DisGeNET disorder analysis

HBG

9/7/2019

R Markdown

Analysis related to disorder genes from DisGeNET. 645 disorders were collected, each of which has gene numbers between [15,193]; for those with more than 200 genes with top 200 scores were chosen; the largest group based on this selection has 193 genes presented in the InWeb PIN.

```
#libraries needed
library(gplots)
```

```
## Warning: package 'gplots' was built under R version 3.5.2
```

```
##
## Attaching package: 'gplots'
```

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

```
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 3.5.2
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

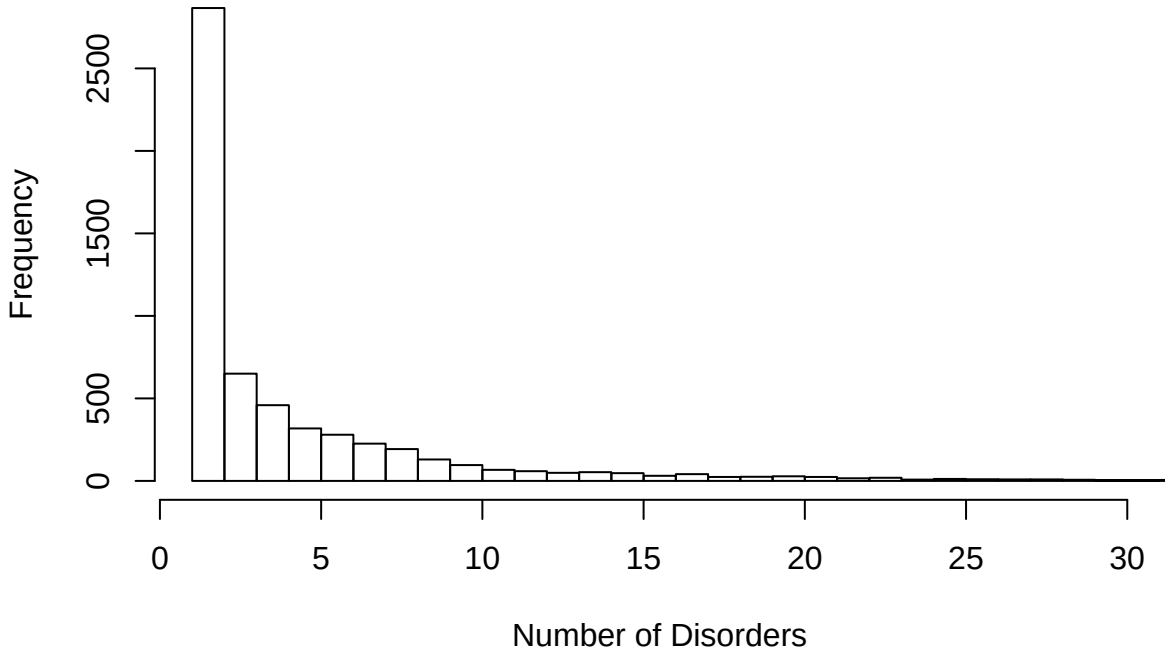
```
## The following object is masked from 'package:base':
##
##     union
```

```
library(Corbi)
```

```
## Warning: package 'Corbi' was built under R version 3.5.2
```

This analysis is based on comparisons between the original InWeb PIN and its 5k null models

```
disorder.summary <- read.csv("refined.disorders.csv", header=T,
                             stringsAsFactors = F)
#only proteins, but no micro RNAs are considered in this analysis
disorder.gene <- read.csv("Discurate.full.non.MIR.csv", header=T, stringsAsFactors = F)
disorder.gene.uniq <- unique(disorder.gene$gene[which(disorder.gene$ID %in% disorder.summary$ID)])
# 5894 uniq genes in all 645 disorders
gene.counts <- read.csv("gene.counts.csv", sep="\t", header=T, stringsAsFactors = F)
#pdf("gene.counts.histogram.pdf", width=5, height=4, paper='special')
hist(gene.counts$disorder.number, breaks=200, xlab="Number of Disorders", xlim=c(1,30), main="")
```




```
#dev.off()

#Genes shared in over 100 disorders
gene.counts$gene[which(gene.counts$disorder.number >100)]
```

```
## [1] "IL1B"  "IL6"   "SOD2"  "TNF"   "TP53"  "PTGS2"
```

```
#Genes shared in over 50 disorders
gene.counts$gene[which(gene.counts$disorder.number >50)]
```

```
##  [1] "APOE"   "BDNF"   "CNR1"   "ACE"    "DRD2"   "ESR1"   "FGFR1"
##  [8] "IL1B"   "IL6"    "MMP9"   "MTHFR"  "NPY"    "SOD2"   "TNF"
## [15] "TP53"   "BCL2"   "HMOX1"  "IGF1"   "INS"    "LEP"    "NOS3"
## [22] "PPARG"  "VEGFA"  "APC"    "BRAF"   "CTNNB1" "EGFR"   "ERBB2"
## [29] "IFNG"   "KRAS"   "MYC"    "NOS2"   "PTGS2"  "STAT3"  "AGT"
## [36] "NGF"    "POMC"   "SOD1"   "ALB"    "IFNA2"  "TGFB1"  "CCL2"
## [43] "MET"    "PIK3CA" "PTEN"   "FOS"    "CAT"    "CSF3"   "CSF2"
```

```
#number of genes that presented in only one disorder
length(which(gene.counts$disorder.number == 1))
```

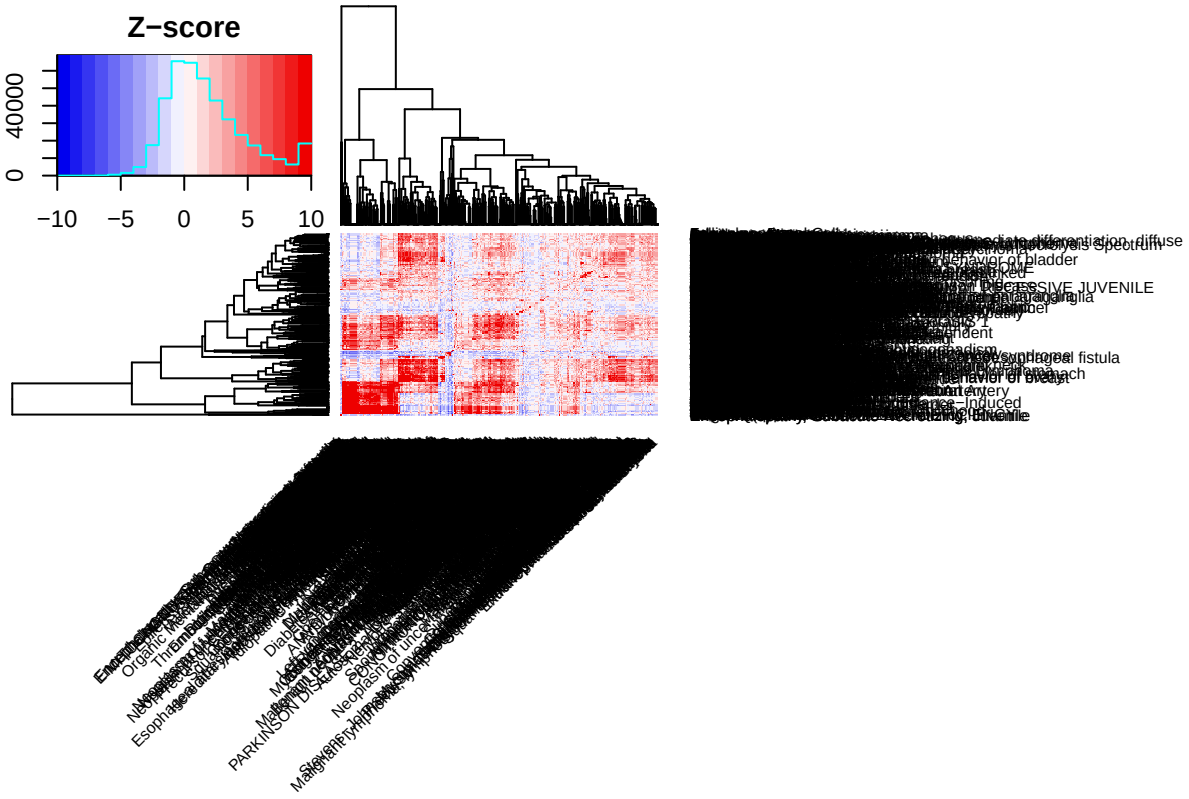
```
## [1] 1791
```

```
disorder.dim <- length(disorder.summary$Name)
```

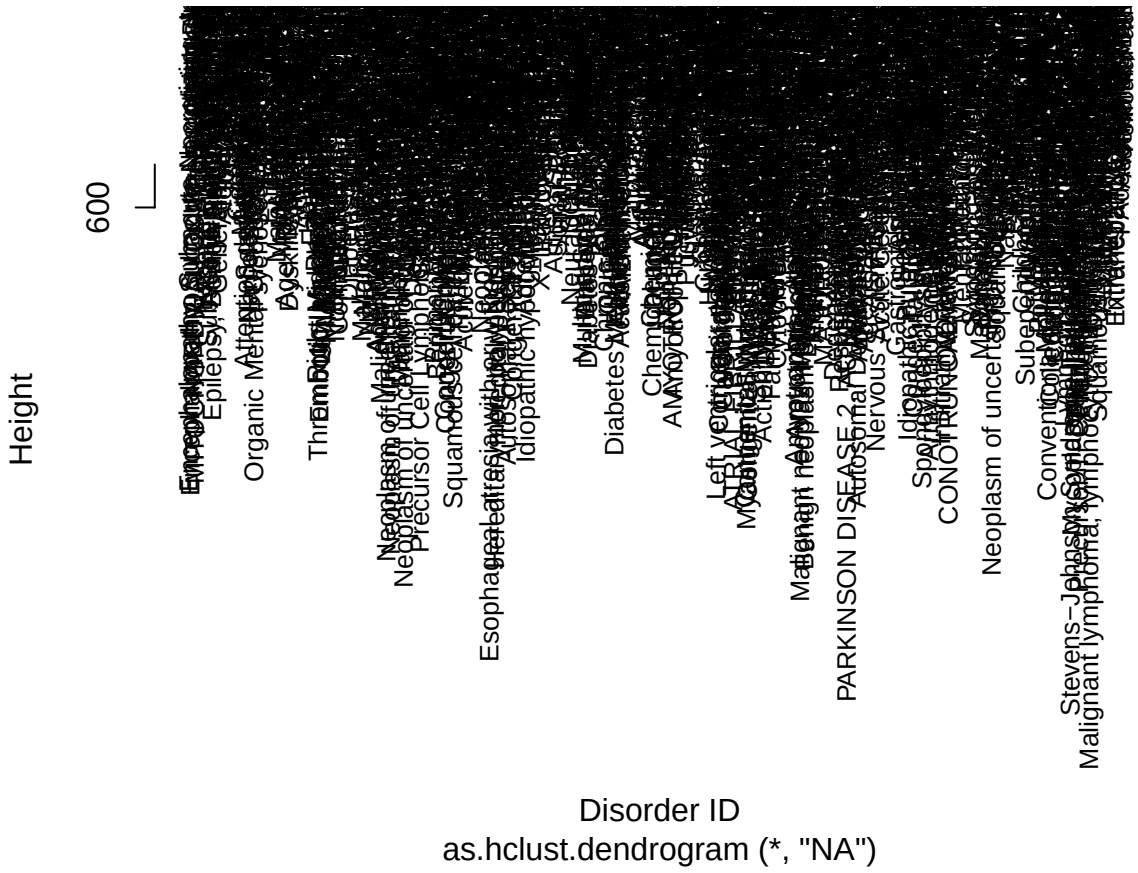
z-scores

```
###the z-scores based on 5k permutations
ddi.dat <- read.csv("ddi.z.5k.csv", header=F, stringsAsFactors = F)
ddi.z <- matrix(unlist(ddi.dat), nrow=disorder.dim, ncol=disorder.dim)

colnames(ddi.z) <- disorder.summary$Name
row.names(ddi.z) <- disorder.summary$Name

## coloring, concentrating on z in [-10,10], blue-white-red coloring
my_palette <- colorRampPalette(c("blue2", "white", "red2"))(n = 20)
```

```
colors = c(seq(-10,10,length=21))

hm <- heatmap.2(ddi.z, col=my_palette, trace='none', breaks=colors,
                key.xlab=NA, key.title="Z-score", key.ylab=NA,
                srtCol=45, adjCol=c(1,0), dendrogram = "both",
                margins=c(14,18.5))
```



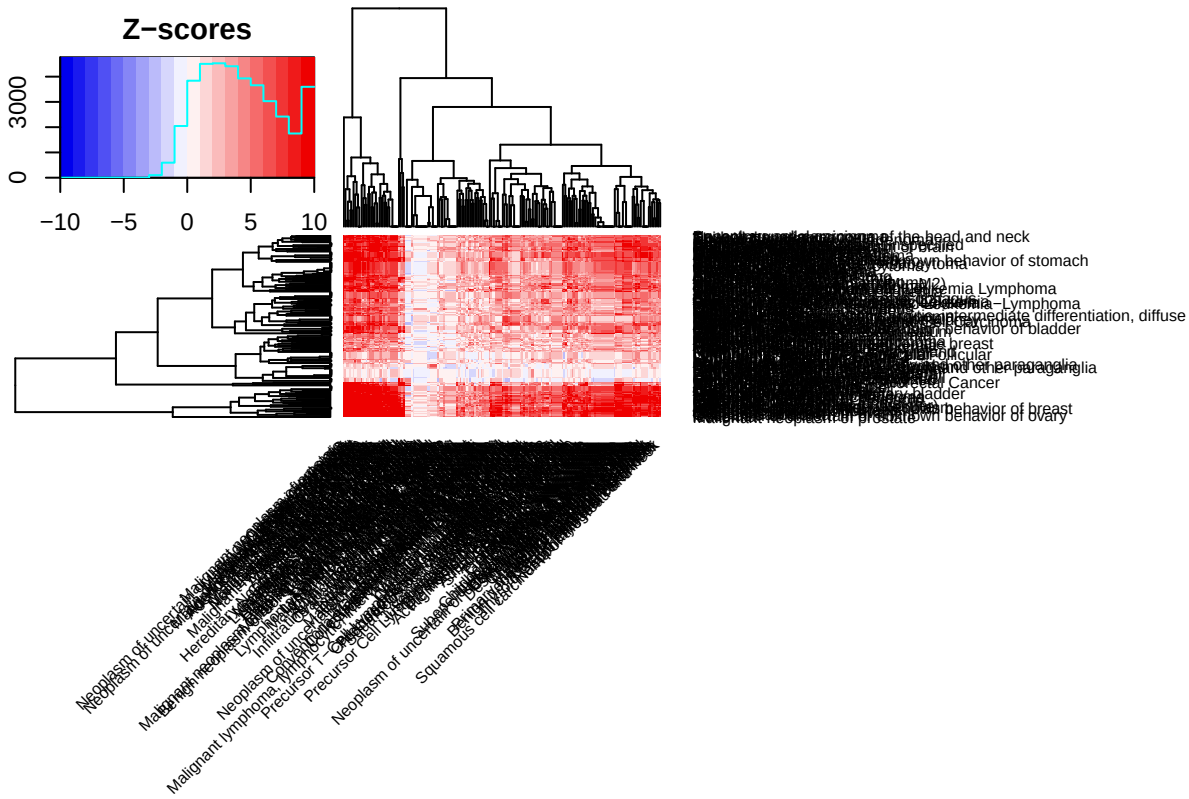

```
hc <- as.hclust(hm$rowDendrogram)
#This tree include all 645 diseases
#pdf("disorder.tree.pdf", height=10, width=80, paper='special')
plot(hc, xlab="Disorder ID", cex=.8)
```



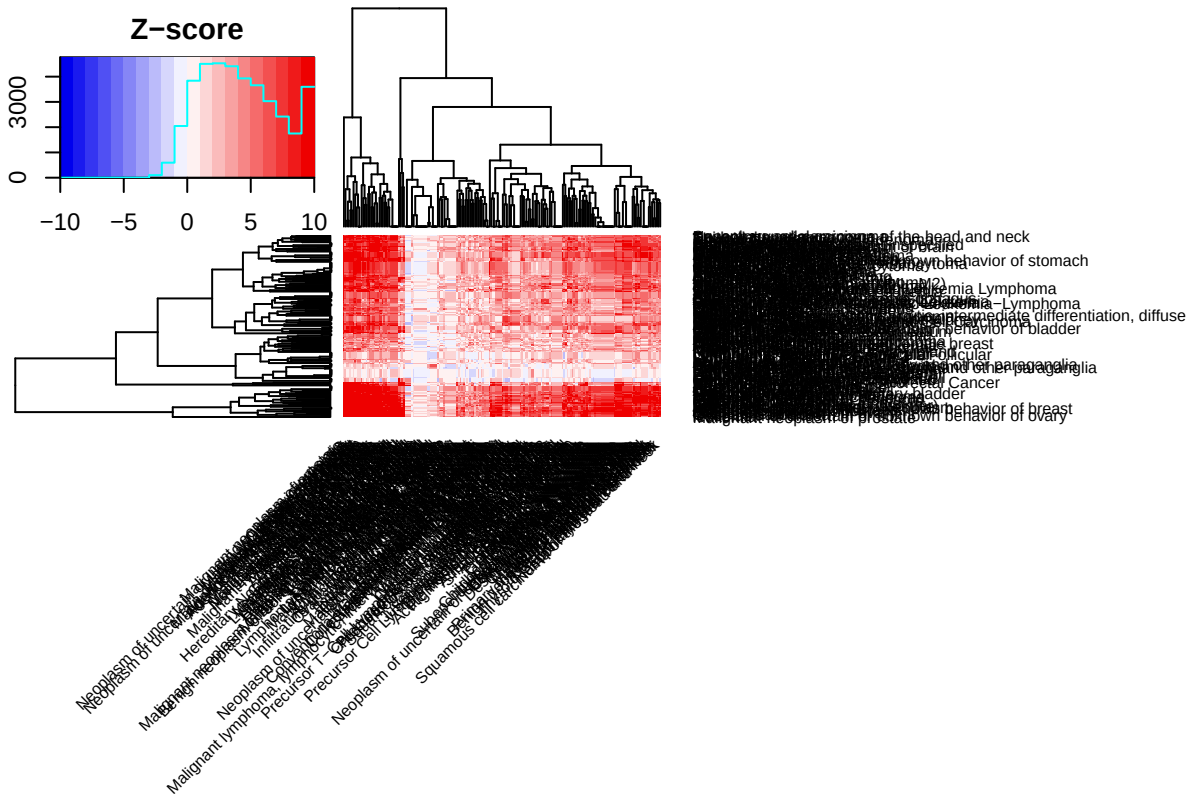

```
#dev.off()

#hierarchy.order <- hc$order
```

select all cancer disorders (neoplastic processes)

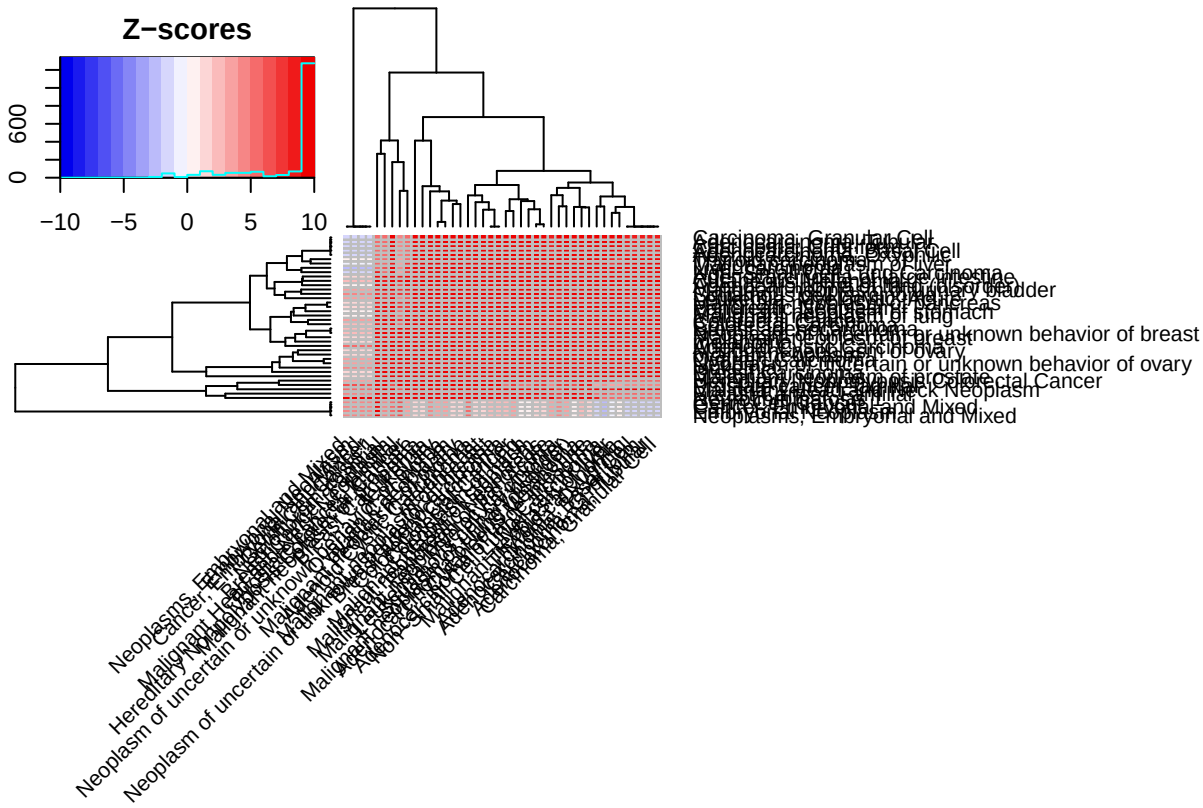
```
name.type <- read.csv("Dis.Name.v.Type-class.csv", header=T, stringsAsFactors = F)
cancers <- which(name.type$Semantic.Type == "Neoplastic Process")
cancer.z.matrix <- submatrix(ddi.z, rows=cancers, cols=cancers)
colnames(cancer.z.matrix) <- disorder.summary$Name[cancers]
row.names(cancer.z.matrix) <- disorder.summary$Name[cancers]
#png("cancers.png", width=13, height=12, units = "in", res=600)
heatmap.2(cancer.z.matrix, col=my_palette, trace='none', breaks=colors,
          key.xlab=NA, key.title="Z-scores", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL,#
          margins=c(14,18.5), srtCol=45, adjCol=c(1,0), dendrogram = "both")
```




```
#dev.off()

#we can select the branch 1 from the hierachy tree (order 1:42)
hm2 <- heatmap.2(cancer.z.matrix, col=my_palette, trace='none', breaks=colors,
                 key.xlab=NA, key.title="Z-score", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL
                 srtCol=45, adjCol=c(1,0), dendrogram = "both",
                 margins=c(14,18.5))
```



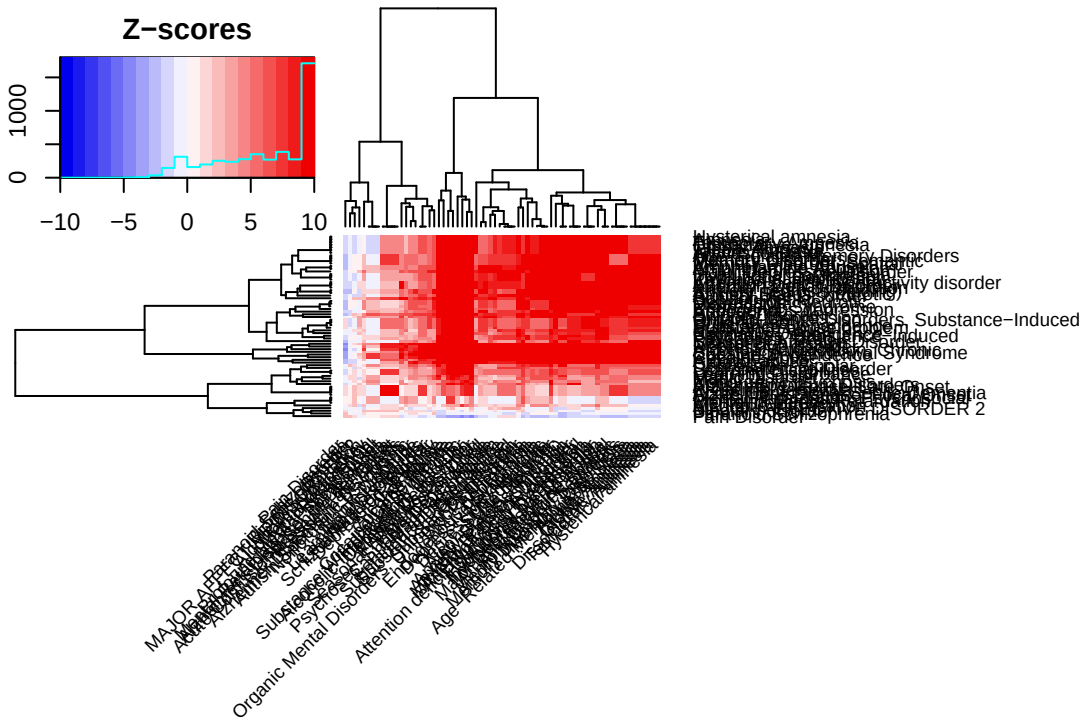

```
hc2 <- as.hclust(hm2$rowDendrogram)

cancer.branch1 <- hc2$order[1:42]
cancer.branch1.matrix <- submatrix(cancer.z.matrix, rows=cancer.branch1, cols = cancer.branch1)
colnames(cancer.branch1.matrix) <- hc2$labels[cancer.branch1]
row.names(cancer.branch1.matrix) <- hc2$labels[cancer.branch1]
#png("cancer.branch1.png", width=13, height=12, units = "in", res=600)
heatmap.2(cancer.branch1.matrix, col=my_palette, trace='none', breaks=colors,
          key.xlab=NA, key.title="Z-scores", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL,#
          margins=c(14,18.5), srtCol=45, adjCol=c(1,0), dendrogram = "both",
          sepcolor="grey", colsep=1:42, rowsep=1:42)
```



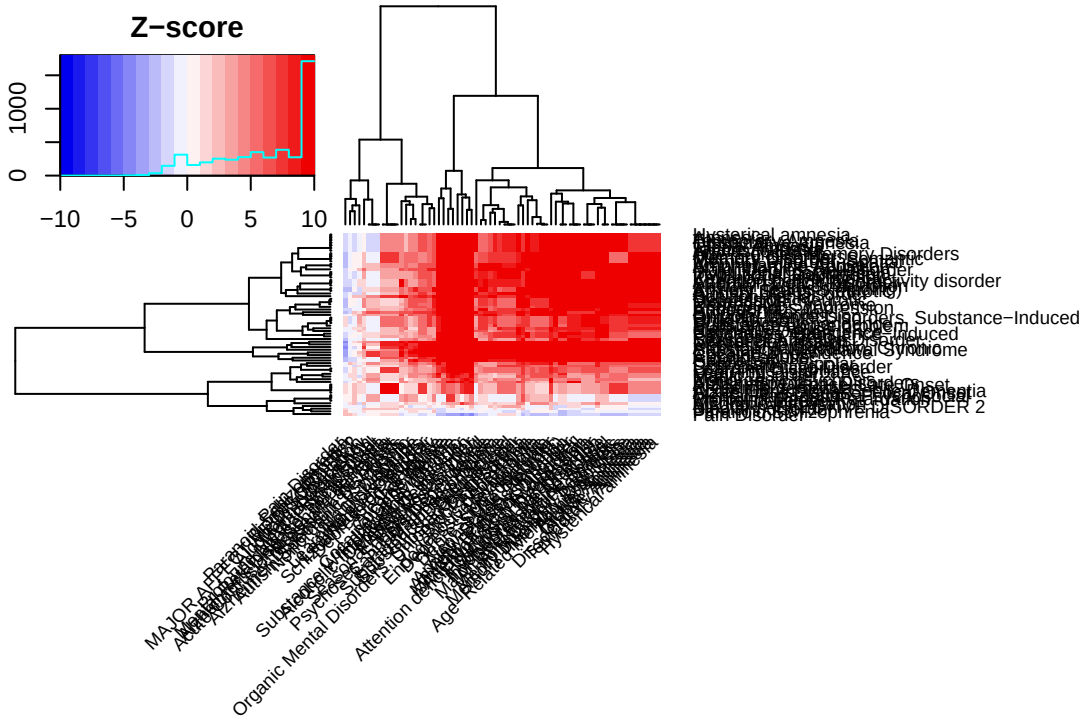

```
#dev.off()
```

mental disorders, "mental or behavioral dysfunction"

```
mental <- which(name.type$Semantic.Type == "Mental or Behavioral Dysfunction")
mental.z.matrix <- submatrix(ddi.z, rows=mental, cols=mental)
colnames(mental.z.matrix) <- disorder.summary$Name[mental]
row.names(mental.z.matrix) <- disorder.summary$Name[mental]

#png("mental.png", width=13, height=12, units = "in", res=600)
heatmap.2(mental.z.matrix, col=my_palette, trace='none', breaks=colors,
          key.xlab=NA, key.title="Z-scores", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL,#
          margins=c(14,18.5), srtCol=45, adjCol=c(1,0), dendrogram = "both")
```




```
#dev.off()

hm3 <- heatmap.2(mental.z.matrix, col=my_palette, trace='none', breaks=colors,
                 key.xlab=NA, key.title="Z-score", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL
                 srtCol=45, adjCol=c(1,0), dendrogram = "both",
                 margins=c(14,18.5))
```



```
hc3 <- as.hclust(hm3$rowDendrogram)

mental.branch2 <- hc3$order[22:68]
mental.branch2.matrix <- submatrix(mental.z.matrix, rows=mental.branch2, cols = mental.branch2)
colnames(mental.branch2.matrix) <- hc3$labels[mental.branch2]
row.names(mental.branch2.matrix) <- hc3$labels[mental.branch2]
png("mental.branch2.png", width=13, height=12, units = "in", res=600)
heatmap.2(mental.branch2.matrix, col=my_palette, trace='none', breaks=colors,
          key.xlab=NA, key.title="Z-scores", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL,#
          margins=c(14,18.5), srtCol=45, adjCol=c(1,0), dendrogram = "both",
          sepcolor="grey", colsep=1:47, rowsep=1:47)
dev.off()
```

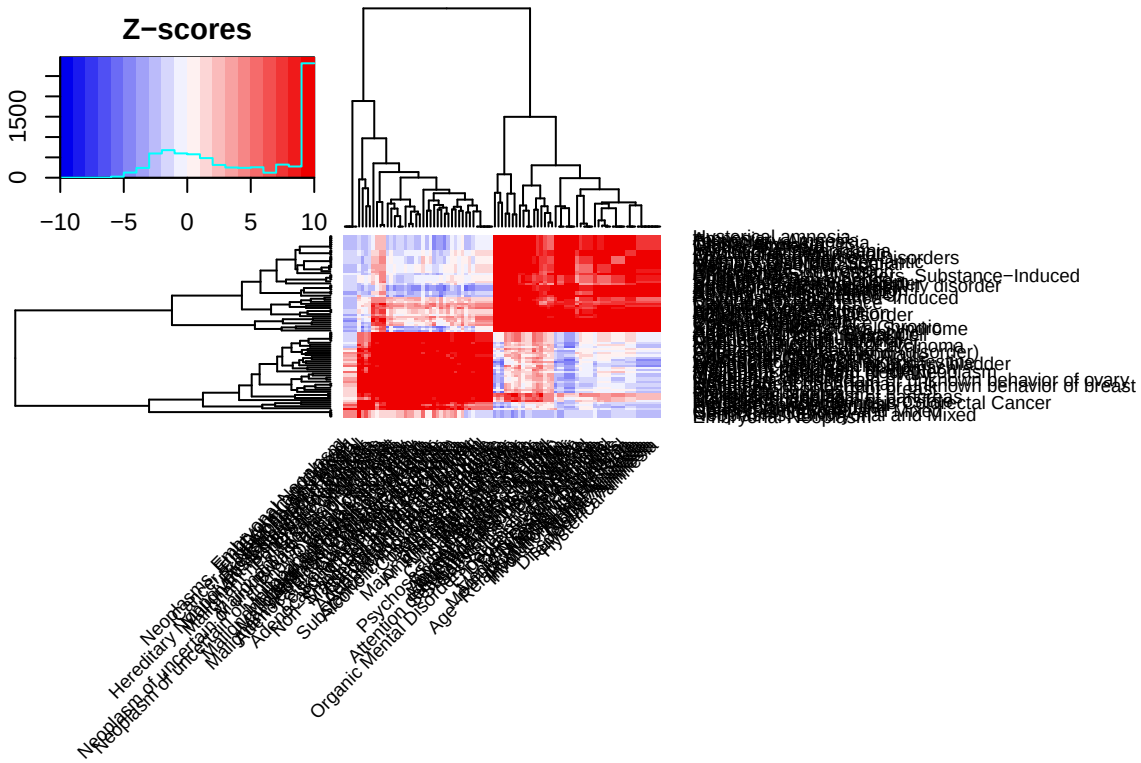
```
## pdf
##   2
```

Now, we are going to put cancer disorder (branch 1) and mental disorder (branch 2) together

```
cancer.mental <- c(hc2$labels[cancer.branch1], hc3$labels[mental.branch2])
length(cancer.mental)
```

```
## [1] 89
```

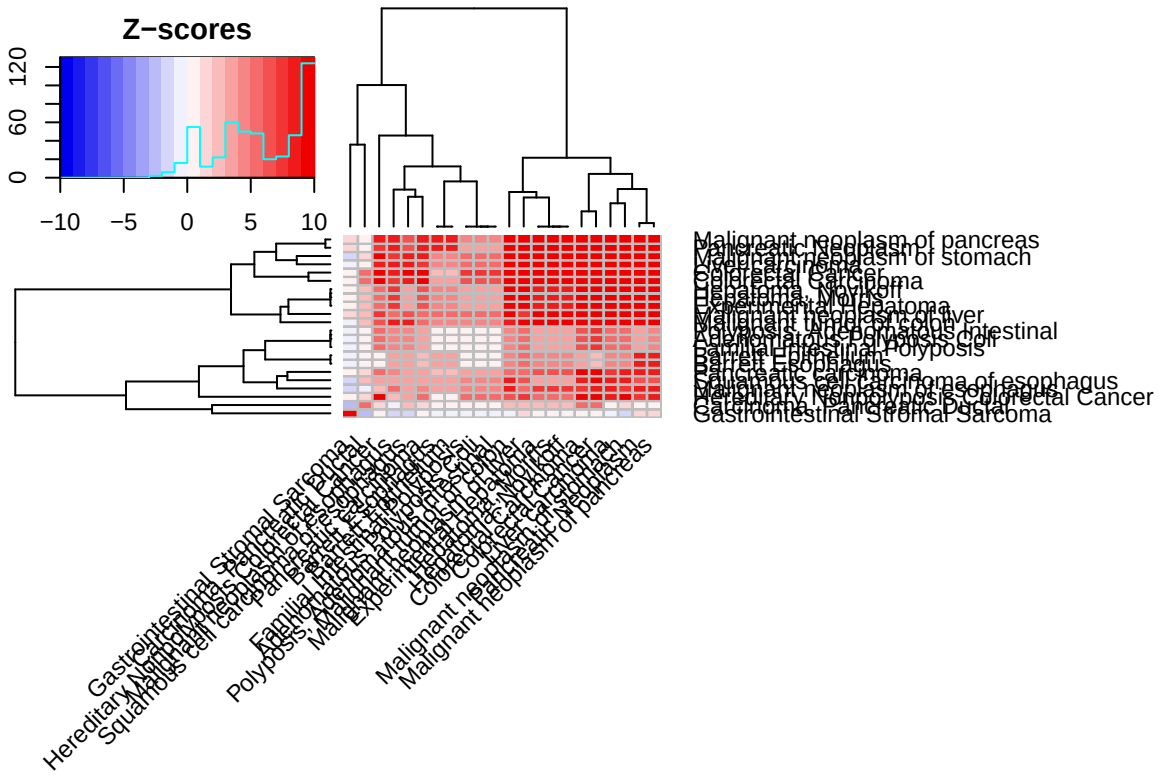
```
#there are 42 cancer plus 47 mental disorders
cm.numbers <- which(disorder.summary$Name %in% cancer.mental)
cm.matrix <- submatrix(ddi.z, rows=cm.numbers, cols=cm.numbers)
colnames(cm.matrix) <- disorder.summary$Name[cm.numbers]
row.names(cm.matrix) <- disorder.summary$Name[cm.numbers]
#png("cancer.mental.png", width=13, height=12, units = "in", res=600)
heatmap.2(cm.matrix, col=my_palette, trace='none', breaks=colors,
          key.xlab=NA, key.title="Z-scores", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL,#
          margins=c(14,18.5), srtCol=45, adjCol=c(1,0), dendrogram = "both")
```




```
#dev.off()
```

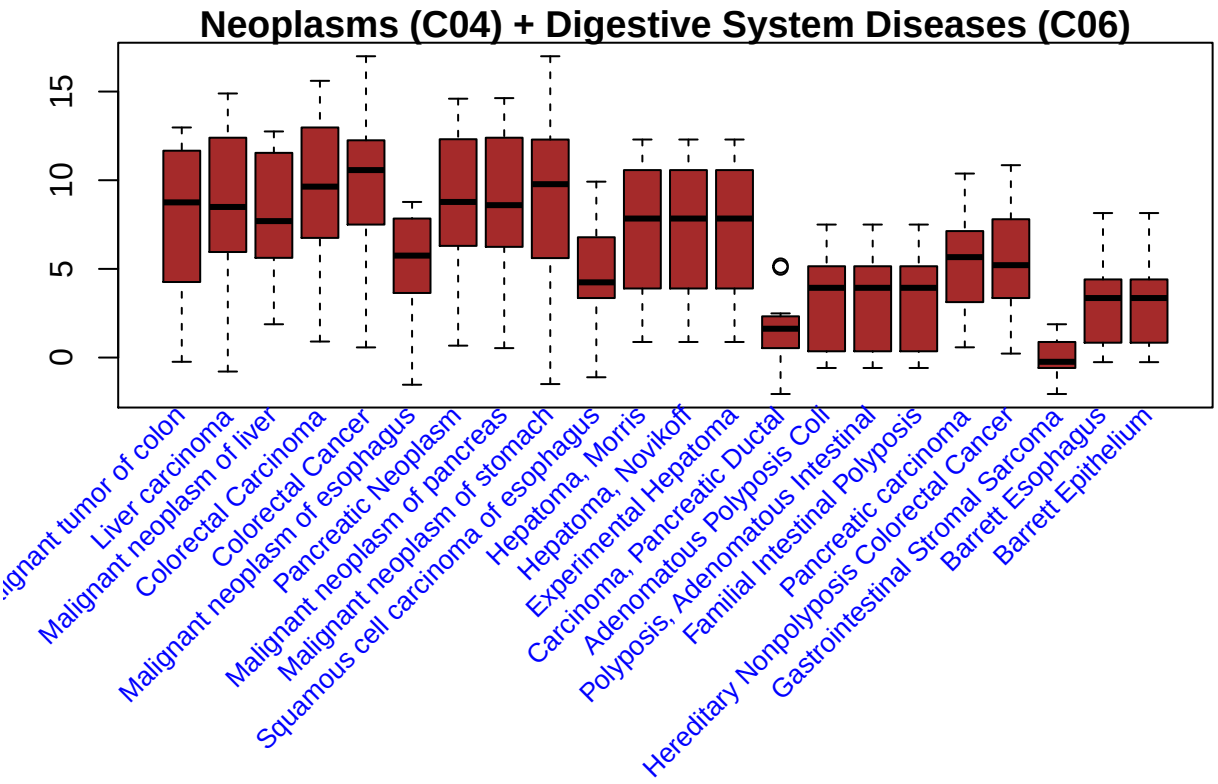
The above analysis are based on the "symentic type" of the disorders as collected in DisGeNET. Most of the disorders (but not all) in DisGeNET also have disorder classes based on the Medical Subject Headings (MeSH) (US library of Medicine). We can also use these headings to select disorders for comparisons.

```
##Disorders in both C04 and C06, cancer in digestive system
# C04: Neoplasms; C06: Digestive System Diseases
# C16: Congenital, Hereditary, and Neonatal Diseases and Abnormalities
# C18: Nutritional and Metabolic Diseases
# C19: Endocrine System Diseases
c04.list <- which(name.type$Disease.class=="C04;C06" |
                  name.type$Disease.class=="C04;C06;C16" |
                  name.type$Disease.class=="C04;C06;C16;C18" |
                  name.type$Disease.class=="C04;C06;C19")
c04.z.matrix <- submatrix(ddi.z, rows=c04.list, cols=c04.list)
colnames(c04.z.matrix) <- disorder.summary$Name[c04.list]
row.names(c04.z.matrix) <- disorder.summary$Name[c04.list]
#png("c04.c06.png", width=13, height=12, units = "in", res=600)
heatmap.2(c04.z.matrix, col=my_palette, trace='none', breaks=colors,
          key.xlab=NA, key.title="Z-scores", key.ylab=NA, key.xtickfun = NULL, key.ytickfun = NULL,#
          margins=c(14,18.5), srtCol=45, adjCol=c(1,0), dendrogram = "both",
          sepcolor="grey", colsep=1:42, rowsep=1:42)
```




```
#dev.off()

# For boxplot without self-interactions (no diagonal elements)
c4.c6.nodiag <- c04.z.matrix
diag(c4.c6.nodiag) <- NA
#pdf("C4.C6.z.boxplot.pdf", width=10, height=6, paper='special')
par(mar=c(12,3,1,1))
boxplot(c4.c6.nodiag, main="Neoplasms (C04) + Digestive System Diseases (C06)", xaxt="n", col='brown')
text(seq(1,22), par("usr")[3]-0.25, srt=45, adj=1, xpd=T,
     col='blue',labels=paste(rownames(c4.c6.nodiag)), cex=0.8, line=10)
```

```
## Warning in text.default(seq(1, 22), par("usr")[3] - 0.25, srt = 45, adj =
## 1, : "line" is not a graphical parameter
```




```
#dev.off()
```

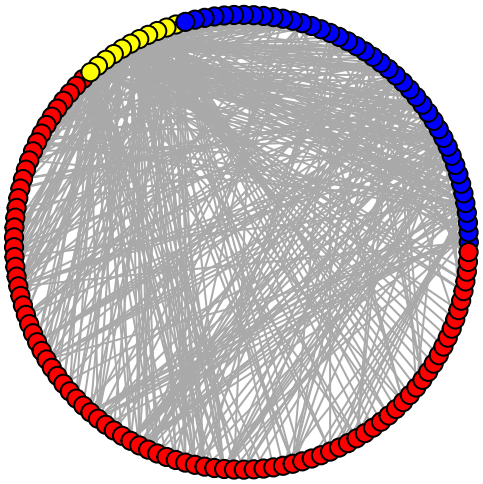
Considering particular disorders, we can calculate the interactions between the gene sets of them.

```
########
## we'll measure two specific sets:
##     C0009402 for "Colorectal Carcinoma"
## vs C0678222 for "Breast Carcinoma"

pin <- read.csv("../../human.pin.csv", header=T, stringsAsFactors = F)

geneA <- pin$geneA
geneB <- pin$geneB

id.a <- c("C0009402")  #colorectal
id.b <- c("C0678222")  #breast
gene.list.a <- disorder.gene$gene[which(disorder.gene$ID %in% id.a)]
gene.list.b <- disorder.gene$gene[which(disorder.gene$ID %in% id.b)]

gene.list.a[which(gene.list.a %in% gene.list.b)]
```

```
##  [1] "AKT1"   "CASP8"  "CDH1"   "CTNNB1" "EP300"  "KRAS"   "MMP1"
##  [8] "PIK3CA" "TP53"   "EXO1"   "CHEK2"
```

```
#"AKT1", "CASP8", "CDH1", "CTNNB1", "EP300", "KRAS", "MMP1", "PIK3CA", "TP53", "EXO1", "CHEK2"
# Is it able to identify the driver gene or distinguish the ONG vs TRG?
ab.int <- which(((geneA %in% gene.list.a) & (geneB %in% gene.list.b)) |
                    ((geneA %in% gene.list.b) & (geneB %in% gene.list.a)))
subA <- geneA[ab.int]
subB <- geneB[ab.int]
```

```
subnet <- data.frame(cbind(subA, subB))
sub.graph <- graph.data.frame(subnet, directed=F)

'%ni%' <- Negate('%in%')

blue.id <- which((as_ids(V(sub.graph)) %in% gene.list.b) &
                   (as_ids(V(sub.graph)) %ni% gene.list.a))
red.id <-  which((as_ids(V(sub.graph)) %in% gene.list.a) &
                   (as_ids(V(sub.graph)) %ni% gene.list.b))
yellow.id <- which((as_ids(V(sub.graph)) %in% gene.list.b) &
                     (as_ids(V(sub.graph)) %in% gene.list.a))
sub.order <- V(sub.graph)[c(blue.id, yellow.id, red.id)]
coords <- layout_in_circle(sub.graph, order = sub.order)

color <- rep("NA", times=length(V(sub.graph)))
color[red.id] <- rep("red", times=length(red.id))
color[blue.id] <- rep("blue", times=length(blue.id))
color[yellow.id] <- rep("yellow", times=length(yellow.id))
V(sub.graph)$color <- color

#pdf("Breast-Colorectal.subnetwork.pdf")
plot.igraph(sub.graph, vertex.color=V(sub.graph)$color,
            vertex.size=8, edge.width=1,vertex.label=NA,
            order=sub.order,layout=coords)
```

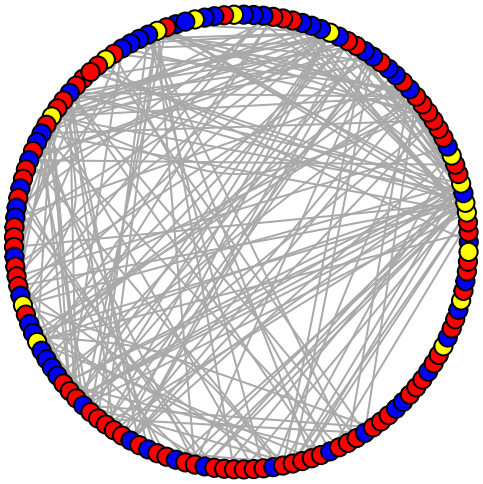


```
#dev.off()
length(E(sub.graph))  #=407
```

```
## [1] 407
```

```
######## ms02
ms02 <- read.csv("../../../ms02star/human/ms02.1.csv", header=T, stringsAsFactors = F)

ms02.geneA <- ms02$id1
ms02.geneB <- ms02$id2

ms02.ab.int <- which(((ms02.geneA %in% gene.list.a) & (ms02.geneB %in% gene.list.b)) |
                   ((ms02.geneA %in% gene.list.b) & (ms02.geneB %in% gene.list.a)))
ms02.subA <- ms02.geneA[ms02.ab.int]
```

```
ms02.subB <- ms02.geneB[ms02.ab.int]

ms02.subnet <- data.frame(cbind(ms02.subA, ms02.subB))
ms02.sub.graph <- graph.data.frame(ms02.subnet, directed=F)

'%ni%' <- Negate('%in%')

ms02.blue.id <- which((as_ids(V(ms02.sub.graph)) %in% gene.list.b) &
                    (as_ids(V(ms02.sub.graph)) %ni% gene.list.a))
ms02.red.id <-  which((as_ids(V(ms02.sub.graph)) %in% gene.list.a) &
                    (as_ids(V(ms02.sub.graph)) %ni% gene.list.b))
ms02.yellow.id <- which((as_ids(V(ms02.sub.graph)) %in% gene.list.b) &
                      (as_ids(V(ms02.sub.graph)) %in% gene.list.a))
ms02.sub.order <- V(ms02.sub.graph)[c(ms02.blue.id, ms02.yellow.id, ms02.red.id)]
ms02.coords <- layout_in_circle(ms02.sub.graph, order = ms02.sub.order)

ms02.color <- rep("NA", times=length(V(ms02.sub.graph)))
ms02.color[ms02.red.id] <- rep("red", times=length(ms02.red.id))
ms02.color[ms02.blue.id] <- rep("blue", times=length(ms02.blue.id))
ms02.color[ms02.yellow.id] <- rep("yellow", times=length(ms02.yellow.id))
V(ms02.sub.graph)$color <- ms02.color

#pdf("ms02.Breast-Colorectal.subnetwork.pdf")
plot.igraph(ms02.sub.graph, vertex.color=V(ms02.sub.graph)$color,
            vertex.size=8, edge.width=1,vertex.label=NA,
            order=sub.order,layout=coords)
```



```
#dev.off()
length(E(ms02.sub.graph))  #=213
```

```
## [1] 213
```

```
which(disorder.summary$ID %in% id.a) #=67
```

```
## [1] 67
```

```
which(disorder.summary$ID %in% id.b) #=38
```

```
## [1] 38
```

```
ddi.z[67,38] # =13.5
```

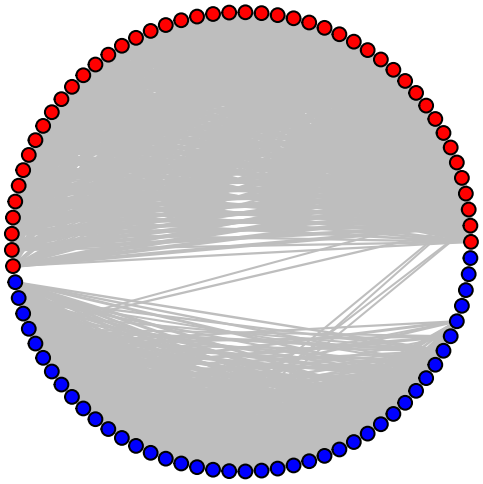
```
## [1] 13.70057
```

Creat a network model for cancer + mental disorder?

```
cm.matrix.nodiag <- cm.matrix
diag(cm.matrix.nodiag) <- NA
cm.net <- graph.adjacency(cm.matrix.nodiag, mode="undirected", weighted=T, diag=F)
E(cm.net)$color <- ifelse(E(cm.net)$weight > 0, "grey", "blue")
coloring <- E(cm.net)$color
E(cm.net)$weight <- ifelse((E(cm.net)$weight > 10),
                     abs(E(cm.net)$weight), -0.5)

red.color <- c(1:3,6,21,23:31,33:37,39:50,56,65,68:74,76:79,87:89)
blue.color <- c(4:5,7:20,22,32,38,51:55,57:64,66:67,75,80:86)
edge.list <- c(red.color, blue.color)
permu.list <- c(1:3,48:49,4,50:63,5,64,6:14,65,15:19,66,20:31,67:71,
                32,72:79,33,80:81,34:40,82,41:44,83:89,45:47)

cm.net.permute <- permute(cm.net, permu.list)
graph.isomorphic(cm.net, cm.net.permute)
```

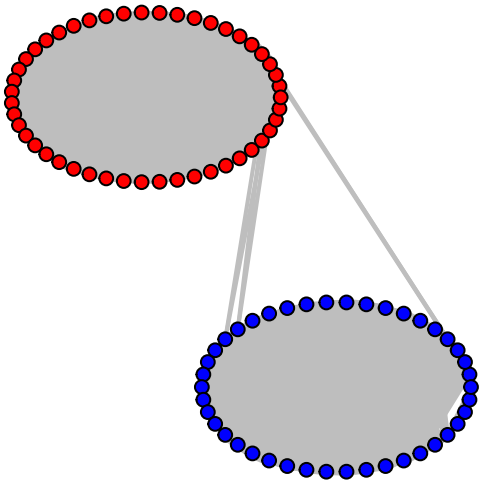
```
## [1] TRUE
```

```
V(cm.net.permute)[1:47]$color <- "red"
V(cm.net.permute)[48:89]$color <- "blue"

groups <- c(rep(1,47), rep(2,42))

plot.igraph(cm.net.permute, vertex.label=NA, layout=layout_in_circle,
            edge.color = coloring, edge.width=E(cm.net)$weight/10,
            vertex.size=6)
```



```
GroupByVertex02 = function(Groups) {
         numGroups = length(unique(Groups))
         GAngle    = (1:numGroups) * 1.5 * pi / numGroups
         Centers   = matrix(c(cos(GAngle), sin(GAngle)), ncol=2)
         x = y = c()
         for(i in 1:numGroups) {
```

```
                curGroup = which(Groups == unique(Groups)[i])
                VAngle = (1:length(curGroup)) * 2 * pi / length(curGroup)
                x = c(x, Centers[i,1] + cos(VAngle) / numGroups )
                y = c(y, Centers[i,2] + sin(VAngle) / numGroups)
        }
        matrix(c(x, y), ncol=2)
}
layout.test <- GroupByVertex02(groups)
plot.igraph(cm.net.permute, vertex.label=NA, layout=layout.test,
            edge.color = coloring, edge.width=E(cm.net)$weight/5,
            vertex.size=6)
```



```
red.angle = (1:49)*2*pi/49
```