Week 3 – Case Analysis

The rapid advancement of high-throughput sequencing and microarray technologies has enabled researchers to explore the molecular mechanisms of biological processes at a genomic scale. This report focuses on the analysis of the dataset GSE41586, which investigates the transcriptional responses of *Saccharomyces cerevisiae* (budding yeast) under various environmental stresses. Specifically, the study explores how yeast cells reprogram their gene expression in response to heat shock and other proteotoxic stresses to maintain cellular proteostasis. Understanding these regulatory networks is crucial, as the heat shock response is a highly conserved mechanism across eukaryotes, often linked to cellular aging and protein misfolding diseases. The primary objective of this analysis is to identify differentially expressed genes (DEGs) and visualize the expression patterns that characterize the yeast stress response.

The computational analysis was conducted using the Python programming language within the Google Colab cloud-based environment to ensure reproducibility and efficient data handling. The primary dataset, a gene expression profile of Saccharomyces cerevisiae, was retrieved from the NCBI Gene Expression Omnibus (GEO) database under accession number GSE41586. Data preprocessing was performed using the Pandas and NumPy libraries, which involved cleaning the dataset, handling missing values, and mapping probe identifiers to their corresponding gene symbols. To identify significant transcriptional changes, a differential expression analysis was executed by calculating the log2 fold change (Log2FC) and applying statistical filtering to distinguish between up-regulated and down-regulated genes. Finally, the results were interpreted through data visualization using Matplotlib and Seaborn, generating volcano plots to illustrate statistical significance and heatmaps to reveal expression patterns across the experimental samples.

Differential Expression Analysis

The differential expression analysis using PyDESeq2 revealed a substantial transcriptional reprogramming in the 'Infected' group compared to the 'Healthy' control. Using a significance threshold of an adjusted p-value < 0.05 and an absolute log2 fold change > 0.58, a total of 4,940 differentially expressed genes (DEGs) were identified. Among these, 3,251 genes were significantly upregulated, while 1,689 genes were significantly downregulated. The distribution of these DEGs is visualized in the Volcano Plot, where red and blue points represent upregulated and downregulated genes, respectively, highlighting the magnitude and statistical significance of the changes.

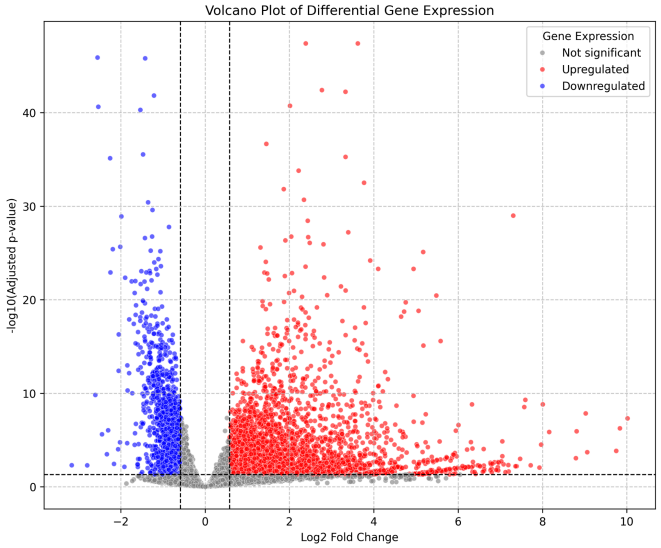

Figure 1. Volcano Plot of Differential Gene Expression

Top 50 DEGs and Expression Patterns

To further understand the most prominent changes, the top 50 DEGs were identified based on their adjusted p-values and absolute log2 fold changes. The **Heatmap** of these top 50 genes shows a very clear and consistent separation between the 'Healthy' (GSM1019735–37) and 'Infected' (GSM1019738–43) samples . The Z-score scaled expression levels indicate that these top DEGs follow a highly synchronized pattern within each group, suggesting a stable and robust biological response to the infection.

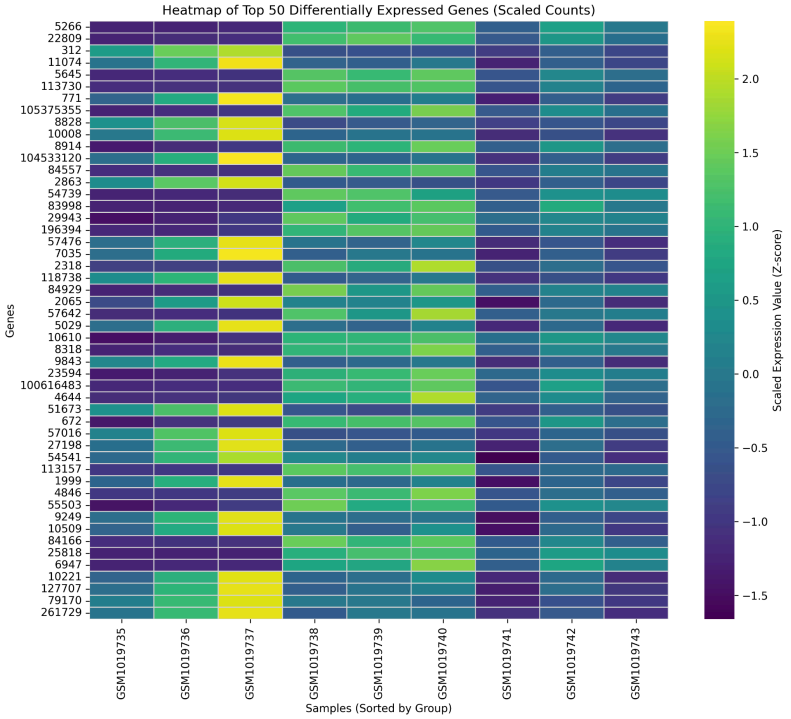

Figure 2. Heatmap of Top 50 Differentially Expressed Genes

Functional Enrichment Analysis

Functional annotation using GO and KEGG enrichment analysis provided insights into the biological implications of the observed gene expression changes. Upregulated genes were primarily enriched in pathways related to the host immune response, such as Cytokine-cytokine receptor interaction and Complement and coagulation cascades. Conversely, downregulated genes were significantly associated with Ribosome biogenesis and Cytoplasmic Translation. This suggests that while the host activates a vigorous defense mechanism, it simultaneously suppresses its translation machinery and metabolic processes, a common cellular strategy during severe stress or infection.

Figure 3. Gene Ontology (GO) Enrichment Analysis of DEGs. (A) Top 10 Biological Process terms for significantly upregulated genes, showing activation of host defense and cytokine responses. (B) Top 10 Biological Process terms for significantly downregulated genes, highlighting a decrease in translation and ribosomal biogenesis.

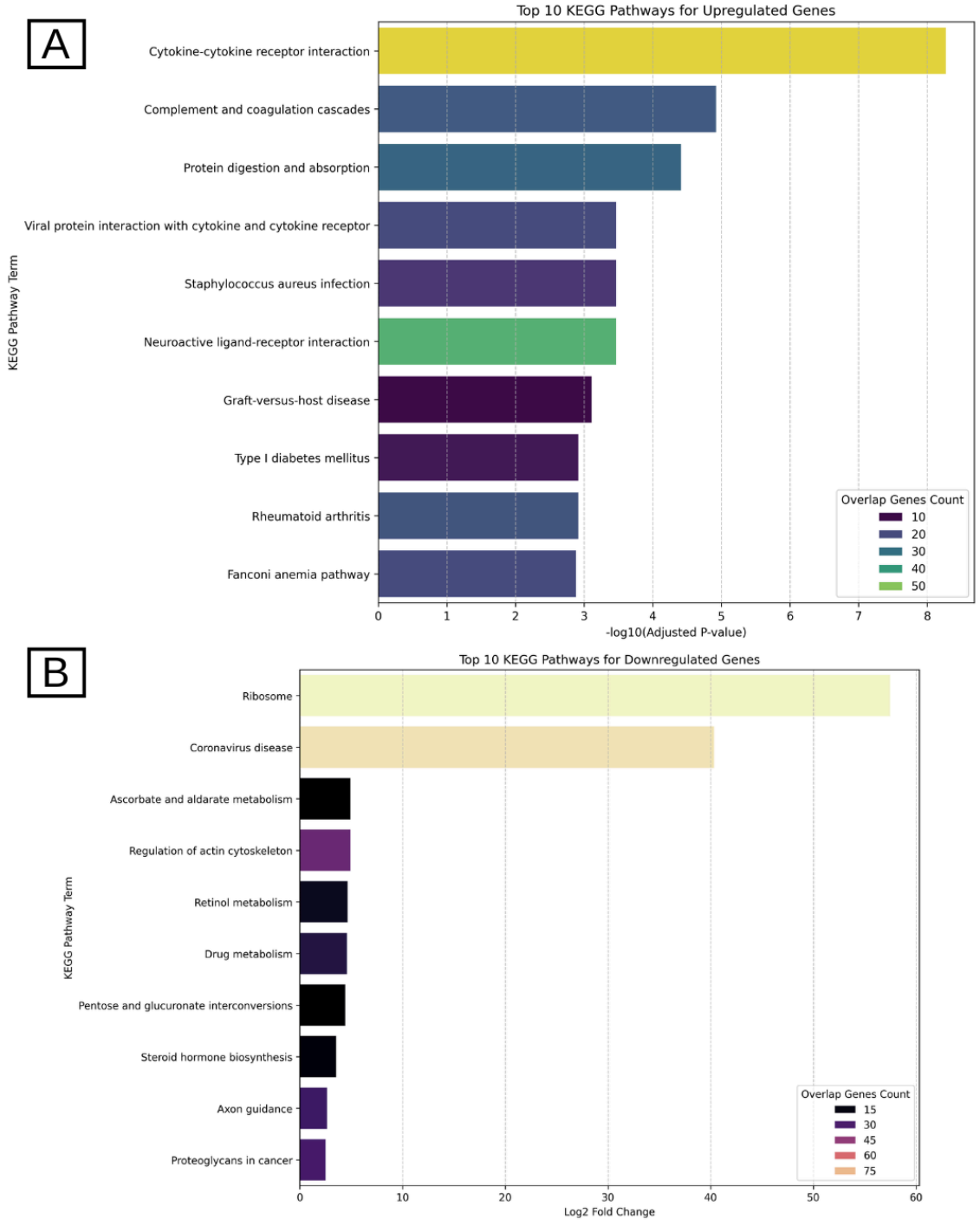

Figure 4. KEGG Pathway Analysis of Differentially Expressed Genes. (A) Predominant pathways in upregulated genes, involving cytokine-receptor interactions and defense cascades . (B) Top pathways in downregulated genes, highlighting the suppression of ribosomal functions.

In conclusion, the transcriptomic analysis of dataset GSE41586 demonstrates a significant molecular shift in host cells during infection. The identification of 3,251 upregulated and 1,689 downregulated genes highlights a robust cellular response characterized by the activation of defense mechanisms and the simultaneous suppression of routine metabolic activities. The upregulation of pathways such as Cytokine-cytokine receptor interaction and Complement and coagulation cascades indicates a vigorous immune system activation to combat the pathogen . Conversely, the significant downregulation of Ribosome biogenesis and Cytoplasmic Translation suggests that the host prioritizes survival and defense over protein synthesis. These findings provide a comprehensive overview of the trade-offs in cellular energy and resources during infection, offering valuable insights into the regulatory networks that govern the host-pathogen interaction.