Yeast do feast

Abstract

After analyzing transcriptome of yeast cells after they begin a fermentation we tried to identify the genes, standing behind the switching of metabolic paths process and some general processes. These are groups are associated with protein translation, yeast reproduction, metabolic paths and chemical stress. It appears to be that after 30 minutes one group of genes switches to another due to changing the substrate.

Introduction

In this project, we studied changes that happened in yeast cells before or during fermentation. Yeasts, one of the most important eukaryotic model organisms, were domesticated more than 6,000 years ago and are still widely used in biotechnology. They are facultative anaerobes, which means they can switch their metabolism depending on the environmental conditions. When there was plenty of glucose and oxygen available, the yeast cells could use both to create significant amounts of ATP, the main cellular currency. They did this through aerobic respiration in the mitochondria (just like we do). However, when there was an oxygen shortage, they switched to fermentation, the process of converting sugars to acids, gases, or alcohol.

In some bacteria (such as Lactobacilli that are used in the production of yogurt and cheese), and in our own muscle cells, fermentation converted sugar into lactic acid and 2 ATP molecules. (That's the same lactic acid that made your muscles sore after working out!) In yeast, however, sugars were converted into ethanol and carbon dioxide (Procopio S., Qian F., Becker T., 2011). In actuality, the process was making energy in the form of ATP and leaving ethanol and CO2 as a by-product. We used this carbon dioxide in baking as a leavening agent, which meant the CO2 caused the dough to expand or rise as the gas formed pockets or bubbles.

Differential expression analysis identifies genes that are expressed differently between different biological conditions, such as healthy versus diseased tissue. This helps researchers understand the molecular mechanisms underlying these differences and identify potential targets for therapeutic intervention. It is particularly important in genomics and transcriptomics research for identifying genes involved in disease progression and drug response. (Morozova O., et al., 2009)

Materials and methods

In materials we used reads from different timestamps of fermentation process. As a reference genome we will use *Saccharomyces cerevisiae*, in the genome database at NCBI (National Center for Biotechnology Information, 3).

To estimate the distribution of transcription levels, we mapped each read to the genome using the hisat2 (Zhang, Y., et al., 2021) tool, which is a fast and sensitive alignment program designed for next-generation sequencing reads. This tool utilizes a graph FM index and a set of small indexes to cover the entire genome, allowing for effective alignment of sequencing reads. We converted the output alignments in SAM format to BAM format using samtools (Petr Danecek et al., 2021).

To analyze the data, we converted the annotation file of the genome from GFF format to GTF format using the gffread (Pertea G., Pertea M., 2020) tool. We used featureCounts to obtain statistics on mapped reads. This program counts mapped reads and takes SAM/BAM files and an annotation file as input, outputting the number of reads assigned to features.

For visualization and normalization of counts, as well as identifying significant changes in gene expression between two groups, we used the Deseq2 tool (Love M. I. et al, 2014). This tool uses negative binomial generalized linear models to test for differential expression, incorporating data-driven prior distributions in estimates of dispersion and logarithmic fold changes.

To extract information about the function of 50 genes that had changes in expression levels, we utilized the Saccharomyces Genome Database. This database provides comprehensive biological information on budding yeast *Saccharomyces cerevisiae*, along with search and analysis tools for exploring this data. The GOTERMMAPPER was used for connecting to this data and grouping the genes due to their function (Hong E. L. et al., 2003).

Results

At first we checked stats of raw data on a different timestamps of fermentation process Table 1:

Table 1 Statistics of reads

file name	length
ferment_0_r1.fastq	9427094
ferment_0_r2.fastq	10360290
ferment_30_r1.fastq	1795557
ferment_30_r1.fastq	6446589

Then with FeatureCount we analized bam files of these reads. Results are described in the `countout.txt` and `countout.txt.summary` files.

Deseq2 provides computed metrics for genes in the `result.txt` file. The Table 2 below displays the results of con-

ducted biological processes search using the top 50 genes obtained from the `result.txt`.

Also deseq2 created `norm-matrix-deseq2.txt` that contain normalised counts that we used to draw a heatmap Figure 1.

Discussion

There is a lot of going on in transcriptome in general all the time. So, in general, 6460 genes in yeast were transcripted differently after 30 minutes. But not all of the were significantly associated with the fermentation starting process.

We took only 50 most significant ones and run GOTERMMAPPER on them. This tool grouped 47 genes as it is shown on the Table 2.

Most of these groups are associated with protein translation: nuclear transport, large and small ribosomal subunits production, amino acid transport, etc. We suppose it happens because the cells need to produce many proteins specifically to convert sugar into ethanol and carbon dioxide. Also, as far as we can tell, yeast started to multiply actively, as there are genes that are associated with recombination, organelle organization and DNA replication were activated. Most proteins are in such groups, we think its due to they are numerous and they are needed for all the general processes, also they are precisely regulated, so we got more genes.

YER062C is the most upregulated gene, encoding Glycerol-1-phosphate phosphohydrolase, which is crucial for glycerol biosynthesis. This is expected since glycerol is the main solute in yeast, and it helps maintain cell hydration during osmotic stress, like in bread dough fermentation. (Joakim, N., et al., 1996)

Table 2 biological processes associated to discovered genes

GO Term	GO Term Usage in Gene List
ribosome biogenesis	20 of 47 genes
protein-containing complex assembly	10 of 47 genes
transmembrane transport	7 of 47 genes
mRNA metabolic process	7 of 47 genes
DNA-templated transcription	5 of 47 genes
carbohydrate metabolic process	4 of 47 genes
amino acid metabolic process	4 of 47 genes
nucleobase-containing small molecule metabolic process	4 of 47 genes
regulation of DNA-templated transcription	3 of 47 genes
tRNA metabolic process	3 of 47 genes
carbohydrate derivative metabolic process	3 of 47 genes
nucleocytoplasmic transport	3 of 47 genes
protein catabolic process	2 of 47 genes
intracellular protein transport	2 of 47 genes
lipid metabolic process	2 of 47 genes
sulfur compound metabolic process	1 of 47 genes
DNA replication	1 of 47 genes
signaling	1 of 47 genes
cytoplasmic translation	1 of 47 genes
generation of precursor metabolites and energy	1 of 47 genes
snRNA metabolic process	1 of 47 genes
protein modification process	1 of 47 genes
DNA recombination	1 of 47 genes

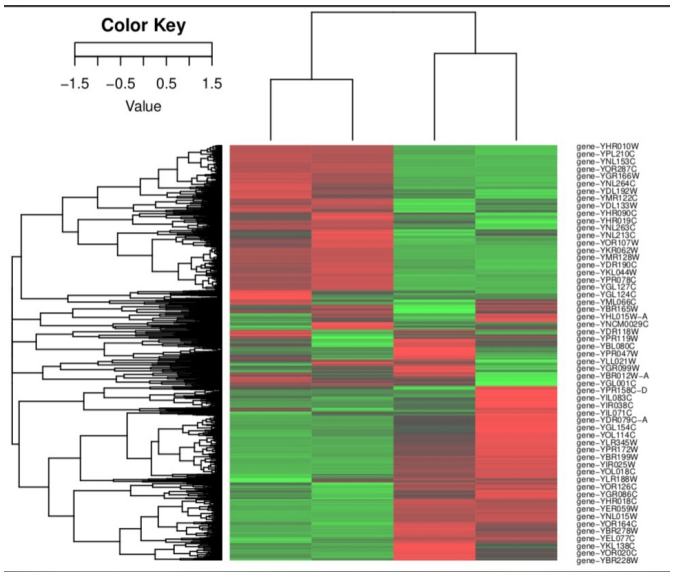

Figure 1 difference in trascriptomes

Acknowledgments

1. Procopio S., Qian F., Becker T. Function and regulation of yeast genes involved in higher alcohol and ester metabolism during beverage fermentation //European Food Research and Technology. – 2011. – . 233. – . 721-729.

2. Morozova O., Hirst M., Marra M. A. Applications of new sequencing technologies for transcriptome analysis //Annual review of genomics and human genetics. – 2009. – . 10. – . 135-151.

3. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2023 March 02]. Available from: https://www.ncbi.nlm.nih.gov/

4. Zhang, Y., Park, C., Bennett, C., Thornton, M. and Kim, D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. Genome Research 31, 1290-1295 (2021).

5. Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, Heng Li GigaScience, Volume 10, Issue 2, February 2021, giab008, https://doi.org/10.1093/gigascience/giab008

6. Pertea G., Pertea M. GFF Utilities: GffRead and GffCompare [version 1; peer review: 2. – 2020.

7. Love M. I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 //Genome biology. – 2014. – . 15. – №. 12. – . 1-21.

8. Hong E. L. et al. Gene Ontology Development and Implementation at the Saccharomyces Genome Database //Proceedings of the Korean Society for Applied Microbiology Conference. – The Korean Society for Microbiology and Biotechnology, 2003. – . 95-102.

9. Joakim N., et al. Purification and Characterization of Two Isoenzymes of DL-Glycerol-3-phosphatase from Saccharomyces cerevisiae