BIOINFORMATICS INSTITUTE

Bio-informatics Practice

March 2, 2024

Breathing Without Oxygen: Analysis of Differential Gene Expression in Yeast Under Hypoxic Conditions

by Artem Vasilev and Tatiana Lisitsa

Abstract

Differential gene expression analysis (DGE) is a crucial aspect of biological research, unraveling the complex interactions among genes and their impact on various physiological processes. This study focuses on understanding the alterations in RNA expression during yeast fermentation, a process with significant implications in biotechnology. Yeast, a fundamental eukaryotic model organism. Employing RNA-sequencing, we analyzed the gene expression profiles of yeast before and during fermentation. The study employed a comprehensive pipeline, including alignment with HISAT2, feature quantification with featureCounts, and identification of differentially expressed genes using the DESeq2 R library. Gene Ontology Slim Term Mapper aided in interpreting the results. A total of 6452 genes were examined, revealing significant differences in 3174 genes, with 1588 downregulated and 1586 upregulated genes. The top 100 highly differentially expressed genes were analyzed, and a subset of 10 genes is presented, showcasing their Gene Ontology annotations. Notable annotations include genes associated with tRNA processing, carbohydrate transport, osmotic stress response, DNA recombination, and regulation of DNA metabolic processes.

Keywords: differential gene expression, DGE, yeasts, *Saccharomyces cerevisiae*, respiration, fermentation

Introduction

Differential gene expression analysis plays a pivotal role in biological research, providing insights into the intricate interactions among genes and their impact on diverse physiological processes, including growth, development, aging, and stress response. This analytical approach is crucial for deciphering the molecular mechanisms underlying these processes. The knowledge gained from studying differential expression is instrumental in the development of novel drugs or the optimization of existing ones, targeting specific genes for inhibition or activation. Further enhancing our comprehension, exploring these dynamics in model organisms proves invaluable, shedding light on conserved biological pathways and facilitating translational applications.

RNA sequencing is a technique wherein cDNA fragments undergo high-throughput sequencing, and short reads are computationally aligned to the genome to pinpoint transcribed regions [1]. Subsequently, statistical testing is applied to assess whether the observed difference in read count for a specific gene is statistically significant. The read count serves as a crucial summary statistic for RNA-Seq, and it was observed to exhibit a linear relationship (to a good approximation) with the abundance of the target transcript [2].

Yeast, one of the most pivotal eukaryotic model organisms, was domesticated over 6,000 years ago for enhanced utilization in the production of bread, beer, and wine. Today, yeast also functions as a versatile cell factory for the production of various fuels, chemicals, food ingredients, and pharmaceuticals. Additionally, *S. cerevisiae* has played a crucial role as a model organism in the study of eukaryotic biology [3]. These organisms exhibit facultative anaerobiosis, demonstrating their ability to adapt metabolism based on environmental conditions. In conditions rich in glucose and oxygen, yeast cells proficiently utilize both to generate substantial ATP, the primary cellular energy currency, through aerobic respiration within the mitochondria, akin to human cellular respiration. However, when faced with oxygen scarcity, yeast cells adeptly transition to fermentation – a metabolic pathway converting sugars into acids, gases, or alcohol.

This investigation aims to elucidate alterations in RNA expression profiles during the fermentation process.

Materials and methods

For our DGE analysis we used two replicates of RNA-seq data from yeast before and during fermentation (SRR941816-SRR941819). As a reference genome and annotation we took *Saccharomyces cerevisiae* from NCBI database (strain S288c, assembly R64). The quality of the reads

was checked with FastQC [4] and MultiQC [5] tools. Given the clear line in Mean Quality Scores and the different number of reads in Sequence Counts, it was concluded that the reads were processed. For the data processing we used pipeline written in the Snakefile using Snakemake [6] (all commands and supplementary materials are available on the GitHub page). This pipeline includes aligning with HISAT2 [7], quantifying the counts of features with featureCounts [8] tool, and finding differentially expressed genes with DESeq2 R library (R 4.3.2) [9]. For some steps we also used samtools [10], basic command line tools like cat, grep and manually created python3 scripts (Python 3.11.8). After data processing Gene Ontology Slim Term Mapper [11] was used to interpret the obtained results.

Results

A total of 6452 gene expressions were analyzed, revealing significant differences (adjusted p-value < 0.05) in 3174 genes. Among these, 1588 genes were downregulated, while 1586 were upregulated. All obtained results are available on the GitHub page.

Subset of 10 genes from top 100 highly differentially expressed is shown in the Table 1. Full tables from GitHub: upregulated_genes.html, downregulated_genes.html (must be downloaded to your local computer to analyze).

Table 1: Gene Ontology annotation for several genes - selected genes

	GO Term (GO ID)	**Annotated Genes**
Upregulated	tRNA processing (GO:0008033)	YOL124C, YOR226C, YPL030W, YPL212C
	carbohydrate transport (GO:0008643)	YDR342C, YDR536W, YHR094C
	response to osmotic stress (GO:0006970)	YER062C, YIL053W
	DNA recombination (GO:0006310)	YGR159C
	response to starvation (GO:0042594)	YGL035C
	regulation of DNA metabolic process (GO:0051052)	YHR085W, YNL182C, YOR359W
Downregulated	carbohydrate metabolic process (GO:0005975)	YKR097W, YNL117W
	transmembrane transport (GO:0055085)	YDR342C
	cellular respiration (GO:0045333)	YMR081C
	monocarboxylic acid metabolic process (GO:0032787)	YNL117W

Heatmaps for all and top 100 highly differentially expressed genes are named: heatmap_all_genes.pdf, heatmap_top100_genes.pdf, respectively.

Discussion

We have chosen 2 genes from upregulated and downregulated groups to assess their role in metabolic pathways.

According to the *Saccharomyces* Genome Database (SGD), the standard name of *YGL035C* gene is *MIG1* and the product of its expression is sequence-specific DNA binding transcription factor involved in the regulation of transcription by RNA polymerase II in response to glucose and starvation. MIG1 involved in glucose repression of the *SUC*, *GAL* and *MAL* genes as well as of the *CAT8* gene. Thus, in the absence of glucose, the expression of this gene increased, and metabolic pathways for obtaining energy without the use of glucose were activated [12].

One of the intriguing genes exhibiting decreased expression is *HXT7* (YDR342C) [13]. *HXT7* encodes a glucose transporter belonging to the glucose-proton symporter family. Glucose transporters play a crucial role in facilitating the transport of glucose across the cell membrane, serving as key regulators in controlling carbohydrate metabolism within the cell.

Similar to other genes in the HXT (hexose transporter) family, *HXT7* is upregulated in response to the presence of glucose in the environment. Under conditions of low glucose concentration, the HXT7 gene can be activated to provide the cell with an additional source of this carbohydrate. In environments associated with bread making, where yeast is employed for dough fermentation, glucose and other carbohydrates serve as essential raw materials for energy production by yeast cells. During the initial fermentation phase when glucose is abundant, genes responsible for glucose transport and metabolism, including *HXT7*, may undergo activation.

However, as the fermentation process advances and the concentration of glucose in the dough diminishes, yeast cells may shift to utilizing alternative carbohydrates, such as maltose or sugars from flour. At this stage, the regulation of genes, including *MIG1* and *HXT7*, may undergo changes in response to alterations in the composition of the medium.

Supplementary materials

GitHub

References

[1] Ugrappa Nagalakshmi et al. "The transcriptional landscape of the yeast genome defined by RNA sequencing". en. In: *Science* 320.5881 (June 2008), pp. 1344–1349.

[2] Ali Mortazavi et al. "Mapping and quantifying mammalian transcriptomes by RNA-Seq". en. In: *Nat. Methods* 5.7 (July 2008), pp. 621–628.

[3] Jens Nielsen. "Yeast systems biology: Model organism and cell factory". en. In: *Biotechnol. J.* 14.9 (Sept. 2019), e1800421.

[4] *FastQC*. June 2015. URL: `https://qubeshub.org/resources/fastqc`.

[5] Philip Ewels et al. "MultiQC: summarize analysis results for multiple tools and samples in a single report". In: *Bioinformatics* 32.19 (June 2016), pp. 3047–3048. ISSN: 1367-4803. DOI: `10.1093/bioinformatics/btw354`. eprint: `https://academic.oup.com/bioinformatics/article-pdf/32/19/3047/49021359/bioinformatics\_32\_19\_3047.pdf`. URL: `https://doi.org/10.1093/bioinformatics/btw354`.

[6] Felix Mölder et al. "Sustainable data analysis with Snakemake". en. In: *F1000Res.* 10 (Jan. 2021), p. 33.

[7] Daehwan Kim et al. "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype". en. In: *Nat. Biotechnol.* 37.8 (Aug. 2019), pp. 907–915.

[8] Yang Liao, Gordon K Smyth, and Wei Shi. "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". en. In: *Bioinformatics* 30.7 (Apr. 2014), pp. 923–930.

[9] Michael I. Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15 (12 2014), p. 550. DOI: 10.1186/s13059-014-0550-8.

[10] Petr Danecek et al. "Twelve years of SAMtools and BCFtools". In: *GigaScience* 10.2 (Feb. 2021). giab008. ISSN: 2047-217X. DOI: 10.1093/gigascience/giab008. eprint: https://academic.oup.com/gigascience/article-pdf/10/2/giab008/36332246/giab008.pdf. URL: https://doi.org/10.1093/gigascience/giab008.

[11] Edith D Wong et al. "Saccharomyces genome database update: server architecture, pangenome nomenclature, and external resources". en. In: *Genetics* 224.1 (May 2023).

[12] Iraj Alipourfard et al. "MIG1 Glucose Repression in Metabolic Processes of Saccharomyces cerevisiae: Genetics to Metabolic Engineering". en. In: *Avicenna J Med Biotechnol* 11.3 (July 2019), pp. 215–220.

[13] E Reifenberger, K Freidel, and M Ciriacy. "Identification of novel HXT genes in Saccharomyces cerevisiae reveals the impact of individual hexose transporters on glycolytic flux". en. In: *Mol. Microbiol.* 16.1 (Apr. 1995), pp. 157–167.