GAGE analysis_mouse_donor treatment_including July 2018 samples

Purpose:

To assess if the differentially expressed murine genes(when donor and treatment are set as factors in the design) are enriched for members of specific pathways. 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.sets.mm)
data(sigmet.idx.mm)
kegg.sets.mm.s = kegg.sets.mm[sigmet.idx.mm]
library(reshape2)
library(stringr)
library(xlsx)
```

Read in the appropriate DGE files

```
data_dir <- "Mouse DGEs_donortreatment"
sampleFiles <- basename(Sys.glob(file.path(data_dir, "*txt")))
sampleNames <- str_replace(sampleFiles, "^[0-9]*-*[0-9]*-*[0-9]*mouse_donor_treatment*","") %>%
  str_replace("_*\\s*analysis_results.txt", "")
sampleNames
```

```
## [1] "Mousegenes-HBV_vs_mock_d28"   "Mousegenes-HBV_vs_mock_d8"
## [3] "Mousegenes-coinf_vs_HBV_d28"  "Mousegenes-coinf_vs_HBV_d8"
## [5] "Mousegenes-coinf_vs_mock_d28" "Mousegenes-coinf_vs_mock_d8"
```

```
##Function to appropriately format files for GAGE analysis and then actually perform
##GAGE analysis
gage_mixed <- function(files) {
  d <- read.delim(files, header = TRUE)
  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
  ##As per GAGE manual suggestion, looking at same.dir = FALSE since genes in pathways
  ##do not tend to just increase or just decrease.
  ef <- gage(na.omit(e), gsets = kegg.sets.mm.s, same.dir = FALSE)
  g <- ef$greater
  gg <- g[mixedorder(rownames(g), decreasing = TRUE),]
}
```

```
##Applying function to DGE files
gage_mixed_list <- lapply(file.path(data_dir, sampleFiles), gage_mixed)
names(gage_mixed_list) <- sampleNames
str(gage_mixed_list)
```

```
## List of 6
##  $ Mousegenes-HBV_vs_mock_d28  : num [1:177, 1:6] 0.895 0.536 0.757 0.421 0.389 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:177] "mmu04977 Vitamin digestion and absorption" "mmu04976 Bile secretion" "mmu049
##   .. ..$ : chr [1:6] "p.geomean" "stat.mean" "p.val" "q.val" ...
##  $ Mousegenes-HBV_vs_mock_d8   : num [1:177, 1:6] 0.8 0.786 0.982 0.621 0.245 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:177] "mmu04977 Vitamin digestion and absorption" "mmu04976 Bile secretion" "mmu049
##   .. ..$ : chr [1:6] "p.geomean" "stat.mean" "p.val" "q.val" ...
##  $ Mousegenes-coinf_vs_HBV_d28 : num [1:177, 1:6] 0.751 0.3145 0.5385 0.2284 0.0983 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:177] "mmu04977 Vitamin digestion and absorption" "mmu04976 Bile secretion" "mmu049
##   .. ..$ : chr [1:6] "p.geomean" "stat.mean" "p.val" "q.val" ...
##  $ Mousegenes-coinf_vs_HBV_d8  : num [1:177, 1:6] 0.66 0.608 0.608 0.834 0.467 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:177] "mmu04977 Vitamin digestion and absorption" "mmu04976 Bile secretion" "mmu049
##   .. ..$ : chr [1:6] "p.geomean" "stat.mean" "p.val" "q.val" ...
##  $ Mousegenes-coinf_vs_mock_d28: num [1:177, 1:6] 0.235 0.596 0.241 0.889 0.346 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:177] "mmu04977 Vitamin digestion and absorption" "mmu04976 Bile secretion" "mmu049
##   .. ..$ : chr [1:6] "p.geomean" "stat.mean" "p.val" "q.val" ...
##  $ Mousegenes-coinf_vs_mock_d8 : num [1:177, 1:6] 0.326 0.26 0.202 0.662 0.444 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:177] "mmu04977 Vitamin digestion and absorption" "mmu04976 Bile secretion" "mmu049
##   .. ..$ : chr [1:6] "p.geomean" "stat.mean" "p.val" "q.val" ...
```

```
##Making output into a data frame
gage_form_mixed <- lapply(gage_mixed_list, data.frame)
gage_mixed_df <- do.call("cbind", gage_form_mixed)%>%
  rownames_to_column(var = "Pathway") %>%
  dplyr::select(Pathway, ends_with("q.val"))
##Excel file of unfiltered pathway analysis
write.xlsx(gage_mixed_df, file.path("GAGE analysis", paste(Sys.Date(),
    "mouse_donortreatment_GAGE analysis_unfiltered.xlsx")))

##Filtering GAGE analysis based on q.val
gage_mixed_df_filtered <-
  dplyr::filter(gage_mixed_df, `Mousegenes-HBV_vs_mock_d28.q.val` <= 0.07 |
                  `Mousegenes-HBV_vs_mock_d8.q.val` <= 0.07 |
                  `Mousegenes-coinf_vs_HBV_d28.q.val` <= 0.07 |
                  `Mousegenes-coinf_vs_HBV_d8.q.val` <= 0.07 |
                  `Mousegenes-coinf_vs_mock_d28.q.val` <= 0.07 |
                  `Mousegenes-coinf_vs_mock_d8.q.val` <= 0.07)

##Reorder the columns
colnames(gage_mixed_df_filtered)
```

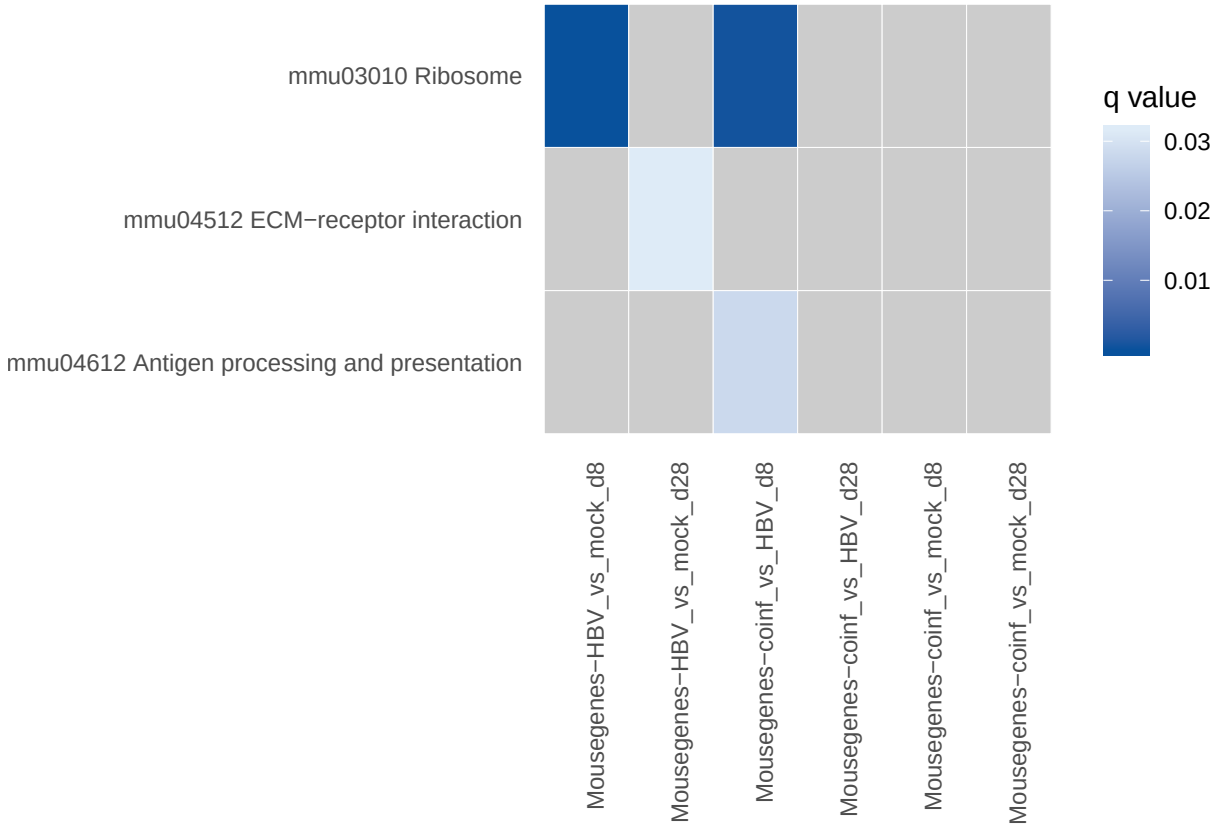
```
## [1] "Pathway"
## [2] "Mousegenes-HBV_vs_mock_d28.q.val"
## [3] "Mousegenes-HBV_vs_mock_d8.q.val"
```

```
## [4] "Mousegenes-coinf_vs_HBV_d28.q.val"
## [5] "Mousegenes-coinf_vs_HBV_d8.q.val"
## [6] "Mousegenes-coinf_vs_mock_d28.q.val"
## [7] "Mousegenes-coinf_vs_mock_d8.q.val"
```

```
gage_mixed_df_filtered <- gage_mixed_df_filtered[,c(1, 3, 2, 5, 4, 7, 6)]

##Make data frame into matrix for heatmap.
gage_mixed_matrix <- as.matrix(gage_mixed_df_filtered[, c(2:7)])
rownames(gage_mixed_matrix) <- gage_mixed_df_filtered[,1]
colnames(gage_mixed_matrix) <- gsub(pattern = ".q.val", replacement = "",
                                    colnames(gage_mixed_matrix))
colnames(gage_mixed_matrix) <-  gsub(pattern = "mousegenes-", replacement = "",
                                     colnames(gage_mixed_matrix))
colnames(gage_mixed_matrix)
```

```
## [1] "Mousegenes-HBV_vs_mock_d8"    "Mousegenes-HBV_vs_mock_d28"
## [3] "Mousegenes-coinf_vs_HBV_d8"   "Mousegenes-coinf_vs_HBV_d28"
## [5] "Mousegenes-coinf_vs_mock_d8"  "Mousegenes-coinf_vs_mock_d28"
```

```
gage_mixed.m <- melt(gage_mixed_matrix)

##Now change all q.values to NA if they are greater than the cutoff of 0.07
##so we can color in the geom_tile all NA values as grey.
gage_mixed.m$value[gage_mixed.m$value > 0.07] <- NA

##If desired, taking the -log10 of the q values.
##gage_mixed.m$value <- -log10(gage_mixed.m$value)

##Make the value column a factor for ggplot purposes only if you want it to be a
##discrete scale versus continuous as it is now (comes up as numeric).
##gage_mixed.m$value <- as.factor(gage_mixed.m$value)

##Function to plot the heatmap of the filtered GAGE matrix.
gage_plotting <- function(gage_matrix_input)
{
matrix_plot <- ggplot(gage_matrix_input, aes(x = Var2, y = Var1, fill = value)) +
geom_tile(aes(fill = value), colour = "white") +
scale_fill_continuous(low = "#08519c", high = "#deebf7", na.value = "grey80") +
theme(axis.text.x = element_text(angle = 90, hjust = 1),
axis.title.x = element_blank(),
axis.title.y = element_blank(),
axis.ticks = element_blank(),
panel.background = element_blank())
print(matrix_plot + labs(fill = "q value"))
ggsave(filename = file.path("GAGE analysis", paste(Sys.Date(), "mouse_donortreatment_GAGE_plot.png")),
       matrix_plot, dpi = 350, height = 7, width = 10)
}

##Apply function to matrix generated above.
gage_plotting(gage_mixed.m)
```




```
##Write Excel file of filtered GAGE output.
write.xlsx(gage_mixed_df_filtered, file.path("GAGE analysis", paste(Sys.Date(),
"mouse_donortreatment_GAGE analysis_filtered.xlsx")))
```

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           stringr_1.3.1
##  [4] reshape2_1.4.3       purrr_0.2.5          ggplot2_3.0.0
##  [7] gplots_3.0.1         gtools_3.8.1         tibble_1.4.2
## [10] dplyr_0.7.6          gageData_2.12.0      gage_2.24.0
## [13] pathview_1.14.0      org.Hs.eg.db_3.4.0   AnnotationDbi_1.36.2
## [16] IRanges_2.8.2        S4Vectors_0.12.2     Biobase_2.34.0
## [19] 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     labeling_0.3
## [25] bit_1.1-14         munsell_0.5.0      pkgconfig_2.0.1
## [28] htmltools_0.3.6    tidyselect_0.2.4   KEGGREST_1.14.1
## [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
```