Croissant Chemistry: RNAseq Analysis of Fermentation in *Saccharomyces cerevisiae*

Aigul Nugmanova[1] and Anya Zhurakovskaya[1]

[1]Bioinformatics Institute

Abstract

The process of fermentation is a critical step in many industries, including food (e.g. croissants) and beverage production. Understanding the underlying genetic mechanisms that regulate fermentation is essential for improving the quality of fermentation processes. In this paper we analyzed RNA sequencing (RNAseq) results and investigated the changes in gene expression of *Saccharomyces cerevisiae* yeasts before and after the start of fermentation. Specifically, we conducted a differential gene expression analysis using DESeq2. Our results revealed large clusters of genes that undergo significant expression changes after the start of fermentation. Furthermore, we identified several genes, up- and downregulated, that play a crucial role in the fermentation process.

Introduction

Understanding the principles and mechanisms through which gene activity coordinates intricate cellular processes in multicellular organisms carries extensive implications for the investigation of all living things [1]. One of the main ways to research these mechanisms is RNA sequencing and differential gene expression analysis. RNA sequencing gives us the reads that provide information about which genes and at what level are expressed in the cell [2]. Differential expression analysis involves obtaining the quantitative indicators of expression in reads and performing statistical analysis to identify quantitative variations in gene expression levels between experimental groups [3].

In this work, we made the gene expression analysis of yeast cells before or during fermentation to see the changes that provide during this process.

Methods

In our analysis, we used RNAseq data acquired from *Saccharomyces cerevisiae* yeast before and after 30 minutes start of fermentation, and for each of those states 2 replicates were created. To map RNAseq reads and understand what genes are expressed, we used reference genome of *Saccharomyces cerevisiae* and its annotation acquired from NCBI Databases.

For differential expression analysis, we used a pipeline that includes several steps. At first, we aligned RNAseq reads and reference genome with HISAT2 [4]. To quantify how many reads were mapped to genomic features in each sample, we used the featureCounts tool[5]. Then we used the DESeq2 R package to calculate the statistics such as log2 fold change and p-value [3].

Based on these results we created a volcano plot and heat map using EnhancedVolcano and ComplexHeatmap R packages [6].

For detailed analysis, we selected 50 genes with the most significant (based on adjusted p-value) change in expression. For this group of genes, we used Gene Ontology Slim Mapper with *Saccharomyces Genome* Database to understand the functionality of selected genes.

Results

FeatureCounts was used to examine 6459 genes, and the number of mapped reads for each sample was determined. Table 1 displays a summary of the mapping outcomes.

Using DESeq2 results we created the heatmap (Figure 1) exhibiting all expressed genes. As we can see samples before and after the start of fermentation clustered together. Moreover, we can see that there are large clusters of genes that after the start of fermentation changed their expression dramatically.

Furthermore, a volcano plot (Figure 2) of all expressed genes was produced using DESeq2. Out of these genes, 2657 exhibited insignificant changes in expression (adjusted p-value < 0.05). After limiting the outcomes to the top 50 genes with the most significant change in expression, we discovered only 2 genes that were downregulated and 48 genes that were upregulated.

Discussion

Analyzing the results of GO terms assigned to groups of 50 selected genes we can see, that most of them are genes involved in rRNA processing, ribosome biogenesis and transcription, which can indicate

Compiled on: April 17, 2023.
Draft manuscript prepared by the author.

Table 1. featureCounts result

Status	Aerobic 1	Aerobic 2	Fermentation 1	Fermentation 2
Assigned	7283573	7983157	1401111	4972361
Unmapped	512971	505207	65072	229677
Read_Type	0	0	0	0
Singleton	0	0	0	0
MappingQuality	0	0	0	0
Chimera	0	0	0	0
FragmentLength	0	0	0	0
Duplicate	0	0	0	0
MultiMapping	1330252	1682079	312460	1202450
Secondary	0	0	0	0
NonSplit	0	0	0	0
NoFeatures	611034	626135	102694	380392
OverlappingLength	0	0	0	0
Ambiguity	35996	36110	4204	15389

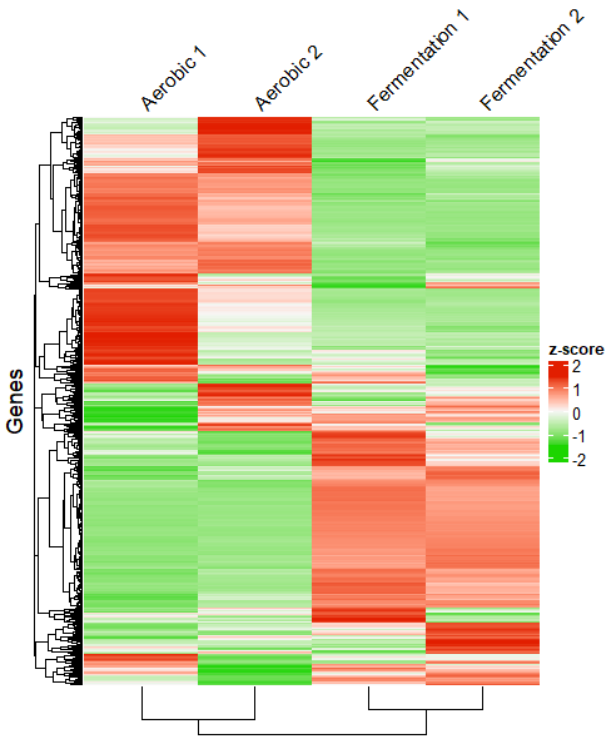

Figure 1. A heatmap of all mapped genes that indicates changes in gene expression

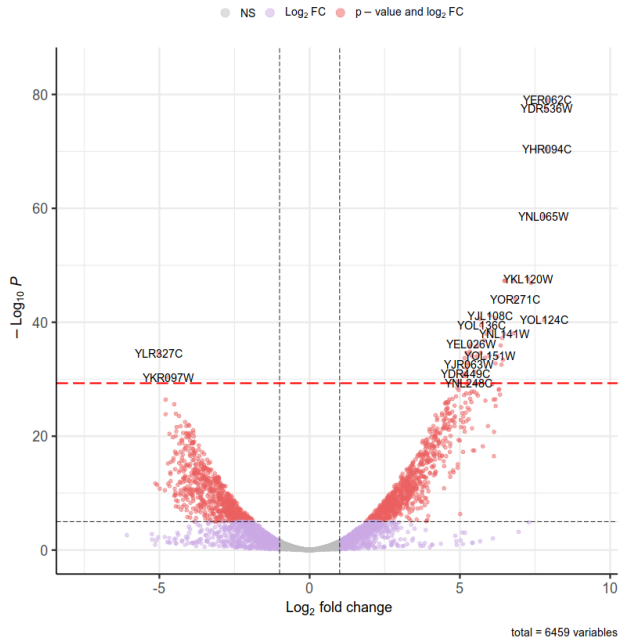

Figure 2. The volcano plot that depicts the relationship between log2 fold change and p-value of differentially expressed *Saccharomyces cerevisiae* genes. The grey dashed line represents p-value = 0.05, while the red dashed line represents the p-value threshold for 50 selected genes. The red dots indicate genes that showed significant changes in expression during fermentation, while the purple dots represent genes whose expression changes were not significant.

increased demand for ribosomes in the cell. We hypothesize that this demand appeared because yeast cells need to activate different metabolic pathways and produce different proteins and ferments for those pathways to work.

One of the two downregulated genes is a PCK1 (phosphoenolpyruvate carboxykinase) gene. This gene functions in a process of gluconeogenesis and interacts with oxaloacetate, which is produced in the Citrate cycle [7]. Since the Citrate cycle requires oxygen, the absence of oxygen disrupts gluconeogenesis. Additionally, we can see that another gene, VID24, which is recognized for its role as a suppressor of gluconeogenesis, is expressed at higher levels in fermenting yeast. This tendency to inhibit gluconeogenesis seems reasonable, as yeasts in the dough are surrounded by glucose, which is an easily available source of energy. Yeasts in dough just do not need to produce glucose from non-carbohydrate sources anymore.

And finally, we see some genes that activate in response to osmotic stress (GPP2, STL1), which can occur when yeast is exposed to high levels of salt or sugar, causing stress alongside heat, cold, and acidity [8, 9].

References

1. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature 2021;596(7871):211–220.
2. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nature reviews genetics 2011;12(2):87–98.
3. Anders S, Huber W. Differential expression analysis for sequence count data. Nature Precedings 2010;p. 1–1.
4. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology 2019;37(8):907–915.
5. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30(7):923–930.
6. Gu Z. Complex heatmap visualization. iMeta 2022;1(3):e43. https://onlinelibrary.wiley.com/doi/abs/10.1002/imt2.43.
7. Zelle RM, Trueheart J, Harrison JC, Pronk JT, van Maris AJA. Phosphoenolpyruvate Carboxykinase as the Sole Anaplerotic Enzyme in Saccharomyces cerevisiae. Applied and Environmental Microbiology 2010;76(16):5383–5389. https://journals.asm.org/doi/abs/10.1128/AEM.01077-10.
8. Hohmann S. Osmotic Stress Signaling and Osmoadaptation in Yeasts. Microbiology and Molecular Biology Reviews 2002;66(2):300–372. https://journals.asm.org/doi/abs/10.1128/MMBR.66.2.300-372.2002.
9. Attfield PV, Kletsas S. Hyperosmotic stress response by strains of bakers' yeasts in high sugar concentration medium. Letters

in Applied Microbiology 2000 10;31(4):323–327. https://doi.org/10.1046/j.1472-765x.2000.00825.x.