RNA-seq. Baking Bread.

Authors: Daria Nikanorova, Anna Rybina

Abstract

In this work, based on RNA-seq data analysis we intended to reveal differentially expressed genes in budding yeasts before and during fermentation. According to our findings, under osmotic stress budding yeasts exhibit enhanced expression level of genes involved in glycerol accumulation (e.g., YER062C and YDR536W) and amino acids export (AQR1). Upregulation of genes related to RNA processing, ribosomal subunits biogenesis and ribosome assembly together with down-regulation of genes responsible for tricarboxylic acid cycle suggest the transition from aerobic metabolism to fermentation which is driven by limited oxygen supply during dough preparation.

Introduction

Differential expression (DE) analysis allows us to identify quantitative changes in expression levels between experimental groups/conditions, characterize which genes are expressed in particular cells/tissues, group up- or down- regulated genes, and determine whether known biological functions or processes are over-represented in such genes. There are three methods that could be applied for analyzing transcriptome data, and in particular DE genes: qPCR (quantitative PCR), microarray, and RNA-seq. While the fromaer approach is mostly used to estimate expression of a few genes with a known sequence, the latter two techniques examine larger numbers of different genes simultaneously.

Recently researchers have been paying increasing emphasis on RNA-seq as it can detect novel transcripts, gene fusions, single nucleotide variants, indels, and other previously unknown changes that arrays cannot reveal. Moreover, RNA-seq could quantify expression across a wider dynamic range with higher specificity and sensitivity, distinguishing molecular isoforms, and identifying even rare and low-abundance transcripts. Problems with cross- and non-specific hybridization become obviated with RNA-seq. However, one should keep in mind that higher sensitivity might provide more statistical biases. Quality of RNA-seq derived data depends on sample preparation process and many other factors resulting in batch-effects. Thus, statistical analysis of RNA-seq data should consider such systematic errors. Another challenges for RNA-seq include removal of abundant rRNA, detecting low-expressed transcripts, paralogous genes, and discerning different isoforms of the same gene.

In this study, we aim to identify genes differentially expressed in budding yeasts before and during fermentation analyzing RNA-seq data.

Methods

Isolated commercial strain of *Saccharomyces cerevisiae* was used for uncovering changes in mRNA levels during bread dough fermentation. Two replicates of samples were taken at 0 and 30 min after the onset of fermentation. RNA was extracted and sequenced with Illumina next-generation technology. *S. cerevisiae* reference genome as well as genome

annotation in GFF format and reference transcriptome were downloaded for the same version (strain S288c, assembly R64). Alignment was carried out by using HISAT 2.0 (Kim et al. 2019). Genome indexes were built with default parameters, which was followed by alignment of obtained RNA-seq single-end reads. Mapped reads were counted for genomic features via featureCounts program from Subread package (Liao, Smyth, and Shi 2014). Finally, differentially expressed genes were analyzed using the DESeq2 package in R (Liao, Smyth, and Shi 2014; Love, Huber, and Anders 2014). Gene ontology was performed using the online version of Gene Ontology Slim Term Mapper (https://www.yeastgenome.org/goSlimMapper). KEGG over-representation test was performed using clusterProfiler package. Visualization of enrichment results was made with enrichplot package. For mapping enriched genes on KEGG pathways, pathview from the pathview package was employed.

Results

The results of alignment are summarized in Table 1. Overall mean alignment rate of 95% indicates appropriate quality of reads. 6760 features and 6420 meta-features were used by featureCounts software from annotation. The results of feature counting are shown in Table 2. Nearly 74% of all alignments were successfully assigned to a particular feature from annotation. Note that a total number of alignments is slightly bigger than total number of reads for each sample, which implies that some reads were aligned more than once on the previous step.

Analysis of differential genes expression is depicted on Fig. 1. Overall, 3187 genes were found to change their expression level significantly during fermentation (FDR-adjusted p-value < 0.05), with 1593 upregulated and 1594 downregulated genes. 50 most significantly up- and downregulated genes were annotated to belong to 40 GO terms (Supplementary Table 1).

Table 1. Summarizing data on alignment results

Sample	Total number of reads	Number of aligned 1 time reads	Overall alignment rate
SRR941816	9043877	7929709	94.25%
SRR941817	9929568	8644596	94.85%
SRR941818	1721675	1507913	96.15%
SRR941819	6172452	5367729	96.20%

Table 2. Summarizing data on featureCounts results

Sample	Total number of reads	Total alignments	Assigned alignments
SRR941816	9043877	9749615	7291724 (74.8%)
SRR941817	9929568	10810481	7987001 (73.9%)

SRR941818	1721675	1880815	1402169 (74.6%)
SRR941819	6172452	6782901	4975470 (73.4%)

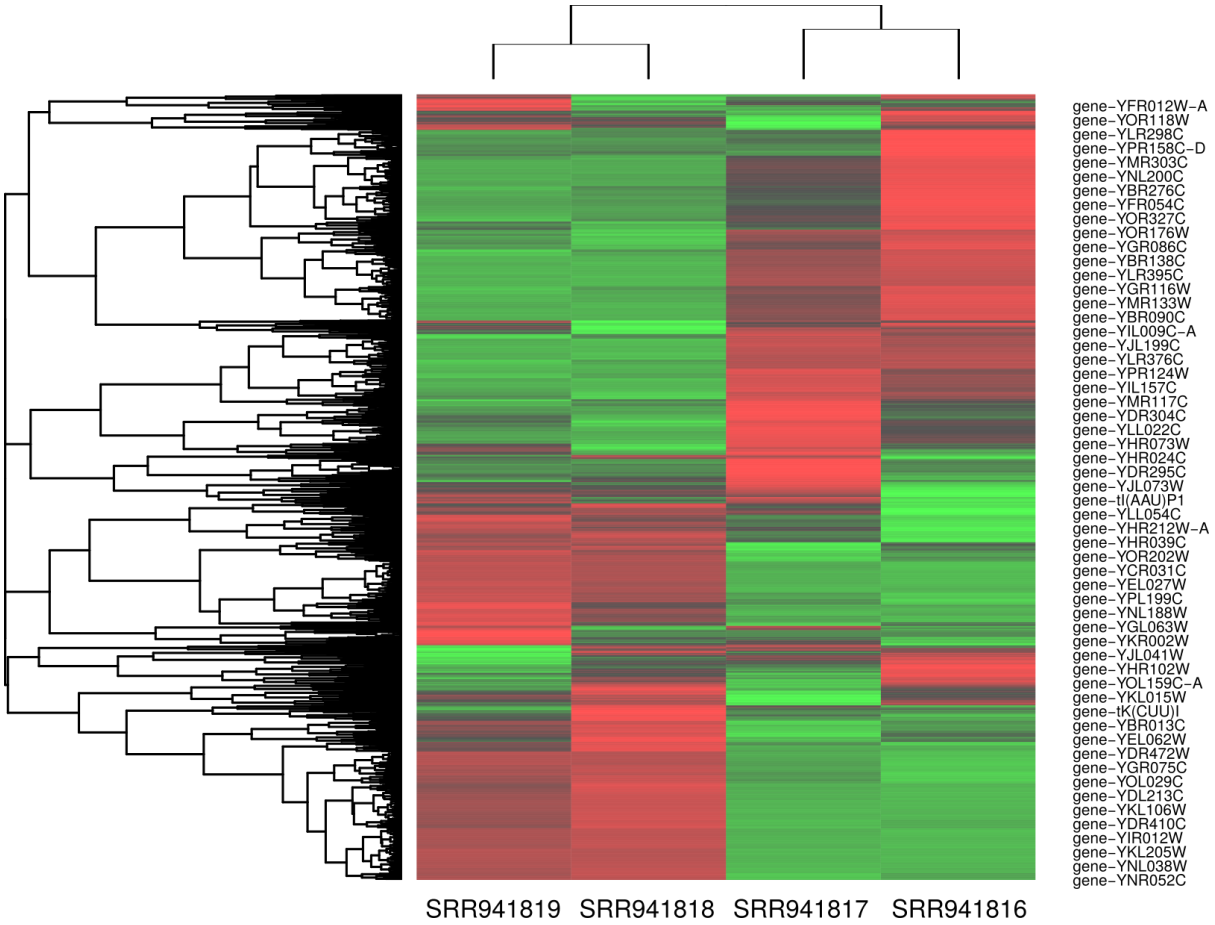

Figure 1. Heatmap of DE genes during fermentation in yeasts

Two replicates of mRNA single-end reads isolated from yeasts before (SRR941816 and SRR941817) and after 30 minutes of fermentation (SRR941818 and SRR941819) were compared. Red indicates upregulated genes, while green highlights downregulated ones.

Discussion

The most dramatically upregulated gene at 30 minutes after fermentation starts is AQR1 (YNL065W) with almost 8 logFC. According to GO, the product of this gene is a plasma membrane export of amino acids. It was previously shown that excretion of amino acids tend to occur when the central cellular metabolism is imbalanced: e.g., when the environment is rich in energy sources but yeast growth is limited (Velasco et al. 2004). We assume that during dough fermentation, osmotic stress may be the most significant growth-restrictive factor resulting in a dramatic sour of AQR1 expression. Facing water shortage while fermenting in dough, yeasts accumulate the osmolyte glycerol in order to save as much water inside cells as possible (Hohmann 2002). Indeed, according to our results, there are several upregulated genes involved in in biosynthesis of glycerol, including the glycerol phosphatase GPP2 (YER062C) and glycerol proton symporter of the plasma

membrane STL1 (YDR536W). All in all, osmotic stress driven by lack of water induces expression of genes involved in glycerol accumulation as well as AQR1 needed for amino acids export.

Of note, there is significant upregulation of 37 genes involved in rRNA processing, ribosomal subunits biogenesis and ribosome assembly at the onset of fermentation (Figure S2). The enrichment of the genes involved in protein synthesis reflects accumulation of enzymes responsible for fermentation rather than aerobic respiration. This assumption is confirmed by considerable decline in expression of genes responsible for gluconeogenesis and cellular respiration. In particular, 7 downregulated genes were annotated to be involved in the tricarboxylic acid (TCA) cycle in mitochondria (Figure S3). These results are consistent with the idea of a shift towards fermentation under lack of oxygen. Interestingly, other cellular processes such as cytoskeleton reorganization, sporulation and cell division are limited, which highlights overall decrease of cellular activity while fermentation.

Together our findings indicate that oxygen lack during dough preparation makes yeasts switch to fermentation, while struggling also with osmotic stress.

Supplementary materials

Detailed pipeline with code is provided here.

References

Hohmann, Stefan. 2002. “Osmotic Stress Signaling and Osmoadaptation in Yeasts.” *Microbiology and Molecular Biology Reviews: MMBR* 66 (2): 300–372.

Kim, Daehwan, Joseph M. Paggi, Chanhee Park, Christopher Bennett, and Steven L. Salzberg. 2019. “Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype.” *Nature Biotechnology* 37 (8): 907–15.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550.

Velasco, Isabel, Sandra Tenreiro, Isabel L. Calderon, and Bruno André. 2004. “Saccharomyces Cerevisiae Aqr1 Is an Internal-Membrane Transporter Involved in Excretion of Amino Acids.” *Eukaryotic Cell* 3 (6): 1492–1503.