**Introduction**

This report presents the results of a differential gene expression analysis comparing two experimental conditions: **A\_vs\_E**，**I\_vs\_E**. The analysis includes the following steps: Data preprocessing and filtering to identify significantly differentially expressed genes. Visualization of results through heatmaps, volcano plots, MA plots, and p-value distributions. Explaining of significant gene expression changes.

**Results**

**1. Volcano Plots**

This Volcano plots (figure1 and figure2) show a total of 994 significantly upregulated genes and 1279 significantly downregulated genes were identified. The distribution of significant genes indicates a broader range of expression changes compared to A\_vs\_E. And the relationship between statistical significance p-value and the magnitude of gene expression change (log2FoldChange). Totally 560 significantly upregulated genes and 916 significantly downregulated genes were identified. Upregulated genes show higher log2FoldChange values (red points), while downregulated genes exhibit lower log2FoldChange values (blue points).

**2. Heatmaps**

Heatmaps (figure 3 and figure 4) were generated for the top 50 most significant genes based on adjusted p-values, highlighting the patterns of upregulation and downregulation in gene expression. The gene in the upper region (red area) showed significant upregulation under experimental condition A. The gene in the lower region (blue area) exhibited significant downregulation under experimental condition E. Samples from experimental conditions A and E clustered into different branches, suggesting clear differences between the conditions. Red genes exhibit a slight upregulation under condition I, whereas blue genes show a slight downregulation under condition E. The color range of the heatmap is narrower compared to A\_vs\_E, indicating a smaller magnitude of log2FoldChange variations.

**3. MA Plots**

MA plots（figure 5 and figure 6） illustrate the relationship between log2FoldChange and mean expression levels (log10 scale). Most genes are distributed around log2FoldChange = 0, with a few outliers showing significant changes. The number of points deviating from 0 is slightly higher in the plot of A\_vs\_E compared to I\_vs\_E, indicating more significant differences in gene expression between the A\_vs\_E conditions.

**4. P-value Distributions**

Histograms of p-values (figure 7 and figure 8) provide information into the distribution of statistical significance across all genes. A clear peak near p-value = 0 indicates a subset of highly significant genes. Multiple number of genes with P-values close to 0, with the most left bar being significantly taller than the others. As the P-value increases, the number of genes progressively diminishes. On the right side of the graph (near 1), there are still some genes, but their count is comparatively low. This distribution indicates many differentially expressed genes between conditions A and E. The number of genes on the left side (with P-value close to 0) remains relatively high, though the bar heights are slightly lower than in A\_vs\_E. As the P-value increases, the number of genes also gradually decreases. In the region of P-value approaching 1, there are more genes compared to A\_vs\_E. It indicates that there is no significant difference in expression between the two conditions.

5. Based on **Log2 Fold Change > 1** and **adjusted P-value < 0.05, as the** generation file :Differential Expression Results.csv shows:

**A\_vs\_E:** the **significant upregulation of gene counts:** 17**, the number of genes significantly down-regulated is** 33**, the average P value is** 1.26e-22**, the mean of Log2 Fold Change is -**2.01**, the range of Log2 Fold Change is between** -11.52and 8.29.

**I\_vs\_E: the significant upregulation of gene counts:** 13**, the number of genes significantly down-regulated is** 37, **the average P value is** 7.19e-28**, the mean of Log2 Fold Change is** -2.3**, the range of Log2 Fold Change is between** -6.39and 3.12**.**

**Discussion**

The comparison between Condition A and Condition E discovers a significant number of differentially expressed genes, showing the substantial biological differences. In the MA plots, demonstrating a substantial change from zero, with numerous genes showing marked upregulation and downregulation. In the P value plots, many genes exhibit practically low P values, highlighting the statistically significant trait of their differential expression. The heatmaps further express this by clearly clustering genes that are significantly upregulated (red) and downregulated (blue). These data suggest that the activation pathways in Condition A likely differ markedly from those observed in Condition E. The number of differentially expressed genes between condition I and condition E is relatively modest. In the MA plots show that many gene variations are close to zero, indicating a more similar expression pattern between the two conditions. In the P-value distribution plot, while there are still some genes with low P-values, their overall count is notably less than that observed in A\_vs\_E. The heatmap reveals a smaller range of expression differences, suggesting that the regulatory changes between conditions I and E are more constrained.

**A\_vs\_E:** The P values are surrounding 0 expresses that there are more significantly differentially expressed genes between conditions A and E, the average of P value is 1.26e-22, the mean of Log2 Fold Change is **-**2.01, the range of Log2 Fold Change is between -11.52 and 8.29. The symmetry of gene expression changes is excellent, and the number of upregulated and downregulated genes is balanced.

**I\_vs\_E:** The significance level of the P value was low entirely, the number of Log2 Fold Change close to 0 for most of genes，the mean of P value is 7.19e-28 , the mean of Log2 Fold Change is-2.3, the range of Log2 Fold Change is between -6.39and 3.12which showing that the gene expression patterns for conditions I and E are more similar.

**Conclusion**

Comparisons of A\_vs\_E revealed significant biological differences, showing upregulation and downregulation changes in many genes, which showing significant differences in activation or repression pathway. However, the comparison of I\_vs\_E showing limited expression patterns and a lower number of significant genes, which reflecting a more similar regulatory mechanism between the two conditions.

**Supplementary Files**

**Figure 1. volcano plot A\_E**图表, 散点图

描述已自动生成

**Figure 2. volcano plot I\_E**

图表, 散点图

描述已自动生成

Figure 3. Heatmaps A\_vs\_E 图片包含 图表

描述已自动生成

Figure 4. Heatmaps I\_vs\_E

图片包含 表格

描述已自动生成

Figure 5. MA Plots A\_vs\_E

图表, 散点图

描述已自动生成

Figure 6. MA plots I\_vs\_E

图表, 散点图

描述已自动生成

Figure 7. P-value Distributions A\_vs\_E

图表, 直方图

描述已自动生成

Figure 8.P-value Distributions I\_vs\_E

图表, 直方图

描述已自动生成

**Significant Gene Lists:**

[**Differential\_Expression\_Results.csv**](file:///Users/phoebe/Desktop/LIFE4138-GENE%20COURSEWORK/Differential_Expression_Results.csv)

[Top\_50\_Genes\_A\_vs\_E.csv](file:///Users/phoebe/Desktop/LIFE4138-GENE%20COURSEWORK/Top_50_Genes_A_vs_E.csv)

[Top\_50\_Genes\_I\_vs\_E.csv](file:///Users/phoebe/Desktop/LIFE4138-GENE%20COURSEWORK/Top_50_Genes_I_vs_E.csv)