

RNAseq for M. avium strain LU439 SmT1 versus SmO2 bacteria - Volcano plot with labelling of the top 10 DEGs

Julia Lienard

2024-05-01

Description

RNAseq analysis was performed on four replicates of SmO2 or SmT1 bacteria of Mycobacterium avium subspecies hominissuis strain LU439. Gene expression analysis was performed using DESeq2 by Giulia Ribeiro and the output table with the detected genes differentially expressed was used to create a Volcano plot

Required packages

```
#library(tidyverse) to install: install.packages("tidyverse")
#library(ggplot2) to install: install.packages("ggplot2")
#library(ggrepel) # to annotate the plot. to install:
↪ install.packages("ggrepel")
```

Loading the data

```
library(readr)

# Input data are the gene expression analysis done using DESeq2 by Giulia
↪ Ribeiro : LU439L2_DESeq2_dataAllAnnot.txt

DESeq2_dataAllAnnot <-
↪ read_delim("~/Desktop/Master/BINP39/RNAseq_visualization/LU439/1_data/2_RNAseq/1_LU439L2",
↪ delim = "\t", escape_double = FALSE,
```

```
trim_ws = TRUE, show_col_types = FALSE)

head(DESeq2_dataAllAnnot)
```

```
# A tibble: 6 x 15
  pgap_ID      sampleA sampleB baseMeanA baseMeanB baseMean log2FoldChange lfcSE
  <chr>        <chr>   <chr>      <dbl>      <dbl>      <dbl>      <dbl> <dbl>
1 pgaptmp_000~ Sm0439~ SmT439~    2037.    2616.    2326.      -0.361  0.208
2 pgaptmp_000~ Sm0439~ SmT439~    2697.    6074.    4386.      -1.17   0.282
3 pgaptmp_000~ Sm0439~ SmT439~     679.     707.     693.     -0.0574 0.170
4 pgaptmp_000~ Sm0439~ SmT439~     405.     586.     495.     -0.529  0.212
5 pgaptmp_000~ Sm0439~ SmT439~   19787.   15808.   17798.      0.324  0.545
6 pgaptmp_000~ Sm0439~ SmT439~    9518.   10092.    9805.     -0.0845 0.393
# i 7 more variables: stat <dbl>, pvalue <dbl>, padj <dbl>, product_PGAP <chr>,
#   Description_Blast2GO <chr>, Length <dbl>, Combined_GO_Names <chr>
```

Plotting

Selecting the data

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v purrr      1.0.2
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
# Create a column that label not significant genes, SmT up and Sm0 up genes
↪ based on cut-offs:
# Create a new column to flag DEGs
DESeq2_dataAllAnnot$SignDEG <- 'Not significant'

# Significant DEGs are defined as absolute log2FoldChange > 1 and pvalue <
↪ 0.05
```

```

DESeq2_dataAllAnnot$SignDEG[DESeq2_dataAllAnnot$log2FoldChange > 1 &
  ↪ DESeq2_dataAllAnnot$padj < 0.05] <- 'SmO upregulated'
DESeq2_dataAllAnnot$SignDEG[DESeq2_dataAllAnnot$log2FoldChange < -1 &
  ↪ DESeq2_dataAllAnnot$padj < 0.05] <- 'SmT upregulated'

# selecting the top 10 DEG for each samples; SmO2 or SmT1 bacteria
top <- 10
topDEG <- bind_rows(DESeq2_dataAllAnnot |>
  filter(SignDEG == 'SmT upregulated') |>
  arrange(desc(abs(log2FoldChange))) |> head(top),
  DESeq2_dataAllAnnot |>
  filter(SignDEG == 'SmO upregulated') |>
  arrange(desc(abs(log2FoldChange))) |> head(top)
)

```

Extracting data for the top 10 DEG as a table

```

topDEG_filt <- topDEG %>% subset(select = c(1,7,12))

write_csv(topDEG_filt, "LU439_lin2_top10DEG.csv")

```

Setting the theme

```

theme_set(theme_classic(base_size = 15) +
  theme(
    axis.title.x = element_text(hjust = 0.5, margin =
      ↪ margin(20,0,0,0), size = 12, color = 'black'),
    axis.title.y = element_text(margin = margin(0,20,0,0), size =
      ↪ 12, color = 'black'),
    plot.title = element_text(hjust = 0.5)
  ))

```

Plotting

```

library(ggplot2)
library(ggrepel) # to annotate the plot

# to get the total number of DEGs for SmT and if needed use this number for
  ↪ the Volcano plot title
DEG_SmT <- sum(DESeq2_dataAllAnnot$SignDEG == 'SmT upregulated')
# to get the total number of DEGs for Sm0 and if needed use this number for
  ↪ the Volcano plot title
DEG_Sm0 <- sum(DESeq2_dataAllAnnot$SignDEG == 'Sm0 upregulated')
# to get the total number of non significant DEGs and if needed use this
  ↪ number for the Volcano plot title
notDEG <- sum(DESeq2_dataAllAnnot$SignDEG == 'Not significant')

# use this tutorial https://ggrepel.slowkow.com/articles/examples.html
# choose colors for plots here:
  ↪ https://sape.inf.usi.ch/quick-reference/ggplot2/colour

volcano <- ggplot(data = DESeq2_dataAllAnnot) +
  geom_point(mapping = aes(x=log2FoldChange,
                          y=-log10(padj),
                          fill=SignDEG),
            color = 'black',
            stroke = 0.3,
            size = 3.5,
            shape = 21) +
  scale_fill_manual(values = c("gray90",
                              "hotpink2",
                              "turquoise")) +
  geom_vline(xintercept=c(-1, 1),
            col = 'gray60',
            linetype = 'dashed') +
  geom_hline(yintercept = c(-log10(0.05)),
            col = 'gray60',
            linetype = 'dashed' ) +
  coord_cartesian(ylim=c(0, 80), xlim=c(-10,10)) +
  scale_x_continuous(breaks = seq(-10, 10, 2)) +
  annotate(geom="text", x=5, y=80,
            label="Sm0 up =513", color="black") +
  annotate(geom="text", x=-5, y=80,
            label="SmT up =588", color="black") +
  labs(title =
        'RNAseq LU439 SmT Lin1 vs Sm0 Lin2',

```

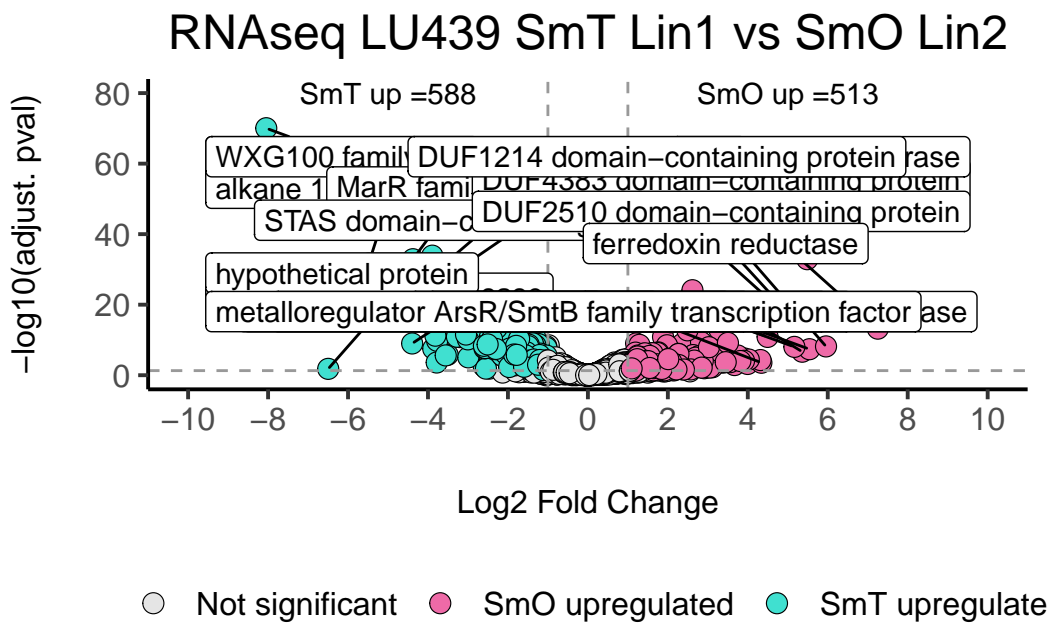
```

    x='Log2 Fold Change',
    y='-log10(adjust. pval)',
    fill = '') +
  guides (fill = guide_legend(position = "bottom"))

volcano <- volcano +
  geom_label_repel(data = topDEG,
    mapping = aes(log2FoldChange,-log10(padj), label =
      ↪ product_PGAP),
    min.segment.length = 0, # to always draw the line from dot
      ↪ to label
    size = 4, # font size label
    max.overlaps = Inf, # even if label overlap, always plot
      ↪ them
    box.padding = 1.5)

volcano

```



```

ggsave(
  "Volcano2_DESeq_LU439L2_annotbig.pdf",
  plot = last_plot(),
  width = 15,

```

```
    height = 15,  
    dpi = 300  
)  
  
ggsave(  
  "Volcano2_DESeq_LU439L2_annot.pdf",  
  plot = last_plot(),  
  width = 6,  
  height = 8,  
  dpi = 300  
)
```