Find Treatment Effect Across Genotypes

Molly Davis & Ann Loraine

2023-10-23

Prelude - define variables

```
assemblies=c("SL4","SL5")
Q = 0.05
lfcThreshold=0
outfname_prefix="results/MvW-temp"
```

Introduction

The following adapted from: https://jira.bioviz.org/browse/IGBF-3460

We observed in previous work that VF36 and the F3H transgenic line have similar gene expression patterns. To understand how the *are* mutant differs from wild-type based on the effect of temperature, we will focus on comparing the *are* mutant to the VF36 cultivar under the temperature condition. This will be done with the use of an interaction term within the design of the analysis. The interaction term answers the question is the temperature condition effect *different* across genotypes?

Goal: Find and interpret how gene expression in *are* differs from gene expression in VF36 based on the temperature condition effect.

Rationale: In the past we have been curious how gene expression is different across genotypes but also the temperature condition. We know that there are differences in expression between genotypes but is there a difference between temperature and genotypes simultaneously.

Results

Load code:

```
source("Common.R")
```

Read counts data for assemblies SL4, SL5:

```
dfs = list()
for (assembly in assemblies) {
  dfs[[assembly]]=getCounts(assembly=assembly,
                            keep_description = T)
}
```

DESeq Function; Fit a model that compares *are* to VF36 at specific time chunk with temperature condition:

```
getDEgenes_time <- function(minute) {
  ddss = list()
  for (assembly in assemblies) {
    d = dfs[[assembly]]
    desc_index = which(names(d)=="description")
    d = d[,-desc_index]
    toks = strsplit(names(d),"\\.")
    genotype=sapply(toks,function(x){x[[1]]})
    temperature=sapply(toks,function(x){x[[2]]})
    time=sapply(toks,function(x){x[[3]]})
    v = genotype%in%c("A","V")&time==minute
    d = d[,v]
    coldata=data.frame(genotype=factor(genotype[v],levels=c("V","A")),
                       temperature=factor(temperature[v], levels = c("28","34")))
    row.names(coldata)=names(d)
    cts = round(as.matrix(d))
    dds = DESeqDataSetFromMatrix(countData=cts,
                                 colData=coldata,
                                  design=~genotype + temperature + genotype:temperature)
    featureData = data.frame(gene=rownames(cts))
    mcols(dds) = DataFrame(mcols(dds), featureData)
    dds = DESeq(dds, minReplicatesForReplace=Inf)
    ddss[[assembly]]=dds
  }
  return(ddss)
}
```

Collect results function:

```
results_time <- function(ddss, minute) {
  rss=list()
  for (assembly in assemblies) {
  dds=ddss[[assembly]]
    rs = results(dds,alpha=Q,lfcThreshold = lfcThreshold,
                 name="genotypeA.temperature34")
    rs = rs[!is.na(rs$padj),]
    rs=cbind(gene=row.names(rs),rs)
    rs["time"] = minute
    rs$description = dfs[[assembly]][rs$gene,"description"]
    if (assembly == "SL5") {
      SL4_gene_names = getSL4GeneNames(rs$description)
    rs$SL4 = SL4_gene_names
    }
    o = order(rs$pvalue)
    rs = rs[o,]
```

```
    rss[[assembly]]=data.frame(rs)
  }
  return(rss)
}
```

15 minute Results

```
min15 <- getDEgenes_time(15)
res15 <- results_time(min15, 15)
```

The previous code chunk collected genes for SL4 and SL5 at 15 minutes that differed by log2 fold-change of 0 or greater between VF36 and *are* with a temperature comparison and where the comparisons had adjusted p value less than or equal to 0.05.

30 minutes Results

```
min30 <- getDEgenes_time(30)
res30 <- results_time(min30, 30)
```

The previous code chunk collected genes for SL4 and SL5 at 30 minutes that differed by log2 fold-change of 0 or greater between VF36 and *are* with a temperature comparison and where the comparisons had adjusted p value less than or equal to 0.05.

45 minutes Results

```
min45 <- getDEgenes_time(45)
res45 <- results_time(min45, 45)
```

The previous code chunk collected genes for SL4 and SL5 at 45 minutes that differed by log2 fold-change of 0 or greater between VF36 and *are* with a temperature comparison and where the comparisons had adjusted p value less than or equal to 0.05.

75 minutes Results

```
min75 <- getDEgenes_time(75)
res75 <- results_time(min75, 75)
```

The previous code chunk collected genes for SL4 and SL5 at 75 minutes that differed by log2 fold-change of 0 or greater between VF36 and *are* with a temperature comparison and where the comparisons had adjusted p value less than or equal to 0.05.

Combine the dataframes

```
combined_results <- rbind(res15, res30, res45, res75)
SL4 <- combined_results[,1]
SL5 <- combined_results[,2]
```

The previous code chunk collected genes for SL4 and SL5 across all times.

View and Save the Results

Examine results by assembly, starting with `SL4`. Also, write a file with results.

```
library(openxlsx)
assembly="SL4"
outfname=paste0(outfname_prefix,"-",assembly,".xlsx")
data_frames <- list("15 minutes" = SL4$res15, "30 minutes" = SL4$res30,
                    "45 minutes" = SL4$res45,"75 minutes" = SL4$res75 )
write.xlsx(data_frames, file = outfname)
```

Results for SL4:

Comparing *are* to VF36, with a temperature condition,

15 minutes: identified 0 differentially expressed genes from assembly SL4 for 15 minute experiments. Results were written to file results/MvW-temp-SL4.xlsx.

30 minutes: identified 6 differentially expressed genes from assembly SL4 for just 30 minute experiments. Results were written to file results/MvW-temp-SL4.xlsx.

45 minutes: identified 0 differentially expressed genes from assembly SL4 for just 45 minute experiments. Results were written to file results/MvW-temp-SL4.xlsx.

75 minutes: identified 0 differentially expressed genes from assembly SL4 for just 75 minute experiments. Results were written to file results/MvW-temp-SL4.xlsx.

```
assembly="SL5"
outfname=paste0(outfname_prefix,"-",assembly,".xlsx")
data_frames <- list("15 minutes" = SL5$res15, "30 minutes" = SL5$res30,
                    "45 minutes" = SL5$res45,"75 minutes" = SL5$res75 )
write.xlsx(data_frames, file = outfname)
```

Results for SL5:

Comparing *are* to VF36, with a temperature condition,

15 minutes: identified 0 differentially expressed genes from assembly SL5 for 15 minute experiments. Results were written to file results/MvW-temp-SL5.xlsx.

30 minutes: identified 6 differentially expressed genes from assembly SL5 for just 30 minute experiments. Results were written to file results/MvW-temp-SL5.xlsx.

45 minutes: identified 0 differentially expressed genes from assembly SL5 for just 45 minute experiments. Results were written to file results/MvW-temp-SL5.xlsx.

75 minutes: identified 0 differentially expressed genes from assembly SL5 for just 75 minute experiments. Results were written to file results/MvW-temp-SL5.xlsx.

Discussion

The linear model allowed us to identify 6 genes whose expression in the temperature differed between VF36 and *are* samples compared with the temperature effect with a log2FC threshold of 0. The log2FC threshold was changed to 0 due to there not being enough DE genes available for analysis. This means we are allowing for a less strict approach when it comes to looking at DE genes that are taken from a more complicated analysis. The analysis design is asking for there to be significance in the difference in genotypes *and* temperatures for there to be DE genes present. While if there was no interaction term present in the design, then we would be asking for genoytype *or* temperature differences. With the interaction, performing the analysis of both factors and pulling the results becomes much easier. Thus, answering the question, is the temperature condition effect *different* across genotypes?

Conclusions

- Six genes were differently expressed at 30 minutes between *are* and VF36 were identified under the effect of temperature with a log2FC of 0.
- Files were written to the "results" directory for the two genotypes (MvW-temp-SL4.xlsx & MvW-temp-SL5.xlsx).

Session Info

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur ... 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/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] stats4    stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] openxlsx_4.2.5.2           EnhancedVolcano_1.16.0
##  [3] ggrepel_0.9.3              ggplot2_3.4.1
##  [5] DESeq2_1.38.3              SummarizedExperiment_1.28.0
##  [7] Biobase_2.58.0             MatrixGenerics_1.10.0
##  [9] matrixStats_0.63.0         GenomicRanges_1.50.2
## [11] GenomeInfoDb_1.34.9        IRanges_2.32.0
```

```
## [13] S4Vectors_0.36.1           BiocGenerics_0.44.0
## [15] edgeR_3.40.2               limma_3.54.2
## [17] readxl_1.4.3               readr_2.1.4
## [19] stringr_1.5.0
##
## loaded via a namespace (and not attached):
##  [1] httr_1.4.7             bit64_4.0.5           blob_1.2.3
##  [4] GenomeInfoDbData_1.2.9 cellranger_1.1.0      yaml_2.3.7
##  [7] pillar_1.9.0           RSQLite_2.3.0         lattice_0.20-45
## [10] glue_1.6.2             digest_0.6.31         RColorBrewer_1.1-3
## [13] XVector_0.38.0         colorspace_2.1-0      htmltools_0.5.6
## [16] Matrix_1.5-3           XML_3.99-0.13         pkgconfig_2.0.3
## [19] zlibbioc_1.44.0        xtable_1.8-4          scales_1.2.1
## [22] tzdb_0.4.0             BiocParallel_1.32.5   tibble_3.2.1
## [25] annotate_1.76.0        KEGGREST_1.38.0       generics_0.1.3
## [28] withr_2.5.0            cachem_1.0.6          cli_3.6.0
## [31] magrittr_2.0.3         crayon_1.5.2          memoise_2.0.1
## [34] evaluate_0.20          fansi_1.0.4           tools_4.2.2
## [37] hms_1.1.3              lifecycle_1.0.3       munsell_0.5.0
## [40] locfit_1.5-9.7         zip_2.2.2             DelayedArray_0.24.0
## [43] AnnotationDbi_1.60.0   Biostrings_2.66.0     compiler_4.2.2
## [46] rlang_1.1.1            grid_4.2.2            RCurl_1.98-1.10
## [49] rstudioapi_0.14        bitops_1.0-7          rmarkdown_2.20
## [52] gtable_0.3.1           codetools_0.2-19      DBI_1.1.3
## [55] R6_2.5.1               dplyr_1.1.3           knitr_1.42
## [58] fastmap_1.1.0          bit_4.0.5             utf8_1.2.3
## [61] stringi_1.7.12         parallel_4.2.2        Rcpp_1.0.10
## [64] vctrs_0.6.3            geneplotter_1.76.0    png_0.1-8
## [67] tidyselect_1.2.0       xfun_0.37
```