

GAGE analysis of various dds outputs

Load required libraries

```
library(plyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize

## 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(stringr)
library(biomaRt)
library(genefilter)
library(DESeq2)
```

```
## Loading required package: S4Vectors
## 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:dplyr':
##
##   combine, intersect, setdiff, union

## 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

##
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:dplyr':
##
##      first, rename

## The following object is masked from 'package:plyr':
##
##      rename

## The following object is masked from 'package:base':
##
##      expand.grid

## Loading required package: IRanges

##
## Attaching package: 'IRanges'

## The following objects are masked from 'package:dplyr':
##
##      collapse, desc, slice

## The following object is masked from 'package:plyr':
##
##      desc

## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Loading required package: SummarizedExperiment
## 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: DelayedArray
## Loading required package: matrixStats

##
## Attaching package: 'matrixStats'

## The following objects are masked from 'package:Biobase':
##
##      anyMissing, rowMedians

## The following objects are masked from 'package:genefilter':
##
##      rowSds, rowVars

```

```

## The following object is masked from 'package:dplyr':
##
##     count
## The following object is masked from 'package:plyr':
##
##     count
## Loading required package: BiocParallel
##
## Attaching package: 'DelayedArray'
## The following objects are masked from 'package:matrixStats':
##
##     colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges
## The following objects are masked from 'package:base':
##
##     aperm, apply
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(RColorBrewer)
library(stringr)
library(viridis)

## Loading required package: viridisLite
library(devtools)
library(ggrepel)
library(reshape2)
library(data.table)

##
## Attaching package: 'data.table'
## The following objects are masked from 'package:reshape2':
##
##     dcast, melt
## The following object is masked from 'package:SummarizedExperiment':
##
##     shift

```

```

## The following object is masked from 'package:GenomicRanges':
##
##     shift
## The following object is masked from 'package:IRanges':
##
##     shift
## The following objects are masked from 'package:S4Vectors':
##
##     first, second
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
library(purrr)

##
## Attaching package: 'purrr'
## The following object is masked from 'package:data.table':
##
##     transpose
## The following object is masked from 'package:DelayedArray':
##
##     simplify
## The following object is masked from 'package:GenomicRanges':
##
##     reduce
## The following object is masked from 'package:IRanges':
##
##     reduce
## The following object is masked from 'package:plyr':
##
##     compact
library(forcats)
library(ape)
library(gtools)
source("http://bioconductor.org/biocLite.R")

## Bioconductor version 3.7 (BiocInstaller 1.30.0), ?biocLite for help
## A newer version of Bioconductor is available for this version of R,
##   ?BiocUpgrade for help
biocLite("org.Hs.eg.db")

## BioC_mirror: https://bioconductor.org
## Using Bioconductor 3.7 (BiocInstaller 1.30.0), R 3.5.2 (2018-12-20).
## Installing package(s) 'org.Hs.eg.db'
## installing the source package 'org.Hs.eg.db'
## Old packages: 'AMR', 'assertthat', 'backports', 'BiocManager', 'blob',
##   'boot', 'broom', 'callr', 'caTools', 'checkmate', 'circlize', 'cli',

```

```
## 'clipr', 'cluster', 'colorspace', 'cowplot', 'curl', 'data.table',
## 'dbplyr', 'dendextend', 'devtools', 'digest', 'dplyr', 'ellipsis',
## 'evaluate', 'foreign', 'formatR', 'fs', 'ggplot2', 'ggrepel', 'git2r',
## 'GlobalOptions', 'glue', 'gplots', 'gtable', 'haven', 'highr', 'Hmisc',
## 'hms', 'htmlTable', 'htmltools', 'htmlwidgets', 'httr', 'KernSmooth',
## 'knitr', 'lambda.r', 'lazyeval', 'markdown', 'MASS', 'Matrix',
## 'matrixStats', 'mgcv', 'microbenchmark', 'mime', 'modelr', 'nlme',
## 'openssl', 'pillar', 'pkgbuild', 'pkgconfig', 'processx', 'progress',
## 'purrr', 'quadprog', 'R6', 'rcmdcheck', 'Rcpp', 'RcppArmadillo',
## 'RCurl', 'readxl', 'remotes', 'reprex', 'rlang', 'rmarkdown', 'rpart',
## 'RSQLite', 'rstudioapi', 'rvest', 'scales', 'selectr', 'stringi',
## 'stringr', 'survival', 'sys', 'tibble', 'tidyr', 'tidyverse', 'tinytex',
## 'usethis', 'whisker', 'xfun', 'XML', 'xml2', 'xtable'
```

```
library(gage)
library(gageData)
data(kegg.gs)
data(go.sets.hs)
data(go.subs.hs)
data(kegg.gs)
go.bp = go.sets.hs[go.subs.hs$BP] ##Only examining GO biological processes terms
```

Purpose

To compare the gene set (specifically KEGG pathways) enrichment within the various DGE profiles generated after mapping reads from the dermal fibroblasts +/- poly(I:C) to the human genome versus their species genome as it currently exists on ENSEMBL.

```
HumanMapped_DGE <- "Expanded_Design_Factor_SpeciesHomologs_Outputs_HumanMapped"
output_dir <- "Expanded_Design_Factor_SpeciesHomologs_Outputs_HumanMapped"
sampleFiles_HumanMapped <- list.files(basename(Sys.glob(file.path(HumanMapped_DGE))),
                                     pattern = "treated dds1 HumanMapped_DGE_results.txt|*related.*.txt")

sampleNames_HumanMapped <- sub('_treated_v_mock|.treatmenttreated.*', '', sampleFiles_HumanMapped) %>%
  sub('\\d+-\\d+-\\d+\\s', '', .) %>%
  sub("HumanMapped.*", "", .) %>%
  sub('species', '', .)
length(sampleFiles_HumanMapped)

## [1] 16

SpeciesMapped_DGE <- "Expanded_Design_Factor_SpeciesHomologs_Outputs_SpeciesMapped"
output_dir <- "Expanded_Design_Factor_SpeciesHomologs_Outputs_SpeciesMapped"

sampleFiles_SpeciesMapped <- list.files(basename(Sys.glob(file.path(SpeciesMapped_DGE))),
                                     pattern = "*treated dds1 DGE_results.txt|*related.*.txt")

sampleNames_SpeciesMapped <- sub('_treated_v_mock|.treatmenttreated.*', '', sampleFiles_SpeciesMapped) %>%
  sub('\\d+-\\d+-\\d+\\s', '', .) %>%
  sub("SpeciesMapped.*", "", .) %>%
  sub('species', '', .)
length(sampleFiles_SpeciesMapped)

## [1] 16
```

```

exptcounts <- function(files) {
  d <- read.delim(files)
  d
}

human_DGEs_read <- llply(file.path(HumanMapped_DGE, sampleFiles_HumanMapped), exptcounts)
names(human_DGEs_read) <- sampleNames_HumanMapped

species_DGEs_read <- llply(file.path(SpeciesMapped_DGE, sampleFiles_SpeciesMapped), exptcounts)
names(species_DGEs_read) <- sampleNames_SpeciesMapped

GAGE analysis and heatmap generation

output_dir <- "GAGE_output"

##GAGE analysis
gage_mixed <- function(input) {
  dd <- dplyr::select(input, 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.gs, same.dir = FALSE)
  g <- ef$greater
  gg <- g[mixedorder(rownames(g), decreasing = TRUE),] %>%
    as.data.frame() %>%
    rownames_to_column(., var = "pathway")
  gg
  ##if applying to a list, need to have "gg" in this line. if applying to a single data frame, then scrip
  ##good as is.
}

GAGE_heatmap <- function(DGEs, name) {
  DGEs_GAGE <- llply(DGEs, gage_mixed) %>%
    llply(., function(x) dplyr::select(x, pathway, q.val)) %>%
    reduce(full_join, by = "pathway") %>%
    mutate(., sum = rowSums(., 2:17)) %>%
    dplyr::arrange(., sum)

  colnames(DGEs_GAGE) <- c("pathway", names(DGEs), "sum")
  write.table(DGEs_GAGE, file = file.path(output_dir, paste0(Sys.Date(),
    name, ".txt")), row.names=FALSE, sep="\t", quote=FALSE)

  DGEs_GAGE_sub <- dplyr::filter_at(DGEs_GAGE, vars(-pathway), any_vars(. <= 0.09))
  write.table(DGEs_GAGE_sub, file = file.path(output_dir, paste0(Sys.Date(),
    name, "Significant.txt")), row.names=FALSE, sep="\t", quote=FALSE)

  ##Now make heatmap
  gage.m <- melt(DGEs_GAGE_sub, id = 'pathway')
  gage.m$value <- -log10(gage.m$value)

```

```

gage.m$pathway <- factor(gage.m$pathway,
  levels = DGEs_GAGE_sub$pathway[order(DGEs_GAGE_sub$sum, decreasing = TRUE)])
gage.m <- gage.m[gage.m$variable != "sum",]
##the more significant, the higher the value.
##Color switch is is at q.val ~ 0.09
pdf(file = file.path(output_dir, paste0(Sys.Date(), name, "Heatmap.pdf")), height = 4, width = 7)
plot <- ggplot(gage.m, aes(gage.m$variable, gage.m$pathway))
p <- plot + geom_tile(color = 'black', aes(fill = gage.m$value)) +
  theme_bw(base_size = 6) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
    axis.text.y = element_text(size = rel(1.5))) +
  scale_fill_gradient2(low = "#e66101", high = "#542788", mid = "white", midpoint = 1.046, limits = c(5
  ylab("") +
  xlab("\n")
print(p)
dev.off()
}

```

```

GAGE_heatmap(human_DGEs_read, "HumanMapped_GAGE_One2OneOrthos_SpeciesBasis")

```

```

## pdf
## 2

```

```

GAGE_heatmap(species_DGEs_read, "SpeciesMapped_GAGE_One2OneOrthos_SpeciesBasis")

```

```

## pdf
## 2

```

Session Info

```

sessionInfo()

```

```

## R version 3.5.2 (2018-12-20)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## 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 gageData_2.20.0
## [3] gage_2.32.1 BiocInstaller_1.30.0
## [5] gtools_3.8.1 ape_5.3
## [7] forcats_0.4.0 purrr_0.2.5
## [9] data.table_1.12.0 reshape2_1.4.3
## [11] ggrepel_0.8.0 usethis_1.4.0
## [13] devtools_2.0.1 viridis_0.5.1
## [15] viridisLite_0.3.0 RColorBrewer_1.1-2

```

```

## [17] ggplot2_3.1.0          gplots_3.0.1
## [19] DESeq2_1.22.2          SummarizedExperiment_1.12.0
## [21] DelayedArray_0.8.0     BiocParallel_1.16.5
## [23] matrixStats_0.54.0     Biobase_2.42.0
## [25] GenomicRanges_1.34.0   GenomeInfoDb_1.18.1
## [27] IRanges_2.16.0         S4Vectors_0.20.1
## [29] BiocGenerics_0.28.0    genefilter_1.64.0
## [31] biomaRt_2.38.0         stringr_1.3.1
## [33] tibble_2.0.1           dplyr_0.7.8
## [35] plyr_1.8.4
##
## loaded via a namespace (and not attached):
## [1] colorspace_1.4-0        rprojroot_1.3-2        htmlTable_1.13.1
## [4] XVector_0.22.0          base64enc_0.1-3        fs_1.2.6
## [7] rstudioapi_0.9.0        remotes_2.0.2          bit64_0.9-7
## [10] AnnotationDbi_1.44.0     splines_3.5.2          geneplotter_1.60.0
## [13] knitr_1.21              pkgload_1.0.2          Formula_1.2-3
## [16] annotate_1.60.0          cluster_2.0.7-1        png_0.1-7
## [19] graph_1.60.0            compiler_3.5.2         httr_1.4.0
## [22] backports_1.1.3         assertthat_0.2.0       Matrix_1.2-15
## [25] lazyeval_0.2.1          cli_1.0.1              acepack_1.4.1
## [28] htmltools_0.3.6         prettyunits_1.0.2      tools_3.5.2
## [31] gtable_0.2.0            glue_1.3.0             GenomeInfoDbData_1.2.0
## [34] Rcpp_1.0.0              Biostrings_2.50.2      gdata_2.18.0
## [37] nlme_3.1-137            xfun_0.4               ps_1.3.0
## [40] XML_3.98-1.16           zlibbioc_1.28.0        scales_1.0.0
## [43] hms_0.4.2              yaml_2.2.0             memoise_1.1.0
## [46] gridExtra_2.3           rpart_4.1-13           latticeExtra_0.6-28
## [49] stringi_1.2.4           RSQlite_2.1.1          desc_1.2.0
## [52] checkmate_1.9.1         caTools_1.17.1.1      pkgbuild_1.0.2
## [55] rlang_0.3.1             pkgconfig_2.0.2        bitops_1.0-6
## [58] evaluate_0.12           lattice_0.20-38        bindr_0.1.1
## [61] labeling_0.3            htmlwidgets_1.3        bit_1.1-14
## [64] tidyselect_0.2.5        processx_3.2.1         magrittr_1.5
## [67] R6_2.3.0               Hmisc_4.1-1           DBI_1.0.0
## [70] pillar_1.3.1            foreign_0.8-71         withr_2.1.2
## [73] KEGGREST_1.22.0         survival_2.43-3        RCurl_1.95-4.11
## [76] nnet_7.3-12            crayon_1.3.4           KernSmooth_2.23-15
## [79] rmarkdown_1.11         progress_1.2.0         locfit_1.5-9.1
## [82] grid_3.5.2             blob_1.1.1            callr_3.1.1
## [85] digest_0.6.18          xtable_1.8-3          munsell_0.5.0
## [88] sessioninfo_1.1.1

```