BIOINFORMATICS INSTITUTE

Differential expression analysis of *Saccharomyces cerevisiae* genes during bread dough fermentation

Anna Malykhina
Bioinformatics Institute, Saint-Petersburg, Russia

Sidorenko Oksana
Bioinformatics Institute, Saint-Petersburg, Russia

March 3, 2023

Abstract

The increasing availability of sequencing of RNA (RNA-seq), which can provide more and more information and details about cellular processes, metabolic pathways, pathogenesis of various diseases, single nucleotide polymorphisms and many other areas, contributes to an increase in scientific interest in these technologies and processes. In this work, we have studied changes in the transcriptome of baker's yeast (*Saccharomyces cerevisiae*) before and during bread dough fermentation. 25 upregulated and 25 downregulated genes with the highest statistically significant changes in the differential expression were selected for analysis and it was found that most of the annotated ones are involved in metabolic pathways (*e.g.* gluconeogenesis and utilization of non-fermentable carbon sources), glucose transport, and response to stress (*e.g.* oxidative stress). We also studied one group of Gene Ontology terms each from mainly upregulated and downregulated (transport processes and metabolism, respectively), and hypothesized that they might be a part of actual changes in yeast glucose transport, osmotic stress and gluconeogenesis during the switch from respiration to fermentation in bread dough.

Keywords: Transcriptome, RNA-seq, *Saccharomyces cerevisiae*, Fermentation, Differential expression

Introduction

RNA-seq analysis based on next-generation sequencing (NGS) data is a highly sensitive and accurate tool for a comprehensive investigation of cells, tissues, organs or whole transcriptome of an organism and an important approach to deciphering the molecular physiology on the different levels.
RNA-seq can be applied to a broad range of scientific questions, such as differential gene expression analysis; difference in transcription rates between environments, conditions, treatments and time; discovery and characterization of transcripts; genome annotation; SNP detection [1] [2] [3] [4] [5] [6].

Differential expression analysis means taking the normalised read count data and performing statistical analysis to discover quantitative changes in gene expression levels between experimental groups [1]. The workflow includes three parts: (a) mapping sequencing reads to a reference genome or transcriptome; (b) quantifying expression levels of individual genes and transcripts; and (c) identifying specific genes and transcripts that are differentially expressed between samples. The resulting sets of differentially expressed genes can be further analyzed, either manually or automatically, for example, for the presence of relevant genes of interest, enrichment of functional gene categories, or overlap with sets of genes or regulatory genomic elements identified in other experiments [7].

The aim of this study was to apply differential expression analysis to investigate changes in the transcriptome of *Saccharomyces cerevisiae* during bread dough fermentation and to isolate a group of genes whose expression level changed significantly as yeast underwent fermentation to make bread rise. We were interested in selecting the 25 downregulated and 25 upregulated ones, identifying the specific processes they are involved in, and explaining the role some of these processes play in the changes that occur in the yeast during fermentation.

Fermentation of sugars into carbon dioxide CO_2, ethanol and secondary metabolites by *Saccharomyces cerevisiae* during bread making leads to leavening of dough and changes in dough rheology [8]. Fermentation is one of two types of metabolism in *Saccharomyces cerevisiae*: similar to numerous other yeast species, it can alternate between fermentative and oxidative metabolism, depending on the availability of glucose and oxygen [9]. *Saccharomyces cerevisiae* consumes sugars as a main carbon source. After glucose enters the yeast cell, glycolysis is used as the pathway to convert it to pyruvate while producing adenosine triphosphate (ATP) along with nicotinamide adenine dinucleotide (NADH) and intermediates. The subsequent respiration of pyruvate can lead to further energy production. It is interesting that with high sugar concentrations, and despite the presence of oxygen, *Saccharomyces cerevisiae* has the tendency to ferment the sugars to ethanol and carbon dioxide [10]. All these processes are the result of massive reprogramming of gene expression.

Methods

Data acquisition

Two replicates of RNA-seq data each from yeast before fermentation and 30 minutes after fermentation were downloaded from:
ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941816/SRR941816.fastq.gz
ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941817/SRR941817.fastq.gz
ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941818/SRR941818.fastq.gz
ftp.sra.ebi.ac.uk/vol1/fastq/SRR941/SRR941819/SRR941819.fastq.gz

The reference genome of *Saccharomyces cerevisiae* (strain S288c, assembly R64) and annotation were downloaded from NCBI database [11] [12]: ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.fna.gz
ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.gff.gz

Workflow

HISAT2 (version 2.2.1) was used for alignment of RNA sequencing reads and reference genome [13]. Files were converted to bam format by samtools (version 1.16.1) [14]. GFF file to GTF format conversion was done by gffread (version 0.11.7) [15]. FeatureCounts function [16] was used from subread package (version 2.0.0). Differentially expressed genes were analyzed by R package Deseq2 (version 1.38.3) calculating log2FoldChange, p-value and adjusted p-value [17]. Gene Ontology (GO) (version 2023-01-01) terms were defined with Gene ONTOLOGY SLIM TERM MAPPER [18]. The detailed genes' functions were investigated through THE *Saccharomyces* GENOME DATABASE (SGD).

Results

4 data files were aligned to reference genome with good overall alignment rate (94.25%, 94.85%, 96.14% and 96.20%, respectively). Annotation was successfully assigned to 74.7%, 73.8%, 74.5% and 73.3% of alignments, accordingly. Expression level was analyzed for 6459 genes. The heatmap for all samples is shown in Figure 1.

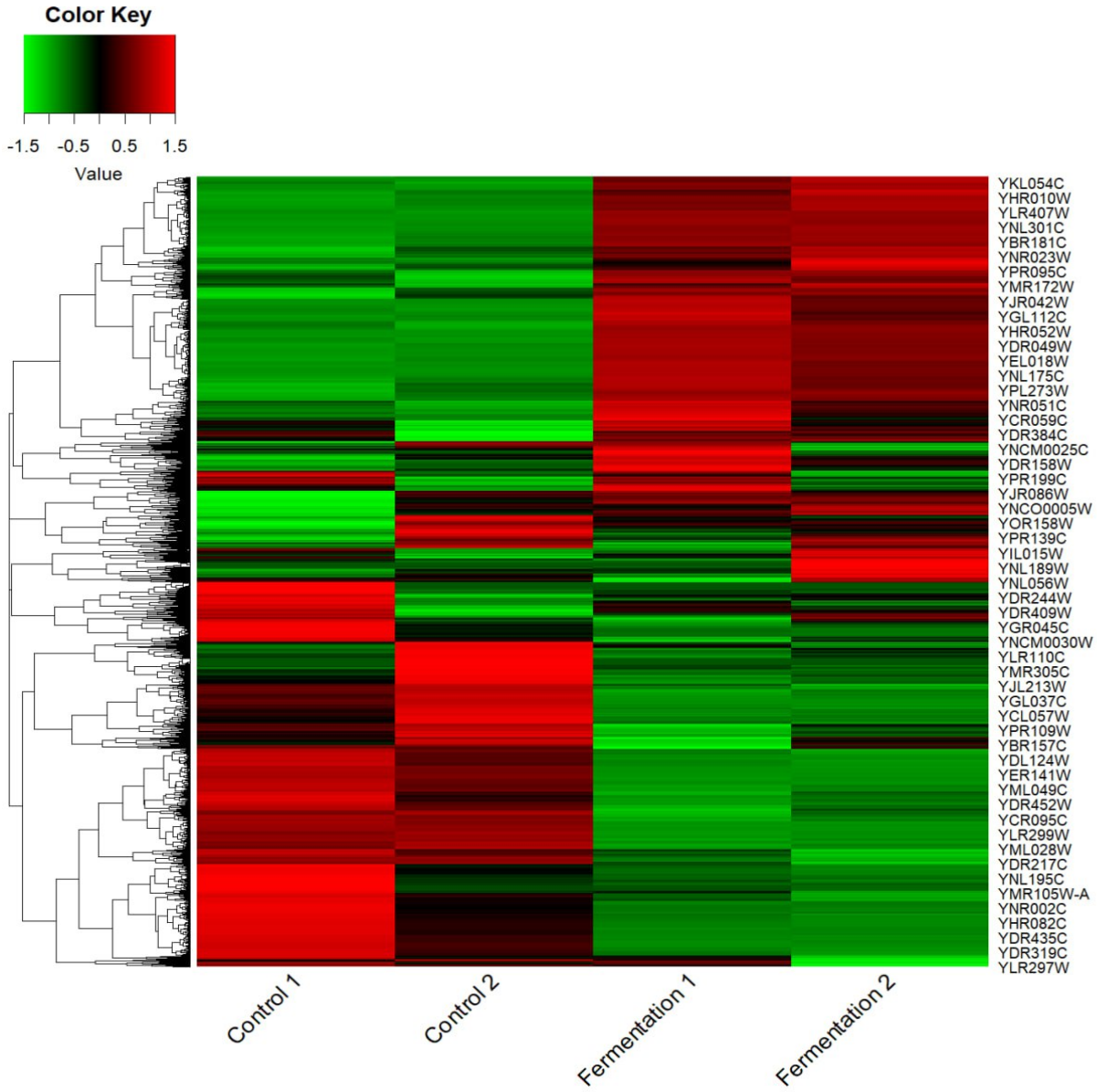

Figure 1: Heatmap for genes expression of *Saccharomyces cerevisiae* in control and fermentation samples.

3279 genes had significantly different expression with adjusted p-value less than 0.05. 1048 genes among them showed the difference of $log_2FoldChange$ more than 1. The maximal upregulation reached $log_2FoldChange$ of 7.9, while maximal downregulation was 6.1. The summary of expression difference is visualized on Figure 2.

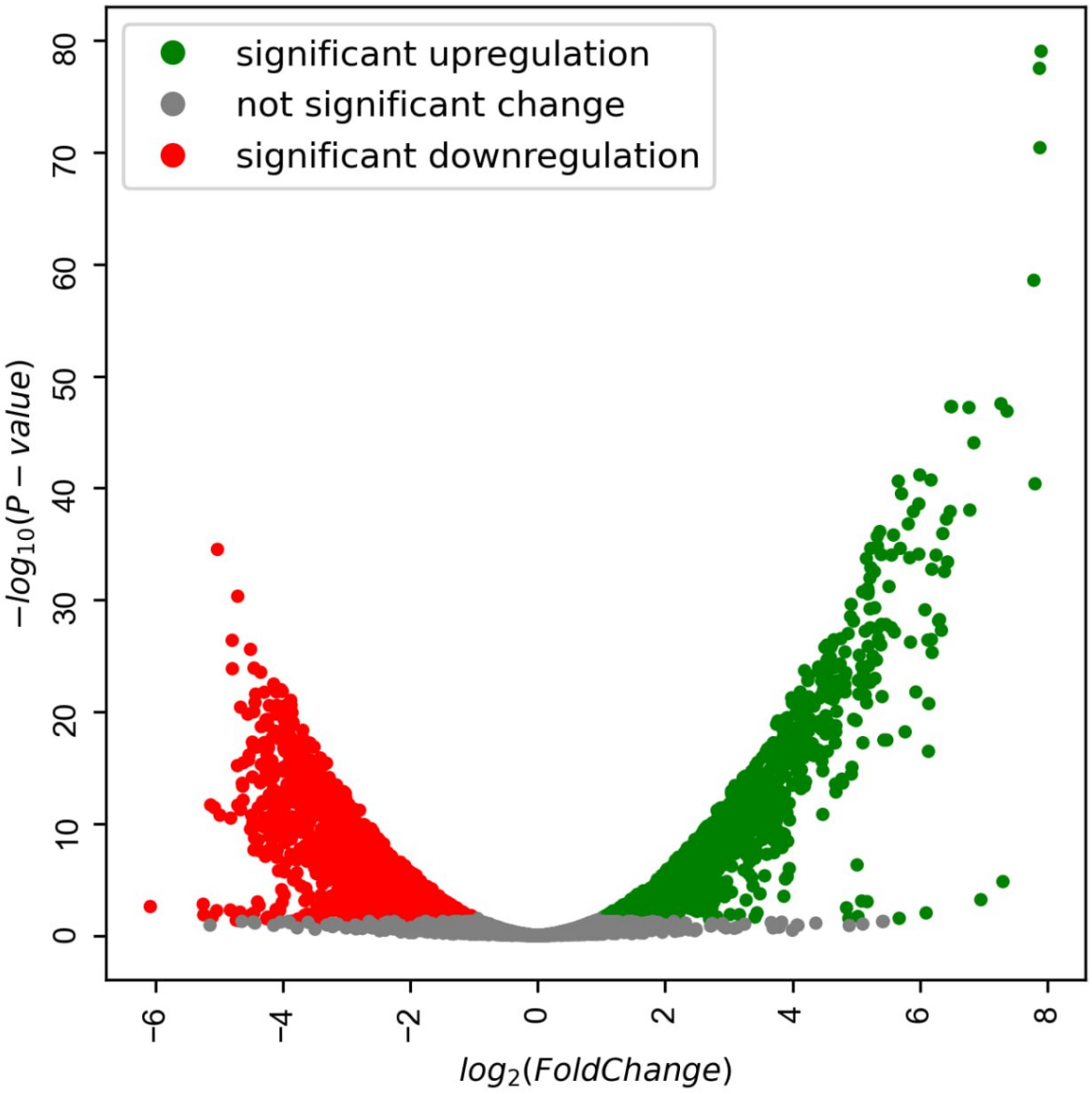

Figure 2: Differently expressed genes of *Saccharomyces cerevisiae* in control and fermentation samples. Adjusted p-value shown. The cut off was done at adjusted p-value < 0.05 and $log_2FoldChange > 1$.

The table of defined GO terms for 50 genes we selected for analysis is presented in Supplementary.

Discussion

Genes' expression heatmap analysis

The heatmap for all samples (Figure 1) shows clustering of genes with similar expression patterns: two replicates within control sample (before fermentation) and two samples within fermentation sample demonstrated similar expression and coordinated behaviour to each other and different expression between samples. The observed general picture looked adequate for the massive reprogramming of gene expression during the switch from respiration to fermentation. Large clusters of upregulated and downregulated genes are distinguished, and most of them changed their behavior to the opposite under changed conditions.

Genes' functions and GO terms analysis

25 upregulated and 25 downregulated genes with the highest statistically significant (p-value and adjusted p-value less than 0.05 threshold) changes in the differential expression were selected for analysis. *De facto* these values in selected genes were much lower than 0.05 threshold. Among these 50 genes 44 were annotated. There were 49 GO terms detected that have changed before and after fermentation. As some genes were annotated to several GO terms, we refined the specific functions of the genes according to The *Saccharomyces* Genome Database (SGD) and previous studies. For example, it has been specified that *YHR094C*, *YDR342C*, and *YDR536W* that were annotated to transmembrane transport (GO:0055085), monoatomic ion transport (GO:0006811) and carbohydrate transport (GO:0008643) were defined as glucose transporters according to SGD data and papers of Smit and Posas [10] [19]. *YNL117W*, *YIL057C*, *YAL054C* that were annotated to different metabolic processes, were involved in utilization of non-fermentable carbon sources according to SGD data and paper of Mattenberger [20].

Results of the analyzed annotation are summed up in Figure 3.

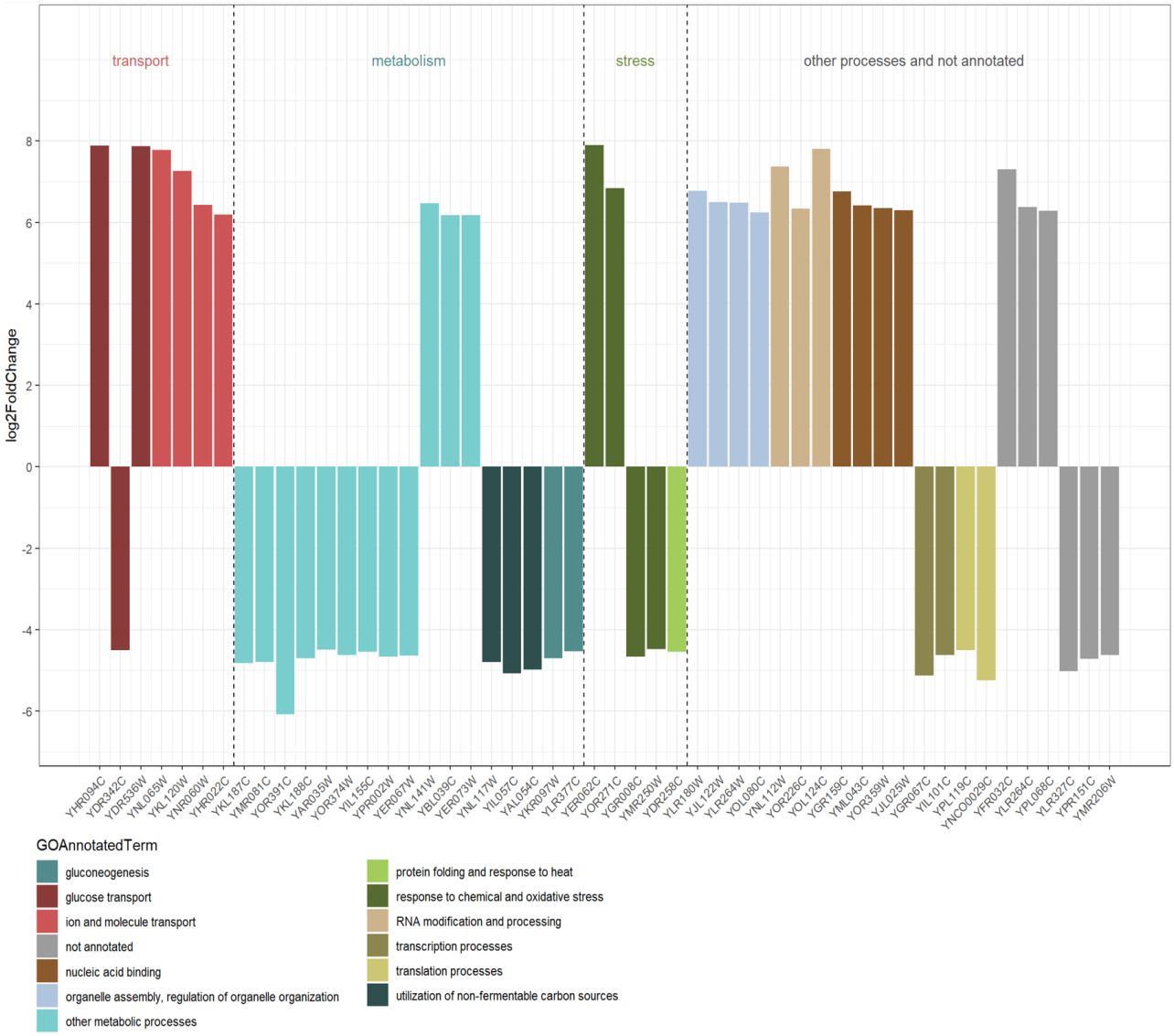

Figure 3: Difference in expression of genes involved in biological processes during bread dough fermentation.

Transport processes are mainly represented by upregulated genes; a subgroup associated with glucose and osmotic stress was distinguished

The barplot in the Figure 3 shows that the genes *YHR094C*, *YDR342C*, *YDR536W*, *YNL065W*, *YKL120W*, *YNR060W*, *YHR022C* involved in transport processes were detected among the most differentially expressed ones under the given conditions. Almost all of these genes were upregulated. Four genes with the highest differential expression among the upregulated genes of these transport subgroup are of most interest.

***YNL065W* (*AQR1*)** codes for a cell membrane transporter that mediates excretion of amino acids under conditions in which, despite an abundant availability of carbon, cell growth is limited by a second factor such as the lack of an essential compound [21]. In bread dough fermentation, osmotic stress is likely to be the growth-restrictive factor that leads to a striking upregulation of *AQR1* [2].

***YHR094C* (*HXT1*), *YDR342C* (*HXT7*),** and ***YDR536W* (*STL1*)** formed a special subgroup among transporters due to their glucose transport function.

STL1 which encodes a homolog of hexose transporters is also annotated by SGD as glycerol proton symporter of the plasma membrane and appears to be the most strongly osmoinduced yeast gene, and thus functionally close to *AQR1* [22] [23]. Stimulation by osmotic stress depends on the high osmolarity glycerol (HOG) pathway and Hot1p transcription factor which is required for normal expression of a subset of the HOG pathway-dependent responses [24]. In addition to osmotic stress, only nutrient starvation seems to stimulate STL1 expression [25] [23].

YHR094C (*HXT1*) and *YDR342C* (*HXT7*) were annotated by SGD as low-affinity glucose transporter of the major facilitator superfamily and high-affinity glucose transporter of the major facilitator superfamily, respectively. Yeast's glucose transporters are hexose transporters display uptake efficiencies consistent with their environmental expression and play physiological roles in addition to feeding the glycolytic pathway. They are encoded by genes *HXT1* through *HXT17* and by *GAL2*, a transporter for galactose that also has a high affinity for glucose [26] [27] [28]. At least one of *HXT1* through to *HXT7* is needed for growth on glucose [29]. A transporter's affinity is determined by the difference in free energy between glucose in solution and glucose bound to the transporter – the larger the free energy difference, the higher is the affinity [28]. This fact explains upregulation of YHR094C (*HXT1*) and downregulation of YDR342C (*HXT7*) showed on the barplot in the Figure 3.

Aslankoohi E, Zhu B, Rezaei MN *et al.* showed that in the first phase of fermentation, the cells need to adapt to the high osmolarity and nutrient concentration of the surrounding dough matrix. This initial adaptation phase is followed by an active fermentation phase, where cells first consume preferred sugars such as glucose and sucrose before switching to maltose. At the end of the short but vigorous fermentation, cells experience starvation and start building up stress resistance [2]. These results explain upregulation of glucose transport genes as well as transport genes involved in response to osmotic stress: unlike the authors of this work, we did not track specific stages of the fermentation process in our study, so we could observe the processes associated with various stages.

A downregulated subgroup of two key regulatory genes in the gluconeogenesis pathway among other metabolism processes was distinguished

YKR097W (*PCK1*) and *YLR377C* (*FBP1*) were annotated by SGD as key enzymes in gluconeogenesis pathway. Gluconeogenesis is a metabolic pathway leading to the generation of glucose from certain non-carbohydrate compounds. To ensure a rapid shift from oxidative phosphorylation to fermentation, the gluconeogenic enzymes undergo allosteric inhibition and covalent modifications followed by proteolysis. The key gluconeogenic enzyme fructose-1,6-bisphosphatase (FBPase) is inhibited by adenosine monophosphate and fructose-2,6-bisphosphate. Inhibition of phosphoenolpyruvate carboxykinase (FBP) and FBPase is necessary to provide a competitive advantage when cells grow on media containing high glucose concentration, by preventing futile cycles of ATP. The fate of the gluconeogenic enzymes whose expression are repressed and their activities are inhibited when yeast cells are treated with glucose, examples strong metabolic changes. This process is termed catabolite repression when it refers to the repression of gene expression or catabolite inhibition when the focus is set on the enzymatic activities. FBPase has long been a subject of catabolite inactivation studies. [30].

Gasmi *et al.* have characterized the gene *YBR239C* encoding a putative zinc cluster protein and it was named *ERT1* (ethanol regulated transcription factor 1). ChIP-chip analysis showed that Ert1 binds to a limited number of targets in the presence of glucose and controls the activity of the *PCK1* and *FBP1*

promoters through a complex interplay with other zinc cluster transcription factors. The strongest enrichment was observed at the promoter of *PCK1* [31]. This fact may be related to coordinated behaviour and similar differential expression of *YKR097W* and *YLR377C* (-4.7 and -4.5 $log_2FoldChange$ value, respectively).

The expression of genes involved in many biological processes in *Saccharomyces cerevisiae* changes dramatically during the switch from respiration to fermentation in bread dough. We have examined in detail only a few of them with high $log_2FoldChange$ values and determined their participation in glucose transport, osmotic stress and gluconeogenesis. And despite the fact that *Saccharomyces cerevisiae* is one of the most studied model organisms, the rapid development and increasing availability of genomic research (including RNA-seq) may lead to further discoveries in the field of yeast adaptation to changes in environmental conditions.

Supplementary

The lab notebook with the detailed pipeline, settings and details can be found at this LINK.

Acknowledgements

We are grateful to our teacher Mikhail Rayko for the idea of the research, provided materials, generous support, constructive comments and helpful discussions.

References

[1] *RNA sequencing — Functional genomics II*. URL: https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/rna-sequencing/ (visited on 03/01/2023).

[2] Elham Aslankoohi et al. "Dynamics of the Saccharomyces cerevisiae Transcriptome during Bread Dough Fermentation". In: *Applied and Environmental Microbiology* 79 (23 Dec. 2013), p. 7325. ISSN: 00992240. DOI: 10.1128/AEM.02649-13. URL: https://pubmed.ncbi.nlm.nih.gov/24056467/.

[3] Andrew Paul Hutchins et al. "Discovery and characterization of new transcripts from RNA-seq data in mouse CD4(+) T cells". In: *Genomics* 100 (5 Nov. 2012), pp. 303–313. ISSN: 1089-8646. DOI: 10.1016/J.YGENO.2012.07.014. URL: https://pubmed.ncbi.nlm.nih.gov/22884873/.

[4] Steven L. Salzberg. "Next-generation genome annotation: We still struggle to get it right". In: *Genome Biology* 20 (1 May 2019), pp. 1–3. ISSN: 1474760X. DOI: 10.1186/S13059-019-1715-2/METRICS. URL: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1715-2.

[5] Zachary N. Harris, Laszlo G. Kovacs, and Jason P. Londo. "RNA-seq-based genome annotation and identification of long-noncoding RNAs in the grapevine cultivar 'Riesling'". In: *BMC Genomics* 18 (1 Dec. 2017). ISSN: 14712164. DOI: 10.1186/S12864-017-4346-6. URL: https://pubmed.ncbi.nlm.nih.gov/29197332/.

[6] Frédéric Jehl et al. "RNA-Seq Data for Reliable SNP Detection and Genotype Calling: Interest for Coding Variant Characterization and Cis-Regulation Analysis by Allele-Specific Expression in Livestock Species". In: *Frontiers in Genetics* 12 (June 2021), p. 1104. ISSN: 16648021. DOI: 10.3389/FGENE.2021.655707/BIBTEX.

[7] Fei Ji and Ruslan I. Sadreyev. "RNA-seq: Basic Bioinformatics Analysis". In: *Current protocols in molecular biology* 124 (1 Oct. 2018), e68. ISSN: 19343647. DOI: 10.1002/CPMB.68. URL: https://pubmed.ncbi.nlm.nih.gov/30222249/.

[8] Mohammad N. Rezaei et al. "Harvesting yeast (Saccharomyces cerevisiae) at different physiological phases significantly affects its functionality in bread dough fermentation". In: *Food Microbiology* 39 (May 2014), pp. 108–115. ISSN: 0740-0020. DOI: 10.1016/J.FM.2013.11.013.

[9] F Rodrigues, P Ludovico, and C Leão. *Biodiversity and ecophysiology of yeasts*. 2006.

[10] Annel Smit. "Maltotriose transport in yeast". In: (2007). URL: https://scholar.sun.ac.za:443/handle/10019.1/21760.

[11] Eric W. Sayers et al. "Database resources of the National Center for Biotechnology Information". In: *Nucleic Acids Research* 50 (D1 Jan. 2022), p. D20. ISSN: 13624962. DOI: 10.1093/NAR/GKAB1112. URL: https://pubmed.ncbi.nlm.nih.gov/19910364/.

[12] *Home - Genome - NCBI*. URL: https://www.ncbi.nlm.nih.gov/genome/ (visited on 03/03/2023).

[13] Daehwan Kim et al. "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype". In: *Nature Biotechnology 2019 37:8* 37 (8 Aug. 2019), pp. 907–915. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0201-4. URL: https://www.nature.com/articles/s41587-019-0201-4.

[14] Petr Danecek et al. "Twelve years of SAMtools and BCFtools". In: *GigaScience* 10 (2 Feb. 2021). ISSN: 2047-217X. DOI: 10.1093/GIGASCIENCE/GIAB008. URL: https://pubmed.ncbi.nlm.nih.gov/33590861/.

[15] Geo Pertea and Mihaela Pertea. "GFF Utilities: GffRead and GffCompare". In: *F1000Research* 9 (2020). ISSN: 20461402. DOI: 10.12688/F1000RESEARCH.23297.2/DOI. URL: https://pubmed.ncbi.nlm.nih.gov/32489650/.

[16] Yang Liao, Gordon K. Smyth, and Wei Shi. "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". In: *Bioinformatics (Oxford, England)* 30 (7 Apr. 2014), pp. 923–930. ISSN: 1367-4811. DOI: 10.1093/BIOINFORMATICS/BTT656. URL: https://pubmed.ncbi.nlm.nih.gov/24227677/.

[17] 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 Dec. 2014). ISSN: 1474760X. DOI: 10.1186/S13059-014-0550-8.

[18] *Gene Ontology Slim Term Mapper — SGD*. URL: https://www.yeastgenome.org/goSlimMapper (visited on 03/03/2023).

[19] Francesc Posas et al. "The Transcriptional Response of Yeast to Saline Stress". In: *Journal of Biological Chemistry* 275 (23 June 2000), pp. 17249–17255. ISSN: 0021-9258. DOI: 10.1074/JBC.M910016199.

[20] Florian Mattenberger et al. "The Role of Ancestral Duplicated Genes in Adaptation to Growth on Lactate, a Non-Fermentable Carbon Source for the Yeast Saccharomyces cerevisiae". In: *International journal of molecular sciences* 22 (22 Nov. 2021). ISSN: 1422-0067. DOI: 10.3390/IJMS222212293. URL: https://pubmed.ncbi.nlm.nih.gov/34830177/.

[21] Isabel Velasco et al. "Saccharomyces cerevisiae Aqr1 is an internal-membrane transporter involved in excretion of amino acids". In: *Eukaryotic cell* 3 (6 Dec. 2004), pp. 1492–1503. ISSN: 1535-9778. DOI: 10.1128/EC.3.6.1492-1503.2004. URL: https://pubmed.ncbi.nlm.nih.gov/15590823/.

[22] *Saccharomyces cerevisiae sugar transporter (STL1) gene, complete cds - Nucleotide - NCBI*. URL: https://www.ncbi.nlm.nih.gov/nuccore/L07492 (visited on 03/03/2023).

[23] Stefan Hohmann. "Osmotic stress signaling and osmoadaptation in yeasts". In: *Microbiology and molecular biology reviews : MMBR* 66 (2 June 2002), pp. 300–372. ISSN: 1092-2172. DOI: 10.1128/MMBR.66.2.300-372.2002. URL: https://pubmed.ncbi.nlm.nih.gov/12040128/.

[24] Martijn Rep et al. “The transcriptional response of Saccharomyces cerevisiae to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes”. In: *The Journal of biological chemistry* 275 (12 Mar. 2000), pp. 8290–8300. ISSN: 0021-9258. DOI: 10.1074/JBC.275.12.8290. URL: https://pubmed.ncbi.nlm.nih.gov/10722658/.

[25] A. P. Gasch et al. “Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes”. In: *Molecular Biology of the Cell* 11 (12 2000), p. 4241. ISSN: 10591524. DOI: 10.1091/MBC.11.12.4241. URL: https://pubmed.ncbi.nlm.nih.gov/11102521/.

[26] Linda F. Bisson, Qingwen Fan, and Gordon A. Walker. “Sugar and Glycerol Transport in Saccharomyces cerevisiae”. In: *Advances in experimental medicine and biology* 892 (2016), pp. 125–168. ISSN: 0065-2598. DOI: 10.1007/978-3-319-25304-6_6. URL: https://pubmed.ncbi.nlm.nih.gov/26721273/.

[27] Andreas Maier et al. “Characterisation of glucose transport in Saccharomyces cerevisiae with plasma membrane vesicles (countertransport) and intact cells (initial uptake) with single Hxt1, Hxt2, Hxt3, Hxt4, Hxt6, Hxt7 or Gal2 transporters”. In: *FEMS yeast research* 2 (4 Dec. 2002), pp. 539–550. ISSN: 1567-1356. DOI: 10.1111/J.1567-1364.2002.TB00121.X. URL: https://pubmed.ncbi.nlm.nih.gov/12702270/.

[28] Luis Fernando Montaño-Gutierrez et al. “A push-pull system of repressors matches levels of glucose transporters to extracellular glucose in budding yeast”. In: (2021). DOI: 10.1101/2021.04.20.440667. URL: https://doi.org/10.1101/2021.04.20.440667.

[29] Roman Wieczorke et al. “Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae”. In: *FEBS letters* 464 (3 Dec. 1999), pp. 123–128. ISSN: 0014-5793. DOI: 10.1016/S0014-5793(99)01698-1. URL: https://pubmed.ncbi.nlm.nih.gov/10618490/.

[30] Olivier Alain Santt. “Gid4p, the regulatory unit of a novel E3 ubiquitin ligase involved in carbohydrate metabolism in yeast”. In: (2009). URL: https://elib.uni-stuttgart.de/bitstream/11682/1235/1/Dissertation.pdf.

[31] Najla Gasmi et al. “The switch from fermentation to respiration in Saccharomyces cerevisiae is regulated by the Ert1 transcriptional activator/repressor”. In: *Genetics* 198 (2 Oct. 2014), pp. 547–560. ISSN: 1943-2631. DOI: 10.1534/GENETICS.114.168609. URL: https://pubmed.ncbi.nlm.nih.gov/25123508/.