Pseudoviolin plots

Purpose

Generating pseudoviolin plots of the differentially expressed genes across all the species used in this study. One plot will be the DGE profiles after mapping the sequencing reads to human genome and the other will be the DGE profiles after mapping the sequencing reads to the species-specific genomes. In both cases, the genes under examination are only those that have a one-to-one human ortholog determined on a species-by-species basis. Since this was done, the human DGEs were also limited to the same set of genes – hence, each NHP species has a "related" human DGE profile even though the reads all came from the same human samples.

```
library(dplyr)
library(stringr)
library(ggplot2)
library(reshape2)
library(plyr)
```

Upload differential gene expression files

```
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
```

```
Mouse <- "Expanded_Design_Factor_Mice"
Mouse_file <-  basename(Sys.glob(file.path(Mouse, "*DGE_results.txt")))
```

Organize the data and select only genes that have a padj <= 0.05 to use for generating the pseudoviolin plots .

```
data_pseudoviolins <- function(files) {
  d <- read.delim(files)
  dd <-dplyr::select(d, log2FoldChange, padj) %>%
    na.omit() %>%
    dplyr::filter(padj <= 0.05) %>%
    dplyr::select(log2FoldChange)
}

Human_DGEs <- llply(file.path(HumanMapped_DGE, sampleFiles_HumanMapped), data_pseudoviolins)
names(Human_DGEs)  <- sampleNames_HumanMapped

Species_DGEs <- llply(file.path(SpeciesMapped_DGE, sampleFiles_SpeciesMapped), data_pseudoviolins)
names(Species_DGEs) <- sampleNames_SpeciesMapped

Mouse_DGEs <- data_pseudoviolins(file.path(Mouse, Mouse_file))

Human_DGEs.m <- melt(Human_DGEs)
```

```
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
```

```
Species_DGEs.m <- melt(Species_DGEs)
```

```
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
## No id variables; using all as measure variables
```

```
Mouse.m <- melt(Mouse_DGEs)
```

```
## No id variables; using all as measure variables
```

```
Human_related.m <- dplyr::filter_at(Human_DGEs.m, vars(L1), any_vars(str_detect(., "related"))) %>%
  add_count(L1)
Human_related.m$L1 <- str_replace_all(Human_related.m$L1, "human_related_", "") %>%
  str_replace(., " ", "") ##the levels of L1 have a space after the name that is unnecessary

Human_NHP.m <- dplyr::filter_at(Human_DGEs.m, vars(L1), all_vars(!str_detect(., "related"))) %>%
  add_count(L1)

##Releveling the species so that the order of the plots is how we want.
Human_NHP.m$L1 <- factor(Human_NHP.m$L1,
                                levels = c("chimpanzee", "bonobo", "gorilla",
                                           "orangutan", "olive_baboon", "rhesus_macaque",
                                           "pigtailed_macaque", "squirrel_monkey"))
Human_related.m$L1 <- factor(Human_related.m$L1,
                                levels = c("chimpanzee", "bonobo", "gorilla",
                                           "orangutan", "olive_baboon", "rhesus_macaque",
                                           "pigtailed_macaque", "squirrel_monkey"))
####################
Species_related.m <- dplyr::filter_at(Species_DGEs.m, vars(L1), any_vars(str_detect(., "related"))) %>%
  add_count(L1)
Species_related.m$L1 <- str_replace_all(Species_related.m$L1, "human_related_", "") %>%
  str_replace(., " ", "") ##the levels of L1 have a space after the name that is unnecessary

Species_NHP.m <- dplyr::filter_at(Species_DGEs.m, vars(L1), all_vars(!str_detect(., "related"))) %>%
  add_count(L1)

Species_NHP.m$L1 <- factor(Species_NHP.m$L1,
                                levels = c("chimpanzee", "bonobo", "gorilla",
                                           "orangutan", "olive_baboon", "rhesus_macaque",
                                           "pigtailed_macaque", "squirrel_monkey"))

Species_related.m$L1 <- factor(Species_related.m$L1,
                                levels = c("chimpanzee", "bonobo", "gorilla",
                                           "orangutan", "olive_baboon", "rhesus_macaque",
                                           "pigtailed_macaque", "squirrel_monkey"))
```

Plotting the human and NHP plots

```
output_dir <- "Pseudoviolin plots output"

##Species genome-mapped DGEs
##NHPs first
species_NHP_plot <- ggplot(Species_NHP.m, aes(x = L1, y = value, fill= L1, color=L1)) + geom_point()
species_NHP_total_plot <- species_NHP_plot + geom_dotplot(binwidth=0.1, binaxis = "y",
                                                          stackdir="center", dotsize = 0.20) +
  xlab("") + ylab ("log2(fold change)") +
  scale_y_continuous(breaks = seq(-10, 20, 5), limits = c(-10, 20)) +
  theme(panel.background = element_rect(fill = "white"),
        panel.grid.major = element_line(colour = "black")) +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank(),
        axis.text.y = element_text(size = rel(2)), axis.title.y = element_text(size = rel(2)),
        legend.position = "none") +
  scale_color_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple")) +
  scale_fill_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple"))
ggsave(filename = file.path(output_dir, paste(Sys.Date(), "SpeciesMappedDGEs_NHPs_pseudoviolin.pdf")),
       plot = species_NHP_total_plot, width = 15, height = 6, device = "pdf")

##Human DGEs paired with each species based off the one-to-one orthologs that each species has with huma
species_human_plot <- ggplot(Species_related.m, aes(x = L1, y = value, fill= L1, color=L1)) +
  geom_point()
species_human_total_plot <- species_human_plot +
  geom_dotplot(binwidth=0.1, binaxis = "y", stackdir="center", dotsize = 0.20) +
  xlab("") + ylab ("log2(fold change)") +
  scale_y_continuous(breaks = seq(-10, 20, 5), limits = c(-10, 20)) +
  theme(panel.background = element_rect(fill = "white"),
        panel.grid.major = element_line(colour = "black")) +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank(),
        axis.text.y = element_text(size = rel(2)), axis.title.y = element_text(size = rel(2)),
        legend.position = "none") +
  scale_color_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple")) +
  scale_fill_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple"))
ggsave(filename = file.path(output_dir,
                            paste(Sys.Date(), "SpeciesMappedDGEs_HumanRelated_pseudoviolin.pdf")),
       plot = species_human_total_plot, width = 15, height = 6, device = "pdf")

##Human genome-mapped DGEs
##NHPs first
human_NHP_plot <- ggplot(Human_NHP.m, aes(x = L1, y = value, fill= L1, color=L1)) + geom_point()
human_NHP_total_plot <- human_NHP_plot + geom_dotplot(binwidth=0.1, binaxis = "y",
                                                      stackdir="center", dotsize = 0.20) +
  xlab("") + ylab ("log2(fold change)") +
  scale_y_continuous(breaks = seq(-10, 20, 5), limits = c(-10, 20)) +
  theme(panel.background = element_rect(fill = "white"),
        panel.grid.major = element_line(colour = "black")) +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank(),
        axis.text.y = element_text(size = rel(2)), axis.title.y = element_text(size = rel(2)),
        legend.position = "none") +
  scale_color_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple")) +
  scale_fill_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple"))
ggsave(filename = file.path(output_dir, paste(Sys.Date(), "HumanMappedDGEs_NHPs_pseudoviolin.pdf")),
       plot = human_NHP_total_plot, width = 15, height = 6, device = "pdf")

##Human DGEs paired with each species based off the one-to-one orthologs that each species has with huma
human_human_plot <- ggplot(Human_related.m, aes(x = L1, y = value, fill= L1, color=L1)) + geom_point()
human_human_total_plot <- human_human_plot + geom_dotplot(binwidth=0.1, binaxis = "y",
                                                          stackdir="center", dotsize = 0.20) +
  xlab("") + ylab ("log2(fold change)") +
  scale_y_continuous(breaks = seq(-10, 20, 5), limits = c(-10, 20)) +
  theme(panel.background = element_rect(fill = "white"),
        panel.grid.major = element_line(colour = "black")) +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank(),
        axis.text.y = element_text(size = rel(2)), axis.title.y = element_text(size = rel(2)),
        legend.position = "none") +
  scale_color_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple")) +
  scale_fill_manual(values= c("#00659c", "#0088ce", "#7dc3e8", "#bee1f4", "darkgreen", "forestgreen",
                               "darkolivegreen3", "purple"))
ggsave(filename = file.path(output_dir,
                            paste(Sys.Date(), "HumanMappedDGEs_HumanRelated_pseudoviolin.pdf")),
       plot = human_human_total_plot, width = 15, height = 6, device = "pdf")

##Mouse DGEs -- since we didn't do any ortholog comparison with human genome, there is just this one
##plot for the Mouse samples.
mouse_plot <- ggplot(Mouse.m, aes(x = variable, y = value, fill= variable, color=variable)) +
  geom_point()
mouse_total_plot <- mouse_plot + geom_dotplot(binwidth=0.1, binaxis = "y",
                                                          stackdir="center", dotsize = 0.20) +
  xlab("") + ylab ("log2(fold change)") +
  scale_y_continuous(breaks = seq(-10, 20, 5), limits = c(-10, 20)) +
  theme(panel.background = element_rect(fill = "white"),
        panel.grid.major = element_line(colour = "black")) +
  theme(axis.ticks.x=element_blank(), axis.text.x = element_blank(),
        axis.text.y = element_text(size = rel(2)), axis.title.y = element_text(size = rel(2)),
        legend.position = "none") +
  scale_color_manual(values= c("magenta")) +
  scale_fill_manual(values= c("magenta"))
ggsave(filename = file.path(output_dir,
                            paste(Sys.Date(), "MouseDGEs_pseudoviolin.pdf")),
       plot = mouse_total_plot, width = 3, height = 6, device = "pdf")
```

```
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] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] bindrcpp_0.2.2 plyr_1.8.4     reshape2_1.4.3 ggplot2_3.1.0
## [5] stringr_1.3.1  dplyr_0.7.8    knitr_1.21
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.0       bindr_0.1.1      magrittr_1.5     munsell_0.5.0
##  [5] tidyselect_0.2.5 colorspace_1.4-0 R6_2.3.0         rlang_0.3.1
##  [9] tools_3.5.2      grid_3.5.2       gtable_0.2.0     xfun_0.4
## [13] withr_2.1.2      htmltools_0.3.6  lazyeval_0.2.1   yaml_2.2.0
## [17] assertthat_0.2.0 digest_0.6.18    tibble_2.0.1     crayon_1.3.4    
## [21] purrr_0.2.5      glue_1.3.0       evaluate_0.12    rmarkdown_1.11  
## [25] stringi_1.2.4    compiler_3.5.2   pillar_1.3.1     scales_1.0.0    
## [29] pkgconfig_2.0.2
```