BIOINFORMATICS INSTITUTE

Portrait of a hero: how yeasts fight and suffer to keep us fed and drunk

Liliia Bogdanova[1, 2 *] and Nikita Vaulin[1, 3, * †]

[1]Bioinformatics Institute, Saint-Petersburg, Russia
[2]Sechenov University, Specialist in Bioengineering and Bioinformatics, Moscow, Russia
[3]Skolkovo Institute of Science and Technology, Moscow, Russia

*Contributed equally.
†vaulin@ro.ru

Abstract

In this work, we use RNA sequencing and differential gene expression analysis to determine which biological and biochemical changes occur in yeast during the transition to fermentation. We show that the fermentation process globally affects everything from RNA maturation to cell cycle regulation. Yeasts cannot continue their happy family life of fissions, but they are forced to stagnate in their development as a carbon dioxide and ethanol factory. But that's why this hard workers are here - in our loaves, in our barrels. And thank them for that!

Key words: *Saccharomyces cerevisiae*, yeats, fermentation, RNA-seq, differential genes expression

Introduction

Saccharomyces cerevisiae is a type of yeast that is commonly used in the food and beverage industry due to its marvelous fermentation [1, 2, 3]. Fermentation is a metabolic process that converts carbohydrates, such as sugars, into alcohol or organic acids using microorganisms like yeasts or bacteria, in the absence of oxygen [2]. Yeast, like shrewd technicians, has tuned its biochemical tools to survive over and over again in the most unexpected conditions. And like the rampage of water, wind, and fire, humans have been able to tame and subdue this unimaginable energy and potential of fermentation [1].

For now such biochemical twist of the wrist is commonly utilized by humans in the production of bread, beer, and wine within the food and beverage industry [1, 4, 5, 6]. In addition to its role in the food industry, *Saccharomyces cerevisiae* has also been used in biotechnology for the production of biofuels and pharmaceuticals [6]. Despite its extensive use in industry, *Saccharomyces cerevisiae* also plays an important role in scientific research as a model organism for studying molecular and cellular biology [3].

Arising at the junction of advanced science and enterprising industry, the genetic engineering of *Saccharomyces cerevisiae* has allowed for the creation of strains with improved productivity and efficiency in various industrial applications [7]. Thus, all new scientific ideas and methods must be born to take control and erupt the efficiency that is potentially ensconced within a single tiny yeast cell. One such tool in the hands of a human is the differential gene expression analysis.

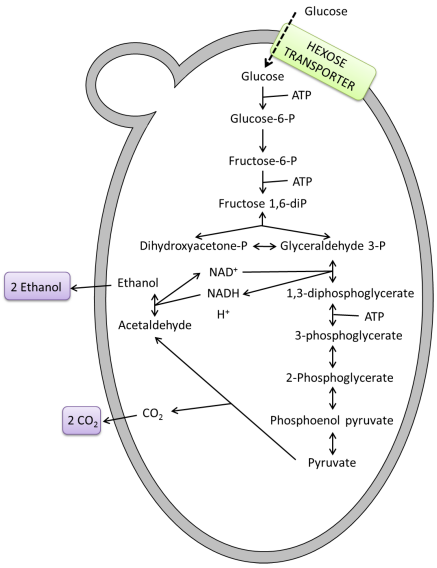


Figure 1. Yeasts fermentation process schema, [2]

Differential gene expression is a process of turning genes on and off in response to environmental or developmental cues [8, 9]. It underlies cell differentiation, tissue specialization and organism development [8, 9] . Differential gene expression can be analyzed through various techniques such as microarray analysis, RNA se-

Compiled on: March 3, 2023.
Don't forget to rest and take care of yourself.

quencing, and qPCR [10, 11, 12].

These methods allow researchers to compare gene expression levels between different samples, and identify genes that are up-regulated or downregulated in the given conditions.

It serves people in many ways: to understand the genetic basis of phenotypic traits and biological processes, to study the external stress response cascades, to identify potentially interesting for science and industry genes, to identify potential drug targets, to study the effects of genetic modifications, to study the molecular-genetic basis of evolution.

In this study, we explore the differential expression of yeast genes during the process of fermentation. This can help us optimize the production of fermented goods such as beer, wine and bread. By identifying the genes involved in key metabolic pathways, we can manipulate these pathways to improve yield and product quality. This also allows us to understand the influence of environmental factors such as temperature, pH and nutrient availability on yeast metabolism.

Materials and methods

RNA-seq data

To find out whether yeast expresses different genes during fermentation than during normal growth, we used the following 4 single-end RNA sequencing reads: SRR941816: fermentation 0 minutes replicate 1, SRR941817: fermentation 0 minutes replicate 2, SRR94181818: fermentation 30 minutes replicate 1, SRR941819: fermentation 30 minutes replicate 2.

Aligning RNA-seq reads to the genomic reference

As the the reference genome we used *Saccharomyces cerevisiae* strain S288c assembly R64 and annotation.

Reference genome indexing was performed with HISAT2 `hisat2-build` command [13]. Reads alignment to the indexed genome was performed also with HISAT2 [13], alignment sorting was performed with samtools package [14]. Example for the SRR941816: `hisat2 -x GCF_000146045.2_R64_genomic -U SRR941816.fastq | samtools sort > SRR941816_out.bam`.

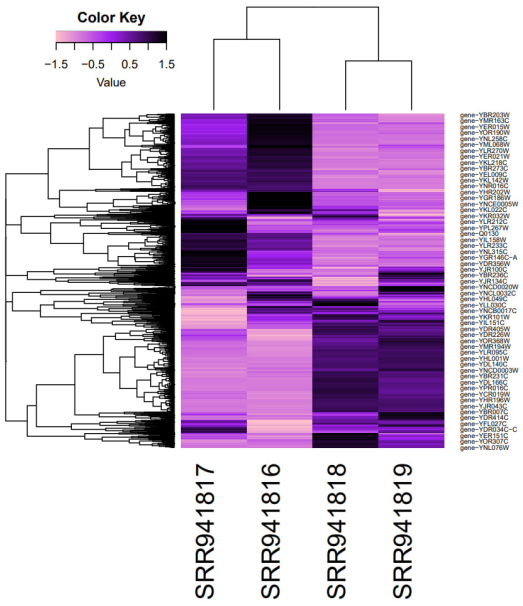


Figure 2. Heatmap of differential gene expression for the control (SRR941816, SRR941817) and fermentation (SRR941818, SR941819) samples

Table 1. RNA-seq alignment statistics

	Fermentation			
	Before		30 min after	
Replicate	1	2	1	2
# Reads	9043877	9929568	1721675	6172452
Aligned 1 time	7929681	8644569	1507903	5367701
Aligned >1 times	594057	773273	147401	570209
Alignment rate	94.25%	94.85%	96.14%	96.20%

Table 2. RNA-seq feature counts statistics

	Fermentation			
	Before		30 min after	
Replicate	1	2	1	2
# Assigned	7283046	7982570	1401053	4972090
Unassigned:				
# Unmapped	520139	511726	66371	234542
# MultiMapping	1299822	1654208	306546	1180669
# NoFeatures	610913	626057	102674	380345
# Ambiguity	35722	35942	4176	15266

Genome features abundances in RNA-seq quantification

The gffread tool was used to convert GFF annotation file to GTF format [15]. The genome feature counts in the RNA-seq reads were obtained with featureCounts function from the Subread package [16]. For further analysis, we selected only 5 columns containing genes ids and their counts in each read `head -n 50 <output from featureCounts> | cut -f 1 | cut -d "-" -f 2`.

Differential genes expression analysis and visualisation

The resulting counts file was processed using DESeq2 [17] R package [18]. DESeq2 performs differential analysis of gene expression based on a negative binomial distribution. Two R scripts for data processing with DESeq2 and heatmap building were kindly provided by our instructor Mike Raiko and can be found in Supplemental materials as well as the our own R script for volcano plot building, gene ontology (GO) and KEGG pathways analysis and visualisation.

Also for GO analysis we took the top 50 significantly changed in expression genes and obtained the gene ontology with web Gene Ontology Slim Term Mapper tool.

Results

After the single-end reads alignment procedure we got 4 alignments files, for each of those alignment statistics are presented in the Table. 1.

Genome features abundances in RNA-seq quantification resulted in 6459 genes, featureCounts statistics are presented in the Table 2. From the 6459 genes 4605 were found to highly change in the expression and for 3180 of those this change was statistically significant with respect to 0.05 adjusted *p-value* threshold. The differential expression estimate presented as a heatmap (Fig. 2) and a volcano plot (Fig. 3) with the highly ($|log_2FC| > 0.6$) and significantly (*adj. p.value* < 0.05) changed in the expression genes with found entries in AnnotationHub snapshot from 2022-10-31 annotated.

Gene ontology analysis results obtained with R script are presented at the Fig. 6c (molecular function), Fig. 6b (biological process) and Fig. 6a (cellular component) and the results obtained for the top 50 significantly changed in expression genes with the web GO Slim Term Mapper are presented in the Table 3.

Table 3. Selected most presented GO terms indentified with web GO Slim Term Mapper

GO Term	Usage in gene list, %	Genome freq. of use, %
Translational elongation	14.58	5.18
DNA repair	12.50	4,09
Chromatin organization	12.50	3.78
DNA recombination	10.42	2.93
Organelle fission	8.33	3.39
Mitotic cell cycle	8.33	4.89
Meiotic cell cycle	8.33	4.38
Regulation of cell cycle	8.33	3.93
Protein phosphorylation	6.25	1.13
DNA damage response	6.25	1.19
rRNA processing	4.17	5.07
RNA modification	4.17	2.73

Discussion

Fermentation induces biological changes in yeasts

For most of the genes accounted in RNA-seq during the alignment there is a markable change in gene expression is observed. For approximately the half of the accounted genes this change was statistically significant, which indicated that the fermentation is a rather important and voluminous process affecting many cellular functions in yeast.

This data is primarily supported by the heatmap presented at the Fig. 2. There are 2 clusters with respect to the RNA-seq reads - controls and fermentation samples clustered together in pairs. It represents that two samples under the study are indeed different but samples replicates share common patterns, as one could expect. It also makes it possible to understand that no errors were made during the process of libraries preparation.

In more details this finding is represented by the volcano plot at the Fig. 3. Significantly changed in expression genes with available annotation are captioned. This figure illustrates how big proportion of overall accounted genes were significantly changed in their expression.

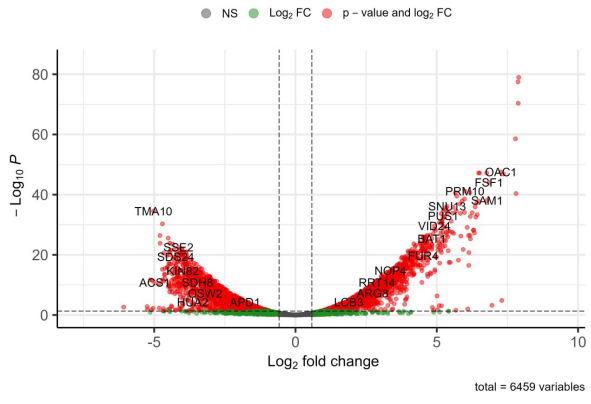


Figure 3. Volcanoplot of differential expression

Gene ontology analysis releals ribosomes and RNA at the epicenter of actions

With the GO Slim Term Mapper web tools we identified gene ontology terms for 50 most significantly changed in their expression genes. According to the Table 3, the main GO terms found were related to DNA: repair, recombination, chromatin organization. The highly represented GO term is translation elongation which is quite obvious. As we expect to see the changes in genes expression levels, firstly it should affect the translation process.

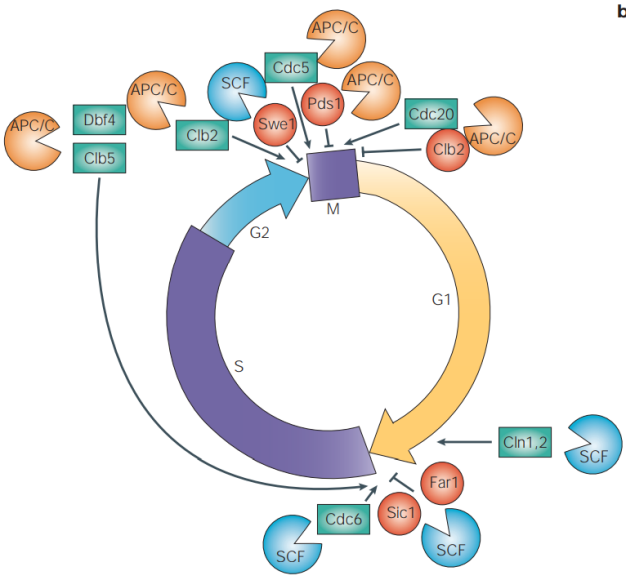


Figure 4. Yeast cell cycle regulation, [19]

We also provided more detailed gene ontology analysis with respect to 3 GO classes: "cellular component", "biological process" and "molecular function". For each of those 10 most significant GO terms are presented at the Fig. 6a, Fig. 6b and Fig. 6c respectfully. For the cellular components most of the genes we linked to the nucleolus and ribosomes. For the molecular functions most of the terms are also related to the ribosomes and RNA molecules (RNA maturation, RNA binding). For the biological process terms the picture is the same ribosomes biogenesis, RNA processing and maturation.

Viae KEGG imperceptae sunt

We analyzed the KEGG pathways that were affected by changes in gene expression. Our analysis revealed only one pathway statistically significantly affected. It is the mechanism of autophagy in yeast. In general, the changes lead to inhibition of autophagy (Fig. 10). We also found strong changes in other pathways, which are however below the threshold of statistical significance (results in Supplemental materials).

Interestingly, the processes of meiosis and cell cycle regulation are also noted to be affected (Fig. 7, Fig. 8. Such proteins as Cdc6, Cdc28, ORC, MCM, and Cdc14 appear to be upregulated. At the same time Cdc20 and Sic1, Cln1/2, SCF, Clb1/2 are downregulated. All of these globally inhibit the yeast cell's transition to division. It opens the G1/S checkpoint, promotes cell transition into the S phase, but blocks direct meiosis and chromosome segregation (Fig. 4)[19].

We also analysed the glycolisys/gluconeogenesis pathway. As one could expect for fermentation, we observe downregulation of different biochemical reaction except for the step of acetaldehyde conversion to ethanol involving NADPH-dependent alcohol dehydrogenase (Fig. 5).

Conclusion

Taken GO's and KEGG's together, yeasts become under the weather

Summarizing our results, we can conclude that the transition to fermentation does lead to enormous biochemical and biological rearrangements within yeast.

First, we note an obvious effect on the cell cycle. Yeast blocks the

passage through division; it prefers to survive fermentation while in the S phase of the cell cycle.

Second, in all GO terms we observe a change in ribosome function, a key effect on RNA including rRNA and mRNA in the cell. All this suggests that fermentation is not just another biochemical reaction cycle. Fermentation fundamentally changes the entire existence and life cycle of yeast. Perhaps every self-respecting yeast has to go through fermentation, and once it does, it will never be the same again.

Thirdly, we see that of all glycolysis (the crown of biochemical pathways) only one part becomes upregulated: ethanol production. Obviously, this is just a byproduct of the depressing need to look for energy in the absence of oxygen. But for us, this byproduct, on a par with carbon dioxide, becomes the key element of the whole story.

That's why yeasts are here - in our loaves, in our barrels. Thank them for that. Thank them for fighting and suffering for us. Skål!.

References

1. Lahue C, Madden AA, Dunn RR, Smukowski Heil C. History and domestication of Saccharomyces cerevisiae in bread baking. Frontiers in genetics 2020;11:584718.
2. Maicas S. The role of yeasts in fermentation processes. Microorganisms 2020;8(8):1142.
3. Botstein D, Chervitz SA, Cherry M. Yeast as a model organism. Science 1997;277(5330):1259–1260.
4. Mendes Ferreira A, Mendes-Faia A. The role of yeasts and lactic acid bacteria on the metabolism of organic acids during winemaking. Foods 2020;9(9):1231.
5. Rai AK, Pandey A, Sahoo D. Biotechnological potential of yeasts in functional food industry. Trends in Food Science & Technology 2019;83:129–137.
6. Freimoser FM, Rueda-Mejia MP, Tilocca B, Migheli Q. Biocontrol yeasts: mechanisms and applications. World Journal of Microbiology and Biotechnology 2019;35:1–19.
7. Kassoumi K, Kousoulou P, Sevastos D, Vamvakas SS, Papadimitriou K, Kapolos J, et al. Fermentation Efficiency of Genetically Modified Yeasts in Grapes Must. Foods 2022;11(3):413.
8. Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome biology 2013;14(9):1–13.
9. Wei Y, Tenzen T, Ji H. Joint analysis of differential gene expression in multiple studies using correlation motifs. Biostatistics 2015;16(1):31–46.
10. Everaert C, Luypaert M, Maag JL, Cheng QX, Dinger ME, Hellemans J, et al. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Scientific reports 2017;7(1):1559.
11. Li J, Hou R, Niu X, Liu R, Wang Q, Wang C, et al. Comparison of microarray and RNA-Seq analysis of mRNA expression in dermal mesenchymal stem cells. Biotechnology letters 2016;38:33–41.
12. McDermaid A, Monier B, Zhao J, Liu B, Ma Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. Briefings in bioinformatics 2019;20(6):2044–2054.
13. Kim D, Langmead B, Salzberg S. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015 Mar;(12):357–360. https://doi.org/10.1038/nmeth.3317.
14. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience 2021 02;10(2). https://doi.org/10.1093/gigascience/giab008, giab008.
15. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Research 2020;(9):304. https://doi.org/10.12688/f1000research.23297.1.
16. Yang L, Gordon S K, Wei S. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Research 2013;(41(10)):108.
17. Michael L, Anders S, Huber W. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 2014;(15):550.
18. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria;, https://www.R-project.org.
19. Reed SI. Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover. Nature reviews Molecular cell biology 2003;4(11):855–864.

Supplemental materials

Working Notes: Link
GitHub repository of the project: Transcriptomics_BI_2023

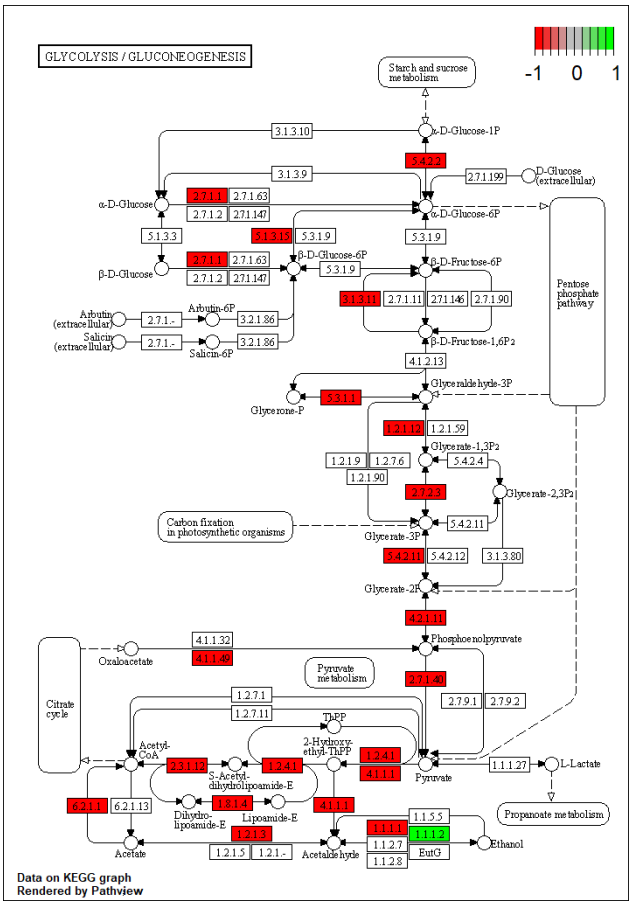
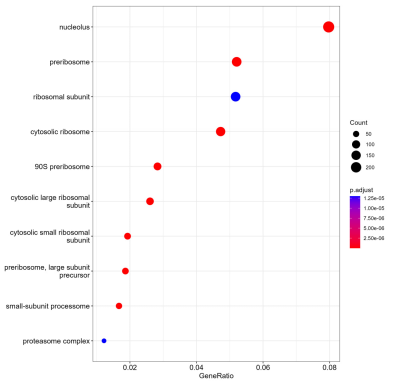
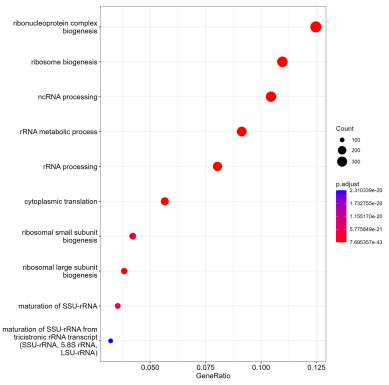


Figure 5. The result of the comparison with the KEGG base, glycolysis pathway

(a) GO cellular components terms



(b) GO biological process terms



(c) GO molecular functions terms

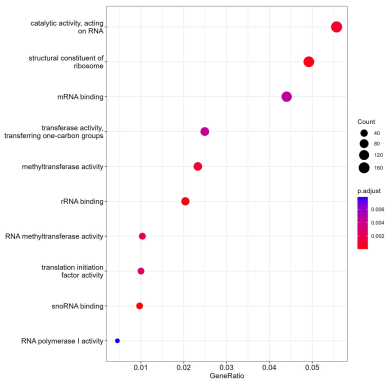


Figure 6. Dotplots for the 10 main GO terms of 3 GO classes

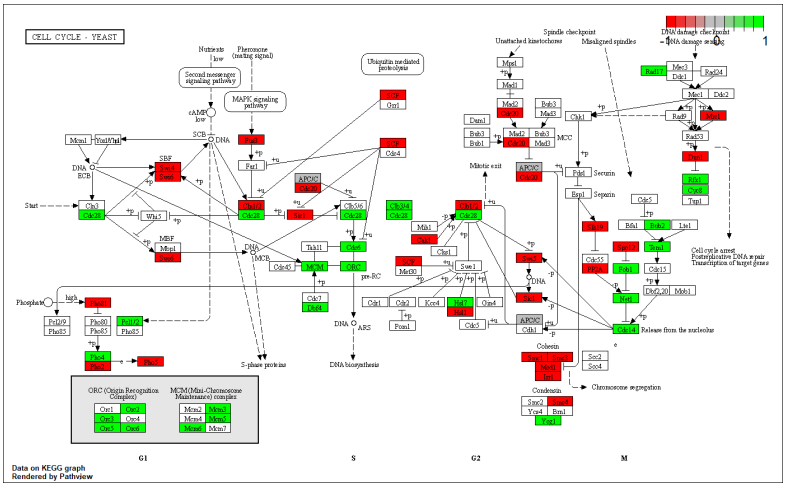


Figure 7. The result of the comparison with the KEGG base, cell cycle

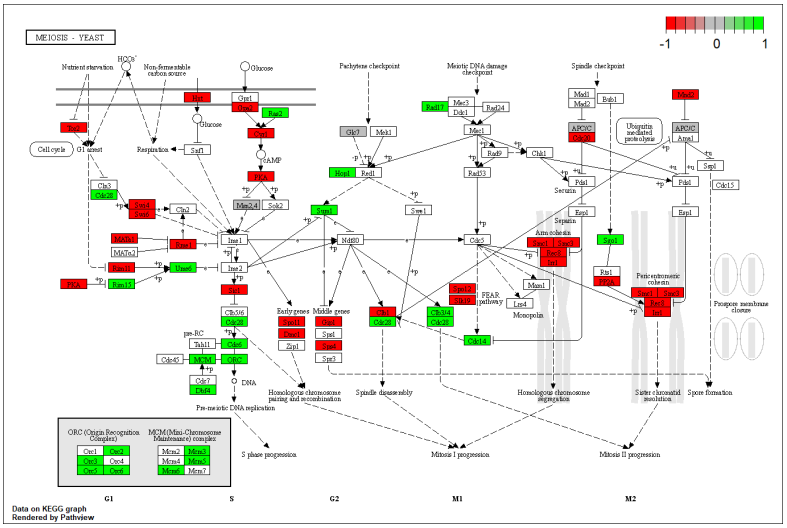


Figure 8. The result of the comparison with the KEGG base, meiosis

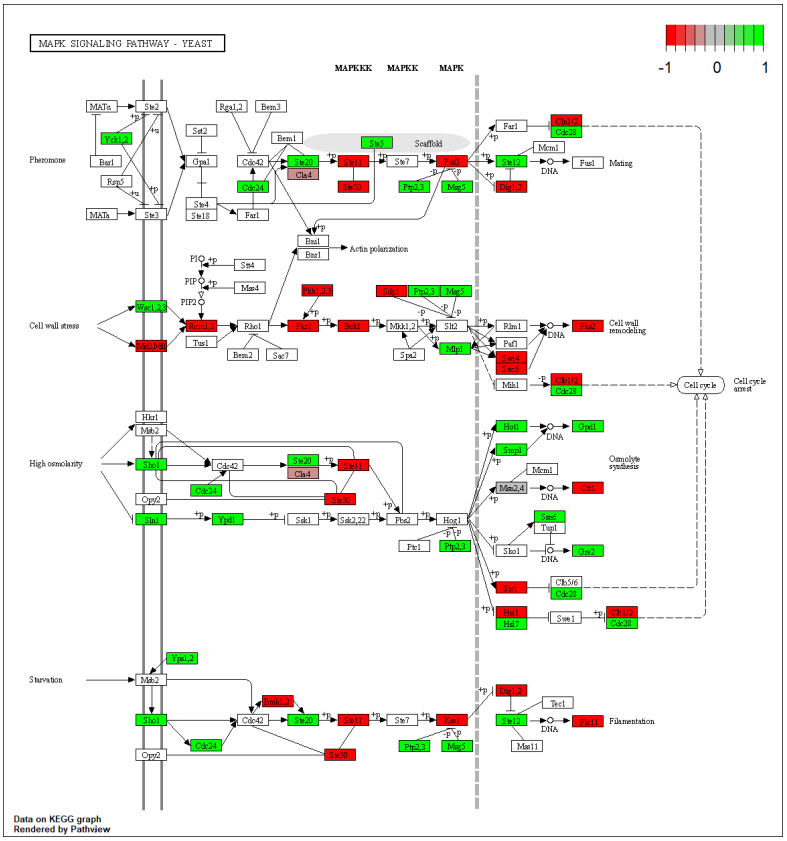


Figure 9. The result of the comparison with the KEGG base, MAPK-signaling

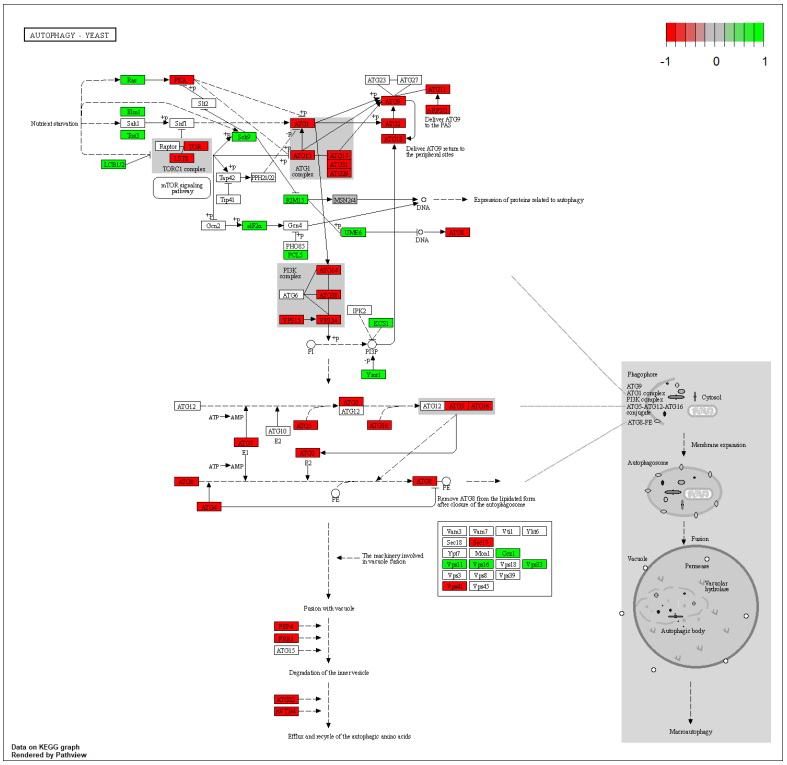


Figure 10. The result of the comparison with the KEGG base, autophagy