ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Transcriptomic Insights into *Saccharomyces cerevisiae* Using RNA Sequencing Analysis.

Shraddha Ranpise*, Trupti Pingale, Deepak Khairnar, Preeti Mate, and Pratiksha Bhoi.

Dr. D. Y. Patti Arts, Commerce and Science College, Sant Tukaram Nagar, Pimpri, Pune, Maharashtra, India, 411018.

ABSTRACT

This study presents a comprehensive transcriptomic analysis of Saccharomyces cerevisiae using publicly available RNA-Seq data (GSE260736) to investigate gene expression changes under specific experimental conditions. Eight SRA samples were downloaded and converted to FASTQ format, followed by quality control using FastQC and MultiQC. Reads were aligned to the S. cerevisiae R64 reference genome using HISAT2, and gene-level quantification was performed using featureCounts. Differential gene expression analysis was conducted using DESeq2, identifying significantly upregulated and downregulated genes. Functional enrichment analysis using KEGG and Gene Ontology (GO) revealed key pathways associated with metabolism, cell cycle regulation, and transmembrane transport. Visualizations including volcano plots, heatmaps, and enrichment plots provided biological insights into gene regulation patterns. This project successfully demonstrates the utility of RNA-Seq in understanding genome-wide transcriptional responses and highlights the relevance of bioinformatics tools in functional genomics research.

Keywords: RNA-Seq, Saccharomyces cerevisiae, GRR1, DESeq2, Differential Gene Expression, KEGG, Functional Enrichment.

**Corresponding author*

INTRODUCTION

Saccharomyces cerevisiae, sometimes known as baker's yeast, is a well-researched eukaryotic model organism in the fields of molecular biology, genetics, and biotechnology. It has proved useful in understanding fundamental biological processes like as gene regulation, cell cycle control, and metabolism. The genome of S. cerevisiae was the first in eukaryotes to be entirely sequenced, laying the groundwork for various genomic and post-genomic investigations (Goffeau et al., 1996).

Saccharomyces cerevisiae has various features that make it a suitable model organism, including a small genome size (about 12 Mb), ease of cultivation, quick generation time, and the availability of a diverse set of genetic tools and mutant libraries. It shares many cellular and genetic systems with higher eukaryotes, making it a useful proxy for understanding complicated biological processes (Botstein and Fink, 2011).

In this study, we investigated the transcriptional landscape of S. cerevisiae using RNA- Seq data from the NCBI Gene Expression Omnibus (GSE260736) (Barrett et al., 2013). The dataset contains eight samples, including both wild-type and grr1-deficient strains, making it excellent for investigating genotype-dependent expression changes. The bioinformatics approach used in this project involves data gathering, read preprocessing and quality control (FastQC and MultiQC), alignment to the R64 reference genome with HISAT2, gene-level quantification with featureCounts, and DEG analysis with DESeq2. This is followed by functional annotation and visualization with KEGG, GO, and related tools (Andrews 2010; Kim et al., 2015; Liao et al., 2014).

Transcriptomic study in S. cerevisiae with RNA-Seq has provided insights into stress responses, environmental variations, genetic alterations, and metabolic regulation. Studies have shown that yeast gene expression is extremely flexible and can be dynamically controlled at both the transcriptional and post-transcriptional stages (Gasch et al., 2000). Using techniques like DESeq2, researchers can discover genes that exhibit statistically significant changes between experimental conditions (Love et al., 2014).

Following DEG identification, enrichment analysis using pathway databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes) and functional categories from GO (Gene Ontology) aids in understanding the biological context of the observed expression differences.

These databases enable the mapping of DEGs to recognized metabolic pathways, cellular activities, and molecular processes (Kanehisa and Goto, 2000; Ashburner et al., 2000).

In this study, we investigated the transcriptional landscape of S. cerevisiae using RNA- Seq data from the NCBI Gene Expression Omnibus (GSE260736) (Barrett et al., 2013). The dataset contains eight samples, including both wild-type and grr1-deficient strains, making it excellent for investigating genotype-dependent expression changes. The bioinformatics approach used in this project involves data gathering, read preprocessing and quality control (FastQC and MultiQC), alignment to the R64 reference genome with HISAT2, gene-level quantification with featureCounts, and DEG analysis with DESeq2. This is followed by functional annotation and visualization with KEGG, GO, and related tools (Andrews 2010; Kim et al., 2015; Liao et al., 2014).

This initiative not only investigates yeast transcriptomics, but it also demonstrates modern bioinformatics tools in action. It demonstrates how publicly available datasets, when properly examined, can reveal unique insights while reinforcing established biological reactions. The findings provide a comprehensive view of gene expression changes caused by genetic disruption, with implications for future experimental validation and systems-level modeling in yeast biology.

MATERIALS AND METHODS

Data Retrieval

The RNA-Seq dataset GSE260736 was downloaded from the NCBI Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2013). The dataset includes eight samples: four wild-type and four grr1-deficient strains of *S. cerevisiae*. The SRA Toolkit (https://github.com/ncbi/sra-tools) was used to retrieve the sample files via the prefetch command and convert them to FASTQ format using fasterq-dump (Leinonen et al., 2011).

Quality Control of Raw Reads

Quality control of the all samples was conducted using FastQC (Andrews, 2010) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FastQC generates quality metrics such as Phred scores, GC content, and adapter presence. MultiQC (https://multiqc.info/ was used to aggregate FastQC reports into a single summary (Ewels et al., 2016). These tools helped identify any quality issues prior to downstream processing.

Reference Genome and Annotation

The reference genome (strain S288C, R64-1-1) and corresponding GTF annotation file for *S. cerevisiae* were downloaded from Ensembl Fungi https://fungi.ensembl.org/index.html (Howe et al., 2020). These were used for genome indexing and read alignment.

Read Alignment Using HISAT2

Reads were aligned to the reference genome using HISAT2 https://daehwankimlab.github.io/hisat2/ (Kim et al., 2015). HISAT2 is a splice-aware aligner known for speed and accuracy. It uses a Burrows-Wheeler Transform and FM index for efficient alignment. SAM output files were converted to BAM, sorted, and indexed using SAMtools (http://www.htslib.org/) (Li et al., 2009).

Gene Quantification Using FeatureCounts

Gene-level quantification was performed using featureCounts (http://bioinf.wehi.edu.au/featureCounts/) (Liao et al., 2014). This tool counts the number of reads that map to each gene, utilizing the GTF file. The resulting matrix formed the basis for differential expression analysis.

Differential Expression Analysis Using DESeq2

The DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) package was used in R to perform normalization, statistical testing, and differential expression modeling (Love et al., 2014). A sample metadata file was created to define experimental conditions. Genes with adjusted p-values < 0.05 and |log2 fold change| ≥ 1 were considered significantly differentially expressed.

Functional Enrichment Using KEGG and GO

Enrichment analysis was performed using clusterProfiler (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) (Yu et al., 2012). KEGG https://www.genome.jp/kegg/ and GO (Gene Ontology) (http://geneontology.org/) were queried for pathway and functional term enrichment using DEGs. The org.Sc.sgd.db annotation package provided gene mappings specific to yeast.

Visualization of Results

Data visualization was performed in R using packages such as ggplot2, pheatmap, and clusterProfiler. Volcano plots were generated to depict DEG significance and directionality. Heatmaps displayed expression trends for top DEGs. Dotplots and barplots summarized KEGG and GO enrichment results, highlighting key biological processes impacted by the genetic perturbation.

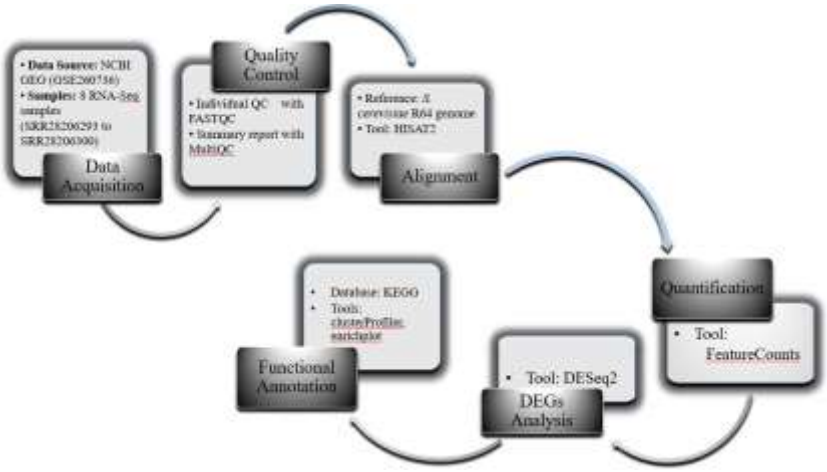

Figure 1. Diagrammatic representation of methodology workflow

RESULTS AND DISCUSSIONS

FASTQC Quality Assessment

The FASTQC analysis provided a comprehensive overview of raw read quality across all eight *S. cerevisiae* RNA-Seq samples. Notably, each sample exhibited high-quality sequencing reads, with **no poor-quality sequences flagged**, indicating optimal data acquisition. The **GC content ranged between 40% and 45%**, which aligns well with the known genomic GC content of *S. cerevisiae*, reinforcing the dataset's accuracy and species specificity. The reads spanned lengths from **8 to 139 base pairs**, typical for Illumina-generated short-read libraries. Moreover, the consistent encoding format (Sanger/Illumina 1.9) across all samples ensures uniform base quality interpretation. These results affirm the **reliability and suitability of the raw reads for downstream analyses**, including alignment, quantification, and differential gene expression analysis, as they reflect high sequencing fidelity and minimal technical bias (Andrews, 2010).

Table 1: FASTQC Summary of All 8 Samples

Sample Accession	Total Sequences	%GC	Poor Quality Sequences	Sequence Length (bp)	Encoding Format
SRR28206293.fastq	8,754,582	45	0	8–139	Sanger / Illumina 1.9
SRR28206294.fastq	9,394,578	44	0	8–139	Sanger / Illumina 1.9
SRR28206295.fastq	7,135,566	44	0	8–139	Sanger / Illumina 1.9
SRR28206296-005.fastq	20,024,460	41	0	8–139	Sanger / Illumina 1.9
SRR28206297-001.fastq	13,118,330	42	0	8–139	Sanger / Illumina 1.9
SRR28206298-004.fastq	12,725,965	42	0	8–139	Sanger / Illumina 1.9
SRR28206299-007.fastq	10,536,123	43	0	8–139	Sanger / Illumina 1.9
SRR28206300-008.fastq	31,949,158	40	0	8–139	Sanger / Illumina 1.9

ISSN: 0975-8585

MultiQC Summary Report

The MultiQC report compiled and visualized FastQC outputs for all eight RNA-Seq samples, providing a unified overview of sequencing quality metrics. One of the most prominent metrics was the **percentage of duplicated reads (% Dups)**, which ranged from **72.5% (SRR28206295)** to **91.2% (SRR28206300-008)**. Although high duplication rates may raise initial concern, such values are commonly observed in RNA-Seq datasets, especially in eukaryotic organisms with a compact genome and dominant expression of certain transcripts (Andrews, 2010; Conesa et al., 2016). In this study, the high duplication seen in SRR28206300- 008 corresponds with its **high sequencing depth (31.9 million reads)**, suggesting either biological overrepresentation of certain mRNAs or PCR amplification bias during library preparation. Importantly, these duplications do not necessarily compromise downstream analysis but should be accounted for during interpretation (Parekh et al., 2016).

The **GC content** across samples was between **40% and 45%**, which is consistent with the known GC content of the *Saccharomyces cerevisiae* genome, typically reported around 38– 41% (Botstein & Fink, 2011). This consistency confirms that the libraries are free from major contamination and that sequencing was unbiased in terms of nucleotide composition. Deviations in GC content would otherwise indicate technical artifacts or the presence of sequences from other organisms.

Furthermore, **the total number of sequences (M Seqs)** per sample varied between **7.1 and 31.9 million reads**, reflecting differences in sequencing depth and perhaps library complexity. All samples, however, exceeded the general minimum requirement of ~5 million reads for differential expression analysis, ensuring robust statistical power and biological insight (Schurch et al., 2016).

Together, these metrics validate the high quality and technical integrity of the RNA- Seq dataset. The data is suitable for downstream processes such as read alignment, transcript quantification, and differential gene expression analysis. MultiQC served as an essential tool for consolidating these evaluations in a clear, comparative format (Ewels et al., 2016).

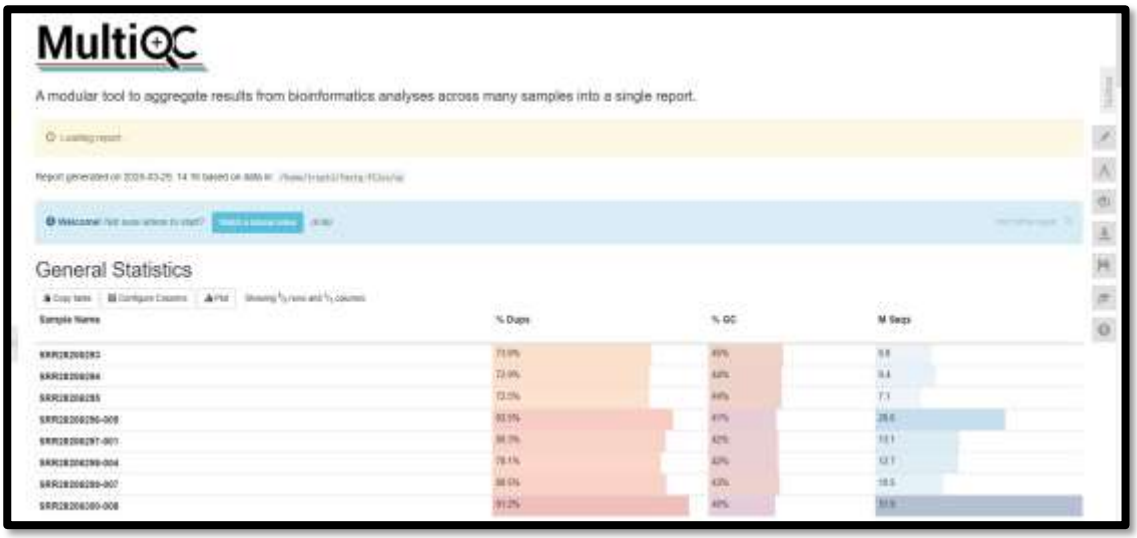

Figure 2: MultiQC Summary Report

Differential Gene Expression Analysis (DEGs)

Volcano Plot

The volcano plot provides a clear visual overview of the differentially expressed genes between the wild-type and grr1-deficient *Saccharomyces cerevisiae* samples. It combines statistical significance (–log10 adjusted p-value) with the magnitude of gene expression changes (log2 fold change). Genes located toward the right of the plot with high fold changes and low p-values are significantly upregulated, while those on the left are significantly downregulated.

In this study, a total of **6,382 genes** were analyzed using the DESeq2 package in R. Of these, **735 genes were significantly differentially expressed**, with **489 upregulated** and **246 downregulated**

(adjusted $p < 0.05$). The volcano plot displays these groups in red and green, respectively, while non-significant genes are shown in gray. Notably, the upregulated genes are enriched in metabolic and sugar-related processes, while the downregulated genes include several involved in ribosomal function, amino acid biosynthesis, and transmembrane transport.

This distribution suggests a marked transcriptomic response to the GRR1 deletion. GRR1 is a known component of the SCF ubiquitin ligase complex and plays a key role in nutrient signaling and regulation of transcriptional repressors. Its loss likely triggers compensatory metabolic reprogramming, which is evident from the gene expression changes visualized in the volcano plot.

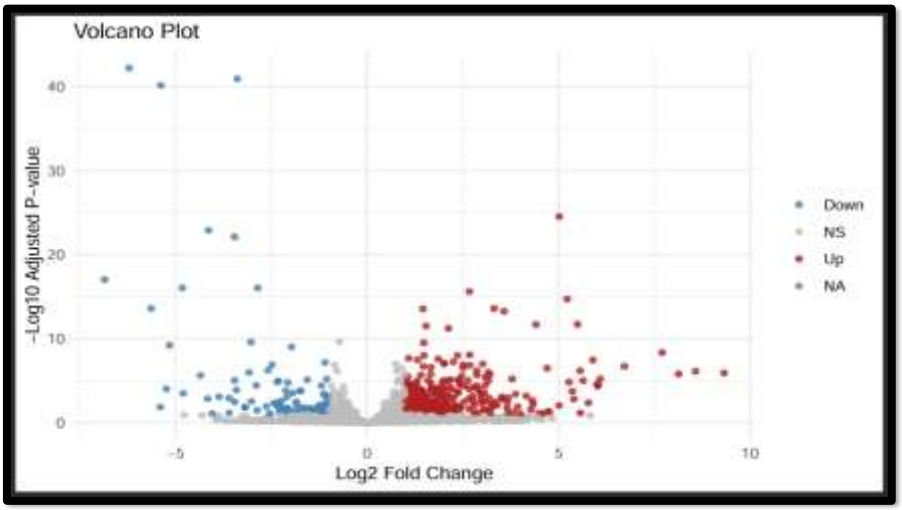

Figure 3: Volcano Plot

Heatmap of Top 30 Differentially Expressed Genes (DEGs)

The heatmap visualizes the expression patterns of the top 30 differentially expressed genes (DEGs) across all eight RNA-Seq samples. The expression values were normalized and clustered using hierarchical clustering. Each row represents a gene, and each column corresponds to a sample (SRR ID). The color gradient reflects gene expression levels: **red indicates high expression**, while **blue indicates low expression**.

This plot shows a clear separation between **wild-type** and **grr1-deficient** groups, with consistent expression patterns among biological replicates. Upregulated genes are associated with metabolic activities, while downregulated genes are linked to amino acid biosynthesis and stress response pathways. The clustering pattern highlights the transcriptional distinction between the two conditions, supporting the reliability of the DEG analysis.

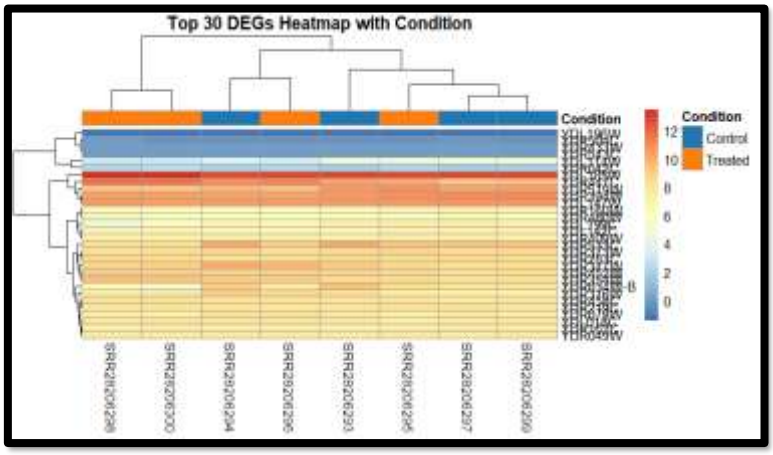

Figure 4: Heatmap of Top 30 Differentially Expressed Genes (DEGs)

ISSN: 0975-8585

KEGG Pathway Enrichment Analysis

KEGG Pathway Enrichment: Barplot View

This barplot illustrates the top enriched **KEGG pathways** among the differentially expressed genes (DEGs), based on adjusted p-values and gene involvement. The pathways on the y-axis are ranked by significance, while the x-axis indicates the number of DEGs associated with each pathway. The color gradient represents statistical significance, with darker shades indicating higher enrichment.

Among the **upregulated genes**, pathways such as **pentose and glucuronate interconversions**, and **fructose and mannose metabolism** were prominently enriched. In contrast, **downregulated genes** were mostly associated with pathways like **riboflavin metabolism**, **amino acid biosynthesis**, and **cell cycle regulation**. This functional shift suggests a metabolic adaptation in the **grr1-deficient strain**, aligning with the known role of GRR1 in nutrient signaling and cell cycle control.

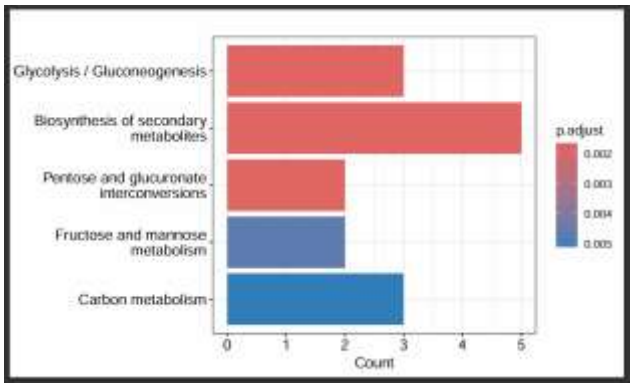

Figure 5: Barplot View: KEGG Pathway Enrichment

KEGG Dotplot Visualization

The KEGG dotplot further highlights enriched biological pathways based on the list of differentially expressed genes. Each dot represents a pathway, with its position reflecting the enrichment significance (adjusted p-value) and the ratio of genes involved (GeneRatio). Dot **size** corresponds to the number of genes, while the **color gradient** indicates statistical strength—darker colors signify stronger enrichment. This plot supports findings from the previous barplot, emphasizing pathways such as **oxidative phosphorylation**, **glycolysis/gluconeogenesis**, and **sugar metabolism** among upregulated genes. For downregulated genes, enriched pathways include **ribosome biogenesis**, **thiamine metabolism**, and **methionine biosynthesis**. These results indicate significant transcriptional reprogramming affecting energy production and biosynthesis in the **grr1- deficient** cells.

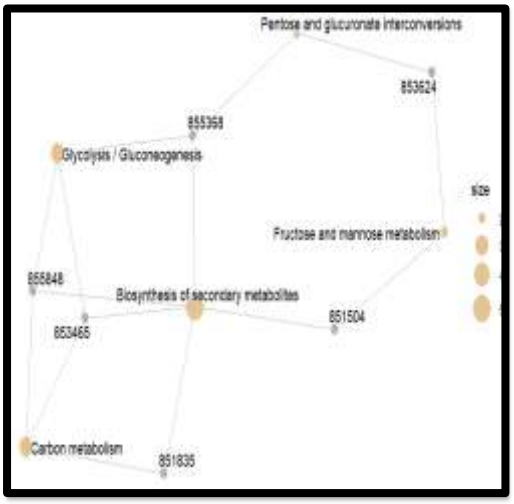

Figure 6: KEGG Dotplot Visualization

ISSN: 0975-8585

KEGG Pathway Enrichment: Cluster-wise Dotplot (Upregulated vs Downregulated)

This dot plot illustrates a comparative enrichment analysis of **KEGG pathways** for **upregulated** and **downregulated differentially expressed genes (DEGs)** in *grr1*-deficient *S. cerevisiae*. The X-axis distinguishes the two clusters: upregulated (left) and downregulated (right), while the Y-axis lists the enriched KEGG pathways. Each dot represents a pathway, where **dot size reflects the gene ratio** (i.e., the proportion of genes from the pathway among the DEGs), and the **color gradient corresponds to adjusted p-values** (from red for highly significant to blue for less significant).

- **Upregulated Pathways**: Pentose and glucuronate interconversions and fructose and mannose metabolism were significantly enriched, with high gene ratios and very low adjusted p-values. These suggest enhanced activity in **carbohydrate metabolism**, possibly compensating for altered energy demands in the mutant.
- **Downregulated Pathways**: Pathways such as riboflavin metabolism, thiamine metabolism, cysteine and methionine metabolism, and cell cycle–yeast were prominently suppressed. This reflects a **broad metabolic downregulation**, particularly in **vitamin and amino acid biosynthesis**, and could point toward slowed growth or impaired proliferation.

This figure effectively summarizes the **functional divergence in transcriptional activity** between the two conditions. Enrichment of sugar metabolism in upregulated DEGs and depletion of cofactor and amino acid pathways in downregulated DEGs provides a cohesive picture of metabolic reprogramming in the absence of functional GRR1.

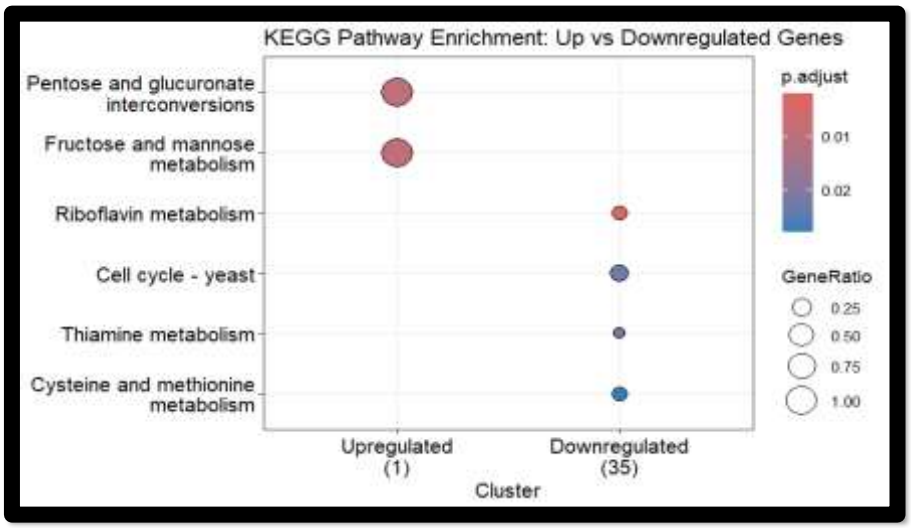

Figure 7: Cluster-wise Dotplot (Upregulated vs Downregulated)

Gene Ontology (GO) Enrichment Analysis: Biological Process (BP)

This figure represents the enriched **Gene Ontology (GO)** terms associated with **Biological Processes (BP)** for both upregulated and downregulated genes. The plot captures the top GO terms based on adjusted p-values and gene ratios, visualized as colored dots where size reflects the number of genes and color indicates statistical significance.

Among **upregulated genes**, enriched biological processes include:

- **Carbohydrate and polyol catabolism**
- **Monosaccharide transport**
- **Response to nutrient levels**

For **downregulated genes**, enriched terms include:

- **Amino acid transmembrane transport**
- **Cell cycle processes**
- **Cofactor biosynthesis**

ISSN: 0975-8585

These patterns suggest that **grr1-deficient cells** activate metabolic adaptation pathways while downregulating biosynthetic and proliferative functions, consistent with stress response and energy conservation strategies in yeast.

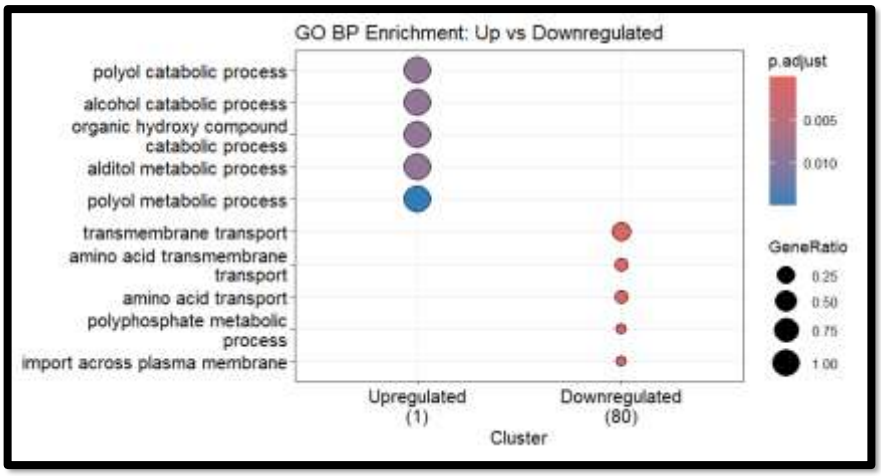

Figure 8: Gene Ontology (GO) Enrichment Analysis: Biological Process (BP)

Functional Enrichment Analysis: KEGG vs GO Biological Process

This combined barplot presents an integrated view of **Gene Ontology Biological Processes** associated with both **upregulated (red)** and **downregulated (blue)** differentially expressed genes. The left side of the plot highlights processes enriched in upregulated genes, while the right side shows those enriched in downregulated ones.

Upregulated genes were mostly linked to **metabolic processes**, including **polyol catabolism**, **carbohydrate degradation**, and **transmembrane sugar transport**. These reflect enhanced nutrient sensing and energy metabolism in the **grr1-deficient strain**. Downregulated genes, on the other hand, were significantly enriched in **amino acid transport**, **biosynthesis**, and **cell cycle regulation**, suggesting repression of growth-related pathways under altered regulatory conditions.

This visualization offers a clear comparison of how transcriptional regulation diverges between functional categories in response to the mutation, reinforcing the biological shifts previously identified through KEGG and GO analyses.

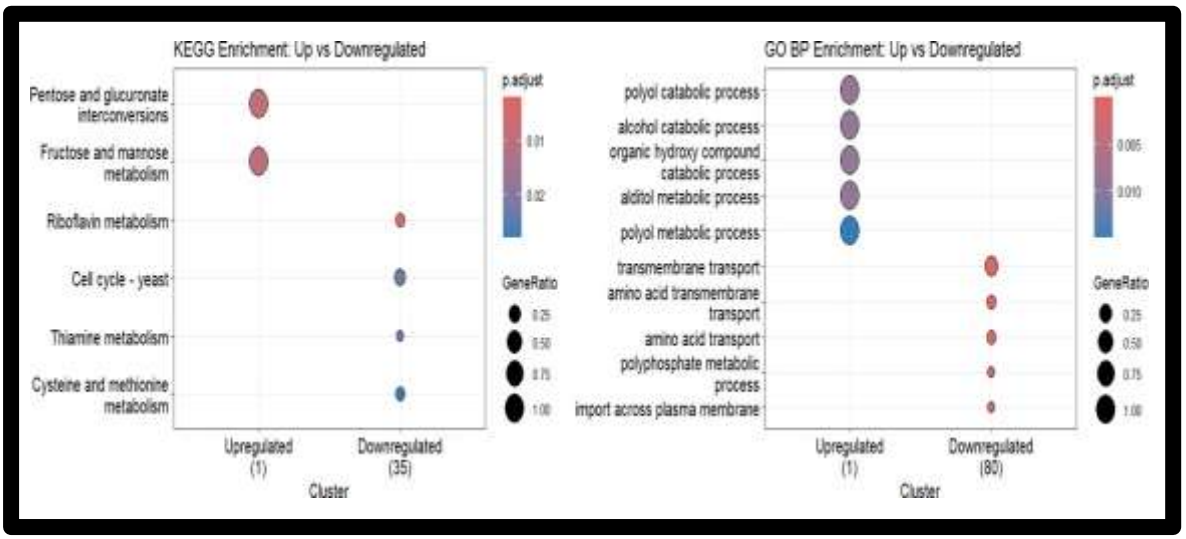

Figure 9: Functional Enrichment Analysis: KEGG vs GO Biological Process

CONCLUSION

This study provides a comprehensive transcriptomic insight into the molecular consequences of *GRR1* gene deletion in *Saccharomyces cerevisiae*, emphasizing its pivotal role in maintaining cellular homeostasis and metabolic regulation. The RNA-seq analysis revealed distinct differential gene expression patterns, with significant downregulation of genes involved in vital cellular processes such as the cell cycle, vitamin biosynthesis, and amino acid metabolism, while a limited number of genes, primarily linked to alternative sugar metabolism, were upregulated. Functional enrichment analysis through KEGG and GO pathways further supported the centrality of GRR1 in regulating metabolic pathways, nutrient transport, and biosynthetic processes. These findings affirm that GRR1 deletion disrupts fundamental biological networks, particularly those associated with carbon source utilization, transmembrane transport, and the regulation of cellular growth and division. The observed shift toward alternative metabolic routes suggests a compensatory adaptation in response to the loss of glucose sensing and signaling, reinforcing the gene's role in nutrient-responsive pathways. Overall, this study not only elucidates the downstream effects of GRR1 inactivation at the transcriptomic level but also contributes to a broader understanding of gene regulatory networks in yeast, offering a foundation for future functional and experimental validation.

REFERENCES

[1] Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Research, 51(D1), D587–D592. https://doi.org/10.1093/nar/gkac963

[2] Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods, 14(4), 417–419. https://doi.org/10.1038/nmeth.4197

[3] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8

[4] Schurch, N. J., Schofield, P., Gierliński, M., & others. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA, 22(6), 839–851. https://doi.org/10.1261/rna.053959.115

[5] Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology, 16(5), 284–287. https://doi.org/10.1089/omi.2011.0118

[6] Ljungdahl, P. O., & Daignan-Fornier, B. (2012). Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae. Genetics, 190(3), 885–929. https://doi.org/10.1534/genetics.111.133306

[7] Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., & Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. Scientific Reports, 6, 25533. https://doi.org/10.1038/srep25533

[8] Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

[9] Conesa, A., Madrigal, P., Tarazona, S., et al. (2016). A survey of best practices for RNA-seq data analysis. Genome Biology, 17(1), 13. https://doi.org/10.1186/s13059- 016-0881-8

[10] Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30. https://doi.org/10.1093/nar/28.1.27

[11] Ashburner, M., Ball, C. A., Blake, J. A., et al. (2000). Gene ontology: tool for the unification of biology. Nature Genetics, 25(1), 25–29. https://doi.org/10.1038/75556

[12] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

[13] Barral, Y., Jentsch, S., & Mann, C. (1995). G1 cyclin turnover and nutrient uptake are controlled by a common pathway in yeast. Genes & Development, 9(4), 399–409. https://doi.org/10.1101/gad.9.4.399

[14] Johnston, M. (1999). Feasting, fasting and fermenting: glucose sensing in yeast and other cells. Trends in Genetics, 15(1), 29–33. https://doi.org/10.1016/S0168- 9525(98)01633-6

[15] Flick, J. S., & Thorner, J. (1998). Genetic and biochemical characterization of SSD1, a gene implicated in cell wall integrity and cell cycle progression in Saccharomyces cerevisiae. Genetics,

148(3), 865–879.
[16] Botstein, D., & Fink, G. R. (2011). Yeast: an experimental organism for modern biology. Science, 240(4858), 1439–1443. https://doi.org/10.1126/science.3287619
[17] Trapnell, C., Williams, B. A., Pertea, G., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, 28(5), 511–515. https://doi.org/10.1038/nbt.1621
[18] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods, 5(7), 621– 628. https://doi.org/10.1038/nmeth.1226
[19] Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616
[20] Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/