**Differential Gene Expression Analysis Report**

Differential gene expression analysis (DGEA) from DESeq was implemented on two groups which were subject to different treatment conditions: AvsD and GvsD. Table 1 and Table 2 presents the summary statistics for all genes from the AvsD and GvsD treatment groups, respectively. For the log2FoldChange data, both treatment groups have a similar mean, standard deviation, minimum and maximum values (Table 1 and Table 2). However, AvsD has an extremely low adjusted pvalue of 1.758e-206 compared to GvsD with a minimum adjusted pvalue of 2.836e-43 (Table 1 and 2, respectively). This implies the presence of a very significantly expressed gene in AvsD. Significant genes being expressed are associated with values that satisfy the following thresholds of adjusted p-value < 0.05 and log2foldchange >= +/-2.00. From 6094 genes (Table 1), 45 upregulated and 148 downregulated genes were identified for the AvsD treatment group. Whereas out of 6094 genes for the GvsD treatment group (Table 2), there were 84 upregulated and 349 downregulated genes.

Table 1 – Statistical summary for the AvsD treatment condition. Values rounded to 3 decimal places.

|  |  |  |  |
| --- | --- | --- | --- |
|  | log2FoldChange | pvalue | Adjusted pvalue |
| Count | 6094 | 6094 | 6094 |
| Mean | -0.105 | 0.235 | 0.295 |
| Standard deviation | 1.429 | 0.298 | 0.324 |
| Min | -25.329 | 2.885e-210 | 1.758e-206 |
| 25% | -0.503 | 0.002 | 0.007 |
| 50% | -0.004 | 0.075 | 0.149 |
| 75% | 0.440 | 0.415 | 0.554 |
| Max | 22.003 | 1.000 | 1.000 |

Table 2 – Statistical summary for the GvsD treatment condition. Values rounded to 3 decimal places.

|  |  |  |  |
| --- | --- | --- | --- |
|  | log2FoldChange | pvalue | Adjusted pvalue |
| Count | 6343 | 6343 | 6343 |
| Mean | -0.177 | 0.2071 | 0.247 |
| Standard deviation | 1.472 | 0.300 | 0.326 |
| Min | -26.055 | 4.472e-47 | 2.836e-43 |
| 25% | -0.771 | 5.503e-05 | 2.200e-04 |
| 50% | -0.046 | 0.026 | 0.051 |
| 75% | 0.573 | 0.336 | 0.448 |
| Max | 20.497 | 1.000 | 1.000 |

A volcano plot was generated for both treatment groups (Figure 1 and Figure 2). Extreme points at either ends of the x-axis are promising candidates for further investigation due to large amounts change in gene expression. However, these values that were below the significant threshold set to identify to top significant genes so were excluded from this category.

Gene Q0055 may be a possible outlier since it has extremely high expression compared to other genes (Figure 1). This would skew the statistical summary generated earlier (Table 1) if this gene is a true outlier. The experiment should be repeated to investigate whether this gene is a true outlier or is an important gene of interest within the biological system (Figure 1). However, this potential outlier has been included for all downstream analysis.

A diagram of a volcano

Description automatically generated with medium confidence

Figure 1 – Volcano plot of AvsD (treatment condition 1). Points where genes are significantly upregulated are coloured in orange. Points where genes are significantly downregulated are coloured in blue. Any genes that are non-significant are coloured in grey. Vertical dashed lines cut the x-axis at log2foldchange >= +/-2.00 where significant up-regulated or down-regulated genes exist. The horizontal dashed line cuts the y-axis at a threshold set to distinguish the top significantly expressed genes.

Compared to AvsD (Figure 1), the GvsD condition does not display outliers and presents fewer top significantly expressed genes (Figure 2).

A graph of a graph with blue and orange dots

Description automatically generated

Figure 2 – Volcano plot of GvD (treatment condition 2). Points where genes are significantly upregulated are coloured in orange. Points where genes are significantly downregulated are coloured in blue. Any genes that are non-significant are coloured in grey. Vertical dashed lines cut the x-axis at log2foldchange >= +/-2.00 where significant up-regulated or down-regulated genes exist. The horizontal dashed line cuts the y-axis at a threshold set to distinguish the top significantly expressed genes.

The MA plot for AvsD (condition 1) is shown in Figure 3. There is a large spread of genes across the x-axis however most of these genes do not have significant log2FoldChange values. There appears to be some clustering of upregulated genes and a tighter clustering of downregulated genes (Figure 3). The MA plot for GvsD (condition 2) is shown in Figure 4. Like the AvsD condition, the GvsD condition presents a large spread of genes across the x-axis where most of these genes do not have significant log2FoldChange values. Compared to AvsD, the GvsD spread is more centralised along the log mean expression. There does not appear to be any clustering of upregulated genes but clustering of downregulated genes is present (Figure 4).

A graph of dots and lines

Description automatically generated with medium confidence

Figure 3 – MA plot of AvD (condition 1). Points where genes are significantly upregulated are coloured in orange. Points where genes are significantly downregulated are coloured in blue. Any genes that are non-significant are coloured in grey.

A graph showing a line of blue and orange dots

Description automatically generated

Figure 4 – MA plot of GvD (condition 2). Points where genes are significantly upregulated are coloured in orange. Points where genes are downregulated are coloured in blue. Any genes that are non-significant are coloured in grey.

Frequencies of the adjusted pvalues for AvsD (condition 1) and GvsD (condition 2) are presented in Figures 5 and 6, respectively. For both conditions, the greatest frequency adjusted pvalues exist between 0 and 0.05, which are considered as significant.

A blue and white graph

Description automatically generated

Figure 5 – Histogram of adjusted p-values of AvD deseq gene data

A blue line with black text

Description automatically generated

Figure 6 – Histogram of adjusted p-values of GvD deseq gene data

A heatmap of the log2FoldChange, based on the genes presented in Table 3, for both conditions is displayed in Figure 7. Condition 1 (AvsD) has more extreme differentially expressed genes compared to condition 2 (GvsD). The genes that are most upregulated and differentially expressed for AvsD (condition 1) are Q0070 and YGL170C. Contrastingly, GvsD has Q0070 downregulated and YGL170C slightly upregulated. The genes that are most downregulated from this list of top differentially expressed genes for AvsD (condition 1) are Q0055, Q0075, and Q0110. These genes are also downregulated in GvsD but by a lesser amount.

The genes that are most upregulated from this list of top differentially expressed genes for GvsD (condition 2) are YOR127W, YJL082W and YKL108W (Figure 7). Interestingly, these genes are upregulated by similar amount in AvsD except for gene YKL108W where it is significantly downregulated. The genes that are most downregulated for GvsD (condition 2) are YNR075W, YKL150W, and YDR233C. YNR075W is downregulated by a similar amount in AvsD. However, the genes YKL150W and YDR233C are less downregulated since the log2FoldChange values are slightly higher in AvsD.

A screenshot of a graph

Description automatically generated

Figure 7 – Heatmap of log2FoldChange for significantly expressed genes for at least one condition. Condition 1 = AvsD. Condition 2 = GvsD.

Table 3 presents the log2FoldChange values and pvalues for the top differentially expressed genes, for at least one condition.

Table 3 – The top significantly expressed genes for at least one condition. Values rounded to 3 dp.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gene ID | log2FoldChange | pvalue | padj | Treatment group |
| Q0055 | -11.099 | 2.885e-210 | 1.758e-206 | Condition\_1 |
| Q0070 | 8.391 | 2.660e-102 | 8.105e-99 | Condition\_1 |
| Q0075 | -9.400 | 9.46e-78 | 1.153e-74 | Condition\_1 |
| Q0110 | -7.456 | 8.916e-60 | 7.762e-57 | Condition\_1 |
| YER009W | -5.444 | 2.277e-42 | 1.261e-39 | Condition\_1 |
| YGL170C | 6.822 | 2.939e-37 | 1.378e-34 | Condition\_1 |
| YHR003C | -5.916 | 4.796e-63 | 4.871e-60 | Condition\_1 |
| YKL108W | -6.163 | 5.422e-87 | 8.260e-84 | Condition\_1 |
| YML087C | -5.163 | 1.171e-35 | 4.757e-33 | Condition\_1 |
| YML123C | -5.91 | 3.575e-38 | 1.815e-35 | Condition\_1 |
| YNL193W | -6.81 | 2.333e-93 | 4.739e-90 | Condition\_1 |
| YNL336W | -2.417 | 5.210e-47 | 3.175e-44 | Condition\_1 |
| YNR075W | -4.844 | 1.736e-36 | 7.556e-34 | Condition\_1 |
| YOR192C-B | -5.06 | 4.492e-56 | 3.422e-53 | Condition\_1 |
| YPL192C | -6.693 | 1.110e-49 | 7.518e-47 | Condition\_1 |
| YDR233C | -3.494 | 4.471e-47 | 2.836e-43 | Condition\_2 |
| YFL009W | -3.032 | 1.059e-40 | 1.679e-37 | Condition\_2 |
| YJL092W | 1.744 | 3.118e-36 | 2.825e-33 | Condition\_2 |
| YKL150W | -4.435 | 4.085e-43 | 1.296e-39 | Condition\_2 |
| YLR089C | -3.06 | 5.217e-40 | 6.619e-37 | Condition\_2 |
| YNCP0026W | -3.103 | 3.020e-41 | 6.386e-38 | Condition\_2 |
| YNR075W | -4.66 | 1.085e-34 | 8.601e-32 | Condition\_2 |
| YOR127W | 2.119 | 8.238e-37 | 8.709e-34 | Condition\_2 |

Figure 8 displays a gene clustering analysis on the top differentially expressed genes for both conditions, based on the genes presented in Table 3. Q0070 and YGL170C genes form their own cluster, separate from the other genes which suggests that these genes share similar expression but are unique from the other genes. This is also true for the YJL092W and YOR127W cluster which is most closely related to the Q0070 and YGL170C cluster. Therefore, compared to the large group of top significantly expressed genes (in orange), the separate branch (in green) may contain critical genes involved in a unique process within the biological system (Figure 8).

A graph with orange lines

Description automatically generated

Figure 8 – Hierarchal clustering of significantly expressed genes for at least one condition. Genes are grouped according to their log2FoldChange values.

Overall, condition 1 (AvsD) presents more extreme significantly expressed genes compared to GvsD. However, both treatment conditions did display significant gene expression. Genes for further investigation within their biological system include YJL092W, YOR127W, Q0070 and YGL170C.