Differential Gene Expression in *Arabidopsis Thaliana* mutants

Manzano, B. P.

Introduction:

Differential gene expression analysis is a powerful tool used to understand what genes are expressed in an organism at a given time. In this analysis, differential gene expression is carried out to compare gene expression between *Arabidopsis Thaliana* mutants and the wild type (Van Aken et al., 2016). The Arabidopsis populations used from this study either had mutations in prohibitin 3 (atphb3) or organellar RNA polymerase (rpotmp) and also either had mutations in ANAC017 (annotated as lowercase anac017) or a wildtype ANAC017 allele (Van Aken et al., 2016). These 4 mutant lines were compared to the wild type line in order to look for differentially expressed genes that may be associated with these mutations.

Methods:

In an R statistical environment, DEseq2 was used to perform differential expression analysis. The following steps were performed in the analysis:

1. Set up global options

   The first chunk of code made sure that echo = TRUE was applied to all subsequent code chunks. Then, the next line set the working directory for all of the subsequent code chunks.

2. Load packages

   The packages DEseq2, pasilla, and apelgm were all load for use in this analysis.

3. Read in Files

The count file E-MTAB-4655-raw-counts.tsv and the meta data file E-MTAB-4655-experiment-design.tsv were read in after being obtained from the paper (Van Aken et al., 2016).

4. Ensure number of rows in column data matches number of columns in count data

   After seeing that the count file had 17 columns and the column data had 15 rows, code was used to ensure these numbers were both 15. First, the gene codes (in the first column) were made into the row names. Then, the first two columns in the count file were removed.

5. Factorization

   Code was used to ensure column data we are interested in is treated as a factor.

6. Creation of DEseqDataSet

   Both the count file with removed columns and the coldata were used to create a DEseqDataSet. The design variable used the input Factor.Value.genotype in order to set up a basic comparison between the different mutant genotypes and the wild type.

7. Releveling:

   To ensure that all mutants were compared to the baseline wild type gene expression, releveling is preformed to set the wild type as the baseline comparison.

8. DEseq2:

   DEseq2 is then run on the DEseqDataSet.

9. Results extraction:

   Results were extracted comparing each of the 4 mutant lines to the wild type line. Results given have an p value of less than 0.05 (alpha = 0.05) and a padj value

(padj < 0.05). This was done to correct for multiple testing errors since a large number of genes were tested.

10. Gene annotation:

    If the results showed at least one significant differentially expressed gene, then the genes were annotated using Vlookup (see README_annotation.txt).

Results:

The only comparison with differentially expressed genes was comparing the atphb3.ANAC017 mutant with the wild type. In this case, one gene was upregulated and 10 were down regulated (Table 1). The rest of the comparisons showed no differentially expressed genes.

Gene_ID	Gene_Name	log2FoldChange	padj
AT1G54040	ESP	1.022026973	0.044601179
AT1G61275	U12	-0.344635365	0.032147876
AT3G48560	ALS	-0.404259049	0.037707464
AT4G16830	RGGA	-0.558247223	0.046631241
AT4G17300	NS1	-0.564644645	0.017103553
AT4G17520	RGGB	-0.565200043	0.017103553
AT4G34980	SBT1.6	-0.653729007	0.016339779
AT4G38970	FBA2	-0.676337921	0.017103553
AT5G01920	STN8	-0.792839674	0.007209097
AT5G54770	THI1	-0.798431754	0.048719836
ATCG00140	ATPH	-1.038040684	2.15E-05

Table 1: Differentially expressed genes in atphb3.ANAC017 mutant vs. wild type comparison.

Discussion:

This low number of differentially expressed genes among these lines suggests that the mutations in the mutants likely do not have a large effect on gene expression on a whole (table 1). Because they are the same species, these lines likely have very similar expression profiles. However, network analysis could be used in the atphb3.ANAC017 mutant vs. wild type comparison in order to find evidence of topological change among networks built for both the

mutant and the wild type (Langfelder and Hovath, 2008). This would provide more detail as to how genetic pathways are different between this mutant and the wildtype.

References:

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559-559. https://doi.org/10.1186/1471-2105-9-559

Van Aken, O., Ford, E., Lister, R., Huang, S., & Millar, A. H. (2016). Retrograde signalling caused by heritable mitochondrial dysfunction is partially mediated by ANAC017 and improves plant performance. *The Plant Journal : For Cell and Molecular Biology*, *88*(4), 542-558. https://doi.org/10.1111/tpj.13276