RNA-seq is the key to investigation the fermentation process of baking bread in yeast

Abstract

This practical work explores how gene expression changes in yeast cells before and during fermentation, using a method known as differential expression analysis. This method measures the differences in gene activity between two conditions, helping to understand how cells respond to changes. By analyzing RNA-seq data from yeast at the start and 30 minutes into fermentation, we aimed to identify which genes become more or less active during this process. The research involved aligning RNA sequences to a reference genome, counting how many times each gene was read, and then using statistical software to find out which genes had significant changes in expression levels. Among the top 50 differentially expressed genes identified, notable findings include the downregulation of two genes, YKR097W and YLR327C, with the former implicated in carbohydrate metabolic processes essential for yeast metabolism. The remaining 48 genes showed increased activity during fermentation and were principally categorized as genes involved in rRNA and ribosome processing, carbohydrate metabolism and membrane transport.

Introduction

Differential expression analysis is a cornerstone of transcriptomics, allowing to identify genes that show statistically significant changes in expression levels between different biological states or conditions. This analysis provides a comprehensive view of gene activity within a cell or tissue and is crucial for understanding the molecular mechanisms of physiological differences, disease states, and responses to treatments. By comparing gene expression profiles across conditions, genes associated with specific phenotypes can be discovered as well as specific biological pathways and mechanisms responsible for this or that phenotypic difference [1].

Differential expression analysis reveals not only the presence of transcripts but also their quantity, enabling the identification of genes that are upregulated or downregulated in response to various exposure or in different disease states. In clinical research, it facilitates the discovery of biomarkers for disease diagnosis, prognosis, and treatment efficacy. For instance, differential expression analysis can identify genes whose expression levels are uniquely altered in cancerous tissues compared to healthy ones, highlighting potential targets for therapy [2].

The two most widely used ways to perform differential expression analysis include the use of microarrays and RNA sequencing. Before the advent of RNA-seq, microarrays were commonly used for differential expression analysis. Although less commonly used now, microarrays are relatively cheap and can still be effective for specific applications, especially when the transcriptome of the organism is well-characterized. As for RNA sequencing, it offers high sensitivity, specificity, and the ability to detect novel transcripts [3-4].

The aim of this practical work was to study changes that happen in yeast cells before or during fermentation with the use of RNA-seq data from yeast before and during fermentation.

Methods

The RNA-seq data from yeast before and during fermentation were used in this study. Each of them was performed in two replicates. SRA numbers SRR941816 and SRR941817 were obtained at the 0th minute of fermentation, SRR941818 and SRR941819

at the 30th minute of fermentation. The R64 assembly of Saccharomyces cerevisiae strain S288c (accession number GCF_000146045.2) was used as a reference genome.

The raw RNA-seq data was aligned with **HISAT2** *v2.2.1* in single-end mode (-U) [5]. The results were processed with **SAMtools** *v1.*13 [6]. The reference annotation was converted from GFF to GTF format using **gffread** *v0.12.7* [7]. **featureCounts** *v2.0.3* was used for read summarization by gene id annotation (-g gene_id) [8]. Differential gene expression analysis was performed using two R-scripts deseq2.r and draw-heatmap.r [9] based on R *v4.2.1* and DESeq2 *v1.38.3* package [10]. The top 50 genes have been annotated in gene ontology terms on [11] (Yeast GO-Slim: Process, SELECT ALL TERMS).

Results

On average, the overall alignment rate in HISAT2 was more than 95% (Table 1). Initially, 6452 genes were found, but after DESeq processing, 3174 remained (Suppl.fig 1). The top 50 genes were visualized on a clustermap (Suppl.fig 2). Two of them were downregulated (YKR097W, YLR327C), the other 48 were upregulated. Only one identifier was unannotated with GO: LR264C. Among the results of the GO annotation were terms related to rRNA and ribosome processing (rRNA processing, ribosomal large and small subunit biogenesis), carbohydrate metabolism (carbohydrate metabolic process), transport (carbohydrate transport, amino acid transport), and others (Suppl.fig 3).

Table 1. Reads alignment rate for 2 controls and 2 samples.

SRA	**reads**	**overall alignment rate**
SRR941816	9 043 877	94.33%
SRR941817	9 929 568	94.91%
SRR941818	1 721 675	96.22%
SRR941819	6 172 452	96.28%

Discussion

Among the top 50 genes that changed their expression in response to exposure, only 2 genes decreased their expression levels. These are YKR097W and YLR327C. According to the GO annotation, the gene YKR097W encodes an enzyme phosphoenolpyruvate carboxykinase PCK1, which is involved in carbohydrate metabolic processes and functions during gluconeogenesis to form phosphoenolpyruvate from oxaloacetate. Overall, the gluconeogenic reactions convert two molecules of pyruvate to a molecule of glucose, with the expenditure of six high-energy phosphate bonds, four from ATP and two from GTP. Expression of genes encoding several of the gluconeogenic enzymes is subject to glucose repression. Glucose repression of PCK1 occurs at very low levels of glucose and is transmitted through multiple signaling pathways [12].

Among the upregulated genes, GO have identified entities involved in the transport of various substances, such as amino acids (GO:0008643) and carbohydrates (GO:0008643). Among such genes STL1 (YDR536W) [13] and AQR1 (YNL065W) [14] were found to be susceptible to activation when cells are stressed. In addition, glucose-regulated genes (STL1

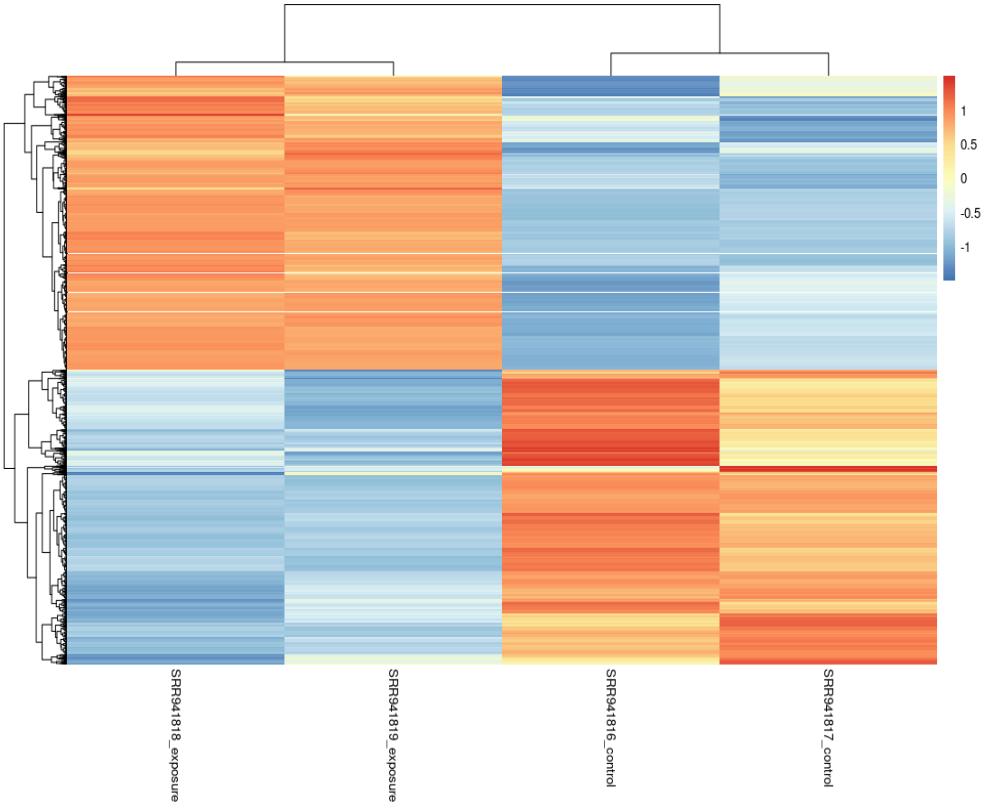
(YDR536W) and HST1 (YHR094C) [15]) also changed their expression. Such an observation is expected, since the processes discussed take place during the fermentation of bread.

References

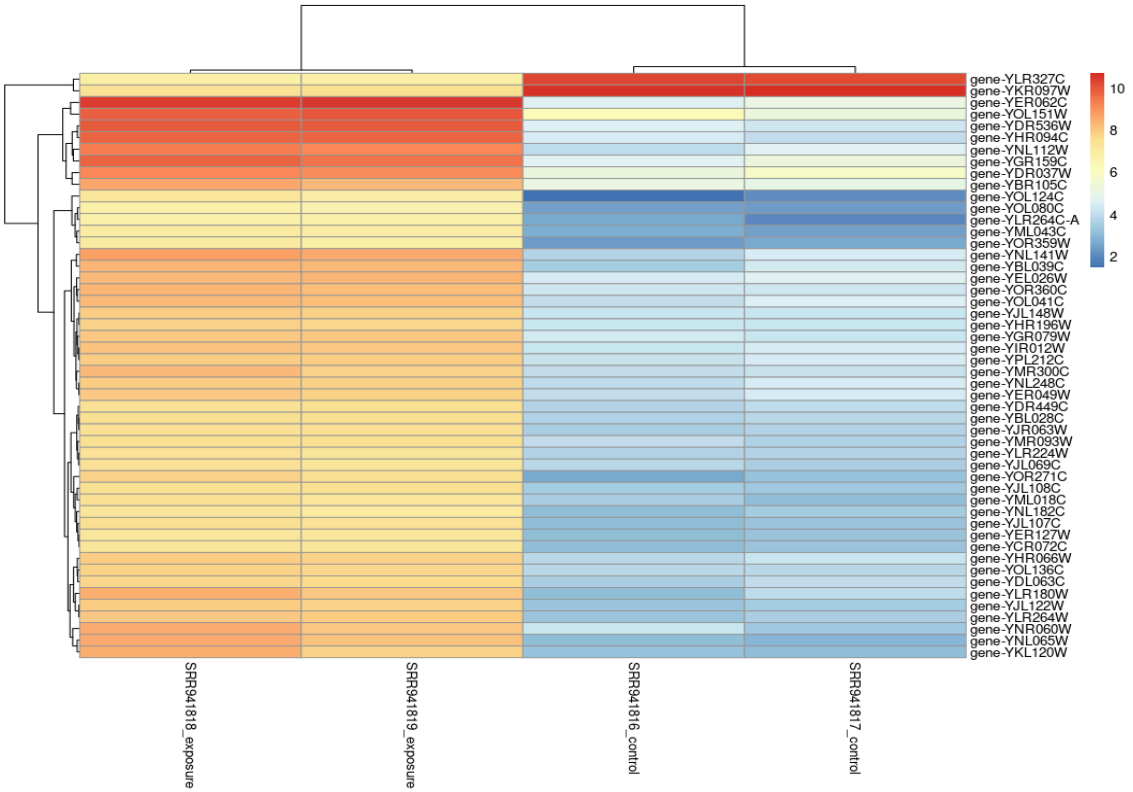
1. Liang Chen, Garry Wong. Transcriptome Informatics.Encyclopedia of Bioinformatics and Computational Biology. Academic Press. 2019. P 324-340 ISBN 9780128114322. https://doi.org/10.1016/B978-0-12-809633-8.20204-5.
2. Hong M, Tao S, Zhang L, Diao LT, Huang X, Huang S, Xie SJ, Xiao ZD, Zhang H. RNA sequencing: new technologies and applications in cancer research. J Hematol Oncol. 2020 Dec 4;13(1):166. doi: 10.1186/s13045-020-01005-x.
3. Kogenaru S, Qing Y, Guo Y, Wang N. RNA-seq and microarray complement each other in transcriptome profiling. BMC Genomics. 2012 Nov 15;13:629. doi: 10.1186/1471-2164-13-629.
4. Corchete, L.A., Rojas, E.A., Alonso-López, D. et al. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. Sci Rep 10, 19737 (2020). https://doi.org/10.1038/s41598-020-76881-x
5. Kim, D., Paggi, J.M., Park, C. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915 (2019). https://doi.org/10.1038/s41587-019-0201-4
6. Twelve years of SAMtools and BCFtools Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, Heng Li GigaScience, Volume 10, Issue 2, February 2021, giab008, https://doi.org/10.1093/gigascience/giab008
7. http://dx.doi.org/10.12688/f1000research.23297.1
8. Liao Y, Smyth GK and Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923-30
9. https://figshare.com/articles/software/Scripts_for_RNA-seq_project/14239304
10. Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology, 15, 550. doi:10.1186/s13059-014-0550-8.
11. http://www.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl
12. Yin Z, et al. (1996) Multiple signalling pathways trigger the exquisite sensitivity of yeast gluconeogenic mRNAs to glucose. Mol Microbiol 20(4):751-64
13. Zhao S, Douglas NW, Heine MJ, Williams GM, Winther-Larsen HC, Meaden PG. The STL1 gene of Saccharomyces cerevisiae is predicted to encode a sugar transporter-like protein. Gene. 1994 Sep 2;146(2):215-9. doi: 10.1016/0378-1119(94)90295-x.
14. Tenreiro S, Nunes PA, Viegas CA. et al. AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in Saccharomyces cerevisiae. Biochem Biophys Res Commun. 2002 Apr 5;292(3):741-8. doi: 10.1006/bbrc.2002.6703
15. Tenreiro S, Nunes PA, Viegas CA. et al. AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in Saccharomyces cerevisiae. Biochem Biophys Res Commun. 2002 Apr 5;292(3):741-8. doi: 10.1006/bbrc.2002.6703

Supplementary materials

Suppl. fig 1. A heatmap with differentially expressed genes



Suppl. fig 2. A heatmap which represents top 50 genes with the lowest adjusted p-value and which were selected for further GO annotation.



Suppl. 3 🗀 6