Plot comparing differential expression of LF82wt vs LF82 mutant and MG1655 mutant vs MG1655 wt

Hernan Lorenzi

02/06/2024

```r
pacman::p_load(AnnotationDbi,pheatmap,EnhancedVolcano,ggpubr,DESeq2,stringr,biomaRt,tidyverse,pcaExplor
```

```r
#
# --- function for PCA plots ---
#
plot_PCA = function(object, color_by="condition",
                    shape_by = 19, ntop=500, size = 3,
                    returnData=FALSE, pcs = c(1,2))
{
  # Check variables are present in object
  intgroup = c(color_by)
  if (shape_by != 19){intgroup <- c(intgroup, shape_by)}
  if (!all(intgroup %in% names(colData(object)))) {
    stop("the argument 'intgroup' should specify columns of colData(dds)")
  }

  # calculate the variance for each gene
  rv <- rowVars(assay(object))

  # select the ntop genes by variance
  select <- order(rv, decreasing=TRUE)[seq_len(min(ntop, length(rv)))]

  # perform a PCA on the data in assay(x) for the selected genes
  pca <- prcomp(t(assay(object)[select,]))

  # the contribution to the total variance for each component
  percentVar <- pca$sdev^2 / sum( pca$sdev^2 )


  intgroup.df <- as.data.frame(colData(object)[, intgroup, drop=FALSE])

  # add the intgroup factors together to create a new grouping factor
  group <- if (length(intgroup) > 1) {
    factor(apply( intgroup.df, 1, paste, collapse=":"))
  } else {
    colData(object)[[intgroup]]
  }

  # assembly the data for the plot
```

```r
  d <- data.frame(PC1=pca$x[,pcs[1]], PC2=pca$x[,pcs[2]], group=group, intgroup.df, name=colnames(object
  colnames(d)[1] <- paste0("PC",pcs[1])
  colnames(d)[2] <- paste0("PC",pcs[2])

  if (returnData) {
    attr(d, "percentVar") <- percentVar[1:2]
    return(d)
  }

  ggplot(data=d, aes_string(x=colnames(d)[1], y=colnames(d)[2], color=color_by, shape=shape_by)) +
    geom_point(size=size) +
    scale_color_lancet() +
    xlab(paste0("PC",pcs[1],": ",round(percentVar[pcs[1]] * 100, digits = 2),"% variance")) + # fixed
    ylab(paste0("PC",pcs[2],": ",round(percentVar[pcs[2]] * 100, digits = 2),"% variance")) + # fixed
    coord_fixed(ratio = (max(d[,1])-min(d[,1]))/(max(d[,2])-min(d[,2])))
}
```

```r
log2fc_lf82_mg <- read.table(file = "./data/Log2FC_LF82_MG.txt", header = T, sep = "\t", row.names = 1)

genes_of_interest <- c("metE", "yobH", "lptM", "yigA", "xerC", "yigB", "ampC", "yciY", "shiA", "acrZ", "

# Generate first plot with all genes
my_title <- "Log2(Fold Change) of differential gene expression\nLF82-WT/LF82-Mut and MG1655-Mut/MG1655-W

df <- log2fc_lf82_mg

# Invert Log2FC for LF82 to reflect WT/Mut rather than the original data Mut/WT
df$Log2FC_LF82 <- df$Log2FC_LF82 * -1

# remove rows where Log2FC LF82/MG == NA
df <- df[!c(is.na(df$Log2FC_MG) | is.na(df$Log2FC_LF82)), ]

df$padj_LF82[is.na(df$padj_LF82)] <- 1
df$padj_MG[is.na(df$padj_MG)] <- 1
df$min_pval <- apply(dplyr::select(df,c("padj_LF82","padj_MG")), 1, function(x){ min(x)})
df <- df[order(df$min_pval, decreasing = T),]
df$significance <- as.factor(apply(dplyr::select(df,c("padj_LF82","padj_MG")), 1, function(x){ ifelse(

myColor <- c("N.S."="gray","Sig both"="orange3","Sig LF82"="green4","Sig MG1655"="yellow3")

p4 <- ggplot(df, aes(x=Log2FC_LF82, y=Log2FC_MG, label = gene_name_MG)) +
  labs(title = my_title) +
  xlab(bquote('LF82 '* ~Log[2]*'(Fold Change)')) +
  ylab(bquote('MG1655 '* ~Log[2]*'(Fold Change)')) +
  geom_point(aes(colour=significance), size = 1, alpha = 1 ) +
  ylim(-3.5,3.5) + #ylim(min(df$Log2FC_MG),max(df$Log2FC_MG)) +
  xlim(-3.5,3.5) + #xlim(min(df$Log2FC_LF82),max(df$Log2FC_LF82)) +
  geom_abline(slope = 1, intercept = 0, col = "black", size=0.5, linetype="dashed") +
  scale_x_continuous(breaks =c(-3,-2,-1,0,1,2,3)) +
  scale_y_continuous(breaks =c(-3,-2,-1,0,1,2,3)) +
  theme_minimal() +
  scale_colour_manual(values = myColor) +
  theme(
```

```
    legend.direction = "vertical",
    legend.position = c(.25, .95),
    legend.justification = c("right", "top"),
    legend.box.just = "right",
    legend.margin = margin(3, 3, 3, 3),
    legend.text = element_text(size = 8),
    legend.title = element_text(face = "bold" ,size = 8, vjust = 0.9))
```

Load libraries

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## Scale for x is already present.
## Adding another scale for x, which will replace the existing scale.
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```

```r
p4 <-  p4 + guides(color = guide_legend(override.aes = list(size = 5)))

# Change legend title
p4$labels$colour = paste("Significance" )

#genes_of_interest
my_list <- genes_of_interest # genes_of_interest.names$gene_name

df.subset <- subset(df, gene_name_LF82 %in% my_list)


p4 <- p4 + geom_point(data = df.subset,
                      aes(x = Log2FC_LF82, y = Log2FC_MG, fill=significance),
                      size = 3,
                      pch=21,
                      colour="black",
                      show.legend = FALSE) +
           scale_fill_manual(values = myColor)

#ggsave2(filename = "Log2FC_dotplot_from_orig_DE_no_labels.pdf", plot = p4, path = "./Plots", width = 1

# Adding labels to genes of interest
p4 <- p4 + geom_text_repel(data = df.subset,
                           aes(label = gene_name_MG),
                           colour = "black",
                           box.padding   = 0.4,
                           point.padding = 0.3,
                           segment.color = 'black',
                           na.rm = TRUE,
                           size = 5,
                           min.segment.length = 0.02,
                           direction = "both",
                           segment.curvature = -0.1,
                           segment.ncp = 3,
```

```
                          segment.angle = 20,
                          max.iter = 1e6,
                          max.overlaps = 20,
                          max.time = 10)

#ggsave2(filename = "Log2FC_dotplot_from_orig_DE.pdf", plot = p4, path = "./Plots", width = 10, height =
p4
```
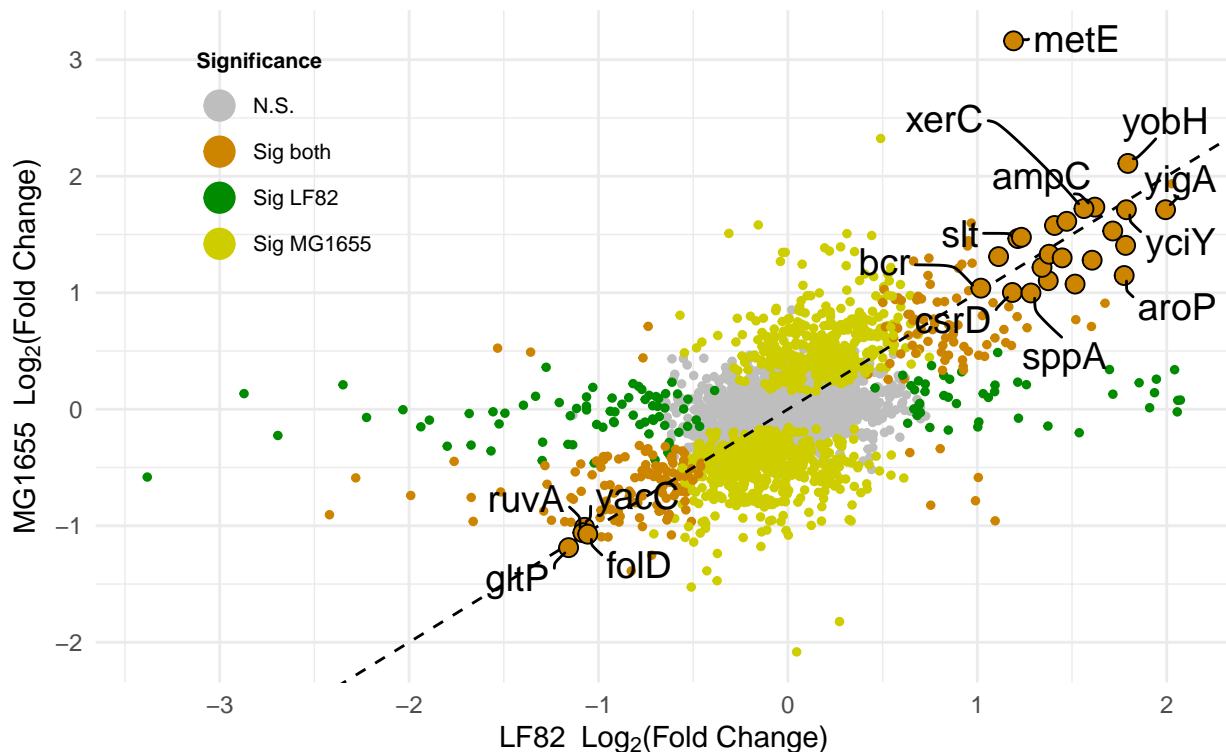
Warning: ggrepel: 12 unlabeled data points (too many overlaps). Consider
increasing max.overlaps

Log2(Fold Change) of differential gene expression
LF82−WT/LF82−Mut and MG1655−Mut/MG1655−WT

```
print(sessionInfo())
```

R version 4.3.1 (2023-06-16)
Platform: x86_64-apple-darwin20 (64-bit)
Running under: macOS Ventura 13.6.3
##
Matrix products: default
BLAS: /Library/Frameworks/R.framework/Versions/4.3-x86_64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.3-x86_64/Resources/lib/libRlapack.dylib; LAPACK v
##
locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
time zone: America/New_York
tzcode source: internal
##
attached base packages:

4

```
## [1] grid      stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] ggraph_2.1.0              broom_1.0.5
##  [3] ggupset_0.3.0            enrichplot_1.22.0
##  [5] cowplot_1.1.1            msigdbr_7.5.1
##  [7] RColorBrewer_1.1-3       viridis_0.6.4
##  [9] viridisLite_0.4.2        ggsci_3.0.0
## [11] GOSemSim_2.28.0          clusterProfiler_4.10.0
## [13] VennDiagram_1.7.3        futile.logger_1.4.3
## [15] pcaExplorer_2.28.0       lubridate_1.9.3
## [17] forcats_1.0.0            dplyr_1.1.4
## [19] purrr_1.0.2              readr_2.1.4
## [21] tidyr_1.3.0              tibble_3.2.1
## [23] tidyverse_2.0.0          biomaRt_2.58.0
## [25] stringr_1.5.1            DESeq2_1.42.0
## [27] SummarizedExperiment_1.32.0 MatrixGenerics_1.14.0
## [29] matrixStats_1.1.0        GenomicRanges_1.54.1
## [31] GenomeInfoDb_1.38.1      ggpubr_0.6.0
## [33] EnhancedVolcano_1.20.0   ggrepel_0.9.4
## [35] ggplot2_3.4.4            pheatmap_1.0.12
## [37] AnnotationDbi_1.64.1     IRanges_2.36.0
## [39] S4Vectors_0.40.2         Biobase_2.62.0
## [41] BiocGenerics_0.48.1
##
## loaded via a namespace (and not attached):
##   [1] fs_1.6.3               bitops_1.0-7          HDO.db_0.99.1
##   [4] httr_1.4.7            webshot_0.5.5         doParallel_1.0.17
##   [7] Rgraphviz_2.46.0     tools_4.3.1           backports_1.4.1
##  [10] utf8_1.2.4           R6_2.5.1              DT_0.30
##  [13] lazyeval_0.2.2       withr_2.5.2           prettyunits_1.2.0
##  [16] gridExtra_2.3        cli_3.6.1             pacman_0.5.1
##  [19] formatR_1.14         TSP_1.2-4             scatterpie_0.2.1
##  [22] topGO_2.54.0         genefilter_1.84.0     yulab.utils_0.1.0
##  [25] gson_0.1.0           DOSE_3.28.1           AnnotationForge_1.44.0
##  [28] limma_3.58.1         rstudioapi_0.15.0     RSQLite_2.3.3
##  [31] gridGraphics_0.5-1   generics_0.1.3        GOstats_2.68.0
##  [34] crosstalk_1.2.1      car_3.1-2             dendextend_1.17.1
##  [37] GO.db_3.18.0         Matrix_1.6-4          fansi_1.0.5
##  [40] abind_1.4-5          lifecycle_1.0.4       yaml_2.3.7
##  [43] carData_3.0-5        qvalue_2.34.0         SparseArray_1.2.2
##  [46] BiocFileCache_2.10.1 blob_1.2.4            promises_1.2.1
##  [49] crayon_1.5.2         shinydashboard_0.7.2  lattice_0.21-8
##  [52] annotate_1.80.0      KEGGREST_1.42.0       pillar_1.9.0
##  [55] knitr_1.45           fgsea_1.28.0          codetools_0.2-19
##  [58] fastmatch_1.1-4      glue_1.6.2            ggfun_0.1.3
##  [61] data.table_1.14.8    treeio_1.27.0.001     vctrs_0.6.5
##  [64] png_0.1-8            gtable_0.3.4          assertthat_0.2.1
##  [67] cachem_1.0.8         xfun_0.41             S4Arrays_1.2.0
##  [70] mime_0.12            tidygraph_1.2.3       survival_3.5-5
##  [73] seriation_1.5.3      iterators_1.0.14      statmod_1.5.0
##  [76] ellipsis_0.3.2       nlme_3.1-162          Category_2.68.0
##  [79] ggtree_3.10.0        bit64_4.0.5           threejs_0.3.3
```

```
##  [82] progress_1.2.3          filelock_1.0.2          colorspace_2.1-0
##  [85] DBI_1.1.3               tidyselect_1.2.0        bit_4.0.5
##  [88] compiler_4.3.1          curl_5.1.0              graph_1.80.0
##  [91] SparseM_1.81            xml2_1.3.6              DelayedArray_0.28.0
##  [94] plotly_4.10.3           shadowtext_0.1.2        scales_1.3.0
##  [97] RBGL_1.78.0             NMF_0.26                rappdirs_0.3.3
## [100] digest_0.6.33           shinyBS_0.61.1          rmarkdown_2.25
## [103] ca_0.71.1               XVector_0.42.0          htmltools_0.5.7
## [106] pkgconfig_2.0.3         base64enc_0.1-3         highr_0.10
## [109] dbplyr_2.4.0            fastmap_1.1.1           rlang_1.1.2
## [112] htmlwidgets_1.6.4       shiny_1.8.0             farver_2.1.1
## [115] jsonlite_1.8.8          BiocParallel_1.36.0     RCurl_1.98-1.13
## [118] magrittr_2.0.3          ggplotify_0.1.2         GenomeInfoDbData_1.2.11
## [121] patchwork_1.1.3         munsell_0.5.0           Rcpp_1.0.11
## [124] babelgene_22.9          ape_5.7-1               stringi_1.8.2
## [127] zlibbioc_1.48.0         MASS_7.3-60             plyr_1.8.9
## [130] parallel_4.3.1          Biostrings_2.70.1       graphlayouts_1.0.2
## [133] splines_4.3.1           hms_1.1.3               locfit_1.5-9.8
## [136] igraph_1.5.1            ggsignif_0.6.4          rngtools_1.5.2
## [139] reshape2_1.4.4          futile.options_1.0.1    XML_3.99-0.16
## [142] evaluate_0.23           lambda.r_1.2.4          BiocManager_1.30.22
## [145] tzdb_0.4.0              foreach_1.5.2           tweenr_2.0.2
## [148] httpuv_1.6.13           polyclip_1.10-6         heatmaply_1.5.0
## [151] gridBase_0.4-7          ggforce_0.4.1           xtable_1.8-4
## [154] tidytree_0.4.5          rstatix_0.7.2           later_1.3.2
## [157] aplot_0.2.2             memoise_2.0.1           registry_0.5-1
## [160] cluster_2.1.4           timechange_0.2.0        shinyAce_0.4.2
## [163] GSEABase_1.64.0
```