

# CM15 set, BP terms frequency & heatmap

*hguo*

*October 7, 2018*

```
library(plyr)
# read the network
network <- read.csv("../Data/yeast.pin.csv",header=TRUE, stringsAsFactors=F)
geneA <- network$geneA
geneB <- network$geneB

##the GO terms for BP
bp.GOcategory.file <- read.csv("../Data/yeast.bp.term.csv",header=TRUE, stringsAsFactors=F)
bp.go.cat <- bp.GOcategory.file$GO.term
bp.dim <- length(bp.go.cat)

# GO term file
bp.GOterm.file <- read.csv("../Data/yeast.bp.gene.term.csv", header=T, stringsAsFactors = F)
bp.GO.gene <- bp.GOterm.file$System
bp.GO.term <- bp.GOterm.file$GO.term

# gene list
list.file <- read.csv("cm15.list.csv", header=T, stringsAsFactors=F)
gene.list <- list.file$gene

glA <- geneB[which(geneA %in% gene.list)]
glB <- geneA[which(geneB %in% gene.list)]
gl.all <- c(glA, glB)
gene.count <- count(gl.all)
orfs <- gene.count$x
freqs <- gene.count$freq

vec <- numeric(length=bp.dim)

for (i in 1:length(orfs)) {
  orf.ith <- orfs[i]
  orf.freq <- freqs[i]
  orf.term <- bp.GO.term[which(bp.GO.gene %in% orf.ith)]
  for (j in 1:length(orf.term)) {
    na <- which(bp.go.cat %in% orf.term[j])
    vec[na] <- vec[na] + orf.freq
  }
}

write.table(vec, file="yeast.cm15.bp.refine.txt", col.names=F, row.names=F, quote=F)
```