Differential RNA expression analysis during fermentation in yeast

Polina Guseva, Anastasiia Dudkovskaia

May 27, 2023

Abstract

The RNA-Seq technique analyses quickly and cheaply the number of RNAs thus allowing to describe of the pool of transcribing genes in a sample. Yeasts such as *Saccharomyces cerevisiae* could not only obtain energy via aerobic respiration but also switch to fermentation in an anaerobic environment. During this shift, the majority of enzymes involved in translation and carbohydrate metabolism (especially for anaerobic pathways) increase their translation. Whereas other enzymes, for example, *PCK1* that launches gluconeogenesis, reduce in amount probably due to the irrelevancy of glucose storing.

Introduction

RNA-Seq is a sequencing technique that uses next-generation sequencing to determine the amount of RNA in a biological sample at a given moment [1]. Mostly it analyzes normalized changes in cellular transcriptome, which is called *differential expression analysis* [2]. For this analysis firstly the RNA samples are fragmented, sequenced, and then mapped to a genome or transcriptome. The expression levels for each gene or isoform are estimated and normalized. After that, the differentially expressed genes (DEGs) are identified using statistical or machine learning methods. Finally, the biological context of these genes is investigated [3]. There are a huge amount of statistical methodologies and computational algorithms that have been developed based on various assumptions about the distribution of the data, which take into account unique features of the RNA-Seq data [4].

Such techniques allow to perform of a comprehensive analysis of metabolic or other pathways [5] showing, for instance, yeasts switching from respiration to fermentation.

In this work, we explore RNA-seq data to research changes that happen in yeast cells in the respiration phase before or during fermentation. By tracking translation levels, genes and gene ontology (GO) terms could be defined as up- or downregulated; their role in metabolism is debated.

Methods

The two replicates of RNA-seq data from *Saccharomyces cerevisiae* are compared before and in 30 minutes of fermentation each [6]. As a reference genome, *S. cerevisiae* S288C (baker's yeast) is used from NDBI (GenBank ID #285798) [7].

First, we aligned out data on the reference genome using `hisat2 [ver 2.2.1]` [8] and sorted by `samtools [ver 1.7-1]` [9]. Second, we used `featureCounts [ver 2.0.0]` [10] to count how many reads align to each gene in a genome annotation. Finally, we found differentially expressed genes with `Deseq2 [ver 1.38.3]` [11].

To analyze processes that are involved in the fermentation, we use `Gene Ontology Slim Term Mapper [ver 0.86]` [12]. For this part, we included only 50 first genes from `Deseq2` output, where genes are ranged by their p-value.

Results

During the processing, there is no substantial loss in numbers. All 4 RNA-Seq runs of *Saccharomyces cerevisiae* have around several million reads each with experiments after 30 minutes containing less (tab.1). The alignment to the reference genome of *S. cerevisiae* is always with coverage of more than 90% (tab.1). As a result of using `featureCounts`, 73.9% of those aligned reads are mapped to recognised genomic features.

Table 1: Amount of reads during processing

	raw data, reads	aligned
0 minutes replica 1	9 043 877	94.25%
0 minutes replica 2	9 929 568	94.85%
30 minutes replica 1	1 721 675	96.14%
30 minutes replica 2	6 172 452	96.20%
total after featureCounts		21 640 122 reads (73.9%)

In total, 6459 genomic features are recognised and half of them do not change significantly during the shift from respiration to fermentation (fig.1, fig.2). Within those that are changed drastically,

again around half of the researched genomic features become more represented among rRNAs, i.e. their translation increases.

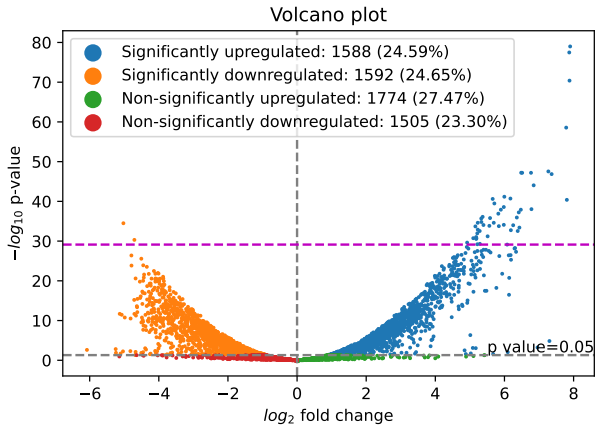

Figure 1: Volcano plot. The dotted magenta line cuts the top 50 recognised genomic features

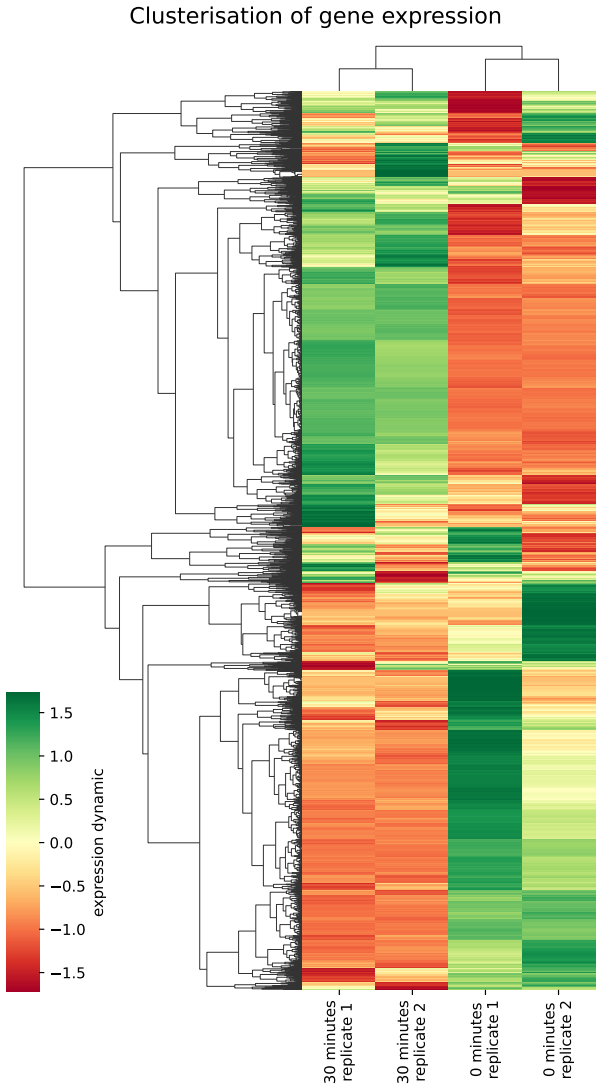

Figure 2: Clustermap of normalized differential expression

The 50 most significantly changed genes are predominantly upregulated with 2 exceptions (fig.1). Among those, gene ontology shows that YLR264C is unannotated and YBL028C, YML018C, TMA10, and YGR079W have no non-root annotations. Other genes are involved primarily in rRNA processing, ribosomal large and small subunit biogenesis, organelle (especially ribosome) assembly, transmembrane transport, transcription by RNA polymerase I, RNA catabolic process, etc (tab.2). The downregulation is detected only for the YKR097W gene from the carbohydrate metabolic process and the YLR327C sequence had no non-root annotations.

Discussion

In general, a quarter of sequenced genes are downregulated and almost the same amount is upregulated (fig.1) leading to half of the gene pool shifting its regulation during the switch from respiration to fermentation. Without respiration, glucose oxidation becomes less effective. Consequently, more enzymes are required [13] to keep up with the energy supply causing a gain in rRNA processing, ribosomal large and small subunit biogenesis, ribosome assembly, etc (tab.2). Therefore, among the top 50 genes that are upregulated, so many enzymes are involved in carbohydrate metabolism (fig.3, tab.2).

On the other hand, among highly downregulated genes in our pool, only one recognised gene is present (tab.2). The YKR097W gene is responsible for encoding phosphoenolpyruvate carboxykinase *PCK1* (4.1.1.32) [14]. This enzyme breaks down oxaloacetate into phosphoenolpyruvate and carbon dioxide [15]. This is the first unique step at the beginning of gluconeogenesis [16], i.e. the metabolic pathway with intention of storing energy. However, right after the switch from respiration to fermentation, the energy is directed to re-adjusting the metabolism to the new conditions rather than storing it. Therefore, the synthesis of this enzyme is downregulated.

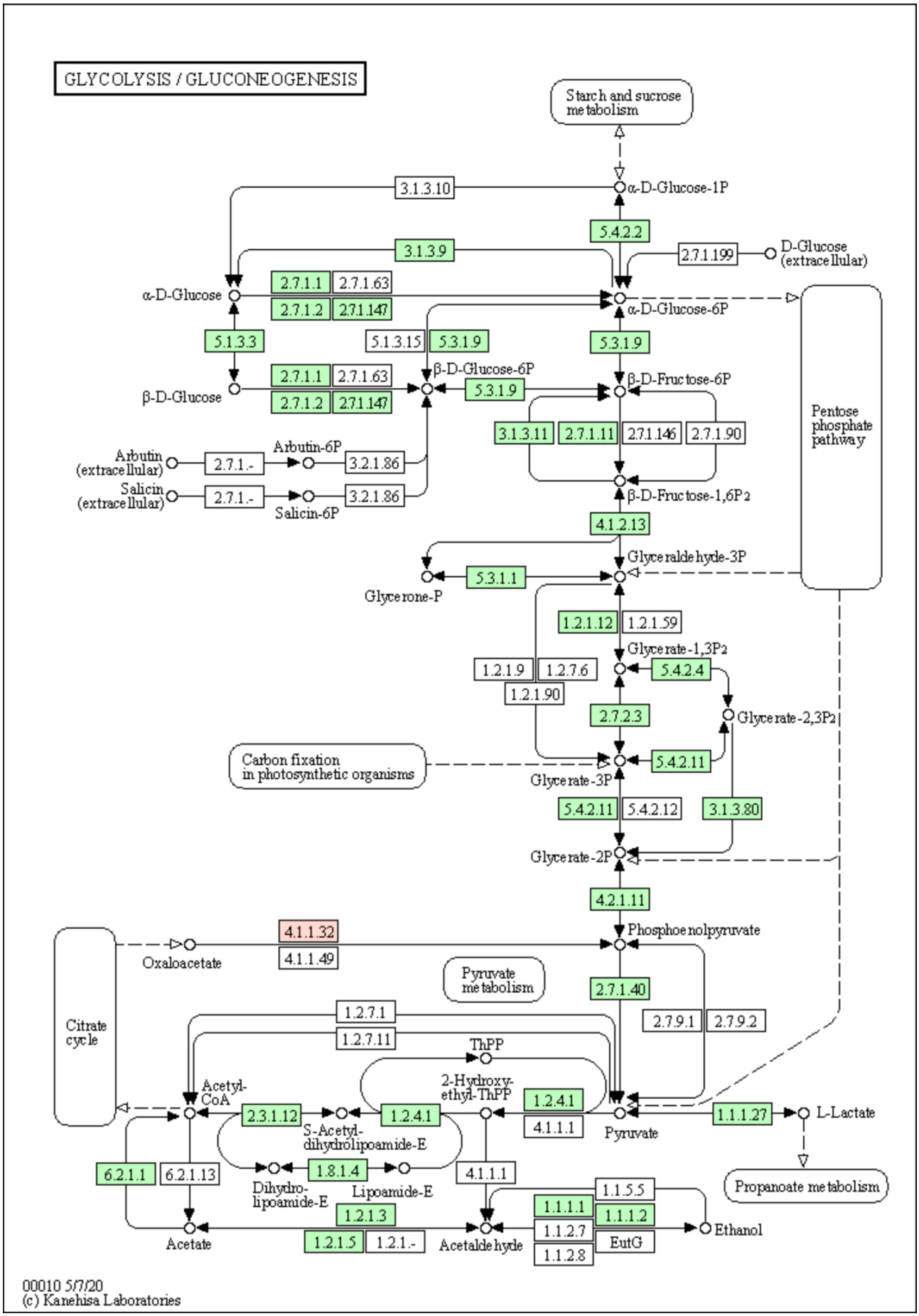

Figure 3: Increase in amount enzymes involved in glycolysis and fermentation (alcoholic, lactate, and acetate). The pathways are made by KEGG

References

[1] S. Marguerat and J. Bähler, "Rna-seq: from technology to biology," *Cellular and molecular life sciences*, vol. 67, pp. 569–579, 2010. doi:10.1007/s00018-009-0180-6.

[2] J. Costa-Silva, D. Domingues, and F. M. Lopes, "RNA-Seq differential expression analysis: An extended review and a software tool," *PloS one*, vol. 12, no. 12, p. e0190152, 2017. doi:10.1371/journal.pone.0190152.

[3] S. K. Whitley, W. T. Horne, and J. K. Kolls, "Research techniques made simple: methodology and clinical applications of RNA sequencing," *Journal of Investigative Dermatology*, vol. 136, no. 8, pp. e77–e82, 2016. doi:10.1016/j.jid.2016.06.003.

[4] A. S. Thind, I. Monga, P. K. Thakur, P. Kumari, K. Dindhoria, M. Krzak, M. Ranson, and B. Ashford, "Demystifying emerging bulk RNA-Seq applications: the application and utility of bioinformatic methodology," *Briefings in bioinformatics*, vol. 22, no. 6, p. bbab259, 2021. doi:10.1093/bib/bbab259.

[5] M. van Dijk, P. Rugbjerg, Y. Nygård, and L. Olsson, "RNA sequencing reveals metabolic and regulatory changes leading to more robust fermentation performance during short-term adaptation of Saccharomyces cerevisiae to lignocellulosic inhibitors," *Biotechnology for Biofuels*, vol. 14, pp. 1–16, 2021. doi:10.1186/s13068-021-02049-y.

[6] NCBI Sequence Read Archive (SRA), "Reads: SRR941816: fermentation 0 minutes replicate 1, SRR941817: fermentation 0 minutes replicate 2, SRR941818: fermentation 30 minutes replicate 1, SRR941819: fermentation 30 minutes replicate 2," 11 2013.

[7] NCBI: Saccharomyces Genome Database, "Gcf_000146045.2_r64: Reference genome, Annotation file," 12 2014.

[8] D. Kim, J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg, "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype," *Nature biotechnology*, vol. 37, no. 8, pp. 907–915, 2019. doi:10.1038/s41587-019-0201-4.

[9] P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, *et al.*, "Twelve years of SAMtools and BCFtools," *Gigascience*, vol. 10, no. 2, p. giab008, 2021. doi:10.1093/gigascience/giab008.

[10] Y. Liao, G. K. Smyth, and W. Shi, "featurecounts: an efficient general purpose program for assigning sequence reads to genomic features," *Bioinformatics*, vol. 30, no. 7, pp. 923–930, 2014. doi:10.1093/bioinformatics/btt656.

[11] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome biology*, vol. 15, no. 12, pp. 1–21, 2014. doi:10.1186/s13059-014-0550-8.

[12] E. I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J. M. Cherry, and G. Sherlock, "GO:: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes," *Bioinformatics*, vol. 20, no. 18, pp. 3710–3715, 2004. doi:10.1093/bioinformatics/bth456.

[13] H. Peter Smits, J. Hauf, S. Müller, T. J. Hobley, F. K. Zimmermann, B. Hahn-Hägerdal, J. Nielsen, and L. Olsson, "Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of Saccharomyces cerevisiae," *Yeast*, vol. 16, no. 14, pp. 1325–1334, 2000. doi:10.1002/1097-0061(200010)16:14%3C1325::AID-YEA627%3E3.0.CO;2-E.

[14] G.-F. Richard, C. Fairhead, and B. Dujon, "Complete transcriptional map of yeast chromosome XI in different life conditions," *Journal of molecular biology*, vol. 268, no. 2, pp. 303–321, 1997. doi:10.1006/jmbi.1997.0973.

[15] P. M. Finnegan and J. N. Burnell, "Isolation and sequence analysis of cDNAs encoding phospho enol pyruvate carboxykinase from the PCK-type C 4 grass Urochloa panicoides," *Plant molecular biology*, vol. 27, pp. 365–376, 1995. doi:10.1007/BF00020190.

[16] J. Ashmore, S. Wagle, and T. Uete, "Studies on gluconeogenesis," *Advances in Enzyme Regulation*, vol. 2, pp. 101–114, 1964. doi:10.1016/S0065-2571(64)80008-X.

Supplementary materials

Table 2: GO Terms from the biological process Ontology

GO Term	**Genes Annotated to the GO Term**
rRNA processing	YDR449C, YEL026W, YER127W, YGR159C, YHR066W, YHR196W, YJL069C, YLR264W, YMR093W, YNL112W, YNL182C, YOL041C, YOL080C
ribosomal large subunit biogenesis	YCR072C, YDL063C, YHR066W, YIR012W, YJL122W, YNL182C, YOL041C, YOL080C
organelle assembly	YCR072C, YGR159C, YHR066W, YIR012W, YLR180W, YLR264W, YNL182C, YOL080C
ribosomal small subunit biogenesis	YDR449C, YEL026W, YER127W, YGR159C, YHR196W, YJL069C, YLR264W, YMR093W
ribosome assembly	YCR072C, YGR159C, YHR066W, YIR012W, YLR264W, YNL182C, YOL080C
transmembrane transport	YDR536W, YHR094C, YJL107C, YJL108C, YKL120W, YNL065W, YOR271C
transcription by RNA polymerase I	YHR196W, YJL148W, YJR063W, YML043C, YMR093W, YNL248C
RNA catabolic process	YER049W, YGR159C, YLR264W, YNL112W, YOR359W
amino acid metabolic process	YBL039C, YDR037W, YLR180W, YMR300C
nucleobase-containing small molecule metabolic process	YBL039C, YMR300C, YNL141W, YOL136C
carbohydrate metabolic process	YBR105C, YER062C, YKR097W, YOL136C
regulation of translation	YER049W, YLR264W, YNL112W, YOR359W
monoatomic ion transport	YDR536W, YHR094C, YNR060W, YOR271C
nuclear transport	YDL063C, YHR196W, YLR264W
protein transport	YBR105C, YDL063C, YOR359W
mRNA processing	YEL026W, YGR159C, YPL212C
proteolysis involved in protein catabolic process	YBR105C, YLR224W
DNA-templated transcription initiation	YML043C, YNL248C
RNA modification	YOL124C, YPL212C
DNA-templated transcription termination	YJR063W, YNL112W
nucleobase-containing compound transport	YHR196W, YLR264W
amino acid transport	YNL065W, YOR271C
carbohydrate transport	YDR536W, YHR094C
RNA splicing	YEL026W, YGR159C
lipid metabolic process	YBL039C, YOL151W
response to chemical	YLR224W, YOR271C
regulation of DNA metabolic process	YNL182C, YOR359W
transcription by RNA polymerase II	YJR063W, YNL112W
tRNA processing	YOL124C, YPL212C
DNA replication	YNL182C
protein targeting	YBR105C
generation of precursor metabolites and energy	YOL136C
cytoplasmic translation	YLR264W

monocarboxylic acid metabolic process	YOL136C
tRNA aminoacylation for protein translation	YDR037W
DNA recombination	YGR159C
response to osmotic stress	YER062C
regulation of organelle organization	YLR180W