

GO term analysis_human_donor time_including July 2018 samples

Purpose:

To assess if the differentially expressed human (also includes HBV and HDV) genes (when donor and time are set as factors in the design) are enriched for members of specific GO terms. This analysis includes the July 2018 samples.

Load required libraries

```
library(pathview)

## Loading required package: org.Hs.eg.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, xtabs
## The following objects are masked from 'package:base':
##
##   anyDuplicated, append, as.data.frame, cbind, colnames,
##   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, rownames, 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)".
## Loading required package: IRanges
## Loading required package: S4Vectors
```

```

##
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:base':
##
##      colMeans, colSums, expand.grid, rowMeans, rowSums
##

## #####
## Pathview is an open source software package distributed under GNU General
## Public License version 3 (GPLv3). Details of GPLv3 is available at
## http://www.gnu.org/licenses/gpl-3.0.html. Particullary, users are required to
## formally cite the original Pathview paper (not just mention it) in publications
## or products. For details, do citation("pathview") within R.
##
## The pathview downloads and uses KEGG data. Non-academic uses may require a KEGG
## license agreement (details at http://www.kegg.jp/kegg/legal.html).
## #####

library(gage)
library(gageData)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:AnnotationDbi':
##
##      select

## The following objects are masked from 'package:IRanges':
##
##      collapse, desc, intersect, setdiff, slice, union

## The following objects are masked from 'package:S4Vectors':
##
##      first, intersect, rename, setdiff, setequal, union

## The following object is masked from 'package:Biobase':
##
##      combine

## The following objects are masked from 'package:BiocGenerics':
##
##      combine, intersect, setdiff, union

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

library(tibble)
library(gtools)
library(gplots)

##

```

```

## Attaching package: 'gplots'

## The following object is masked from 'package:IRanges':
##
##      space

## The following object is masked from 'package:S4Vectors':
##
##      space

## The following object is masked from 'package:stats':
##
##      lowess
library(ggplot2)
library(purrr)

##
## Attaching package: 'purrr'

## The following objects are masked from 'package:IRanges':
##
##      reduce, simplify
data(kegg.gs)
data(go.sets.hs)
data(go.subs.hs)
go.bp = go.sets.hs[go.subs.hs$BP] ##Only examining GO biological processes terms
library(reshape2)
library(stringr)
library(viridis)

## Loading required package: viridisLite
library(xlsx)

Read in the appropriate DGE files
data_dir <- "Human DGEs_donortime"
sampleFiles <- basename(Sys.glob(file.path(data_dir, "*txt")))
sampleNames <- str_replace(sampleFiles, "[0-9]*-[0-9]*-[0-9]*human_donor_time*", "") %>%
  str_replace("_*\\s*analysis_results.txt", "")
sampleNames

## [1] "humangenesis-d28_vs_d8_HBV" "humangenesis-d28_vs_d8_cointf"
## [3] "humangenesis-d28_vs_d8_mock"

##Function to appropriately format files for GO term analysis and then perform
##GO term analysis using GAGE package

##Assessing enrichment for GO terms that are downregulated amongst DGEs
GO_bp_less <- function(files) {
  d <- read.delim(files, header = TRUE)
  ##d <- d[-grep("AAB", d[,1]),]
  dd <- dplyr::select(d, log2FoldChange, padj, ENTREZID) %>%
    na.omit() %>%
    dplyr::select(log2FoldChange, ENTREZID) %>%
    distinct(ENTREZID, .keep_all = TRUE)
  dd$ENTREZID <- gsub(pattern = ",.*", replacement = "", dd$ENTREZID)

```

```

e = dd$log2FoldChange
names(e) = dd$ENTREZID
ef <- gage(na.omit(e), gsets = go.bp, same.dir = TRUE)
g <- ef$less
gg <- g[mixedorder(rownames(g), decreasing = TRUE),]
}

##Assessing enrichment for GO terms that are upregulated amongst DGEs
GO_bp_greater <- function(files) {
  d <- read.delim(files, header = TRUE)
  ##d <- d[-grep("AAB", d[,1]),]
  dd <- dplyr::select(d, log2FoldChange, padj, ENTREZID) %>%
  na.omit() %>%
  dplyr::select(log2FoldChange, ENTREZID) %>%
  distinct(ENTREZID, .keep_all = TRUE)
  dd$ENTREZID <- gsub(pattern = ".*", replacement = "", dd$ENTREZID)
  e = dd$log2FoldChange
  names(e) = dd$ENTREZID
  ef <- gage(na.omit(e), gsets = go.bp, same.dir = TRUE)
  g <- ef$greater
  gg <- g[mixedorder(rownames(g), decreasing = TRUE),]
}

##applying function
GO_less <- lapply(file.path(data_dir, sampleFiles), GO_bp_less)
names(GO_less) <- sampleNames

##making output into a data frame
form_GO_less <- lapply(GO_less, data.frame)
GO_less_df <- do.call("cbind", form_GO_less)%>%
  rownames_to_column(var = "Pathway") %>%
  dplyr::select(Pathway, ends_with("q.val"))

##filtering data frame to include GO terms where at least one group
##has a padjval <= 0.05
GO_less_df <- GO_less_df[as.logical((rowSums(is.na(GO_less_df))-
  ncol(GO_less_df))),] %>%
  dplyr::filter_all(any_vars(.<= 0.05))
head(GO_less_df)

```

```

##                                     Pathway
## 1 GO:1901615 organic hydroxy compound metabolic process
## 2      GO:1901606 alpha-amino acid catabolic process
## 3      GO:1901605 alpha-amino acid metabolic process
## 4      GO:0072329 monocarboxylic acid catabolic process
## 5      GO:0071941 nitrogen cycle metabolic process
## 6      GO:0071466 cellular response to xenobiotic stimulus
## humangenex-d28_vs_d8_HBV.q.val humangenex-d28_vs_d8_coinf.q.val
## 1                      1.886120e-04                0.9815381
## 2                      4.436008e-05                0.9815381
## 3                      6.456727e-06                0.9815381
## 4                      3.723026e-06                0.9815381
## 5                      1.680450e-02                0.9815381
## 6                      3.767713e-08                0.9815381

```

```
## humangenex-d28_vs_d8_mock.q.val
## 1 0.039613328
## 2 0.007349637
## 3 0.004869017
## 4 0.001666165
## 5 0.162148207
## 6 0.001005973

##csv of data frame
write.csv(GO_less_df, file = file.path(data_dir, paste(Sys.Date(),
  "donortime_GO biological processes lesser.csv")))

##applying function
GO_greater <- lapply(file.path(data_dir, sampleFiles), GO_bp_greater)
names(GO_greater) <- sampleNames

##making output into a data frame
form_GO_greater <- lapply(GO_greater, data.frame)
GO_greater_df <- do.call("cbind", form_GO_greater)%>%
  rownames_to_column(var = "Pathway") %>%
  dplyr::select(Pathway, ends_with("q.val"))

##filtering data frame to include GO terms where at least one group
##has a padjval <= 0.05
GO_greater_df <- GO_greater_df[as.logical((rowSums(is.na(GO_greater_df))-
  ncol(GO_greater_df))),] %>%
  dplyr::filter_all(any_vars(.<= 0.05))
head(GO_greater_df)
```

```
##
## 1 GO:2000147 positive regulation of cell motility
## 2 GO:2000146 negative regulation of cell motility
## 3 GO:2000145 regulation of cell motility
## 4 GO:1901701 cellular response to oxygen-containing compound
## 5 GO:1901342 regulation of vasculature development
## 6 GO:0072599 establishment of protein localization to endoplasmic reticulum
## humangenex-d28_vs_d8_HBV.q.val humangenex-d28_vs_d8_coinf.q.val
## 1 0.0212262745 0.9412814
## 2 0.0441540026 0.9412814
## 3 0.0001073218 0.9412814
## 4 0.0196536089 0.9412814
## 5 0.0177052810 0.9412814
## 6 0.3150402736 0.9412814
## humangenex-d28_vs_d8_mock.q.val
## 1 5.707096e-01
## 2 5.498094e-01
## 3 1.989167e-01
## 4 2.016429e-01
## 5 5.231209e-01
## 6 8.092183e-06

##csv of data frame
write.csv(GO_greater_df, file = file.path(data_dir, paste(Sys.Date(),
  "donortime_GO biological processes greater.csv")))
```

Session Info

sessionInfo()

```
## R version 3.3.3 (2017-03-06)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] parallel stats4      stats      graphics  grDevices  utils      datasets
## [8] methods    base
##
## other attached packages:
## [1] bindrcpp_0.2.2      xlsx_0.6.1          viridis_0.5.1
## [4] viridisLite_0.3.0   stringr_1.3.1        reshape2_1.4.3
## [7] purrr_0.2.5         ggplot2_3.0.0        gplots_3.0.1
## [10] gtools_3.8.1        tibble_1.4.2         dplyr_0.7.6
## [13] gageData_2.12.0     gage_2.24.0          pathview_1.14.0
## [16] org.Hs.eg.db_3.4.0  AnnotationDbi_1.36.2 IRanges_2.8.2
## [19] S4Vectors_0.12.2    Biobase_2.34.0       BiocGenerics_0.20.0
##
## loaded via a namespace (and not attached):
## [1] KEGGgraph_1.32.0    Rcpp_0.12.18         xlsxjars_0.6.1
## [4] png_0.1-7           Biostrings_2.42.1    assertthat_0.2.0
## [7] rprojroot_1.3-2     digest_0.6.15        R6_2.2.2
## [10] plyr_1.8.4          backports_1.1.2      RSQLite_2.1.1
## [13] evaluate_0.11       httr_1.3.1           pillar_1.3.0
## [16] zlibbioc_1.20.0     rlang_0.2.1          lazyeval_0.2.1
## [19] rstudioapi_0.7      gdata_2.18.0         Rgraphviz_2.18.0
## [22] blob_1.1.1          rmarkdown_1.10       bit_1.1-14
## [25] munsell_0.5.0       pkgconfig_2.0.1      htmltools_0.3.6
## [28] tidyselect_0.2.4    KEGGREST_1.14.1      gridExtra_2.3
## [31] XML_3.98-1.12       withr_2.1.2          crayon_1.3.4
## [34] bitops_1.0-6        grid_3.3.3           gtable_0.2.0
## [37] DBI_1.0.0           magrittr_1.5         scales_0.5.0
## [40] graph_1.52.0        KernSmooth_2.23-15   stringi_1.2.4
## [43] XVector_0.14.1      tools_3.3.3          bit64_0.9-7
## [46] glue_1.3.0          yaml_2.2.0           colorspace_1.3-2
## [49] caTools_1.17.1.1    rJava_0.9-10         memoise_1.1.0
## [52] knitr_1.20          bindr_0.1.1
```