MS02star annotation, Human PD-1

hguo

October 8, 2018

```r
## read the original network
network <- read.csv("../Data/human.pin.csv", header=T, stringsAsFactors=F)
geneA <- network$geneA
geneB <- network$geneB

## the GO terms for biological processes (BP)
## In this slim, each GO terms was shared by at least 20 genes
GOcategory.file <- read.csv("../Data/human.bp.term.csv",header=TRUE, stringsAsFactors=F)
bp.go.cat <- GOcategory.file$GO.term
bp.dim <- length(bp.go.cat)

# Gene-GO term file
# Note in human PIN the standard gene names have been used
GOterm.file <- read.csv("../Data/human.bp.gene.term.csv", header=T, stringsAsFactors=F)
bp.GO.gene <- GOterm.file$gene   #it should be changed to System for yeast pin
bp.GO.term <- GOterm.file$GO.term

#gene list, only a single gene is specified here for annotation
gene.list <- c("PDCD1")

vec <- numeric(length=bp.dim)

for (i in 1:length(gene.list)) {
    orf <- as.character(gene.list[i])
    intA <- geneB[which(geneA %in% orf)]
    for (j in 1:length(intA)) {
        bpA <- bp.GO.term[which(bp.GO.gene %in% intA[j])]
        for (k in 1:length(bpA)) {
           na <- which(bp.go.cat %in% bpA[k])
           vec[na] <- vec[na] + 1
        }
    }

    intB <- geneA[which(geneB %in% orf)]
    for (s in 1:length(intB)) {
        bpB <- bp.GO.term[which(bp.GO.gene %in% intB[s])]
        for (t in 1:length(bpB)) {
           nb <- which(bp.go.cat %in% bpB[t])
           vec[nb] <- vec[nb] + 1
        }
    }
}

write.table(vec, file="human.pdcd1.bp.txt", col.names=F, row.names=F, quote=F)

# Now the ms02star permutations
for (p in 1:100) {
```

```r
permutation.file <- paste("../ms02star/human/", "ms02.", p, ".csv", sep="")
permutation <- read.csv(permutation.file, header=T, stringsAsFactors = F)
geneA <- permutation$id1
geneB <- permutation$id2

vecp <-  numeric(length = bp.dim)
for (i in 1:length(gene.list)) {
    orf <- as.character(gene.list[i])
    intA <- geneB[which(geneA %in% orf)]
    for (j in 1:length(intA)) {
        bpA <- bp.GO.term[which(bp.GO.gene %in% intA[j])]
        for (k in 1:length(bpA)) {
            na <- which(bp.go.cat %in% bpA[k])
            vecp[na] <- vecp[na] + 1
        }
    }

    intB <- geneA[which(geneB %in% orf)]
    for (s in 1:length(intB)) {
        bpB <- bp.GO.term[which(bp.GO.gene %in% intB[s])]
        for (t in 1:length(bpB)) {
            nb <- which(bp.go.cat %in% bpB[t])
            vecp[nb] <- vecp[nb] + 1
        }
    }
}

output <- paste("ms02.human", "/", "pdcd1", "/", "ms02.", p, ".bp.matrix.csv", sep="")

write.table(vecp, file = output, col.names=F, row.names=F, quote=F)
}
```

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

conn.dim <- 1

hspin <- matrix(as.numeric(unlist(read.table("human.pdcd1.bp.txt", header=F, sep=","))), nrow=bp.dim, n
obs <- c(hspin)

perm <- c()
for (i in 1:100) {
    name <- paste("ms02.human", "/", "pdcd1", "/", "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="human.pdcd1,bp.z.csv", sep=",", row.names=F, col.names=F, quote=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

##
```

```r
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:0001932"
##         "GO:0043407"
##         "GO:0048008"
##         "GO:0006958"
##         "GO:0002244"
##         "GO:0006911"
##         "GO:0009755"
##         "GO:0042267"
##         "GO:0050829"
##         "GO:0070372"
##         "GO:0006470"
##         "GO:0021549"
##         "GO:0007163"
##         "GO:0034332"
##         "GO:0046627"
##         "GO:0071222"
##         "GO:0006910"
##         "GO:0048709"
##         "GO:0048873"
##         "GO:0035335"
##         "GO:0050731"
```

```
##         "GO:0051260"
##         "GO:0006641"
##         "GO:0032753"
##         "GO:0045879"
##         "GO:0050871"
##         "GO:0046854"
##         "GO:0001816"
##         "GO:0046676"
##         "GO:0048839"
##         "GO:0051209"
##         "GO:0051924"
##         "GO:0060325"
##         "GO:0050900"
##         "GO:0030168"
##         "GO:0071345"
##         "GO:0006882"
##         "GO:0042110"
##         "GO:0050890"
##         "GO:0007169"
##         "GO:0050853"
##         "GO:0007166"
##         "GO:0032729"
##         "GO:0032689"
##         "GO:0050690"
##         "GO:0042130"
##         "GO:0050776"
##         "GO:0031295"
##         "GO:0006955"
##         "GO:0042102"
##         "GO:0030217"
##         "GO:0002250"
##         "GO:0019886"
##         "GO:0060333"
##         "GO:0050852"
##         "GO:0070207"
##         GO:0001932
## enriched "GO.Term"
##         "regulation of protein phosphorylation"
##         "negative regulation of MAP kinase activity"
##         "platelet-derived growth factor receptor signaling pathway"
##         "complement activation, classical pathway"
##         "hematopoietic progenitor cell differentiation"
##         "phagocytosis, engulfment"
##         "hormone-mediated signaling pathway"
##         "natural killer cell mediated cytotoxicity"
##         "defense response to Gram-negative bacterium"
##         "regulation of ERK1 and ERK2 cascade"
##         "protein dephosphorylation"
##         "cerebellum development"
##         "establishment or maintenance of cell polarity"
##         "adherens junction organization"
##         "negative regulation of insulin receptor signaling pathway"
##         "cellular response to lipopolysaccharide"
##         "phagocytosis, recognition"
```

```
##         "oligodendrocyte differentiation"
##         "homeostasis of number of cells within a tissue"
##         "peptidyl-tyrosine dephosphorylation"
##         "positive regulation of peptidyl-tyrosine phosphorylation"
##         "protein homooligomerization"
##         "triglyceride metabolic process"
##         "positive regulation of interleukin-4 production"
##         "negative regulation of smoothened signaling pathway"
##         "positive regulation of B cell activation"
##         "phosphatidylinositol phosphorylation"
##         "cytokine production"
##         "negative regulation of insulin secretion"
##         "inner ear development"
##         "release of sequestered calcium ion into cytosol"
##         "regulation of calcium ion transport"
##         "face morphogenesis"
##         "leukocyte migration"
##         "platelet activation"
##         "cellular response to cytokine stimulus"
##         "cellular zinc ion homeostasis"
##         "T cell activation"
##         "cognition"
##         "transmembrane receptor protein tyrosine kinase signaling pathway"
##         "B cell receptor signaling pathway"
##         "cell surface receptor signaling pathway"
##         "positive regulation of interferon-gamma production"
##         "negative regulation of interferon-gamma production"
##         "regulation of defense response to virus by virus"
##         "negative regulation of T cell proliferation"
##         "regulation of immune response"
##         "T cell costimulation"
##         "immune response"
##         "positive regulation of T cell proliferation"
##         "T cell differentiation"
##         "adaptive immune response"
##         "antigen processing and presentation of exogenous peptide antigen via MHC class II"
##         "interferon-gamma-mediated signaling pathway"
##         "T cell receptor signaling pathway"
##         "protein homotrimerization"
##
## enriched "Z-score"
##         "3.164"
##         "3.164"
##         "3.164"
##         "3.172"
##         "3.374"
##         "3.374"
##         "3.374"
##         "3.374"
##         "3.374"
##         "3.374"
##         "3.384"
##         "3.627"
##         "3.938"
```

```
##         "3.938"
##         "3.938"
##         "3.967"
##         "4.337"
##         "4.337"
##         "4.337"
##         "4.357"
##         "4.477"
##         "4.607"
##         "4.874"
##         "4.874"
##         "4.874"
##         "4.874"
##         "5.229"
##         "5.658"
##         "5.658"
##         "5.658"
##         "5.658"
##         "5.658"
##         "5.658"
##         "5.823"
##         "6.438"
##         "6.641"
##         "6.965"
##         "9.496"
##         "9.9"
##         "10.117"
##         "10.709"
##         "11.368"
##         "11.426"
##         "11.49"
##         "12.187"
##         "13.468"
##         "17.663"
##         "18.216"
##         "18.712"
##         "18.73"
##         "20.697"
##         "20.74"
##         "24.028"
##         "25.793"
##         "40.757"
##         "Inf"
```

```r
write.table(enriched, file="human.pdcd1.bp.enriched.csv", row.names=F, col.names=F, quote=F, sep="\t")

###No suppressed terms have been found
#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)

#write.table(sup, file="human.pdcd1.bp.suppressed.csv", row.names=F, col.names=F, quote=F, sep="\t")
###

#Note that there may be "inf" Z-scores owing to lack of sampling (i.e., zero in standard deviations)
#We can also extract the GO-terms for the gene for comparison

terms <- bp.GO.term[which(bp.GO.gene %in% orf)]
extract <- c("GO.ID", "GO.Term")
for (i in 1:length(terms)) {
  go.id <- as.character(terms[i])
  go.con <- Term(GOID(go.id))
  extract <- rbind(extract, c(go.id, go.con))
}

print(extract)
```

```
##                    GO:0006915
## extract "GO.ID"       "GO.Term"
##         "GO:0006915"  "apoptotic process"
##         "GO:0006959"  "humoral immune response"
##         "GO:0007275"  "multicellular organism development"
##         "GO:0031295"  "T cell costimulation"
##         "GO:0043066"  "negative regulation of apoptotic process"
```

```
write.table(extract, file="human.pdcd1.bp.original.csv", row.names=F, col.names=F, quote=F, sep="\t")
```