

P vs NP, differential measurement

HBG

11/5/2018

This script use the differential measure (z-score) between tumor/non-tumor cells

```
### this block perform z-score calculations
library("microbenchmark")
library("matrixStats")

## the GO terms for biological processes (BP)
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)
conn.dim <- 20
hspin <- matrix(as.numeric(unlist(read.table("human.rnf43.PvNP.bp.txt", header=F, sep=","))), nrow=bp.d
obs <- c(hspin)

perm <- c()
for (i in 1:100) {
  name <- paste("ms02.human", "/", "rnf43.PvNP", "/", "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.pvnp <- matrix(zscore, nrow=bp.dim, ncol=conn.dim)

#write.table(z, file="human.rnf43.PvNP.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

z.pvnp <- t(z.pvnp)

rownames(z.pvnp) <- c("5%", "10%", "15%", "20%", "25%", "30%", "35%", "40%", "45%", "50%",
  "55%", "60%", "65%", "70%", "75%", "80%", "85%", "90%", "95%", "100%")
```

```

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

colors = c(seq(min(z.pvnp),-10.1,length=100),seq(-9.9,9.9,length=100),seq(10.1,max(z.pvnp),length=100))
my_palette <- colorRampPalette(c("blue2", "white", "red2"))(n = 299)

png(filename = "17.PvsNP.z.hm.png",width=6, height=5.5, res=1200, unit="in")
heatmap.2(z.pvnp, col=my_palette, breaks=colors, dendrogram='none', Rowv=F, Colv=F,
  trace='none', offsetRow = 0, offsetCol = 0,
  xlab="Biological Process Terms", ylab="PDAC vs Normal Cells, Quantiles",
  margins = c(2,3.5), key.title = "Color Bar", key.xlab="Z-score", key.ylab=NA,
  revC = T,
  labCol = NA, #labRow =,
  #srtCol=45, adjCol=c(1,0),
  #lmat=rbind(c(0,3,4), c(2,1,0)), lwid=c(1.5,4,2),
  scale="none", symbreaks=T, symm=F, symkey = F)
dev.off()

## pdf
## 2

#correlations to quantile 20 (95%-100%) genes
y.val <- c()
for (i in 1:20) {
  y.val2 <- c()
  for (j in 1:20) {
    cor.val <- cor.test(z.pvnp[j,], z.pvnp[i,])
    y.val2 <- rbind(y.val2, cor.val$estimate)
  }
  y.val <- cbind(y.val, y.val2)
}

rownames(y.val) <- c("5%", "10%", "15%", "20%", "25%", "30%", "35%", "40%", "45%", "50%",
  "55%", "60%", "65%", "70%", "75%", "80%", "85%", "90%", "95%", "100%")
colnames(y.val) <- c("5%", "10%", "15%", "20%", "25%", "30%", "35%", "40%", "45%", "50%",
  "55%", "60%", "65%", "70%", "75%", "80%", "85%", "90%", "95%", "100%")
my_palette2 <- colorRampPalette(c("blue2", "white", "red2"))(9)
png(filename = "17.PvsNP.z.hm.cor.matrix.png",width=6, height=5.5, res=1200, unit="in")
heatmap.2(y.val, col=my_palette2, dendrogram='none', Rowv=F, Colv=F,
  trace='none', offsetRow = 0, offsetCol = 0,
  xlab="Quantiles of Bayes Factors", ylab="Quantiles of Bayes Factors",
  margins = c(3.5,3.5), key.title = "Color Bar", key.xlab="Pearson Correlation", key.ylab=NA,
  revC = T,
  #labCol = NA, #labRow =,
  #srtCol=45, adjCol=c(1,0),
  #lmat=rbind(c(0,3,4), c(2,1,0)), lwid=c(1.5,4,2),
  scale="none", symbreaks=T, symm=F, symkey = F)
dev.off()

## pdf
## 2

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
## 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
##

```

```

hspin <- matrix(as.numeric(unlist(read.table("human.rnf43.all.bp.txt", header=F, sep=","))), nrow=bp.dim,
obs <- c(hspin)

perm <- c()
for (i in 1:100) {
  name <- paste("ms02.human", "/", "rnf43.heatmap", "/", "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.p <- matrix(zscore, nrow=bp.dim, ncol=conn.dim)
z.p <- t(z.p)
rownames(z.p) <- c("5%", "10%", "15%", "20%", "25%", "30%", "35%", "40%", "45%", "50%",
"55%", "60%", "65%", "70%", "75%", "80%", "85%", "90%", "95%", "100%")
colnames(z.p) <- bp.go.cat

hspin <- matrix(as.numeric(unlist(read.table("human.rnf43.NP.bp.txt", header=F, sep=","))), nrow=bp.dim,
obs <- c(hspin)

perm <- c()
for (i in 1:100) {
  name <- paste("ms02.human", "/", "rnf43.NP", "/", "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.np <- matrix(zscore, nrow=bp.dim, ncol=conn.dim)
z.np <- t(z.np)
rownames(z.np) <- c("5%", "10%", "15%", "20%", "25%", "30%", "35%", "40%", "45%", "50%",
"55%", "60%", "65%", "70%", "75%", "80%", "85%", "90%", "95%", "100%")
colnames(z.np) <- bp.go.cat

enrich <- names(z.pvnp[20, which(z.pvnp[20,] > 8)])
suppress <- names(z.pvnp[20, which(z.pvnp[20,] < -8)])
enriched.terms <- c("G0.ID", "G0.Term", "Z.Tumor", "Z.Normal", "Z.Differential")
for (i in 1:length(enrich)) {
  id <- as.character(enrich[i])
  term <- Term(GOID(id))
  seri <- which(bp.go.cat %in% id)
  z.tumor <- z.p[20,seri]
  z.normal <- z.np[20,seri]
  z.diff <- z.pvnp[20,seri]
  enriched.terms <- rbind(enriched.terms, c(id, term, z.tumor, z.normal, z.diff))
}

```

```

}
enriched.terms

##
## enriched.terms "GO.ID"
## "GO:0006886"
## "GO:0006888"
## "GO:0006890"
## "GO:0008286"
## "GO:0015031"
## "GO:0015721"
## "GO:0015991"
## "GO:0016241"
## "GO:0019216"
## "GO:0034220"
## "GO:0043401"
## "GO:0045165"
## "GO:0045893"
## "GO:0060070"
## "GO:0061024"
## "GO:1904837"
## GO:0006886
## enriched.terms "GO.Term"
## "intracellular protein transport"
## "ER to Golgi vesicle-mediated transport"
## "retrograde vesicle-mediated transport, Golgi to ER"
## "insulin receptor signaling pathway"
## "protein transport"
## "bile acid and bile salt transport"
## "ATP hydrolysis coupled proton transport"
## "regulation of macroautophagy"
## "regulation of lipid metabolic process"
## "ion transmembrane transport"
## "steroid hormone mediated signaling pathway"
## "cell fate commitment"
## "positive regulation of transcription, DNA-templated"
## "canonical Wnt signaling pathway"
## "membrane organization"
## "beta-catenin-TCF complex assembly"
##
## enriched.terms "Z.Tumor" "Z.Normal" "Z.Differential"
## "-5.797" "-12.979" "9.586"
## "0.447" "-5.762" "10.926"
## "5.891" "1.758" "8.685"
## "-4" "-9.397" "11.969"
## "-4.142" "-8.97" "10.599"
## "-0.11" "-0.828" "8.931"
## "6.554" "2.113" "8.124"
## "3.269" "-1.876" "8.35"
## "-8.752" "-11.681" "8.712"
## "-4.452" "-9.392" "9.001"
## "-5.947" "-7.538" "8.016"
## "-7.727" "-9.86" "9.338"
## "-11.599" "-11.275" "9.389"

```

```
##          "-8.112"  "-11.894"  "11.8"
##          "-6.643"  "-9.363"   "8.481"
##          "-0.481"  "-1.023"   "8.535"

write.table(enriched.terms, file="17.PvNP.bp.enriched.csv", sep="\t", col.names = F, row.names = F, quo

suppressed.terms <- c("GO.ID", "GO.Term", "Z.Tumor", "Z.Normal", "Z.Differential")
for (i in 1:length(suppress)) {
  id <- as.character(suppress[i])
  term <- Term(GO.ID(id))
  seri <- which(bp.go.cat %in% id)
  z.tumor <- z.p[20,seri]
  z.normal <- z.np[20,seri]
  z.diff <- z.pvnp[20,seri]
  suppressed.terms <- rbind(suppressed.terms, c(id, term, z.tumor, z.normal, z.diff))
}
suppressed.terms

##
## suppressed.terms "GO.ID"
##                "GO:0000184"
##                "GO:0000245"
##                "GO:0000375"
##                "GO:0000381"
##                "GO:0000398"
##                "GO:0006351"
##                "GO:0006357"
##                "GO:0006364"
##                "GO:0006368"
##                "GO:0006369"
##                "GO:0006370"
##                "GO:0006405"
##                "GO:0006406"
##                "GO:0006412"
##                "GO:0006413"
##                "GO:0006614"
##                "GO:0007186"
##                "GO:0008380"
##                "GO:0016070"
##                "GO:0031124"
##                GO:0000184
## suppressed.terms "GO.Term"
##                "nuclear-transcribed mRNA catabolic process, nonsense-mediated decay"
##                "spliceosomal complex assembly"
##                "RNA splicing, via transesterification reactions"
##                "regulation of alternative mRNA splicing, via spliceosome"
##                "mRNA splicing, via spliceosome"
##                "transcription, DNA-templated"
##                "regulation of transcription by RNA polymerase II"
##                "rRNA processing"
##                "transcription elongation from RNA polymerase II promoter"
##                "termination of RNA polymerase II transcription"
##                "7-methylguanosine mRNA capping"
##                "RNA export from nucleus"
##                "mRNA export from nucleus"
```

```

##          "translation"
##          "translational initiation"
##          "SRP-dependent cotranslational protein targeting to membrane"
##          "G protein-coupled receptor signaling pathway"
##          "RNA splicing"
##          "RNA metabolic process"
##          "mRNA 3'-end processing"
##
## suppressed.terms "Z.Tumor" "Z.Normal" "Z.Differential"
##          "14.405" "16.471" "-11.518"
##          "21.839" "26.653" "-9.467"
##          "26.615" "33.108" "-8.647"
##          "25.057" "30.565" "-8.374"
##          "37.487" "45.921" "-25.573"
##          "-37.573" "-35.083" "-11.756"
##          "-30.787" "-29.925" "-10.889"
##          "29.684" "33.13" "-9.211"
##          "30.011" "36.357" "-9.318"
##          "28.085" "36.305" "-9.622"
##          "26.45" "33.574" "-8.672"
##          "27.319" "34.43" "-11.203"
##          "28.886" "35.153" "-13.816"
##          "14.239" "15.477" "-8.634"
##          "13.262" "14.146" "-8.171"
##          "11.217" "13.523" "-10.395"
##          "-48.543" "-51.947" "-9.709"
##          "33.127" "40.361" "-11.357"
##          "27.901" "33.063" "-10.41"
##          "31.361" "40.541" "-15.223"

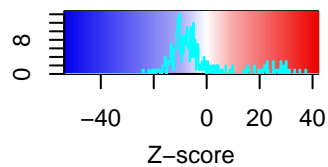
```

```
write.table(suppressed.terms, file="17.PvNP.bp.suppressed.csv", sep="\t", col.names = F, row.names = F,
```

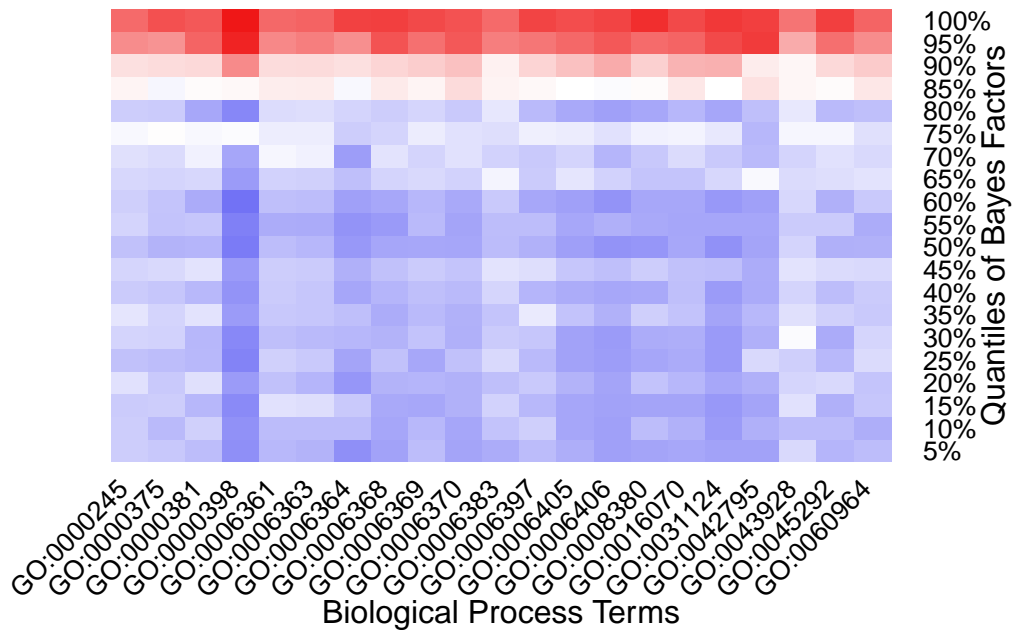
```

# PDAC cells
p.pos <- z.p[, which(z.p[20,] > 20)]
heatmap.2(p.pos, col=my_palette, breaks=colors, dendrogram='none',
  trace='none', Rowv=F, Colv=F, revC=T, #labCol=NA,
  xlab="Biological Process Terms", ylab="Quantiles of Bayes Factors",
  margins = c(5.5,4), key.title = NA, key.xlab="Z-score", key.ylab=NA,
  scale="none", symbreaks=T, symm=F, symkey = F,
  srtCol=45, adjCol=c(1,0),
  main="Tumor cells, top 5% enriched")

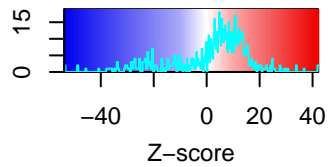
```



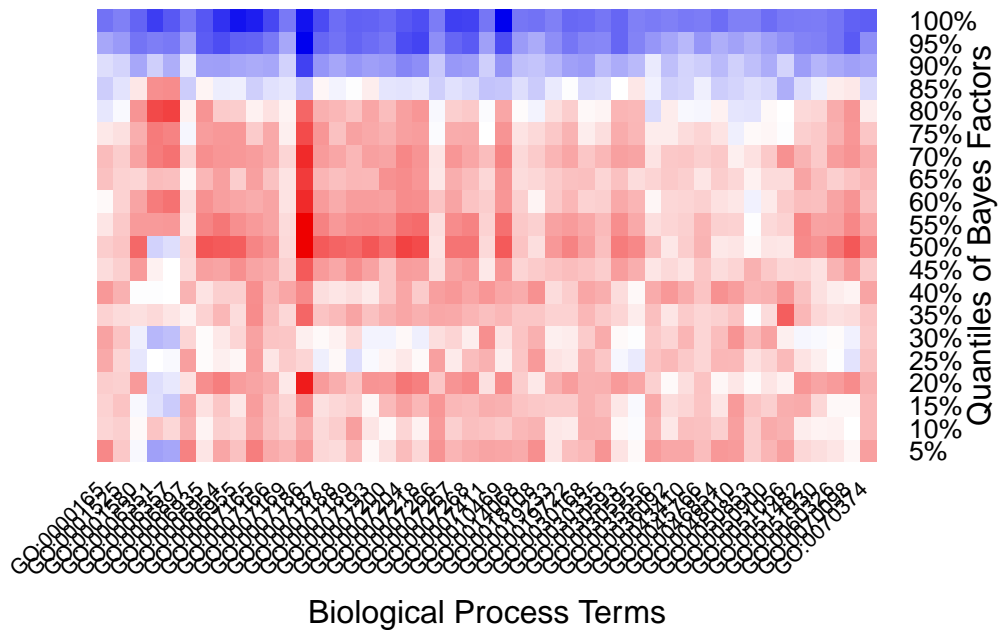
Tumor cells, top 5% enriched



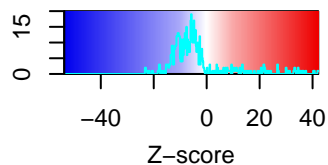
```
p.neg <- z.p[, which(z.p[20,] < -20)]
heatmap.2(p.neg, col=my_palette, breaks=colors, dendrogram='none',
  trace='none', Rowv=F, Colv=F, revC=T, #labCol=NA,
  xlab="Biological Process Terms", ylab="Quantiles of Bayes Factors",
  margins = c(5.5,4), key.title = NA, key.xlab="Z-score", key.ylab=NA,
  scale="none", symbreaks=T, symm=F, symkey = F,
  srtCol=45, adjCol=c(1,0),
  main="Tumor cells, top 5% suppressed")
```

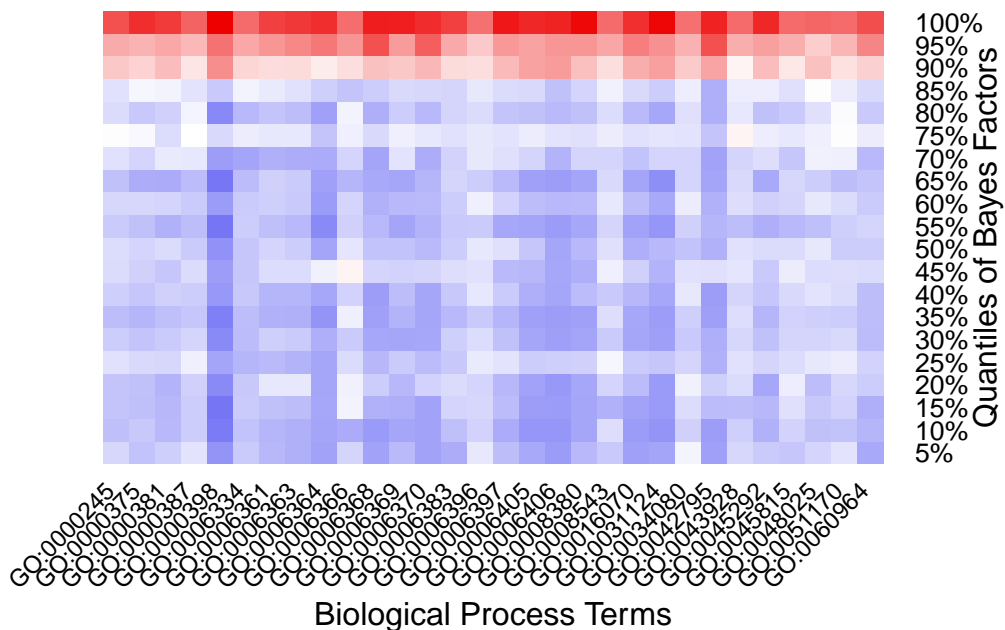
Tumor cells, top 5% suppressed



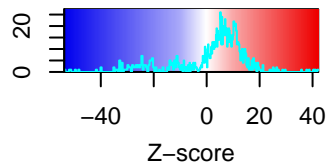
```
# Normal Cells
np.pos <- z.np[, which(z.np[20,] > 20)]
heatmap.2(np.pos, col=my_palette, breaks=colors, dendrogram='none',
  trace='none', Rowv=F, Colv=F, revC=T, #labCol=NA,
  xlab="Biological Process Terms", ylab="Quantiles of Bayes Factors",
  margins = c(5.5,4), key.title = NA, key.xlab="Z-score", key.ylab=NA,
  scale="none", symbreaks=T, symm=F, symkey = F,
  srtCol=45, adjCol=c(1,0),
  main="Normal cells, top 5% enriched")
```



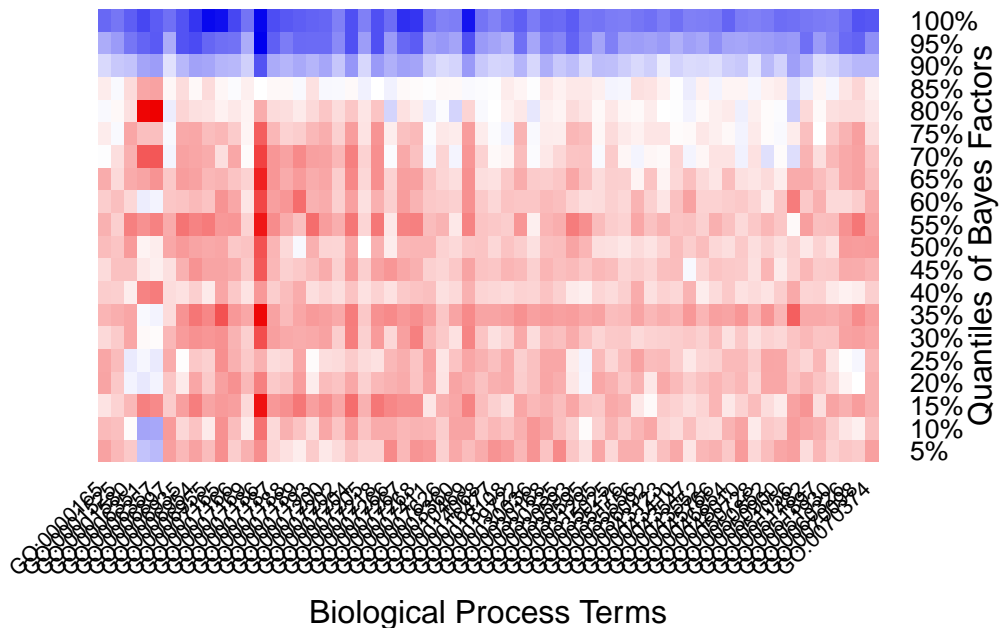
Normal cells, top 5% enriched



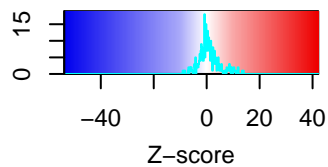
```
np.neg <- z.np[, which(z.np[20,] < -20)]
heatmap.2(np.neg, col=my_palette, breaks=colors, dendrogram='none',
  trace='none', Rowv=F, Colv=F, revC=T, #labCol=NA,
  xlab="Biological Process Terms", ylab="Quantiles of Bayes Factors",
  margins = c(5.5,4), key.title = NA, key.xlab="Z-score", key.ylab=NA,
  scale="none", symbreaks=T, symm=F, symkey = F,
  srtCol=45, adjCol=c(1,0),
  main="Normal cells, top 5% suppressed")
```



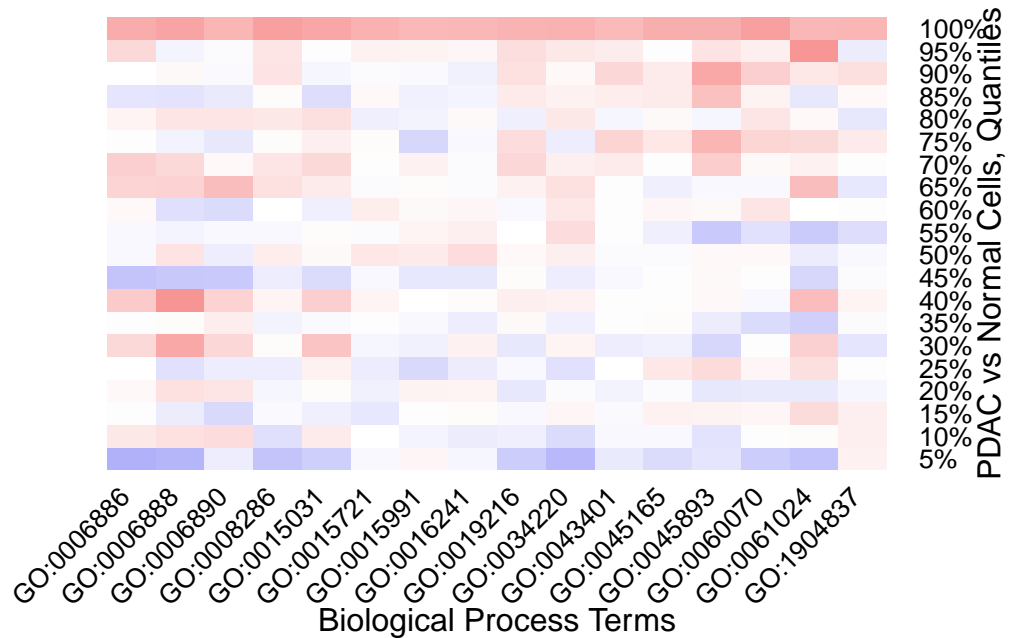
Normal cells, top 5% suppressed



```
# Differential z-scores
pvnp.pos <- z.pvnp[, which(z.pvnp[20,] > 8)]
heatmap.2(pvnp.pos, col=my_palette, breaks=colors, dendrogram='none',
  trace='none', Rowv=F, Colv=F, revC=T, #labCol=NA,
  xlab="Biological Process Terms", ylab="PDAC vs Normal Cells, Quantiles",
  margins = c(5.5,4), key.title = NA, key.xlab="Z-score", key.ylab=NA,
  scale="none", symbreaks=T, symm=F, symkey = F,
  srtCol=45, adjCol=c(1,0),
  main="Tumor vs normal cells, Enriched")
```

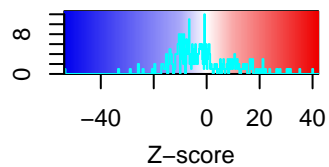


Tumor vs normal cells, Enriched



```
pvpnp.neg <- z.pvpnp[, which(z.pvpnp[20,] < -8)]

heatmap.2(pvpnp.neg, col=my_palette, breaks=colors, dendrogram='none',
  trace='none', Rowv=F, Colv=F, revC=T, #labCol= pvpnp.neg.ids,
  xlab="Biological Process Terms", ylab="PDAC vs Normal Cells, Quantiles",
  margins = c(5.5,4), key.title = NA, key.xlab="Z-score", key.ylab=NA,
  scale="none", symbreaks=T, symm=F, symkey = F,
  srtCol=45, adjCol=c(1,0),
  main="Tumor vs normal cells, suppressed")
```



Tumor vs normal cells, suppressed

