

# Z-score heat map, yeast CM15 set, BP terms

HBG

10/7/2018

```
library("microbenchmark")
library("matrixStats")

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)

conn.dim <- 1

biogrid <- matrix(as.numeric(unlist(read.table("yeast.cm15.bp.txt", header = F, sep = ","))), nrow=bp.d
obs <- c(biogrid)

perm <- c()
for (i in 1:50) {
  name <- paste("ms02.yeast", "/", "cm15", "/", "ms02.", i, ".bp.matrix.csv", sep = "")
  mat <- matrix(as.numeric(unlist(read.table(name, header=F, sep=","))), nrow=bp.dim, ncol=conn.dim)
  perm <- rbind(perm, c(mat))
}

mean <- colMeans(perm)
std <- colSds(perm)

zscore <- round((obs - mean)/std, 3)

z <- matrix(zscore, nrow=bp.dim, ncol=conn.dim)

write.table(z, file="yeast.cm15.bp.z.csv", sep = ",", row.names=F, col.names=F)

library('gplots')

##
## Attaching package: 'gplots'
## The following object is masked from 'package:stats':
##
##      lowess
library('GO.db')

## Loading required package: AnnotationDbi
## Loading required package: stats4
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
```

```

##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB
## 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,
##   colMeans, colnames, colSums, dirname, do.call, duplicated,
##   eval, evalq, Filter, Find, get, grep, grepl, intersect,
##   is.unsorted, lapply, lengths, Map, mapply, match, mget, order,
##   paste, pmax, pmax.int, pmin, pmin.int, Position, rank, rbind,
##   Reduce, rowMeans, rownames, rowSums, sapply, setdiff, sort,
##   table, tapply, union, unique, unsplit, which, 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)".
##
## Attaching package: 'Biobase'
## The following objects are masked from 'package:matrixStats':
##
##   anyMissing, rowMedians
## Loading required package: IRanges
## Warning: package 'IRanges' was built under R version 3.5.1
## Loading required package: S4Vectors
##
## Attaching package: 'S4Vectors'
## The following object is masked from 'package:gplots':
##
##   space
## The following object is masked from 'package:base':
##
##   expand.grid
##
order <- order(z)
bp.go.cat <- bp.go.cat[order]
z <- z[order]

z <- t(z)
colnames(z) <- bp.go.cat

```

```

enriched.list <- bp.go.cat[which(z >= 3)]
enriched <- c("GO.ID", "GO.Term", "Z-score")
for (i in 1:length(enriched.list)) {
  term <- Term(GOID(as.character(enriched.list[i])))
  enriched <- rbind(enriched, c(as.character(enriched.list[i]),term,
                                z[which(bp.go.cat %in% enriched.list[i])]))
}

print(enriched)

```

```

##
## enriched "GO.ID"
##          "GO:0006487"
##          "GO:0006364"
##          "GO:0006325"
##          "GO:0007030"
##          "GO:0006417"
##          "GO:0017148"
##          "GO:0006413"
##          "GO:0032197"
##          "GO:0006099"
##          "GO:0000055"
##          "GO:0002181"
##          "GO:0006886"
##          "GO:0006511"
##          "GO:0016573"
##          "GO:0030163"
##          "GO:0000463"
##          "GO:0000027"
##          "GO:0042254"
##          "GO:0006508"
##          "GO:0016579"
##          "GO:0034622"
##          "GO:0042273"
##          GO:0006487
## enriched "GO.Term"
##          "protein N-linked glycosylation"
##          "rRNA processing"
##          "chromatin organization"
##          "Golgi organization"
##          "regulation of translation"
##          "negative regulation of translation"
##          "translational initiation"
##          "transposition, RNA-mediated"
##          "tricarboxylic acid cycle"
##          "ribosomal large subunit export from nucleus"
##          "cytoplasmic translation"
##          "intracellular protein transport"
##          "ubiquitin-dependent protein catabolic process"
##          "histone acetylation"
##          "protein catabolic process"
##          "maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)"
##          "ribosomal large subunit assembly"
##          "ribosome biogenesis"

```

```
##      "proteolysis"
##      "protein deubiquitination"
##      "cellular protein-containing complex assembly"
##      "ribosomal large subunit biogenesis"
##
## enriched "Z-score"
##      "3.017"
##      "3.09"
##      "3.136"
##      "3.198"
##      "3.231"
##      "3.274"
##      "3.496"
##      "3.694"
##      "3.783"
##      "3.849"
##      "3.982"
##      "4.091"
##      "4.201"
##      "4.861"
##      "5.087"
##      "5.284"
##      "5.326"
##      "5.817"
##      "6.075"
##      "6.297"
##      "6.325"
##      "7.612"
```

```
write.table(enriched, file="yeast.cm15.bp.enriched.csv", row.names=F, col.names=F, quote=F, sep="\t")
```

```
sup.list <- bp.go.cat[which(z <= -3)]
sup <- c("GO.ID", "GO.Term", "Z-score")
for (i in 1:length(sup.list)) {
  term <- Term(GOID(as.character(sup.list[i])))
  sup <- rbind(sup, c(as.character(sup.list[i]),term,
                      z[which(bp.go.cat %in% sup.list[i])]))
}
```

```
print(sup)
```

```
##
## sup "GO.ID"
##      "GO:0008380"
##      "GO:0000398"
##      "GO:0007049"
##      "GO:0006281"
##      "GO:0000727"
##      "GO:0006397"
##      "GO:0006289"
##      "GO:0051301"
##      "GO:0006270"
##      "GO:0006974"
##      "GO:0007059"
##      "GO:0043044"
```

```

##      "GO:0007131"
##      "GO:0045944"
##      "GO:0000086"
##      "GO:0006310"
##      "GO:0046777"
##      "GO:0006468"
##      "GO:0006298"
##      "GO:0016310"
##      "GO:0032508"
##      GO:0008380
## sup "GO.Term"                                "Z-score"
##      "RNA splicing"                          "-7.112"
##      "mRNA splicing, via spliceosome"         "-6"
##      "cell cycle"                            "-4.612"
##      "DNA repair"                            "-4.49"
##      "double-strand break repair via break-induced replication" "-4.482"
##      "mRNA processing"                       "-4.102"
##      "nucleotide-excision repair"             "-3.956"
##      "cell division"                         "-3.953"
##      "DNA replication initiation"             "-3.937"
##      "cellular response to DNA damage stimulus" "-3.698"
##      "chromosome segregation"                 "-3.591"
##      "ATP-dependent chromatin remodeling"     "-3.591"
##      "reciprocal meiotic recombination"       "-3.373"
##      "positive regulation of transcription by RNA polymerase II" "-3.364"
##      "G2/M transition of mitotic cell cycle"  "-3.35"
##      "DNA recombination"                     "-3.31"
##      "protein autophosphorylation"            "-3.173"
##      "protein phosphorylation"               "-3.096"
##      "mismatch repair"                       "-3.083"
##      "phosphorylation"                       "-3.066"
##      "DNA duplex unwinding"                  "-3.016"

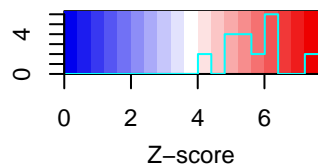
write.table(sup, file="yeast.cm15.bp.suppressed.csv", row.names=F, col.names=F, quote=F, sep="\t")

# plot the top 10 enriched BP terms
top10.z <- matrix(z[(length(z)-9):length(z)], ncol=1)
top10.bp.id <- bp.go.cat[(length(z)-9):length(z)]
rownames(top10.z) <- bp.go.cat[(length(z)-9):length(z)]

colors = c(seq(0,max(z),length=20))
my_palette <- colorRampPalette(c("blue2","white", "red2"))(n = 19)

#png(filename = "yeast.cm15.bp.enriched.png",width=10, height=4, res=1200, unit="in")
heatmap.2(cbind(top10.z, top10.z), trace="none", Colv=NA, Rowv=NA, dendrogram="none",
          col=my_palette, breaks=colors, revC=T, key.title=NA, key.xlab="Z-score", key.ylab=NA,
          labCol="", labRow="", cellnote=cbind(rownames(top10.z), rownames(top10.z)),notecol=1,
          colsep=1:ncol(z), rowsep=1:nrow(z), sepcolor = "lightgrey")

```



GO:0042273	GO:0042273
GO:0034622	GO:0034622
GO:0016579	GO:0016579
GO:0006508	GO:0006508
GO:0042254	GO:0042254
GO:0000027	GO:0000027
GO:0000463	GO:0000463
GO:0030163	GO:0030163
GO:0016573	GO:0016573
GO:0006511	GO:0006511

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

```
term <- c("GO.ID", "GO.Term", "Z-score")
for (i in 1:10) {
  top10.term <- Term(GOID(as.character(top10.bp.id[i])))
  term <- rbind(term, c(top10.bp.id[i], top10.term, top10.z[i]))
}
```

```
print(term)
```

```
##
## term "GO.ID"
##      "GO:0006511"
##      "GO:0016573"
##      "GO:0030163"
##      "GO:0000463"
##      "GO:0000027"
##      "GO:0042254"
##      "GO:0006508"
##      "GO:0016579"
##      "GO:0034622"
##      "GO:0042273"
##      GO:0006511
## term "GO.Term"
##      "ubiquitin-dependent protein catabolic process"
##      "histone acetylation"
##      "protein catabolic process"
##      "maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)"
##      "ribosomal large subunit assembly"
##      "ribosome biogenesis"
##      "proteolysis"
##      "protein deubiquitination"
```

```
##      "cellular protein-containing complex assembly"
##      "ribosomal large subunit biogenesis"
##
## term "Z-score"
##      "4.201"
##      "4.861"
##      "5.087"
##      "5.284"
##      "5.326"
##      "5.817"
##      "6.075"
##      "6.297"
##      "6.325"
##      "7.612"
```

```
#heatmap.2(cbind(top10.z, top10.z), trace="none", Colv=NA, Rowv=NA, dendrogram="none",
#           col=my_palette, breaks=colors, revC=T, key.title=NA, key.xlab="Z-score", key.ylab=NA,
#           labCol="", labRow="", cellnote=cbind(rownames(top10.z), term), notecol=1,
#           colsep=1:ncol(z), rowsep=1:nrow(z), sepcolor = "lightgrey")
```