Part 2

Gyaban Essilfie-Bondzie

2023-04-26

#Dseq Data Construction

```
# load library
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
```

```
library(DESeq2)
```

```
## Loading required package: S4Vectors
## Loading required package: stats4
## Loading required package: BiocGenerics
##
## Attaching package: 'BiocGenerics'
##
## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union
##
## 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, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, 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.max, which.min
##
##
## Attaching package: 'S4Vectors'
##
## The following objects are masked from 'package:lubridate':
##
##     second, second<-
##
## The following objects are masked from 'package:dplyr':
##
##     first, rename
##
## The following object is masked from 'package:tidyr':
##
##     expand
##
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname
##
## Loading required package: IRanges
##
## Attaching package: 'IRanges'
##
## The following object is masked from 'package:lubridate':
##
##     %within%
##
## The following objects are masked from 'package:dplyr':
##
##     collapse, desc, slice
##
## The following object is masked from 'package:purrr':
##
##     reduce
##
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Loading required package: SummarizedExperiment
## Loading required package: MatrixGenerics
## Loading required package: matrixStats
##
## Attaching package: 'matrixStats'
##
```

```
## The following object is masked from 'package:dplyr':
##
##     count
##
##
## Attaching package: 'MatrixGenerics'
##
## The following objects are masked from 'package:matrixStats':
##
##     colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
##     colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##     colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##     colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##     colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##     colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##     colWeightedMeans, colWeightedMedians, colWeightedSds,
##     colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##     rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##     rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##     rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##     rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##     rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##     rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##     rowWeightedSds, rowWeightedVars
##
## 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")'.
##
##
## Attaching package: 'Biobase'
##
## The following object is masked from 'package:MatrixGenerics':
##
##     rowMedians
##
## The following objects are masked from 'package:matrixStats':
##
##     anyMissing, rowMedians
```

```
# read count data
rna_cts <- read.csv("Marra2014_count_table_spleen.tsv", sep = "\t", row.names
= "gene_id")
rna_cts <- as.matrix(rna_cts)
colnames(rna_cts)
```

```
##  [1] "D1" "D2" "D3" "D4" "C1" "C2" "C3" "C4" "H1" "H2" "H3" "H4"
```

```
# create a dataframe of sample info
col_data <- data.frame(
  row.names=colnames(rna_cts),
  condition=c("desert", "desert", "desert", "desert", "desert", "desert",
"desert", "desert","mesic","mesic","mesic", "mesic"),
  species=c("Dipodomys", "Dipodomys", "Dipodomys", "Dipodomys",
"Chaetodipus", "Chaetodipus", "Chaetodipus", "Chaetodipus", "Heteromys",
"Heteromys", "Heteromys", "Heteromys")
)

# Construct a DESeq data set
de_obj <- DESeqDataSetFromMatrix(
  countData = rna_cts,
  colData = col_data,
  design = ~ condition)

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables
in
## design formula are characters, converting to factors

# Run differential gene expression analysis
dds <- DESeq(de_obj)

## estimating size factors
## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## -- note: fitType='parametric', but the dispersion trend was not well
captured by the
##    function: y = a/x + b, and a local regression fit was automatically
substituted.
##    specify fitType='local' or 'mean' to avoid this message next time.
## final dispersion estimates
## fitting model and testing
## -- replacing outliers and refitting for 32 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)
## estimating dispersions
## fitting model and testing

# get results
res <- results(dds)
```

Making new df and critgenes set ordered by up and down genes

```
#make res into dataframe
resplot <- as.data.frame(res)


#add column to data frame to show up,down and significance
resplot$diffexpressed <- "NO"
```

```
# if log2Foldchange > 0.6 and pvalue < 0.05, set as "UP"
resplot$diffexpressed[resplot$log2FoldChange > 0.6 & resplot$pvalue < 0.05]
<- "UP"
# if log2Foldchange < -0.6 and pvalue < 0.05, set as "DOWN"
resplot$diffexpressed[resplot$log2FoldChange < -0.6 & resplot$pvalue < 0.05]
<- "DOWN"

#make row.names into a column called gene_name
resplot$gene_name <- row.names(resplot)

#Make critgenes
upgenes <- resplot %>% filter(diffexpressed == "UP") %>% arrange(pvalue) %>%
head(10)
downgenes <- resplot %>% filter(diffexpressed == "DOWN") %>% arrange(pvalue)
%>% head(10)
critgenes <- rbind(upgenes, downgenes)

#get gene names
critgenes$gene_name
```

```
##  [1] "mypt2" "f135a" "cers2" "pgap2" "arsk"  "pld1"  "sphk2" "glcm"
"niban"
## [10] "armx1" "nu5m"  "brox"  "m3k12" "cb043" "kld10" "gna1"  "tom34"
"rn214"
## [19] "ube2f" "ci123"
```

Make Volcano Plot with labels

```
library(EnhancedVolcano)
```

```
## Loading required package: ggrepel
```

```
#make names into a column
matchlist <- c('mypt2' ,'f135a' ,'cers2' ,'pgap2' ,'arsk' ,'pld1' ,'sphk2'
,'glcm' ,'niban' ,'armx1' ,'nu5m' ,'brox'
,'m3k12','cb043','kld10','gna1','tom34','rn214','ube2f','ci123')

#Making the Plot
EnhancedVolcano(resplot,
                lab = resplot$gene_name,
                x = 'log2FoldChange',
                y = 'pvalue',
                xlim=c(-12,12),
                ylim=c(0,85),
                selectLab = matchlist, #adds labels to sig genes
                pCutoff = 10e-14,
                FCcutoff = 2.0,
                pointSize = 2.0,
                labSize = 3.0,
                shape = c(4, 35, 17, 18),
                colAlpha = 1,
```

```
                legendPosition = 'right',
                legendLabSize = 14,
                legendIconSize = 5.0)
```

```
## Warning: The `guide` argument in `scale_*()` cannot be `FALSE`. This was
deprecated in
## ggplot2 3.3.4.
## ℹ Please use "none" instead.
## ℹ The deprecated feature was likely used in the EnhancedVolcano package.
##   Please report the issue to the authors.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```