

1: Differential Gene Expression Analysis Report (oat)

how the effect of resistance on gene expression changes over time

Note: two samples were recognized as outliers and removed from the experiment since they did not belong to the correct cluster (or groups) (24h, 198_5/6_sus) and (1d, 198_5/1_res)

Introduction

In this report, we present the results of a Differential Gene Expression (DGE) analysis conducted for the oat experiment. The analysis done using DESeq2 package in R. We used a specific model for the analysis to understand how gene expression is affected by desired factors and their interactions.

Experimental Design

The choice of the model for the analysis plays a crucial role in understanding the relationship between the variables in the experiment. In our case, we employed the model $\sim \text{resistance} + \text{timepoint} + \text{resistance: timepoint}$. This model was selected for the following reasons:

Main Effects: By including res and timepoint as main effects in the model, we account for the impact of the response variable (res) and the timepoint variable on gene expression independently.

Interaction Term: **whether the effect of resistance on gene expression is time dependent.** For example, it can detect genes for which resistance has a significant effect only at certain timepoints. This interaction term is vital when we suspect that the effect of the response variable may vary over time.

In between PCA and Heatmap of samples (their distance) were provided to show how they are clustered and for the recognition of outliers. In this experiment two outliers were identified which were already removed from the data set. Furthermore you can find more detailed graph such as library size distribution, variance stabilizing plot (regarding normalization) and dispersion plot.

Analysis Results

Significant genes:

After conducting the DGE analysis, we identified a set of significant genes. These genes are those whose expression levels differ significantly between different experimental conditions and timepoints. We divided these significant genes into two categories,

1. Upregulated Genes: Genes with increased expression in specific conditions or timepoints compared to others.

2. Downregulated Genes: Genes with decreased expression in specific conditions or timepoints compared to others.

The lists of significant genes, both upregulated and downregulated, have been saved as CSV files for further examination. This information can be found in the accompanying files.

MA plot (Bland–Altman plot)

We have also generated an MA plot to visualize the distribution of significant genes between two individuals or two groups. (Here between resistance and susceptible). The plot shows the log-Fold Change for each gene against its average expression across all samples in the two conditions being contrasted.

Heatmap of Top 30 Significant Genes

To gain further insight into the gene expression patterns, we created a heatmap of the top 30 significant genes. Heatmaps are useful for visualizing how the expression of these genes varies across different conditions and timepoints. This heatmap provides a clear representation of the clustering and expression patterns of the most influential genes in our analysis. These visualizations assist in the identification of common trends or patterns among the significant genes. This can aid in the interpretation of underlying biological processes or pathways that are affected by the experimental conditions and timepoints. Furthermore, hierarchical clustering of these genes are also available.

Visualization of Gene Expression Changes Over Time

For each of the 30 significant genes identified in our analysis, we generated individual ggplot graphs to illustrate how gene expression levels change across the five timepoints. These visualizations provide a comprehensive view of the temporal dynamics of gene expression and offer insights into the behavior of these genes under different experimental conditions. (all graphs are available in the folder of 01_1significant_gene_plots which is the subfolder of 01-DEGs-analysis-timeseries)

Conclusion

The chosen model, including both main effects and interaction terms, provides a comprehensive view of how these factors impact gene expression. In general 53 genes were

significantly expressed ($pvalue < 0.05$, $absolute(lfc) > 1$) which 6 were upregulated and the rest down regulated. (CSV files were provided)

The accompanying CSV files and visualizations are available for further investigation and discussion. We believe that the results of this analysis will contribute to a better understanding of the underlying biology of the experimental conditions. You can find more information and relate figures on the folder of timeseries analysis)

2: Differential Gene Expression Analysis Report (oat)

In this case, we employed the model $\sim \text{resistance} + \text{timepoint}$. This model was selected for the following reasons.

- It allows you to assess how the resistance factor (e.g., different treatment groups) influences gene expression, irrespective of the timepoint.
- It also enables you to examine how time affects gene expression across all resistance group.
- The model doesn't account for any specific interaction effect between resistance and time; it treats the two factors as independent contributors to gene expression.

For this model all required information and graphs are stored in DEGS-main effect folder.

First the number of all significant genes was calculated ($pvalue < 0.05$) which was 1326 genes, Then the differentially expressed ones including up and down regulated ones were identified which are 191 and 187 genes respectively. All information can be found in 02-DEGs-analysis-main effects)

3: Differential Gene Expression Analysis Report (oat)

considering only the "resistance" factor without including the "time" factor or their interaction.

$\sim \text{resistance}$

Like the two previous designs I tried to identify the number of the differentially expressed genes and also the upregulated (139 genes) and downregulated ones (106 genes) were stored in csv files in 03-DEGs-analysis-resistance effect folder

- This model focuses exclusively on understanding the impact of the resistance factor on gene expression, without considering variations related to time.
- It provides a straightforward analysis if you believe that the "time" factor does not play a significant role in gene expression changes in the context of your study.
- It may be suitable when you want to assess the influence of resistance treatment across all timepoints without considering time-specific effects.

Differences between res and sus for each timepoint individually

Finally in this experiment each timepoint was extracted separately and differential expression analysis was done for each.

In the DEGs of each timepoint separately folder you will see 5 sub-folders with the names of the results_0h,6h,12h,1d and 2d. For each timepoint the significant differentially expressed genes under two resistance and susceptible situations were recognized and related figures (e.g. heatmap of the top 30 significant genes, and MA plot) were also provided.