BIO312G: Final Project (Report)

Arun Krishnaraj - ak37738

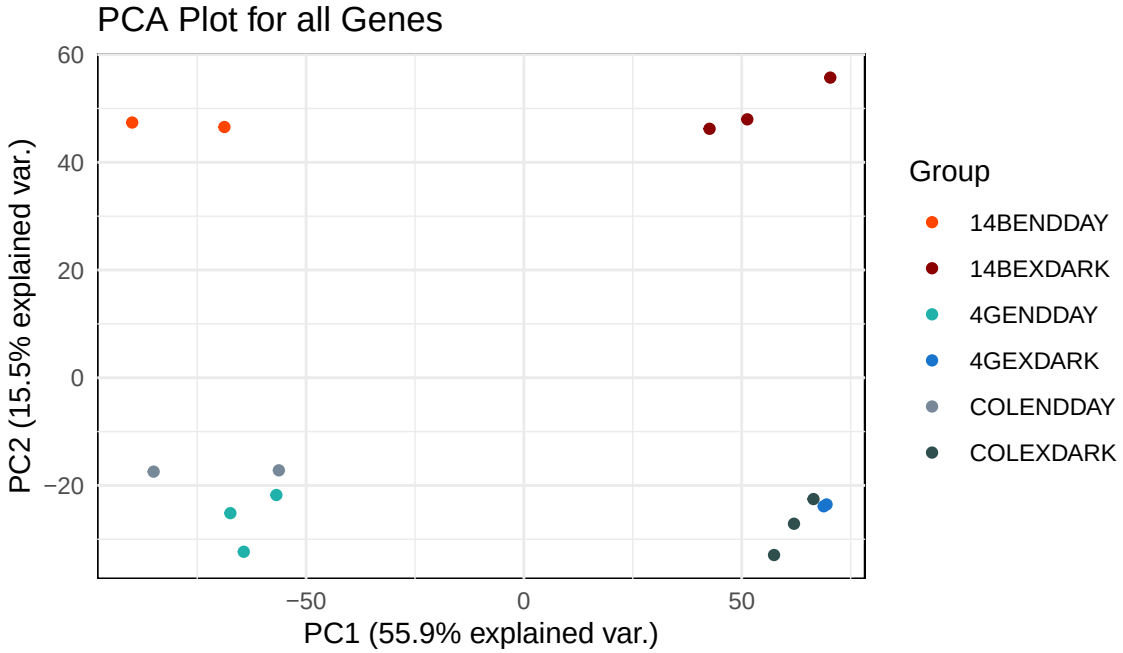
In this project, I'll examine differences in plant gene expression across genotype and light-treatment variants; I seek to determine if significant differential gene expression can be attributed to genotype-timepoint interaction. A full code file is available for review as `BIO321-RNAseq-code.R`, from which plots were imported. RNA sequence expression data was obtained for multiple *Arabidopsis thaliana* genotypes:

- **Col**: wild-type *Arabidopsis thaliana* line
- **14B**: null mutant line for translation initiation factors eIF4G1 and eIF4G2
- **4G**: null mutant line for translation initiation factor eIF4G

Samples were collected at two distinct light conditions: "End Day" (light) and "Ex Dark" (dark). We are interested in understanding how differences in gene expression between light condition vary across the 3 Arabidopsis genotypes.

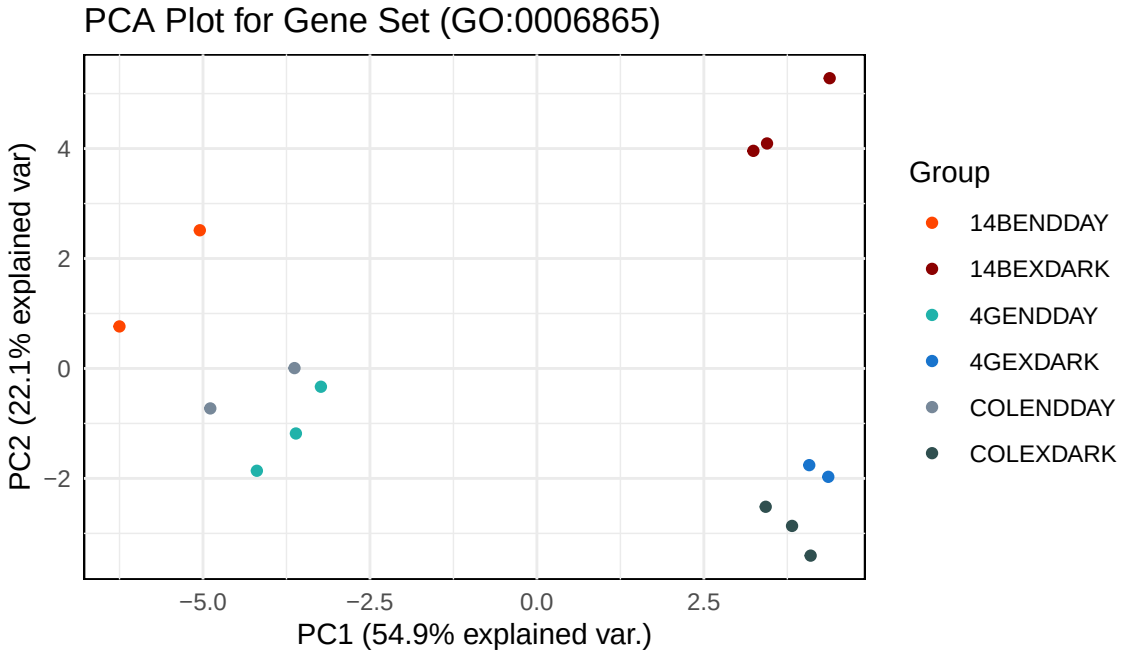
First, a `DESeq` object was created from the expression count data, designed to test the effect of time-genotype interaction on gene expression. Using the likelihood ratio test, we identify any genes that show changes in expression across time-genotype interaction groups. At a desired false discovery rate (FDR) of ≤ 0.10, 1,429 genes were found to be differentially expressed across interaction groups; these significant genotype-timepoint interaction genes were found using the built-in Benjamini-Hochberg adjustment to `DESeq` p-values. Using the adjusted p-values reported by `DESeq` and an FDR cutoff of 10%, we expect to see 142.9 false discoveries of significant interaction terms.

Expression count data was then normalized with DESeq and $\log_2$-transformed with an offset of 1. PCA was performed on the transformed count data to reduce dimensionality, the results of which are shown here:



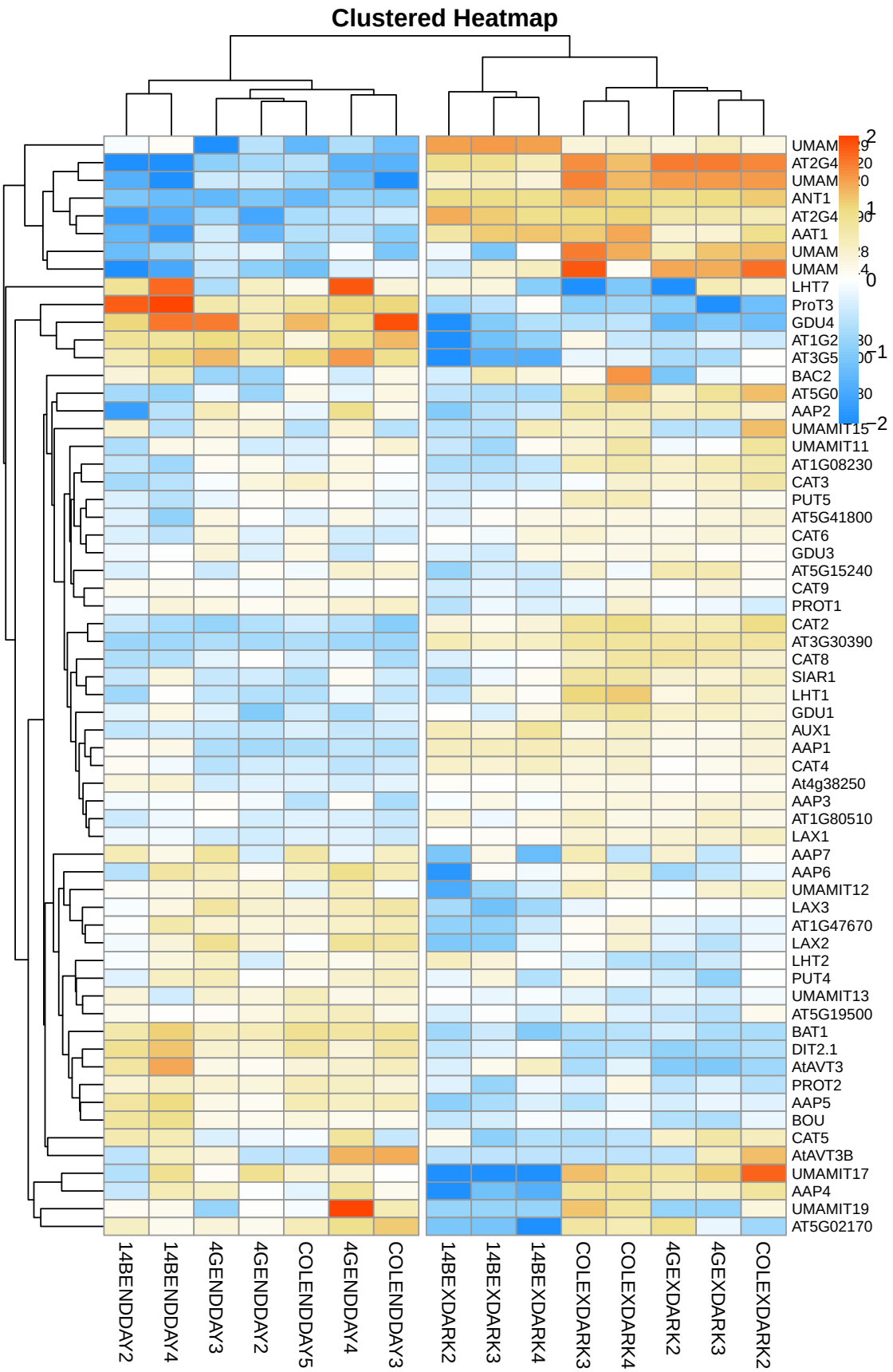
PC1 highly separates samples by light treatment group across all 3 Arabidopsis genotypes: light treatment groups tend to score low on PC1, while dark treatment groups score highly for PC1. PC2 highly separates 14B mutant from both 4G mutant and Wild-type samples across light treatment groups. 14B mutants score highly on PC2, while 4G and Wild samples score low. The first and second PCs in this model explain 55.9% and 15.5% of overall gene expression variance respectively. Overall, the PCA plot indicates that differential expression may occur as a result of genotype-light interactions.

The specific gene set given by ontology ID GO:0006865 was selected for further analysis. Information on selected genes was obtained, and is available for reference as `Final-Project-Gene-desc.tsv`. PCA was again performed on the transformed count data for the gene set of interest, the results of which are shown here:



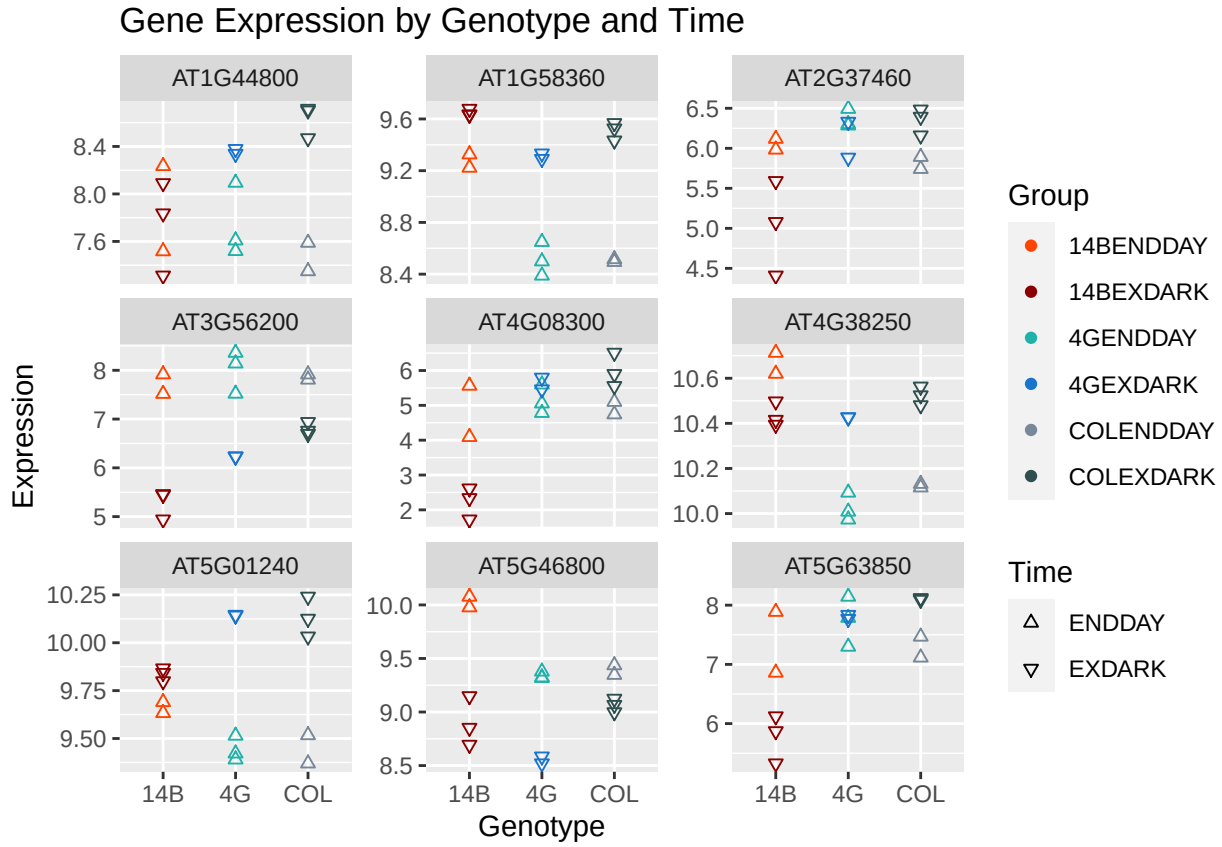
Compared to the full-gene PCA plot, this PCA plot provides similar separation along PC1; dark treatment samples are again found to be high for PC1 and light samples are found to be low for PC1. Separation along PC2 is somewhat changed from the full gene-set: dark treatment samples are well separated across genotype, while light samples have worse separation across genotype. Overall, this change suggests that for the gene-set of interest, the interaction of light and genotype have significant effects on expression; separation from the control groups is maximized by changing both light treatment and genotype.

Transformed count data for the gene set was used to generate a clustered heatmap of differences from average gene expression level.


Clustered Heatmap
UMAM
AT2G4
UMAM
ANT1
AT2G4
AAT1
UMAM
UMAM
LHT7
ProT3
GDU4
AT1G2
AT3G5
BAC2
AT5G0
AAP2
UMAMIT15
UMAMIT11
AT1G08230
CAT3
PUT5
AT5G41800
CAT6
GDU3
AT5G15240
CAT9
PROT1
CAT2
AT3G30390
CAT8
SIAR1
LHT1
GDU1
AUX1
AAP1
CAT4
At4g38250
AAP3
AT1G80510
LAX1
AAP7
AAP6
UMAMIT12
LAX3
AT1G47670
LAX2
LHT2
PUT4
UMAMIT13
AT5G19500
BAT1
DIT2.1
AtAVT3
PROT2
AAP5
BOU
CAT5
AtAVT3B
UMAMIT17
AAP4
UMAMIT19
AT5G02170
2
1
0
−1
−2
14BENDDAY2
14BENDDAY4
4GENDDAY3
4GENDDAY2
COLENDDAY5
4GENDDAY4
COLENDDAY3
14BEXDARK2
14BEXDARK3
14BEXDARK4
COLEXDARK3
COLEXDARK4
4GEXDARK2
4GEXDARK3
COLEXDARK2

The samples cluster highly by light-treatment, shown by the vertical dendrogram; the two largest clusters contain all the light and dark treatment samples respectively. When we consider the two largest gene clusters, we can note a highly cold area of the heatmap on the light treatment cluster for the upper gene cluster. This gene cluster has strongly warm cells for the dark treatment cluster; this gene cluster seems to be highly differentially expressed between light and dark treatments. Several other areas of more mild temperature inversion between light and dark clusters can be seen throughout the heatmap.

Transformed expression count data was then filtered to include only the 9 most significant genes from the gene set, based on p-value obtained from `DESeq`. The expression levels of these 9 genes across light-treatment and genotype are shown below



The change in level of expression for genes AT1G58360, AT3G56200, AT5G01240, and AT5G46800 across timepoints are consistent in direction between the 3 genotype samples; that is the relative expression levels for light and dark samples are consistent across genotypes. However, other genes like AT4G08300, and AT4G38250 show strong variance of expression difference between timepoints among the samples. Additionally, genes like AT1G44800, and AT5G01240 show no clear difference in expression between timepoints for wild-type samples, but show more significant differences for mutant genotypes. Overall, this expression data suggests that gene expression differences between timepoint varies both in direction and magnitude between genotypes; this behavior suggests the significance of genotype-timepoint interaction on gene expression levels for the given gene-set. The effect of light-treatment on gene expression appears to be changed in both translation initiation factor null-mutant lines, though the change is somewhat inconsistent between mutant lines. Further work may include additional genes for analysis, or investigate relative contributions of genotype-timepoint main effects and interaction effect on differential gene expression.