Practice

BIOINFORMATICS INSTITUTE

Finding differentially expressed genes in yeast during fermentation

Ekaterina Grigorieva[1, 2] and Anastasia Bukreeva[1, 3]

[1]Bioinformatics Institute
[2]Moscow Institute of Physics and Technology
[3]Lomonosov Moscow State University

Abstract

RNA-seq is the main approach for studying differential gene expression. In this work, we analysed RNA-seq obtained on yeast cells at different metabolic stages to see how gene expression changes before and after the beginning of the fermentation process.

Keywords: differential gene expression analysis, differentially expressed genes, bioinformatics tools, visualization and interpretation, R/Bioconductor package

Introduction

Differential gene expression (DEG) is the foundation for the existence of the entire diversity of cell types. Goodwin *et al.* (2016) The differential activity of genes is based on reversible (epigenomic) changes in individual chromatin regions with preservation of the continuous structure (corresponding to the epigenetic variability of the genome: condensation, decondensation of chromatin, DNA methylation) and irreversible changes in chromatin regions, which may be associated with partial restructuring of the genome.
DEG analysis is used for a wide range of tasks. RNA-seq is the most common approach for analyzing differential gene expression. Differential expression analysis is taking normalized read count data and performing statistical analysis to detect quantitative changes in expression levels between experimental groups. Pertea *et al.* (2016) DGE analysis can provide significant insight into the genetic mechanisms in organisms that contribute to phenotypic differences, including models of cell differentiation, detection of tumor origin and investigation of many other pathological and physiological processes. Tang *et al.* (2018)

Materials and methods

Input data

In this study we used two replicates of RNA-seq data from yeast before fer- mentation (SRR941816, SRR941817) and 30 minutes after the beginning of fermentation (SRR941818, SRR941819) to study changes of gene expression during fermentation. The *Saccharomyces cerevisiae* genome strain S288c assembly R64 and its annotation were used as a reference.

Used tools

In this study we used the following tools:

- HISAT2 (v2.2.1) was used to align RNA-seq data to reference genome Kim *et al.* (2019)
- Gffread (v0.12.7) was used to convert files from GFF to GTF format Pertea and Pertea (2020)
- FeautureCounts (v2.0.1) was used for gene expression quantifying and concatenation results Bates *et al.* (2014)
- Deseq2 (v1.34.0) R package was used to find differentially expressed genes Ross-Innes *et al.* (2012)
- GO SlimMapper was used to find gene ontology terms Engel *et al.* (2014)

More detailed information can be found in the attached laboratory notebook.

Results

As a result of the work, we obtained the levels of gene expression before and after 30 minutes of fermentation and selected differentially expressed genes. The heatmap built from these data is shown in the Figure 1.

We decided that genes were differentially expressed during fermentation if their adjusted p-value was less then 0.05 and their expression level changed by more than 1.5-fold. Up-regulated genes were those whose log2FoldChange was bigger than $log_2\frac{3}{2}$ and down-regulated log2FoldChange was less than -$log_2\frac{3}{2}$ (Figure 2). The number of up-regulated genes was 1588 and number of down-regulated genes - 1592, so the total number of genes with differential expression was 3180.

Then we chose 50 up-regulated and 50 down-regulated differentially expressed genes with minimum adjusted p-value (Figure 3) and found out the functions they perform. We used gene ontology terms. Having selected the first 50 up-regulated and 50 down-regulated genes with the smallest adjusted p-value, we obtained the gene ontology Table 1.

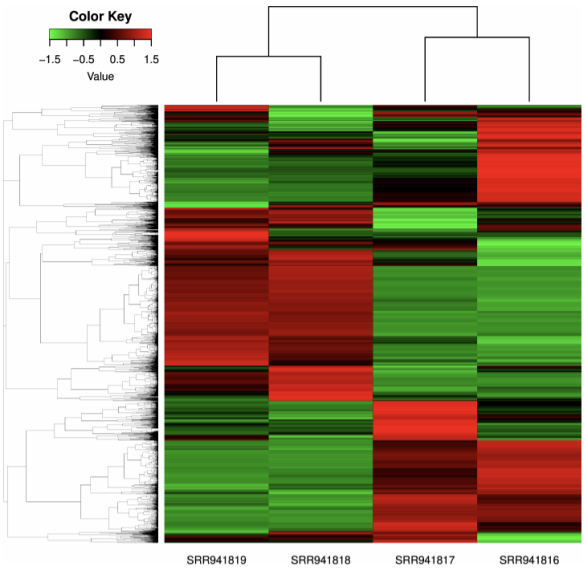

Figure 1 Heatmap of differentially expressed genes expression before and after 30 minutes of fermentation

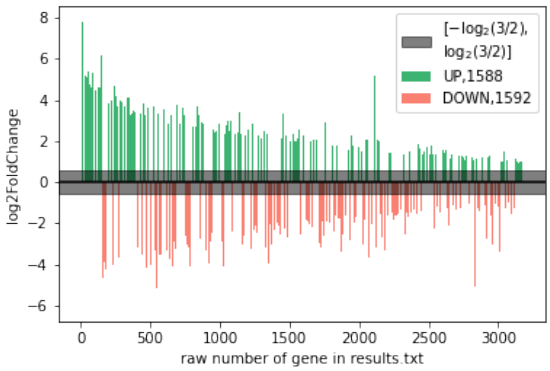

Figure 2 Log2FoldChange of differentially expressed genes

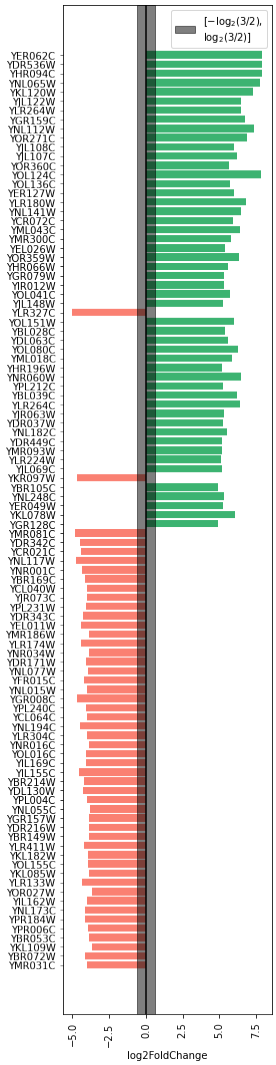

Figure 3 Log2FoldChange of chosen differentially expressed genes, top-50 up-regulated and top-50 down-regulated genes

Discussion

According to the results obtained, we see that the RNA profile changes quite dramatically during the fermentation process. Genes that showed a high level of expression before fermentation are significantly suppressed after 30 minutes of fermentation. Genes that initially showed a low level, on the contrary, increased after active fermentation. The activity of genes related to protein synthesis increases. This is due to the fact that cells are switching to a new type of energy production, which stimulates the activity and production of a new set of proteins.

Cluster changing of gene expression under different physiological conditions has also been described in Jeffries et al, which is also confirmed by our results. Jeffries and Van Vleet (2009)

Acknowledgments

We thank Mike Rayko for supervising and supporting us, providing guidelines and recommendations.

Literature cited

Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823. .

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS *et al*. 2014. The reference genome sequence of saccharomyces cerevisiae: then and now. G3: Genes, Genomes, Genetics. 4:389–398.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics. 17:333–351.

Jeffries TW, Van Vleet JRH. 2009. Pichia stipitis genomics, transcriptomics, and gene clusters. FEMS yeast research. 9:793–807.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. Nature biotechnology. 37:907–915.

Pertea G, Pertea M. 2020. Gff utilities: Gffread and gffcompare [version 2; peer. .

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of rna-seq experiments

with hisat, stringtie and ballgown. Nature protocols. 11:1650–1667.

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR *et al*. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 481:389–393.

Tang W, Wan S, Yang Z, Teschendorff AE, Zou Q. 2018. Tumor origin detection with tissue-specific mirna and dna methylation markers. Bioinformatics. 34:398–406.

Supplementary data

The code, raw data and all results are available at https://github.com/grigorievaekaterina/Project_6

Table 1 Gene Ontology results

TERM	ANNOTATED_GENES	UP	DOWN	DECISION
rRNA processing	YDR449C, YEL026W, YER127W, YGR128C, YGR159C, YHR066W, YHR196W, YJL069C, YKL078W, YMR093W, YNL112W, YNL182C, YOL041C, YOL080C	14	0	UP
carbohydrate metabolic process	YBR105C, YBR149W, YCL040W, YEL011W, YER062C, YFR015C, YIL155C, YIL162W, YKR097W, YNL117W, YOL136C, YPR184W	3	9	DOWN
ribosomal small subunit biogenesis	YDR449C, YEL026W, YER127W, YGR128C, YGR159C, YHR196W, YJL069C, YKL078W, YMR093W	9	0	UP
lipid metabolic process	YBL039C, YGR157W, YJR073C, YKL182W, YLR133W, YNR016C, YOL151W, YPL231W, YPR006C	2	7	DOWN
ribosomal large subunit biogenesis	YCR072C, YDL063C, YHR066W, YIR012W, YJL122W, YNL182C, YOL041C, YOL080C	8	0	UP
transcription by RNA polymerase I	YGR128C, YHR196W, YJL148W, YJR063W, YML043C, YMR093W, YNL248C	7	0	UP
monocarboxylic acid metabolic process	YCL040W, YKL182W, YNL117W, YNR016C, YOL136C, YPL231W, YPR006C	1	6	DOWN
response to chemical	YDR216W, YGR008C, YKL109W, YLR224W, YNL173C, YOR271C	2	4	DOWN
nucleobase-containing small molecule metabolic process	YBL039C, YCL040W, YMR300C, YNL141W, YNR001C, YOL136C	4	2	UP
ribosome assembly	YCR072C, YGR159C, YHR066W, YIR012W, YNL182C, YOL080C	6	0	UP
transmembrane transport	YCL040W, YDR342C, YDR343C, YDR536W, YHR094C, YNL065W	3	3	DOWN
protein folding	YBR072W, YBR169C, YMR186W, YOR027W, YPL240C	0	5	DOWN
nucleobase-containing compound transport	YGR128C, YHR196W, YLR264W, YNL055C, YNR034W	3	2	UP
carbohydrate transport	YCL040W, YDR342C, YDR343C, YDR536W, YHR094C	2	3	DOWN
generation of precursor metabolites and energy	YCL040W, YEL011W, YFR015C, YOL136C, YPR184W	1	4	DOWN
ion transport	YKL120W, YNL055C, YNL065W, YNR060W	3	1	UP
cellular respiration	YKL085W, YKL109W, YLR304C, YMR081C	0	4	DOWN
RNA catabolic process	YER049W, YLR264W, YNL112W, YOR359W	4	0	UP
regulation of translation	YER049W, YLR264W, YNL112W, YOR359W	4	0	UP
transcription by RNA polymerase II	YDR216W, YJR063W, YKL109W, YNL112W	2	2	DOWN
protein phosphorylation	YDL130W, YNL055C, YOL016C, YPL004C	0	4	DOWN
mitochondrion organization	YLR304C, YNL055C, YOR027W, YPL240C	0	4	DOWN
regulation of DNA metabolic process	YNL182C, YOR359W, YPL240C	2	1	UP
regulation of protein modification process	YDL130W, YNL055C, YPL004C	0	3	DOWN
proteolysis involved in cellular protein catabolic process	YBR105C, YBR169C, YLR224W	2	1	UP
protein targeting	YBR105C, YOR027W, YPL240C	1	2	DOWN
response to heat	YBR072W, YMR186W, YPL004C	0	3	DOWN
DNA-templated transcription, elongation	YJL148W, YNL248C	2	0	UP
regulation of organelle organization	YLR180W, YPL240C	1	1	DOWN
response to osmotic stress	YER062C, YPL240C	1	1	DOWN
peptidyl-amino acid modification	YER049W, YNL077W	1	1	DOWN
cellular amino acid metabolic process	YCL064C, YLR180W	1	1	DOWN
RNA modification	YOL124C, YPL212C	2	0	UP
cytoplasmic translation	YDL130W, YLR264W	1	1	DOWN
telomere organization	YMR186W, YPL240C	0	2	DOWN
tRNA processing	YOL124C, YPL212C	2	0	UP
endocytosis	YBR214W, YPL004C	0	2	DOWN
DNA-templated transcription, termination	YJR063W, YNL112W	2	0	UP
chromatin organization	YDR216W	0	1	DOWN
protein modification by small protein conjugation or removal	YNL077W	0	1	DOWN
meiotic cell cycle	YNL194C	0	1	DOWN
cytoskeleton organization	YDR171W	0	1	DOWN
DNA recombination	YGR159C	1	0	UP
sporulation	YNL194C	0	1	DOWN
DNA replication	YNL182C	1	0	UP
organelle assembly	YLR180W	1	0	UP
cell wall organization or biogenesis	YOL155C	0	1	DOWN
tRNA aminoacylation for protein translation	YDR037W	1	0	UP
protein maturation	YPL240C	0	1	DOWN
oligosaccharide metabolic process	YIL162W	0	1	DOWN
mRNA processing	YEL026W	1	0	UP
amino acid transport	YNL065W	1	0	UP
organelle fusion	YNL015W	0	1	DOWN
RNA splicing	YEL026W	1	0	UP
peroxisome organization	YDR216W	0	1	DOWN