Yeast metabolims changes during fermentation: RNA-seq insights

Kalinichenko O., Traktirov D.

1 Abstract

In addition to classical aerobic respiration, yeasts can use an anaerobic one called fermentation. We used RNA-seq data collected before and during fermentation to determine differentially expressed genes, assuming that they may affect this process. We also investigated the functions of the obtained genes to determine their influence.

2 Introduction

Yeasts are eukariots which have been used by humans for centuries. They can breathe in two different ways: aerobic and anaerobic ones [1]. When there is not enough oxygen, they switch to anaerobic respiration called fermentation. Unlike many bacteria and human muscle cells, yeasts produce ethanol and carbon dioxide as the side products of fermentation. This effect has been used by humans, for example, to make bubbles in bread.

The aim of our research is to find what genes are involved in fermentation in yeasts and how they function. To do this we find differentially expressed genes. We look at transcriptome in *Saccharomyces cerevisiae* before and after fermentation during bread cooking and find genes which have significantly different levels of expression (in brief, the amount of RNA transcribed from these genes). We investigate the functions of these genes and make suppositions regarding their role in fermentation.

3 Methods

We used RNA-seq data in `fastq` format. There are two replicates before fermentation (SRR941816 and SRR941817) and two replicates 30 minutes during fermentation (SRR941818 and SRR941819) from NCBISequence Read Archive [2].

We also used *S. cerevisiae* (strain S288c, assembly R64) reference [3] in `fasta` format and its annotation in `gff` format [4] from NCBI.

We applied `HISAT2` to align reads to reference. Then, we converted `gff` format to `gtf` using gffread [6]. After that we obtained gene counts with `featureCounts` [7]. We left only columns necessary for further analysis with `cut`. After that, we found differentially expressed genes with `Deseq2` [8] in `R` language [9]. We used `draw-heatmap.r` in `R` [9] for drawing a heatmap.

We took 50 most significantly differentially expressed genes (with lowest adjusted p-values) and analysed them using the Saccharomyces Genome Database [10].

4 Results

First, the reference genome was indexed and reads were aligned to this genome (*S. cerevisiae*, strain S288c, assembly R64). The alignment results are shown in Table 1.

Table 1: Main alignment information

	total reads	aligned times			Overall alignment rate
		0	1	>1	
SRR941816	9043877	5.75%	87.68%	6.57%	94.25%
SRR941817	9929568	5.15%	87.06%	7.79%	94.85%
SRR941818	1721675	3.85%	87.58%	8.56%	96.15%
SRR941819	6172452	3.80%	86.96%	9.24%	96.20%

After using `featureCounts` we got (74.7%, 73.8%, 74.5% and 73.3%) successfully aligned reads for all four RNA-seq files. As a final result of `featureCounts` step we obtained a total of 6458 genes, for which it was possible to calculate expression levels.

After calculating changes in expression levels of this genes with the help of R scripts, we found significant changes in expression levels of 3180 genes (adjusted p-value < 0.05): expression of 1613 genes is upregulated, whereas expression of 1567 genes is downregulated. The heatmap of the expression levels of all genes can be seen in Figure 1.

It is possible to analyse all proteins with significant changes in expression. However, we decided to analyse only top50 genes (50 genes with the lowest adjusted p-values). From a list of 50 identifiers one identifier (ID) represented a currently unannotated gene (YLR264C), and 7 IDs respresented valid gene names that either could not be mapped to terms in the current GO slim set or were currently annotated to the root node for the slim set being used (YOR360C, YGR079W, YBL028C, YLR327C, YJL108C, YML018C, YJL107C). For other 42 genes it was possible to find GO Term(s).

5 Discussion

In this research we compared the expression levels of different genes before fermentation and 30 minutes after it started.

There was a different number of expressed genes before and after fermentation. Before it there were around 4709 genes, and after 30 minutes there were only 1759 ones (some from the previous list of 4709 genes, and some new ones).

In particular, we found that YBL039C and YOL151W are downregulated during fermentation, and database search showed that they are responsible for lipid metabolic process. This suggests thay may play an important role in feremntation. We suppose that the swith to fermentation may prevent the cell from using too many lipids, as much less energy is produced.

Also, we determined that YKR097W, YNL117W and YDR342C became upregulated after 30 minutes of fermentation. Their GO-terms show that they are involved in carbohydrate metabolic process and carbohydrate transport, which suggests they may play an important role in anaerobic respiration. In particular, we know that cells emit CO_2 during fermentation, so a much better carbohydrate transport is needed.

It is interesting that some other genes with the same GO-terms had lower expression after fermentation. We suppose that there are different roles in the aforementioned pathways which may cause fermentation.

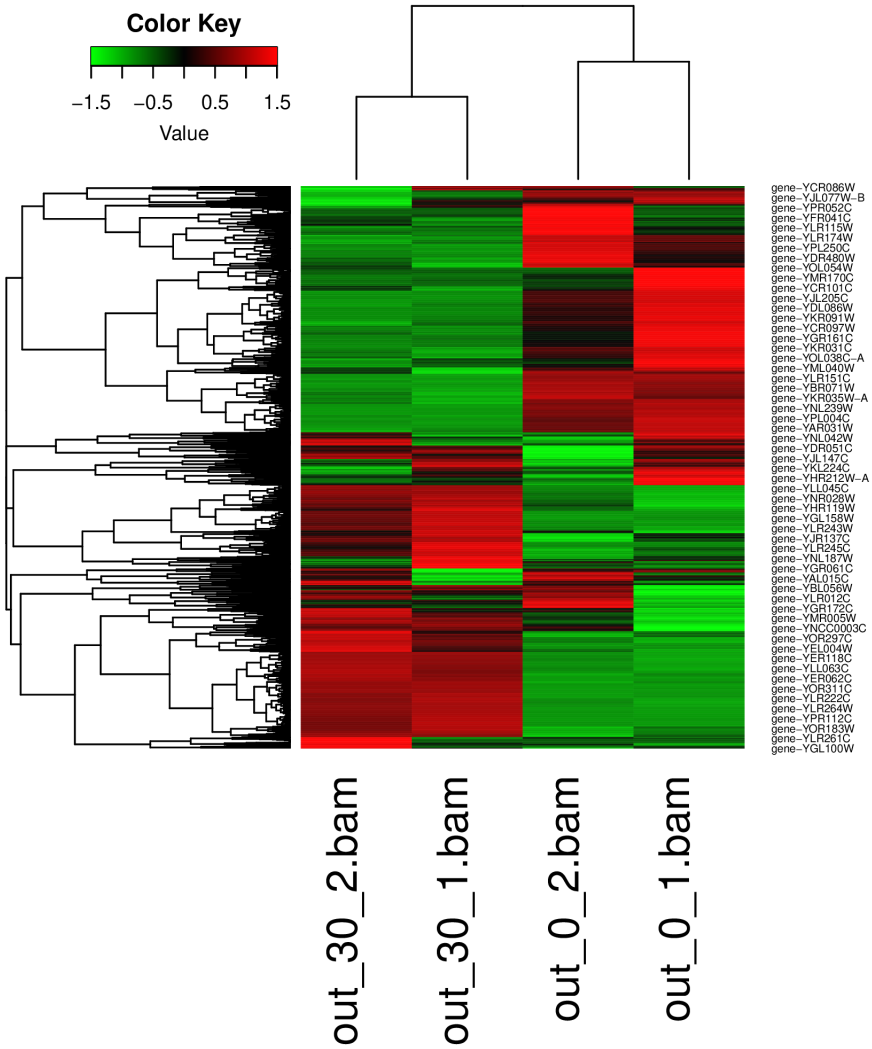

Figure 1: Heatmap of expressed genes for all 4 alignments.

References

[1] Otterstedt, Karin et al. "Switching the mode of metabolism in the yeast Saccharomyces cerevisiae." EMBO reports vol. 5,5 (2004): 532-7. doi:10.1038/sj.embor.7400132

[2] https://www.ncbi.nlm.nih.gov/sra

[3] ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.fna

[4] ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.gff.

[5] https://daehwankimlab.github.io/hisat2/

[6] https://github.com/gpertea/gffread

[7] https://academic.oup.com/bioinformatics/article/30/7/923/232889

[8] https://bioconductor.org/packages/release/bioc/html/DESeq2.html

[9] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

[10] https://www.yeastgenome.org/goSlimMapper