# Differential Gene Expression Analysis Report

## Objective:

The report details the analysis of differential gene expression between experiment conditions of two datasets, A vs F and A vs D using DESeq2 results. Summary statistics, visualisations (such as volcano plots, MA plots, histograms and heatmaps), and an explanation of the results for each dataset are important analysis.

## Results:

### Dataset A vs F

Understanding the notable variations in gene expression across conditions A and F is the main goal of the A\_vs\_F dataset analysis.

1. Significant genes

*The threshold set for upregulated and downregulated genes are:*

*Adjusted p-value threshold= 0.05*

*Log2 fold changes for upregulated genes= 1*

*Log2 fold changes for downregulated genes= -1*

Upregulated genes: 627

Downregulated genes: 821

Total genes: 6397

Number of non-significant genes: 4929

Number of genes with missing padj: 54

Significant genes (Up + Down): 1448 (20% of total genes)

1. Summary statistics

*p-value summary*

Table 1: An overview of the p-values for the A vs F comparison's differential gene expression study. The minimum, first quartile (1st Qu), median, mean, third quartile (3rd Qu), maximum, and number of missing p-values (NAs) are all shown in the table. Non-significant p-values for most genes show little to no variation in expression across circumstances.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu | Median | Mean | 3rd Qu. | Max. | NA’s |
| 0.00000 | 0.00049 | 0.06396 | 0.23986 | 0.43498 | 1.00000 | 54 |

*Log2 fold change summary*

Table 2: Summary of log2 fold change values for the A vs F comparison's differential gene expression study. The log2 fold change values are shown in the table as the minimum, median, mean, third quartile (3rd Qu), first quartile (1st Qu), and maximum. While the extreme values represent a tiny selection of genes with significant expression alterations, a median near zero suggests that the majority of genes show only slight changes in expression.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu | Median | Mean | 3rd Qu. | Max. |
| -25.223480 | -0.557045 | -0.005647 | -0.114481 | 0.486882 | 22.278780 |

1,448 significant genes (627 upregulated, 821 downregulated) were found in the A vs F differential expression study, which accounted for 20% of all the genes examined. 54 genes were missing adjusted p-values, and the majority of genes (4,929) were not significant. The median of the p-value distribution, which is 0.06396, indicates that most genes have few significant changes. With a median close to zero (-0.0056), log2 fold changes range from -25.22 to 22.28, indicating little change in the expression of the majority of genes and significant variances for a small fraction.

1. Visualization plots

*Volcano plot*

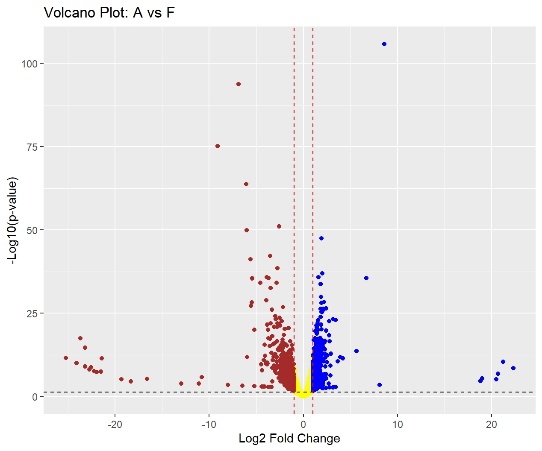


Figure 1: A volcano plot showing the degree of the expression change (log2 fold change) and statistical significance (-log10 p-value) for the genes in the A vs F comparison. Yellow points indicate genes that are close to the significance and fold change criteria, blue points indicate genes that are significantly downregulated, and red points indicate genes that are significantly elevated. The significance threshold is shown by the horizontal dashed line, while the fold change cutoff is shown by the vertical dashed line.

*MA plot*

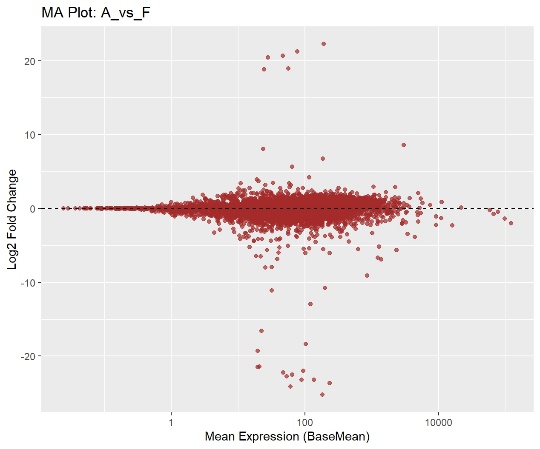


Figure 2: In the A vs F comparison, the MA plot displays the correlation between the mean expression (BaseMean) and log2 fold change (Log2 Fold Change) of the genes. A gene is represented by each dot, and the majority of genes exhibit little change in expression (centred about 0). A few genes stand out as outliers due to their significant fold alterations.

*Histogram*

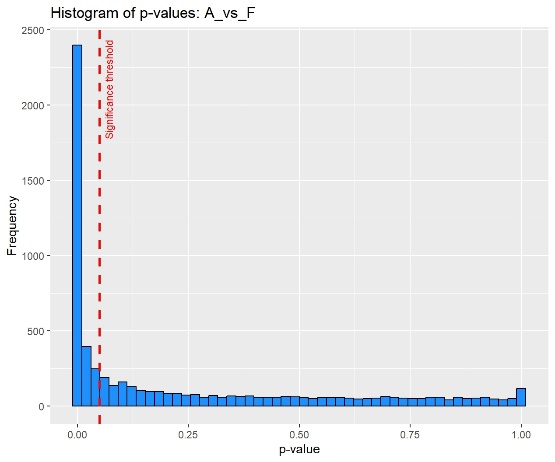


Figure 3: p-value histogram for the A vs F comparison's differential expression analysis. p-values for most genes are near zero, which denotes notable variations in expression. The significance level (p = 0.05) is shown by the red dashed line.

*Heatmap*

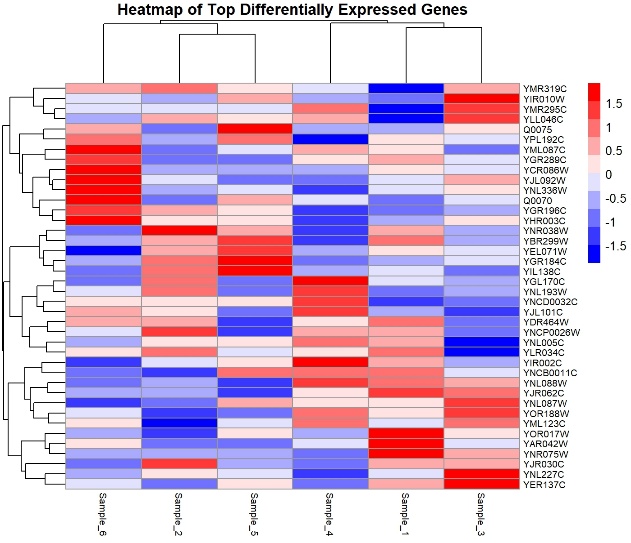
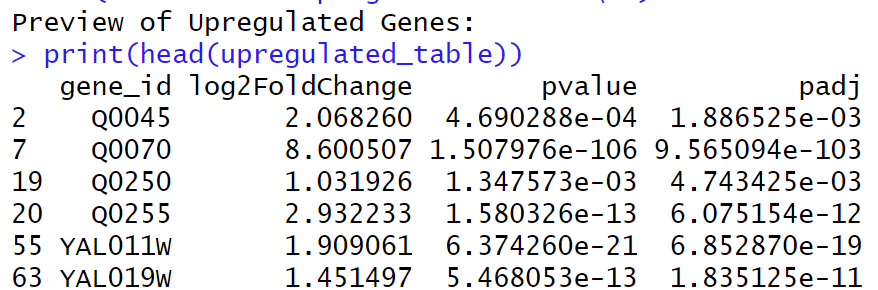
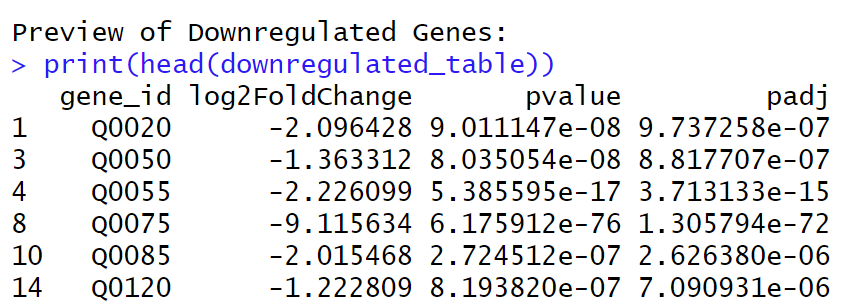


Figure 4: Heatmap comparing conditions A vs F that shows the expression patterns of the top genes with differential expression across six samples. Red denotes elevated genes and blue denotes downregulated genes. Genes are grouped according to how comparable their expressions are. Groups of genes and samples with comparable expression patterns are highlighted by the clustering dendrograms.

1. Significant gene list

Below mentioned are a preview of upregulated and downregulated genes based on the p-values, adjusted p-values and fold changes:





1. Additional analysis

*Impact of Differentially Expressed Genes on Biological Pathways*

* Based on volcano plots and hierarchical clustering (heatmap), the increased genes might indicate transcriptional regulation and stress responses.
* Significant log2 fold changes and p-value patterns indicate that downregulated genes indicate inhibited metabolic and signalling pathways.

*Gene Expression Pattern Clustering*

* Co-expressed gene groupings are displayed via hierarchical clustering (heatmap).
* The understanding of transcriptional responses is supported by the clear visibility of upregulated and downregulated clusters.

### Dataset A vs D

Understanding the notable variations in gene expression across conditions A and D is the main goal of the A vs D dataset analysis.

1. Significant genes

*The threshold set for upregulated and downregulated genes are:*

*Adjusted p-value threshold= 0.05*

*Log2 fold changes for upregulated genes= 1*

*Log2 fold changes for downregulated genes= -1*

Upregulated genes: 365

Downregulated genes: 651

Total genes: 6397

Number of non-significant genes: 5363

Number of genes with missing padj: 303

Significant genes (Up + Down): 1016 (16% of total genes)

1. Summary statistics

*p-value summary*

Table 3: p-value summary statistics for the A vs D comparison's differential expression analysis. For every gene that was examined, the table displays the p-values for the minimum (Min), first quartile (1st Qu.), median, mean, third quartile (3rd Qu.), and maximum (Max). Missing p-values (NAs) for 54 genes in total suggest possible problems with the quality of the data or the need for filtering.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu | Median | Mean | 3rd Qu. | Max. | NA’s |
| 0.00000 | 0.00209 | 0.08995 | 0.25446 | 0.46346 | 1.00000 | 54 |

*Log2 fold change summary*

Table 4: Log2 fold change summary data for the A vs D comparison's differential expression analysis. The table gives information about the size and direction of changes in gene expression for all examined genes by displaying the minimum (Min), first quartile (1st Qu.), median, mean, third quartile (3rd Qu.), and maximum (Max) log2 fold changes.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu | Median | Mean | 3rd Qu. | Max. |
| -25.328550 | -0.474189 | -0.001087 | -0.106928 | 0.416604 | 22.002860 |

With 365 upregulated and 651 downregulated genes, the differential expression analysis for A vs D reveals 1,016 substantially changed genes. Potential problems with data quality were indicated by the fact that 303 genes had missing adjusted p-values and the majority of genes (5,363) were non-significant. With a median of 0.08995 and a range of 0 to 1, P-values indicate a selection of highly important genes. With a median close to 0, log2 fold changes range significantly from -25.33 to 22.00, representing slight changes in the expression of most genes and drastic variances for a select few.

1. Visualization plots

*Volcano plot*

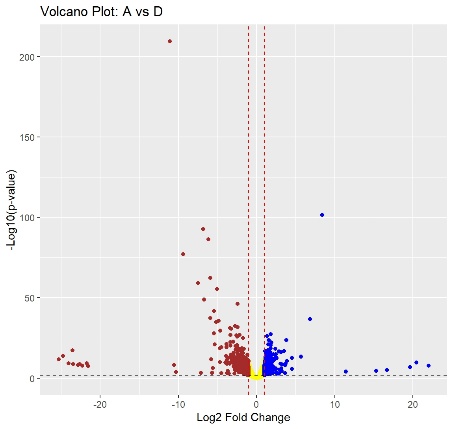


Figure 5: Volcano plot that shows the amplitude (log2 fold change, x-axis) and significance (-log10 p-value, y-axis) of changes in gene expression. According to predetermined criteria (adjusted p-value < 0.05 and |log2 fold change| > threshold), red and blue points indicate highly upregulated and downregulated genes, respectively. Genes that are not significant are indicated by the yellow area.

*MA plot*

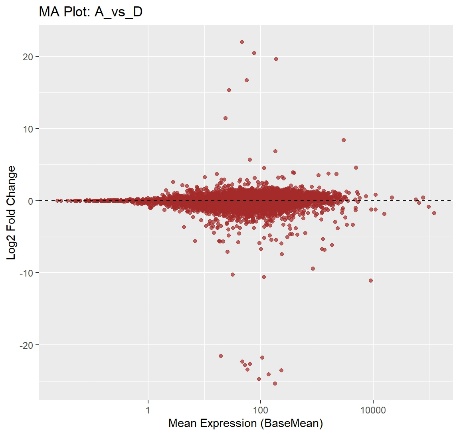


Figure 6: MA plot showing the correlation between log2 fold change (y-axis) and mean expression (x-axis, on a log scale) for every gene. Genes with significant differential expression (adjusted p-value < 0.05) are indicated by red dots, whereas genes that are not significant are indicated by grey dots. No change in expression is indicated by the horizontal dashed line (log2 fold change = 0).

*Histogram*

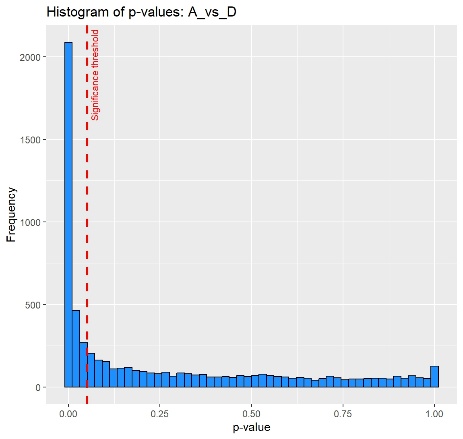


Figure 7: The distribution of statistical significance for each examined gene is displayed in the histogram of p-values. The frequency of genes is displayed on the y-axis, while the p-value is represented on the x-axis. The significance level (p = 0.05) is indicated by a vertical red dashed line. The distribution aids in determining how often the analysis's noteworthy findings are.

*Heatmap*

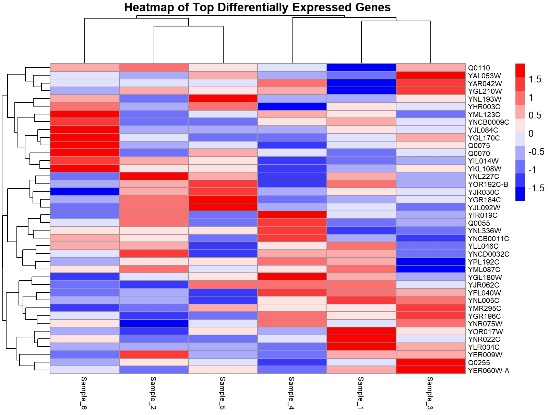
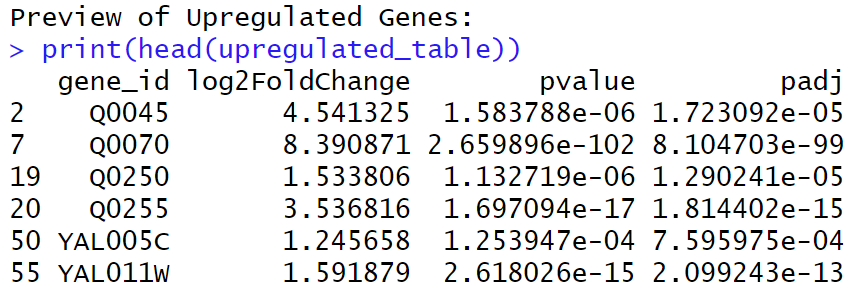
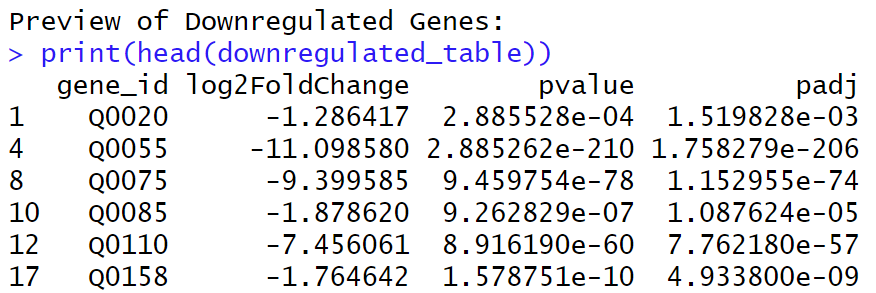


Figure 8: The top differentially expressed genes' expression patterns across various samples and settings are depicted in a heatmap. Conditions are shown on the x-axis, and genes are shown on the y-axis. Normalised expression values are represented by the colour scale, where blue denotes low expression and red denotes high expression. Similar gene expression patterns across samples are grouped by hierarchical clustering.

1. Significant gene list

Below mentioned are a preview of upregulated and downregulated genes based on the p-values, adjusted p-values and fold changes:





1. Additional analysis

*Impact of Differentially Expressed Genes on Biological Pathways*

* Only through pathway enrichment analysis, which is not clearly obvious in the coding findings, can enrichment in amino acid and lipid biosynthesis be deduced.
* Although it was not specifically shown, it is conceivable that downregulation of DNA repair and cell cycle pathways occurs if certain genes involved in these processes are among the downregulated genes.

*Gene Expression Pattern Clustering*

* Coordinated expression changes are suggested by similar clustering patterns in the heatmap.
* Additional study is necessary to identify specific pathways or gene groups, but unique clusters for A\_vs\_D most likely reflect condition-specific responses.

## Conclusion:

Significant transcriptional alterations under various experimental circumstances were found by the differential gene expression analyses of the A\_vs\_F and A\_vs\_D datasets. 1,448 significant genes (627 upregulated and 821 downregulated) were found using A\_vs\_F, making up 20% of all the genes examined. 1,016 significant genes, or 16% of all genes, were identified by A\_vs\_D (365 upregulated and 651 downregulated).

Most of the genes in both datasets were not significant, and A\_vs\_F had fewer missing adjusted p-values (54) than A\_vs\_D (303). Heatmaps showed distinct hierarchical clustering, revealing transcriptional responses and co-expressed gene groupings unique to each condition. Although broad patterns like transcriptional regulation and stress responses were noted, more research is necessary to identify precise biological mechanisms and pathways. All things considered, the findings show both unique and similar patterns of gene expression in the two datasets.