clusterProfiler':
##
##     rename
```

```
## The following object is masked from 'package:utils':
##
##     findMatches
```

```
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname
```

```
##
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:R.oo':
##
##     trim
```

```
## The following object is masked from 'package:clusterProfiler':
##
##     slice
```

```
## The following object is masked from 'package:grDevices':
##
##     windows
```

```
##
## Attaching package: 'AnnotationDbi'
```

```
## The following object is masked from 'package:clusterProfiler':
##
##     select
```

```
##
```

```
library(enrichplot)
library(ggplot2)
library(stringr)
library(tibble)
library(ggupset)
library(DOSE)
library(AnnotationDbi)
library(pathview)
```

```
## ########################################################################################
## Pathview is an open source software package distributed under GNU General
## Public License version 3 (GPLv3). Details of GPLv3 is available at
## http://www.gnu.org/licenses/gpl-3.0.html. Particullary, users are required to
## formally cite the original Pathview paper (not just mention it) in publications
## or products. For details, do citation("pathview") within R.
##
## The pathview downloads and uses KEGG data. Non-academic uses may require a KEGG
## license agreement (details at http://www.kegg.jp/kegg/legal.html).
## ########################################################################################
```

Read data

```
setwd('E:/R-Programming-Practices/Functional Enrichment/Gene Ontology Terms')
GENEs<- data.frame(read.csv('GBM.csv'))
```

Convert Symbol to ENTREZID

```
Converted <- bitr(unique(GENEs$Genes), fromType = "SYMBOL",
                  toType = c("ENTREZID"),
                  OrgDb = org.Hs.eg.db)
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
## Warning in bitr(unique(GENEs$Genes), fromType = "SYMBOL", toType =
## c("ENTREZID"), : 1.62% of input gene IDs are fail to map...
```

```
head(Converted)
```

```
##     SYMBOL ENTREZID
## 1     YAE1    57002
## 2     CBX3    11335
## 3    PSMA2     5683
## 4    SSBP1     6742
## 5    BUD31     8896
## 6 STARD3NL    83930
```

Prepare data for Functional Enrichment Analysis

```
Gene_List <- Converted$ENTREZID
Gene_List = sort(Gene_List, decreasing = TRUE)
head(Gene_List)
```

```
## [1] "9997" "9994" "9993" "9992" "9991" "9990"
```

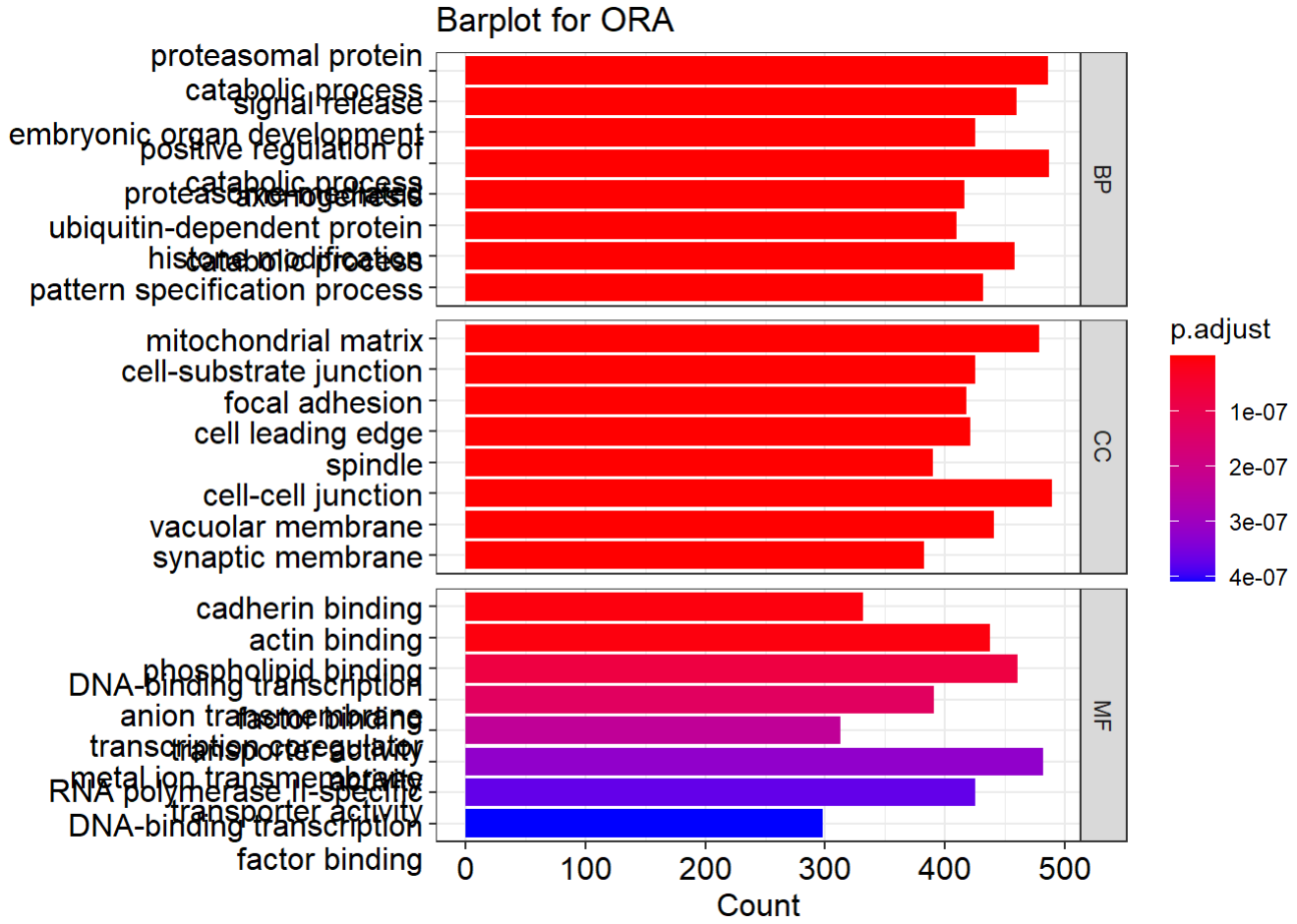
#N.B. Sorting to deacreasing order is required for enrichment analysis

Overrepresentation (ORA) Analysis altogether

```
All_result <- enrichGO(gene     = Gene_List,
                       OrgDb         = org.Hs.eg.db,
                       keyType       = 'ENTREZID',
                       ont           = "ALL",
                       pAdjustMethod = "BH",
                       pvalueCutoff  = 0.01,
                       qvalueCutoff  = 0.01)
```

Visualization

```
barplot(All_result, split = "ONTOLOGY")+facet_grid(ONTOLOGY~., scale = "free")+
                                                    ggtitle("Barplot for ORA")
```

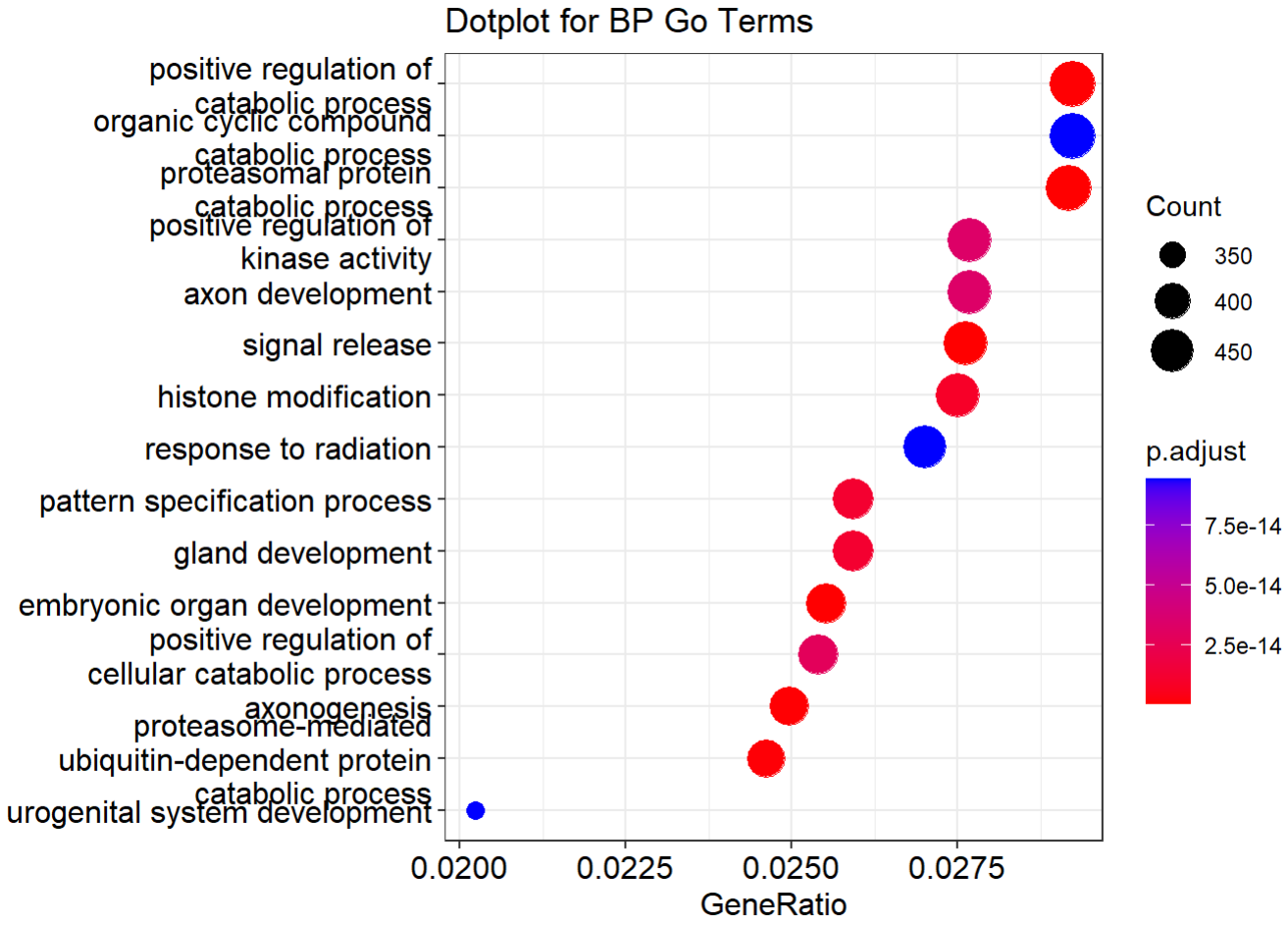


Biological processes gene ontlogy terms

```
BP_result <- enrichGO(gene     = Gene_List,
                      OrgDb         = org.Hs.eg.db,
                      keyType       = 'ENTREZID',
                      ont           = "BP", #Should be changed during MF and CC
                      pAdjustMethod = "BH",
                      pvalueCutoff  = 0.01,
                      qvalueCutoff  = 0.01)
```

Visualization

```
dotplot(BP_result, showCategory=15) + ggtitle("Dotplot for BP Go Terms")
```



Integrate log2FC for ridge plot

#In this step we need prepare our data similar to geneList (package='DOSE') #Let's read new data and omit any NA values

```
DEGs<-read.csv('DEGs.csv')
DEGs<- na.omit(DEGs)
```

Convert gene symbols to ENTREZID

```
Converted <- bitr(unique(DEGs$Genes), fromType = "SYMBOL",
                  toType = c("ENTREZID"),
                  OrgDb = org.Hs.eg.db)
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
head(Converted)
```

```
##   SYMBOL ENTREZID
## 1 CDC25A      993
## 2  HJURP    55355
## 3  NCAPG    64151
## 4   RRM2     6241
## 5  CCNA2      890
## 6   BUB1      699
```

Prepare data

```
gene_list<- DEGs$logFC
names(gene_list)<- Converted$ENTREZID
gene_list<- sort(gene_list, decreasing = T)
```

Let's perform disease ontology (DO) analysis

```
DO<- gseDO(gene_list, pvalueCutoff = 1)
```

```
## preparing geneSet collections...
```

```
## GSEA analysis...
```

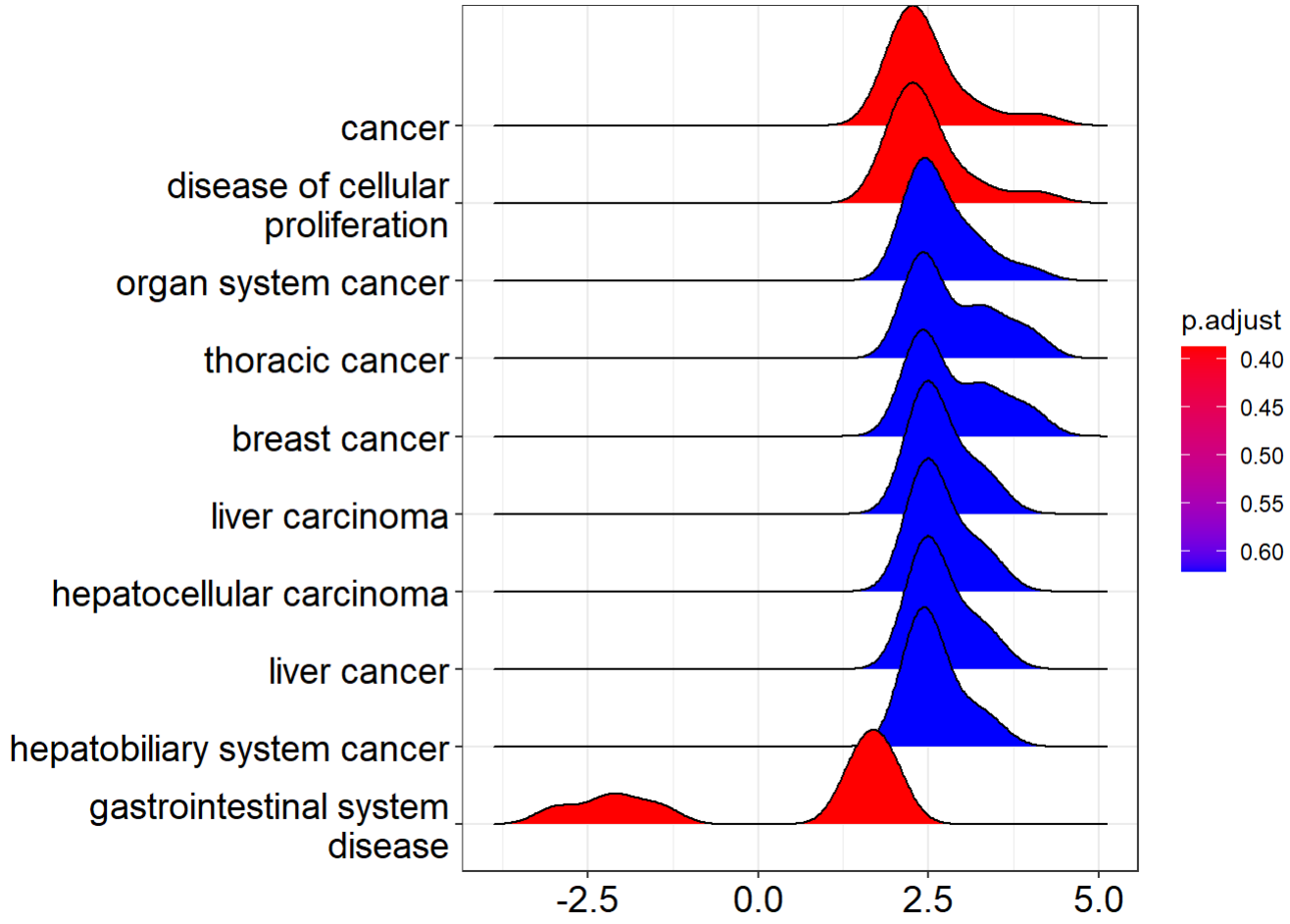
```
## Warning in preparePathwaysAndStats(pathways, stats, minSize, maxSize, gseaParam, : There are
ties in the preranked stats (1.59% of the list).
## The order of those tied genes will be arbitrary, which may produce unexpected results.
```

```
## leading edge analysis...
```

```
## done...
```

```
ridgeplot(DO, showCategory = 10)
```

```
## Picking joint bandwidth of 0.305
```



Disease gene network (DGN) Analysis

Prepare data

```
dgn_data<-names(gene_list)[abs(gene_list) > 2]
dgn_data<-sort(dgn_data, decreasing = T)
```

Perform DGN Analysis

```
dgn_res <- enrichDGN(dgn_data, pvalueCutoff = 0.05)
head(dgn_res)
```

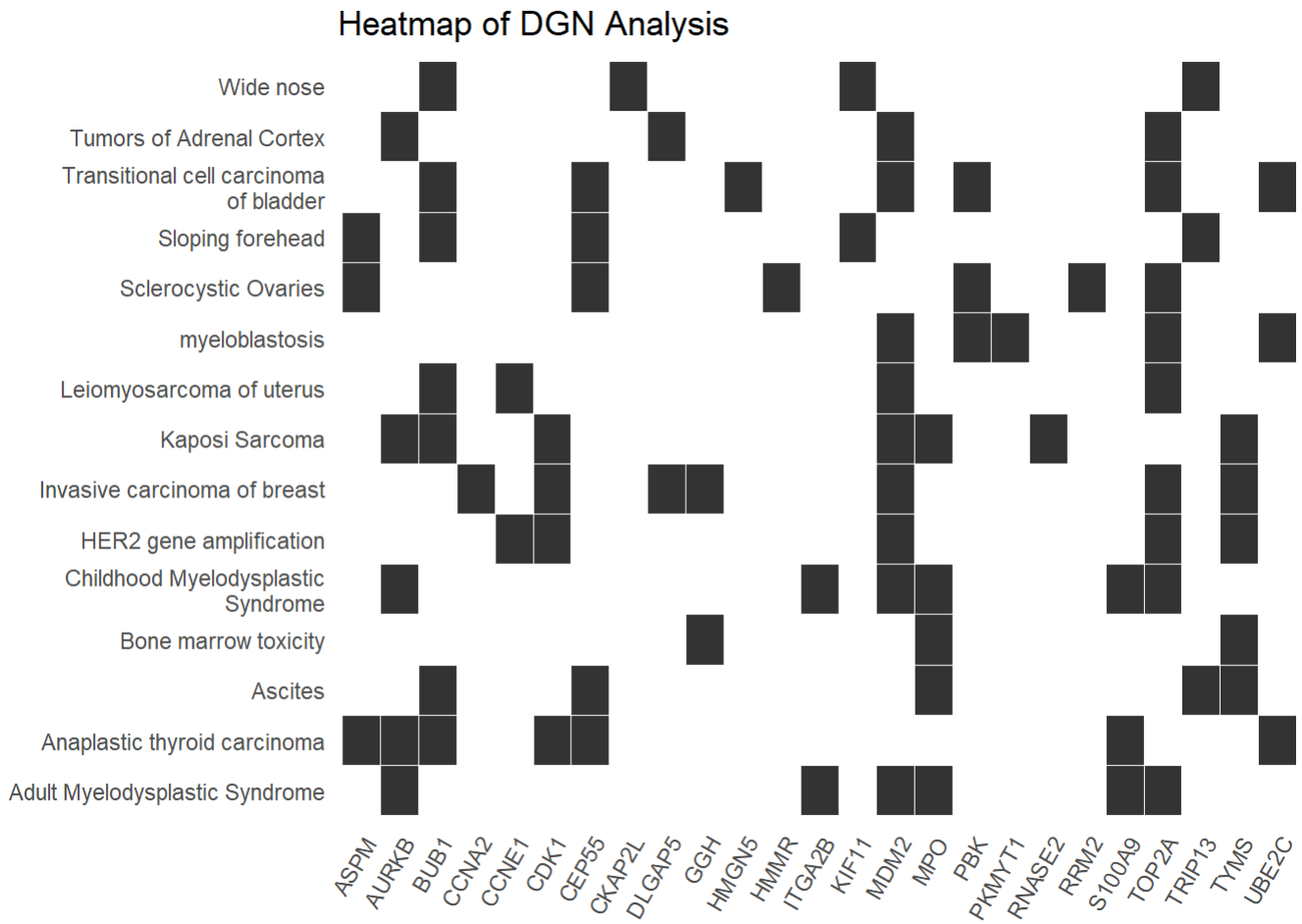
```
##               ID                          Description GeneRatio   BgRatio
## C0679427 C0679427                        myeloblastosis      5/39  72/21671
## C1136382 C1136382                  Sclerocystic Ovaries      6/39 144/21671
## C0948168 C0948168                  Bone marrow toxicity      3/39  13/21671
## C0279680 C0279680 Transitional cell carcinoma of bladder      7/39 333/21671
## C0238461 C0238461           Anaplastic thyroid carcinoma      7/39 392/21671
## C1857679 C1857679                       Sloping forehead      5/39 149/21671
##                pvalue     p.adjust       qvalue
## C0679427 1.853585e-07 0.0001858053 0.0001290191
## C1136382 2.111424e-07 0.0001858053 0.0001290191
## C0948168 1.522037e-06 0.0008424595 0.0005849854
## C0279680 1.914681e-06 0.0008424595 0.0005849854
## C0238461 5.607826e-06 0.0019739549 0.0013706708
## C1857679 6.849921e-06 0.0020057071 0.0013927189
##                                      geneID Count
## C0679427            9088/7153/55872/4193/11065     5
## C1136382     7153/6241/55872/55165/3161/259266     6
## C0948168                        8836/7298/4353     3
## C0279680 79366/7153/699/55872/55165/4193/11065     7
## C0238461  983/9212/699/6280/55165/259266/11065     7
## C1857679            9319/699/55165/3832/259266     5
```

Convert gene id to name in the enrichment object

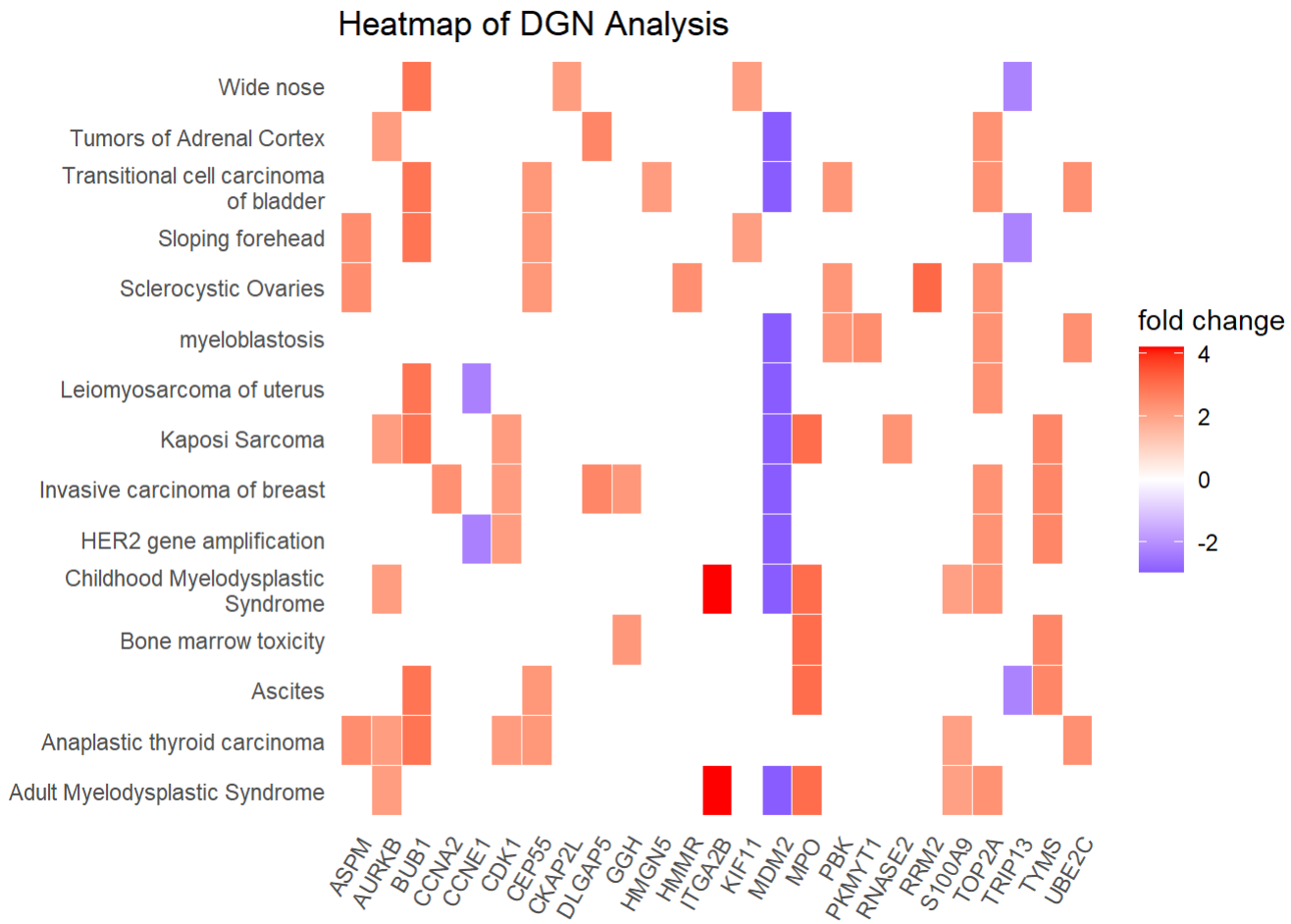
```
dgn_res <- setReadable(dgn_res, 'org.Hs.eg.db', 'ENTREZID')
```

Plot

```
Plot1<- heatplot(dgn_res, showCategory = 15) + ggtitle("Heatmap of DGN Analysis")
Plot1
```

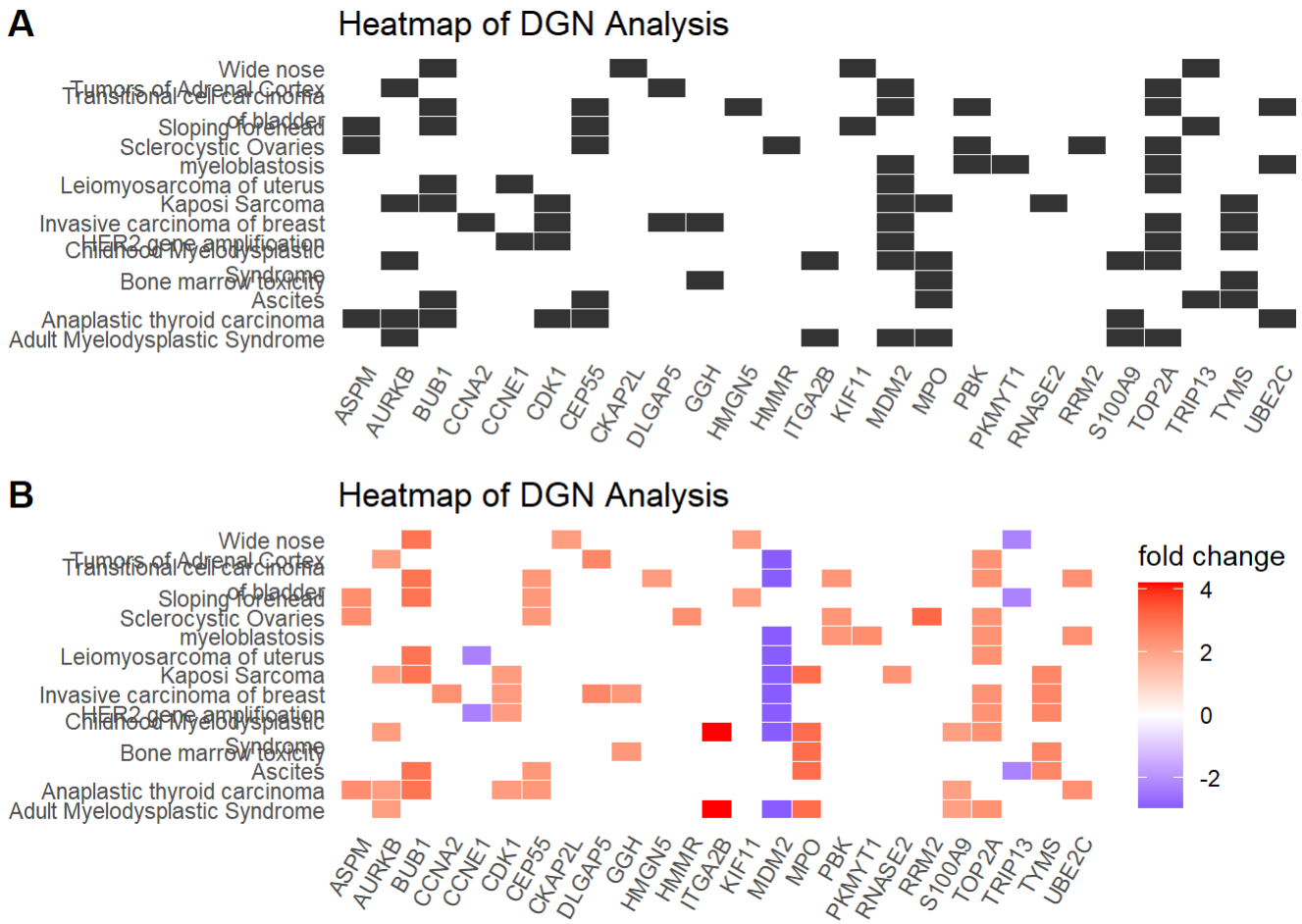



```
Plot2<- heatplot(dgn_res, showCategory = 15, foldChange=gene_list) +
        ggtitle("Heatmap of DGN Analysis")
Plot2
```



Combine two plots together

```
cowplot::plot_grid(Plot1, Plot2, ncol=1, labels=LETTERS[1:2])
```

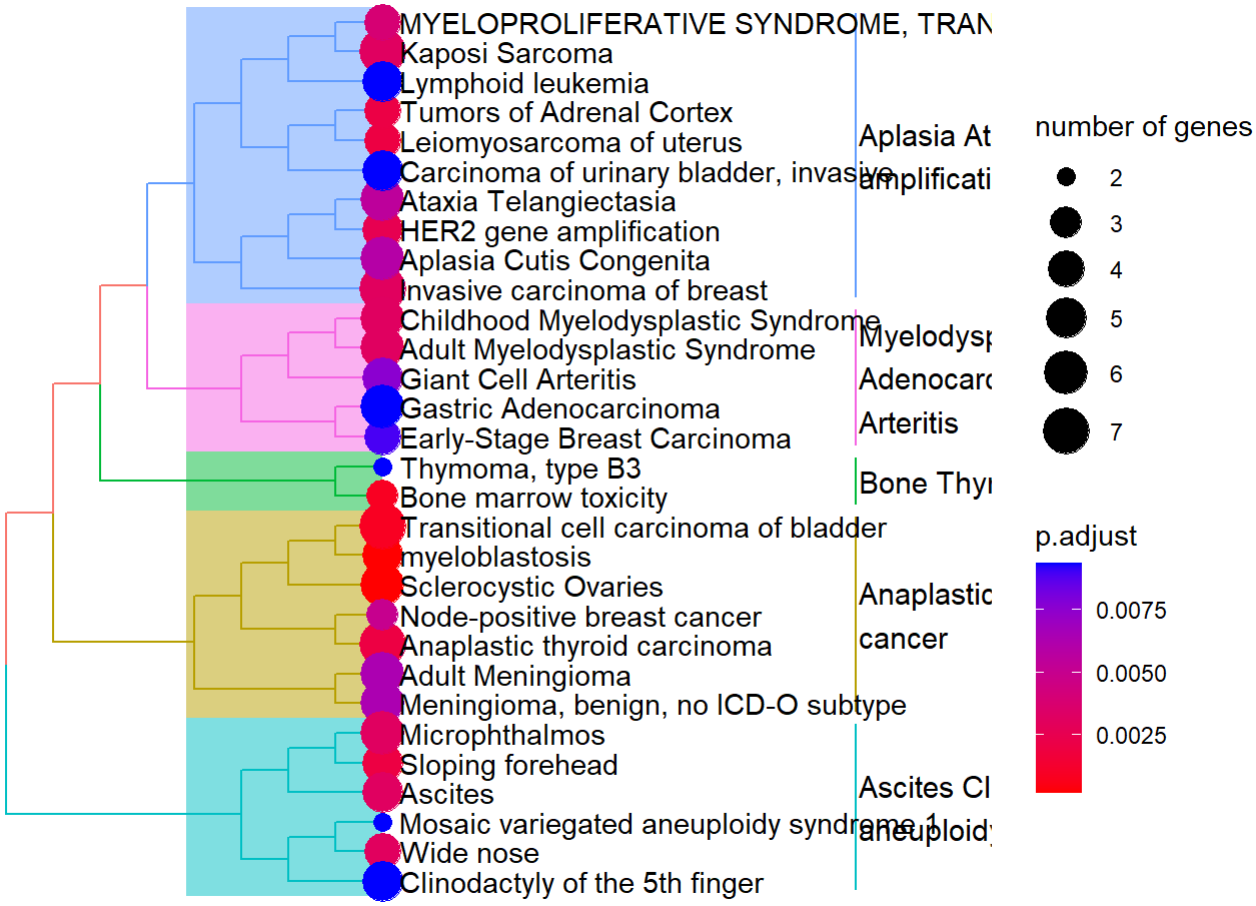


Tree plot

```
tree_data<- pairwise_termsim(dgn_res)
Tree_plot<-treeplot(tree_data)
```

```
## Warning in stat_tree(data = data, mapping = mapping, geom = "segment", position = position, :
Ignoring unknown parameters: `hang`
## Ignoring unknown parameters: `hang`
```

```
Tree_plot
```



Cnet plot

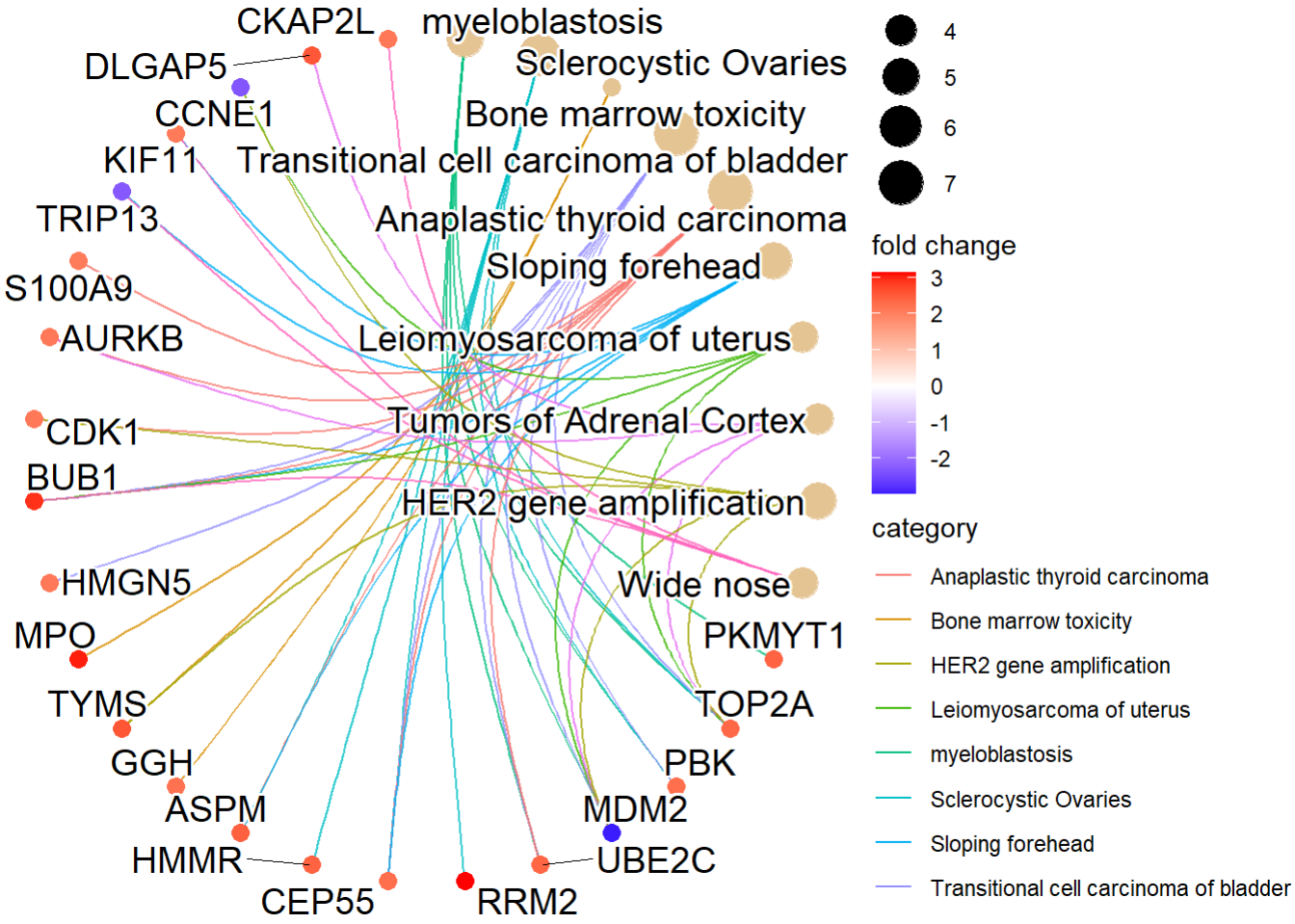
```
Cnet_plot <- cnetplot(dgn_res, foldChange=gene_list, circular = TRUE,
                      showCategory=10, colorEdge = TRUE)
```

```
## Warning in cnetplot.enrichResult(x, ...): Use 'color.params = list(foldChange = your_value)'
instead of 'foldChange'.
##  The foldChange parameter will be removed in the next version.
```

```
## Warning in cnetplot.enrichResult(x, ...): Use 'color.params = list(edge = your_value)' instea
d of 'colorEdge'.
##  The colorEdge parameter will be removed in the next version.
```

```
## Scale for size is already present.
## Adding another scale for size, which will replace the existing scale.
```

```
Cnet_plot
```



#N.B. Cnetplot, Tree plot and Heatplot work best for small number of genes.

#Reference: 'http://yulab-smu.top/clusterProfiler-book/index.html (http://yulab-smu.top/clusterProfiler-book/index.html)'