Mining transcriptomic data to identify *Saccharomyces cerevisiae* signatures related to improved and repressed ethanol production under fermentation

Sima Sazegari [1], Ali Niazi [2], Zahra Zinati * [3], Mohammad Hadi Eskandari[4]

1. Institute of Biotechnology, Shiraz University, Shiraz, Iran, email: sazegari.s@gmail.com
2. Institute of Biotechnology, Shiraz University, Shiraz, Iran, email: niazi@shirazu.ac.ir
3. Assistant Professor, Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University, Shiraz, Iran. Corresponding author email*: zahrazinati@shirazu.ac.ir, Tel: +987136139940, Fax: +9853546476
4. Department of Food Science and Technology, Shiraz University, Shiraz, Iran, email: eskandar@shirazu.ac.ir

Abstract

Saccharomyces cerevisiae is known for its outstanding ability to produce ethanol in industry. Identifying the dynamic of gene expression in *S. cerevisiae* in response to fermentation is required for the establishment of any ethanol production improvement program. The goal of this study was to identify the discriminative genes between improved and repressed ethanol production as well as clarifying the molecular responses to this process through mining the transcriptomic data. Through 11 machine learning based algorithms from RapidMiner employed on available microarray datasets related to yeast fermentation performance under Mg^{2+} and Cu^{2+} supplementation, 172 probe sets were identified by at least 5 AWAs. Some have been identified as being involved in carbohydrate metabolism, oxidative phosphorylation, and ethanol fermentation. Principal component analysis (PCA) and heatmap clustering were also validated the top-ranked selective probe sets. According to decision tree models, 17 roots with 100% performance were identified. *OLI1* and *CYC3* were identified as the roots with the best performance, demonstrated by the most weighting algorithms and linked to top two significant enriched pathways including porphyrin biosynthesis and oxidative phosphorylation. *ADH5* and *PDA1* are also recognized as differential top-ranked genes that contribute to ethanol production. According to the regulatory clustering analysis, *Tup1* has a significant effect on the top-ranked target genes *CYC3* and *ADH5* genes. This study provides a basic understanding of the *S. cerevisiae* cell molecular mechanism and responses to two different medium conditions (Mg^{2+} and Cu^{2+}) during the fermentation process.

Key words: *Saccharomyces cerevisiae,* fermentation, Microarray analysis, Machine learning, Principal component analysis, Hierarchical clustering

Introduction

In research and industry, *Saccharomyces cerevisiae* is used as one of the main microorganisms for bio-ethanol production. In addition to its high ethanol production capability, its stability for anaerobic fermentation and low pH tolerance facilitates its use in industry for ethanol production [1]. In terms of molecular biology, the genetics of *S. cerevisiae* is known, the genome has been sequenced, and many

genes have been functionally annotated and characterized [2,3], so genetic manipulation of this organism is well developed [4]. There are different *S. cerevisiae* industrial strains used for bioethanol production. Molecular study of industrial strains with the aim of providing insight for improved ethanol production, is of great interest due to their importance for large-scale production. *S. cerevisiae JP1* is one of the dominant strains in fermentation industry since it exhibits high temperature tolerance, stability under low pH and high fermentation rate [5]. Several researches have been conducted on the *S. cerevisiae* metabolic engineering to generate efficient ethanol producing strains [6,7]. Suji et al [8], for example used the *PHO13* deletion in conjunction with *LAD1* and *ALX1* heterologous expression to improve *S. cerevisiae* for arabinose consumption, resulting in a 3.5-fold increase in specific ethanol productivity. Furthermore, transcriptomic studies have revealed the role of genes in ethanol fermentation. Under fermentation, gene expression analysis revealed the presence of stress-response and energy-related genes in *S. cerevisiae* supplemented with Mg^{2+} [9]. Upregulation of transketolase and transaldolase genes have been reported through transcriptome analysis of engineered *S. cerevisiae* under fermentation of arabinose sugar [10]. Identifying the molecular basis and dynamics of gene expression profiles related to yeast response in improved bioethanol production conditions is critical for developing new manipulated strains with increased ethanol yield. It also shed light on the mechanisms that yeast uses to improve production.

Metal supplements are effective in the yeast metabolic pathways that produce ethanol. Among these, zinc, magnesium, manganese, and copper have been extensively researched and shown to have regulatory effects on ethanol production [11, 12, 13]. Mg^{2+} ion is involved in phosphorylation, DNA and protein synthesis, as well as cell membrane rigidity and proliferation, and it has the potential to increase ethanol accumulation through fermentation [14, 9]. Furthermore, Mg^{2+} may improve the *S. cerevisiae* tolerance to high ethanol concentration during glucose and xylose fermentation [15, 16]. Mg^{2+} medium supplementation, in particular, resulted in a 29% increase in ethanol production by regulating the expression of cell wall and membrane related genes using *S. cerevisiae* [16]. Copper is also known as a critical element for yeast biological functions, particularly in its ion form Cu^{++}. Some essential activities, such as cytochrome c oxidase, a component of oxidative phosphorylation, and

65 superoxide dismutase are dependent on Cu^{2+} [17]. Copper stress, on the other hand, caused by an excess
66 of copper, can result in ROS generation and DNA damage. At high concentrations, it also has a negative
67 impact on cell membrane stability and enzyme activity. [18]. A high copper concentration (1.5 mM)
68 inhibited cell growth, glucose and fructose consumption during fermentation by *S. cerevisiae* [19]. Few
69 studies have been conducted to investigate the effect of Cu^{2+} on the physiology and fermentation ability
70 of *S. cerevisiae* cell. As a result, despite their lack of research, Mg and Cu have the potential to modulate
71 the gene expression network involved in the fermentation process.

72 It would be possible to identify the critical genes and clarify the mechanisms involved in the ethanol
73 production process using bioinformatics-based analysis of the *S. cerevisiae* expression dataset.
74 Computational approaches for identifying key genes involved in the fermentation process could
75 elucidate the transcriptomic dynamics of yeast ethanol fermentation and reveal expression signatures
76 that could be underutilized for improved production. RapidMiner is one of the most useful and widely
77 used mining tools for data analysis [20]. Machine learning algorithms, both supervised and
78 unsupervised models, are widely used in gene expression data analysis and gene identification [21,22].
79 Different gene selection algorithms, such as Information Gain, Information Gain Ratio, rule induction,
80 SVM, and PCA, are widely used in gene expression analysis using RapidMiner. Cheng et al [23] used
81 RapidMiner to preform four machine learning weighting models on gene expression datasets related to
82 Huntington's disease, including decision tree, rule induction, random forest, and generalized linear
83 algorithms, in order to identify contributing genes to this disorder. In another study, Zinati et al [24]
84 used ten different weighting algorithms to identify the genes that differentiate between sour and acidic
85 lemon taste.

86 Valuable publicly available data on *S. cerevisiae* genome-wide expression experiments could be used
87 for functional genomic analysis through machine learning. Machine learning algorithms' discriminative
88 ability aids in revealing the underlying biological process in microarray data analysis [20]. In light of
89 the availability of such useful primary data sets and the potential of RapidMiner as an efficient tool for
90 biological data analysis, we used available microarray expression dataset related to *S. cerevisiae*
91 supplemented with Copper and Magnesium metal components under fermentation to investigate the

underlying molecular basis of fermentation used by *S. cerevisiae*. The goal of this study was to identify the critical genes contribute to discriminate the improved (Mg^{2+} treatment) and low ethanol production (Cu^{2+} treatment at toxic concentration), as well as to elucidate the transcriptomic response of *S. cerevisiae* under these two conditions. *S. cerevisiae* transcriptome analysis using data mining and machine learning by both supervised and unsupervised models was used in this study as a novel approach to identify the underlying gene regulation mechanisms that can be used to optimize fermentation performance.

Materials and Methods

Data Collection

For this study available microarray datasets related to yeast fermentation performance under Mg^{2+} (500 mg/L) or Cu^{2+} (1 mg/L) supplementation was used. Microarray data of the industrial yeast *S. cerevisiae* JP1 strain downloaded from the GEO repository of the NCBI database (GEO number: GSE75803) was used. To meet the research objective, the probe sets with significant differential expression (concomitant Adj. $p < 0.05$ and $B \geq 3$) were chosen for this study.

Data cleaning

We used RapidMiner software (RapidMiner Studio 7.6) [25] to enter the 6300-differential expressed probe sets as numerical features, as well as high and low bioethanol as class features. For better processing, inefficient or redundant probe sets with less than or equal to a given standard deviation (SD) threshold (0.1), as well as correlated probe sets (correlation ≥ 0.95), were carefully removed from the dataset. The resulting list, which only contained efficient probe sets, was designated as the Final Cleaned (FCdb) database.

Attribute weighting algorithms

Eleven attribute weighting algorithms with cut-off ≥ 0.7 were used in the FCdb to identify the most effective probe sets contributing to discriminate ethanol content. Weights close to 1 indicate that a specific probe set in ethanol content is more important. The main probe sets were those determined by

the majority of AWAs (intersection of the weighing method). The attribute weighting algorithms used in this investigation, as well as the statistical background description for each one, are as follows (RapidMiner Studio 7.6):

Weight by Information Gain and Information Gain Ratio

This algorithm is a well-established superior method for gene selection in microarray data analysis [26,27]. In this method, the attributes (probe sets) are weighted according to their class label (high or low ethanol production).

Weight by Rule

Based on a single rule and the relationship between attributes (genes) and considering the errors, the weight of each attribute is measured through rule algorithm [28] and is used as a selective method for microarray analysis.

Weight by Deviation, Weight by Correlation and Weight by Chi Squared Statistic

The standard deviation of attributes is used as a weighting parameter in the deviation weighting method. The correlation method, on the other hand, weighs the label attributes based on the correlation. In addition, for labeling the attributes, we used the Chi Squared Statistic weighting algorithm, which takes the Chi squared into account.

Weight by Gini Index and Weight by Uncertainty

Due to the label attribute in this model, the weight of attributes is determined by measuring the Gini coefficient as an inequality index of sample data. According to each attribute, the lower the Gini index of the attribute, the more equal dispersion among attributes is considered. The weight for uncertainty model, on the other hand, is determined using the symmetrical uncertainty due to the class attribute.

Weight by Relief

This model is one of the most reliable algorithms for weighting genes because it is based on the determination of values between probe sets of the same and different classes in a short distance.

Weight by SVM

SVM is one of the most powerful classification models for gene expression analysis [21]. The SVM method weighs attributes using the coefficients of the normal vector of a linear SVM.

Weight by PCA

This model performs attribute weighting due to the class attribute based on the component number parameter of PCA and the value of the components.

Decision tree models

Eleven new datasets were generated using the entire probe sets with weight >0.70. They were annotated based on the models used for attribute weighting (Relief, Information gain, Uncertainty, Information gain ratio, Chi Squared, Rule, Correlation, Deviation, SVM and PCA, Gini index). Random Tree, Decision Tree, Decision Stump, and Random Forest were the tree induction models used for 12 datasets (FCdb and 11 datasets produced by specific weighting algorithms). Each model had four criteria (Gini Index, Gain Ratio Information Gain, and Accuracy). We used a ten-fold validation algorithm with appropriate sampling to create trees with RapidMiner. The performance of the model was evaluated and used to compare various models based on the accuracy of each model in identifying the target variable (high and low bioethanol content) and according to the attribute variables (normal expression of the probe set). Performance is expressed as a measure of model accuracy in this case. We calculated the accuracy by dividing the number of correct predictions by the total number of samples. The value of the attribute accuracy that is expected to be the same as the value of the labeled attribute is referred to as the correct prediction. These models were used with a minimum gain of 0.1 to obtain a split and a maximum tree depth of 20. For pruning 0.25 confidence level was considered with a pessimistic error calculation.

Unsupervised analysis of the top ranked probe sets derived by supervised AWAs

Unsupervised principal component analysis (PCA) and hierarchical clustering heatmap were used to evaluate the power of top-ranked probe sets which differentiate the fermentation under different

supplementation treatments. For unsupervised analysis, a web-based tool Clustvis (https://biit.cs.ut.ee/clustvis/) was used [30]. The PCA analysis was carried out in the PCA Methods R package using unit variance scaling on rows and Singular Value Decomposition (SVD) with the imputation method. The clustering heatmap was created with the pheatmap R package (version 0.7.7). The clustering heatmap was constructed using correlation, Pearson correlation subtracted from 1, and the average distance of all possible pairs [31].

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis

The pathway enrichment analysis was carried out using YeastEnrichr (https://maayanlab.cloud/YeastEnrichr/) [30,31]. The biochemical pathways related to key probe sets were identified using the KEGG2019 database. Pathways with p-value < 0.1 were considered significant.

Exploring for transcription factors among top-ranked genes and Regulator cluster analysis

We used yeastract database (http://www.yeastract.com/formrankbyhomotf.php) to identify transcription factors (TFs) among the 172 probe sets identified by at least 5 attribute weighting algorithms [32]. The TFs and their target genes were identified using this tool based on DNA binding sites and expression evidence. Furthermore, we used the regulator DB database (http://wyrickbioinfo2.smb.wsu.edu/cgi-bin/RegulatorDB/cgi/home.pl) to run regulator cluster diagram to determine the regulatory effect of the identified TFs on the target genes [33,34]. It provides data on mutant regulator expression for selected regulators and target genes.

Results

Ranking probe sets by AWAs

After cleaning 6300 probe sets by RapidMiner, we obtained 1813 probe sets. Eleven AWAs were used to identify informative probe sets. Following AWAs analysis, 172 probe sets were identified by at least 5 attribute weighting algorithms (Supplementary File, sheet S1). Furthermore, there were distinct probe

191 sets classified by at least five algorithms that respond discriminatively to supplement treatment and/or
192 are particularly related to ethanol production during fermentation. Sheet S2 of the Supplementary File
193 contains the probe sets as well as the AWAs used to identify the probe sets. Some of the informative
194 probe sets were recognized to be involved in carbohydrate metabolism, TCA cycle, oxidative
195 phosphorylation, and ethanol fermentation while others were related to stress responses, cell membrane
196 structure, and cell growth which could be indirectly effective in ethanol production. Some of the top
197 informative genes are presented in (Table 1).

probe sets	Standard Gene Name	AWAs Names	AWAs number	Gene Name
A_06_P4554	MRP8	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, PCA, Relief, Rule, SVM, Uncertainty	10	Uncharacterized, response to stress
A_06_P1016	OLI1	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, PCA, Relief, Rule, SVM, Uncertainty	10	ATP synthase subunit 9, mitochondrial;OLI1;ortholog
A_06_P1397	ADH5	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, PCA, Rule, SVM, Uncertainty	9	Alcohol dehydrogenase 5;ADH5;ortholog
A_06_P1238	PKC1	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, PCA, Rule, SVM	8	Protein serine/threonine kinase
A_06_P3384	GTR2	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, PCA, Rule, SVM, Uncertainty	9	GTP-binding protein
A_06_P1063	CYC3	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, Relief, Rule, SVM, Uncertainty	9	Cytochrome c heme lyase
A_06_P1003	COX1	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, Rule, SVM, Uncertainty	8	cytochrome c oxidase
A_06_P2810	PDA1	Chi Square Statistic, Correlation, Gini Index, Information Gain, Information Gain ratio, Rule, SVM	7	Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial;PDA1;ortholog
A_06_P2931	QCR6	Chi Square Statistic, Correlation, PCA, Rule, SVM	5	Cytochrome b-c1 complex subunit 6;QCR6;ortholog
A_06_P6820	ALD6	Chi Square Statistic, PCA, Rule, Uncertainty	4	Magnesium-activated aldehyde dehydrogenase, cytosolic;ALD6;ortholog

198 **Table 1.** Some of the informative probe sets identified by at least AWAs.

199

200

201 **Decision Tree models**

202 The decision tree models were used to achieve pattern recognition between important genes as well as
203 with the genes with the highest distinguishing power. The lowest and highest performances were 0%
204 and 100%, respectively (Supplementary File, sheet S3). There were 17 probe sets with 100%
205 performance in the roots of decision tree models (Table 2).

206 **Table 2.** Decision Tree models roots identified as exhibited 100% performance.

elements	GENE_SYMBOL	AWS	DESCRIPTION
A_06_P1016	OLI1	10	BioProcess=ATP synthesis coupled proton transport
A_06_P2475	CWC21	9	BioProcess=biological_process unknown
A_06_P1002	ORF:Q0017	9	BioProcess=biological_process unknown
A_06_P1034	CYS3	9	BioProcess=sulfur amino acid metabolism*
A_06_P1131	HTB2	8	BioProcess=chromatin assembly/disassembly
A_06_P1068	KRE23	9	BioProcess=biological_process unknown
A_06_P2984	CGR1	9	BioProcess=rRNA processing*
A_06_P1298	ORF:YBR051W	9	BioProcess=biological_process unknown
A_06_P1524	ORF:YBR270C	8	BioProcess=biological_process unknown
A_06_P3287	ORF:YGR067C	9	BioProcess=biological_process unknown
A_06_P1063	CYC3	9	BioProcess=not yet annotated
A_06_P2051	RPS13	9	BioProcess=protein biosynthesis
A_06_P1967	OST4	10	BioProcess=not yet annotated
A_06_P1049	ORF:YAL027W	9	BioProcess=biological_process unknown
A_06_P1003	COX1	8	BioProcess=aerobic respiration
A_06_P2023	ORF:YDR036C	7	BioProcess=biological_process unknown
A_06_P1023	ORF:Q0297	8	BioProcess=biological_process unknown

207

Unsupervised Analysis

As a complementary confirmation, the 172 top ranked probe sets were validated using PCA and hierarchical clustering heatmap, identified with supervised attribute weighting models. According to the results, the 172 significant probe sets could accurately differentiate between two different fermentation conditions, thus confirming the significance and accuracy of the identified probe sets (Fig. 1). In particular, the captured variances with the first two components on all recognized 6031 probe sets and informative 172 probe sets were up to 74% and 50%, respectively. Furthermore, it could efficiently separate informative 172 probe sets under Cu^{2+} or Mg^{2+} supplementation in the hierarchical clustering heat map, (Fig. 1). In total, 64 probe sets were up and down regulated by Mg^{2+} and Cu^{++}, while 108 probe sets were up and down regulated through Cu and Mg supplementation, respectively (Fig. 2).

Fig. 1. Two-dimensional plot related to the first two principal components. GSM1968101, GSM1968110, GSM1968100 and GSM1968108 are samples related to Mg^{2+} supplementation. GSM1968106, GSM1968114, GSM1968103 and GSM1968112 are samples related to Cu^{2+} supplementation.

Fig. 2. The heatmap related to 172 probe sets which were recognized by at least 5 attribute weighting algorithms (AWAs). Each row corresponds to the different samples including Mg^{2+} (high ethanol production) and Cu^{2+} supplementation (repressed ethanol production). Columns exhibits hierarchically clustered probe sets. The normalized intensity expressions of probe sets were shown as a color scale. The up and down-expression levels were represented as red and blue scales, respectively.

Pathway enrichment analysis of genes

Significant enriched pathways such as Porphyrin metabolism, Oxidative phosphorylation, Glycolysis, Amino sugar and nucleotide sugar metabolism, Cell cycle, Meiosis and Citrate cycle (TCA cycle) were identified using the KEGG enrichment analysis. The enriched pathways and the related genes are presented in Table 3.

232 **Table 3.** KEGG enrichment analysis of 172 probe sets. The significant pathways with adjusted P-
233 value < 0.1 are represented.

Term	Adjusted P-value	Genes
Porphyrin and chlorophyll metabolism	0.000582985	HEM2;HEM12;CYC3;YFH1
Oxidative phosphorylation	0.007415084	OLI1;QCR6;ATP6;COX1;ATP2
Endocytosis	0.007415084	CAP1;APL3;LAS17;ARC15;VPS25
RNA degradation	0.025376563	POP2;RRP42;SSQ1;CCR4
Meiosis	0.049279656	CLN3;HMRA2;MSN4;APC9;TPD3
Autophagy	0.049279656	KCS1;VPS8;MSN4;PEP4
Ubiquitin mediated proteolysis	0.049279656	UBC13;UBC6;APC9
Protein processing in endoplasmic reticulum	0.049279656	OST4;UBC6;PDI1;SSE2
Glycolysis / Gluconeogenesis	0.064176371	PDA1;PGM2;ADH5
Galactose metabolism	0.072088565	GAL7;PGM2
Phosphatidylinositol signaling system	0.072088565	KCS1;PKC1
Amino sugar and nucleotide sugar metabolism	0.075323668	GAL7;PGM2
Spliceosome	0.075323668	PRP43;ECM2;PRP8
MAPK signaling pathway	0.072088565	TUP1;MKC7;MSN4;PKC1
Pentose phosphate pathway	0.075323668	SOL4;PGM2
Alanine, aspartate and glutamate metabolism	0.075323668	GDH3;NIT3
Cell cycle	0.075323668	CLN3;TUP1;APC9;TPD3
Citrate cycle (TCA cycle)	0.075323668	PDA1;LSC2
Ribosome biogenesis in eukaryotes	0.075323668	UTP15;CKB1;RIO1
Glycine, serine and threonine metabolism	0.075323668	SER1;CYS3

234

235 **Identification of transcription factors and their targets**

236 Among the 172 informative probe sets identified by yeastract analysis were seven transcription factors:
237 *YGR067C, HAP4, NRG2, TUP1, TOS8, MSN4, and PDC2*. Surprisingly, the targets of the identified
238 transcription factors were discovered among the 172 genes identified by RapidMiner analysis and
239 ranked by at least 5 algorithms (Supplementary File, sheet S5). These findings support the AWAs'
240 ability to correctly identify top-ranked probe sets. Furthermore, regulator clustering related to TFs and
241 their targets (both ranked by at least eight algorithms) was performed to demonstrate the effect of top-
242 ranked TFs on top-ranked target genes based on the transcription factors mutants. The results showed
243 that *Hap4p, Tup1,* and *TOS8* mutants resulted in different ratios of up and down-regulation of target
244 genes (Fig. 3). Although *Hap4* and *TOS8* resulted in down or up regulation of target genes, their effect

on none of target genes was significant. According to the findings, the Tup1 transcription factor has the greatest impact on the target genes expression. The Tup1 knocked out mutant significantly induce the expression of *CYC3 (YAL039C),* while causing highest level of down regulation of *YBL11C*, *ADH5 (YBR145W)*.

Fig. 3. The regulatory clustering heatmap related to genes targeted by identified transcription factors *Hap4p* and *Tup1* and *TOS8*. The cluster is represented as the log mRNA ratio of each target gene in each regulator mutant.

Discussion

In this study, machine learning and decision tree models were used to analyze the transcriptome of *S. cerevisiae* during the fermentation process in two conditions: repressed ethanol production and high ethanol production supplemented with Cu^{2+} and Mg^{2+}. Indeed, for the most accurate prediction methods, we used both supervised and unsupervised models. In summary, we used 11 supervised models to achieve high accuracy results. In addition, a PCA analysis as an unsupervised model and a hierarchical clustering heat map were used to validate the 172 top-ranked probe sets identified by supervised-based models. Furthermore, we used pathway enrichment, transcription factor and regulatory analysis to validate the machine learning analysis results (Fig. 4). According to RapidMiner-assisted analysis, some probe sets were identified as playing a distinct role in ethanol production. Nonetheless, it should be noted that the function of some identified probe sets has not yet been clarified, despite the fact that they may be critical in ethanol production. *ADH5* or Alcohol dehydrogenase, which was weighted by 9 algorithms and classified in Glycolysis / Gluconeogenesis by KEGG enrichment analysis, contributes to ethanol production by reducing acetaldehyde to ethanol [35]. *OLI1* is distinguished by ten algorithms and is rich in Oxidative phosphorylation term, which encodes F0-ATP synthase subunit c and generates ATP in yeast mitochondria [36]. Metal ions, such as Cu^{++}, are known to have a negative effect on mitochondrial respiratory components, as it slowed the respiratory chain in *PC12* and liver cells at toxic doses [37,38]. That is most likely the main reason for the down regulation of *OLI1*, which is an important component of the oxidative phosphorylation pathway when exposed to toxic Cu. RNA-seq analysis revealed that this gene was enriched as a significant gene between the wild and high glucose

tolerant mutant strains of *S. cerevisiae* [39]. In addition to this gene, *COX1* has an AWA weight of 8 and is involved in the final electron chain reaction in the respiratory system [40]. It encodes one of the cytochrome c oxidase subunits and, like OLI1, has been shown to be repressed by Cu^{2+} treatment. *PDAI* encodes alpha subunit of pyruvate dehydrogenase and converts the pyruvate to acetyl-CoA through oxidative decarboxylation [41]. This gene was found to be enriched in the Glycolysis/Gluconeogenesis pathway by seven weighting algorithms used in this study. *PDAI* directs the pyruvate metabolism to Acetyl- COA in mitochondria to provide the TCA cycle substrate. In other words, directing the pyruvate to TCA cycle *PDAI* keeps pyruvate from being consumed in the fermentation process or ethanol production. *PDAI* was down regulated in Mg-containing medium, which accounts for improved ethanol production, and was upregulated in the repressed fermentation condition, by Cu. *QCR6* is a subunit of cytochrome bc1 complex and contributes to oxidative phosphorylation. Cytochrome C is known to be activated by Cu metal ion [42]. *QCR6* was up regulated, as expected, by Cu supplementation. Similarly, in Pichia stipites, cytochrome bc1 disruption resulted in increased ethanol production. [43]. Granados-Arvizu et al [44] also concluded that cytochrome bc_1 complex repression would be a promising way to enhance ethanol production in *Saccharomyces stipitis*. *ALD6* or Aldehyde dehydrogenases is activated by Mg and have a distinct role in the formation of acetate from pyruvate in an alternate pyruvate dehydrogenase bypass pathway [45]. ALD6 expression was found to be increased with Mg supplementation, which corresponded to the activation of this enzyme by Mg++. Since it consumes the acetaldehyde source that ADH enzymes can use to produce ethanol, deleting *ALD6* via Crisper/CAS 9 genome editing resulted in increased ethanol production [46].

Fig. 4. The schematic illustrates the methodology of the study with summarized results.

Based on decision tree analysis, 17 identified roots performed flawlessly, some of which have unknown molecular functions and have yet to be characterized. Surprisingly, *OLI1* and *CYC3* were identified by the highest attribute weighting algorithms (10 and 9), were enriched in the second most important biochemical pathway, and were also identified as decision tree model roots with 100% performance. *COX1* is also shown as a complete root, but it is identified using 8 weighting algorithms. As previously stated, *OLI1* is an F0-ATP synthase subunit c that contributes to the electron transport chain. *CYC3* is

299 also known as Cytochrome c heme lyase and has a strong sensitivity to ethanol. Indeed, the null mutants
300 for this gene showed ethanol sensitivity. Both of these genes are involved in ATP generation and are
301 up regulated in Mg supplemented medium. Nonetheless, it has been established that Mg^{2+} has an effect
302 on energy metabolism and ATP production in the cell [47].

303 Confirming the results of the AWAs and decision tree models analysis through RapidMiner, the Kegg
304 pathway analysis showed that two most significant terms, porphyrin biosynthesis and oxidative
305 phosphorylation, were enriched in *CYC3* and *OLI1* and *COX1*. Cell cycle and division, as well as
306 Ribosome biogenesis, are identified as significant terms in the KEGG pathway enrichment. They may
307 have an impact on ethanol production even though they do not directly contribute to the fermentation
308 bioprocess. For example, in addition to its role in yeast cell growth and proliferation, which affects
309 ethanol production, ribosome biogenesis is predicted to be associated with fermentation, and some
310 related genes, such as *SFP1* are thought to be involved in glycolysis control as well [35,48].
311 Nonetheless, significant phosphatidylinositol signaling and MAPK signaling pathways identified in this
312 study by enrichment analysis were reported to be responsible for cell proliferation/growth regulation
313 and critical for stress responses [49,50]. *PKC1* which was attributed by 8 algorithms and remarkably
314 enriched in phosphatidylinositol signaling system is a serine/threonine kinase which is suggested to
315 have role in response to copper toxicity since it was upregulated in Cu^{2+} supplementation or reduced
316 ethanol production according to heatmap clustering. Confirming this finding, Zhou et al [49] reported
317 that 5-hydroxymethyl-2-furaldehyde, which is toxic to industrial fermentative *S. cerevisiae* strain,
318 increases the expression of *PKC1* gene. Furthermore, according to AWAs analysis, some genes are
319 involved in stress responses, cell growth and proliferation, protein synthesis, fatty acids and lipid
320 metabolism, all of which may contribute to ethanol production efficiency. *MRP8* was assigned by ten
321 algorithms as a response to cell wall stress, and its expression has been reported to be induced under
322 stress conditions [51]. Its function, however, is unknown. Cu supplementation induces the expression
323 of this gene in response to the stress condition caused by copper. *GTR2*, a GTPase subunit, was weighted
324 using nine algorithms. It is suggested in this study that it contributes to tolerance response to CU

inhibitor because it was up regulated by copper. As an implication for this result, the null mutant related to *GTR* gene showed decreased resistance to Zn metal at inhibitory amount [52].

According to the crucial role of TFs in gene expression regulation and to confirm the results obtained from attribute weighting algorithms analysis, the TFs and their targets were explored among 172 probe sets. According to the regulatory clustering analysis, *Tup1* has a significant effect on the top-ranked target genes. *Tup1* is a transcriptional repressor in *S. cerevisiae* has the ability to repress target genes via various molecular mechanisms, and it contributes to carbon catabolite repression of transcription by glucose [53, 54]. Regarding the results of this study on regulatory clustering analysis, the *Tup1* mutant caused decreased expression in some of the target genes and up regulation in others. In other words, the deletion of *Tup1* resulted in downregulation of *YBL111C,* whose biological function is unknown and *YBR145W* (*ADH5*) at most. The *ADH5* gene has also been identified as the top-ranked gene with 9 AWAs through RapidMiner analysis. On the other hand, the *TUP1* knock out resulted in significant upregulation of *YAL039C* (*CYC3*). Indeed, the *CYC3* gene, which was confirmed by the greatest number of AWAs and a decision tree model, was also shown to be a top target of the transcription factor involved in ethanol production responses in this study. *Hap4* is a transcription factor involved in the regulation of the respiratory genes' expression and ethanol tolerance. The role of *TUP1* and *HAP4* in glucose fermentation have been studied and recently confirmed in thermotolerant yeast, *Ogataea polymorpha* [54]. Moreover, the overexpression of *HAP4* gene caused enhanced glucose consumption and ethanol production in *S. cerevisiae* [55,56]. In this study, the HAP4 gene was also identified as top-ranked gene attributed by nine AWAs. Although the results confirm its involvement in the identified probe sets regulation, it does not demonstrate significant up or down-regulation effect on the target genes. Overall, *OLI1, CYC3, COX1 and ADH5* were ranked as the most critical genes in the differentiation of two improved and repressed ethanol production conditions because they were the most frequently identified genes across analyses. These important findings shed light on the complex pathways and regulatory responses that genes use to contribute to ethanol production. However, additional experimental analysis could fully clarify the results. Overall, the findings of this study could

be used to further investigate the possibility of improving ethanol through overexpression or knock out strategies. Furthermore, additional experimental testing to confirm the findings is strongly advised.

Abbreviations

(PCA) Principal component analysis, (KEGG) Kyoto Encyclopedia of Genes and Genomes, *ADH5* (Alcohol dehydrogenase), (CYC3) Cytochrome c heme lyase

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files}.

Competing interests

The authors declare that they have no competing interests

Funding

No funding

Authors' contributions

SS, ZZ and AN contributed to the study conception and design. ZZ and SS analyzed the Data. The manuscript was written by SS. ME, ZZ and AN contribute to scientific revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Supplementary data

The Supplementary File, sheet S1 and S2. The list of the probe sets along with the attribute weighting algorithms (AWAs) through which the probe sets were identified.

The Supplementary File, sheet S3 and S4. The decision tree models (performance and roots).

The Supplementary File, sheet S5. The list of identified transcription factors and their targets.

References

[1] Moysés DN, Reis VCB, Almeida JRMD, Moraes LMPD, Torres FAG (2016) Xylose fermentation by *Saccharomyces cerevisiae*: challenges and prospects, Int. J. Mol. Sci. 17, 207. https://doi.org/10.3390/ijms17030207.

[2] Louis E (2011) *Saccharomyces cerevisiae*: gene annotation and genome variability, state of the art through comparative genomics, Methods Mol. Biol. 759: 31-40. https://doi.org/10.1007/978-1-61779-173-4_2.

[3] Proux-Wéra E, Armisén D, Byrne KP, Wolfe KH (2012) A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach, BMC bioinform. 13, 237. https://doi.org/10.1186.

[4] Gohil N, Panchasara H, Patel S, Ramírez-García R, Singh V (2017) Book review: recent advances in yeast metabolic engineering, Front. Bioeng. Biotechnol. 5, 71. https://doi.org/10.3389/fbioe.2017.00071.

[5] da Silva Filho EA, de Melo HF, Antunes DF, dos Santos SK, do Monte Resende A, Simoes, DA, de Morais MAJr (2005) Isolation by genetic and physiological characteristics of a fuel-ethanol fermentative Saccharomyces cerevisiae strain with potential for genetic manipulation, J Ind Microbiol Biotechnol 32: 481–486. https://doi.org/10.1007/s10295-005-0027-6.

[6] Kobayashi Y, Sahara T, Ohgiya S, Kamagata Y, Fujimori KE (2018) Systematic optimization of gene expression of pentose phosphate pathway enhances ethanol production from a glucose/xylose mixed medium in a recombinant *Saccharomyces cerevisiae*, AMB Express. 8 139. https://doi.10.1186/s13568-018-0670-8.

[7] Liu K, Yuan X, Liang L, Fang J, Chen Y, He W, Xue T (2019) Using CRISPR/Cas9 for multiplex genome engineering to optimize the ethanol metabolic pathway in *Saccharomyces cerevisiae*, Biochem. Eng. J. 145: 120-126a. https://doi.org/10.1016/j.bej.2019.02.017.

[8] Jeong S, Shon D, Liu JC, Kim KH, Shin M, Kim SR (2019) Deletion of PHO13 improves aerobic l-arabinose fermentation in engineered *Saccharomyces cerevisiae*, J. Ind. Microbiol. Biotechnol. 46: 1725-1731. https://doi.org/10.1007/s10295-019-02233-y.

[9] de Souza RB, Silva RK, Ferreira DS, Junior SDSLP, de Barros Pita W, de Morais Junior MA (2016) Magnesium ions in yeast: setting free the metabolism from glucose catabolite repression, Metallomics. 8: 1193-1203. https://doi.org/10.1039/c6mt00157b.

[10] Wisselink HW, Cipollina C, Oud B, Crimi B, Heijnen JJJT, Pronk JJ, Van Maris AJ (2010) Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered *Saccharomyces cerevisiae*, Metab. Eng. 12: 537-551. https://doi.org/10.1016/j.ymben.2010.08.003.

[11] Pejin JD, Mojović LV, Pejin DJ, Kocić-Tanackov SD, Savić DS, Nikolić SB, Djukić-Vuković, AP (2015) Bioethanol production from triticale by simultaneous saccharification and fermentation with magnesium or calcium ions addition, Fuel. 142: 58-64. (2015). https://doi.org/10.1016/j.fuel.2014.10.077.

[12] Taloria D, Samanta S, Das S, Pututunda C (2012) Increase in bioethanol production by random UV mutagenesis of *S. cerevisiae* and by addition of zinc ions in the alcohol production media, APCBEE Procedia. 2: 43-49. https://doi.org/10.1016/j.apcbee.2012.06.009.

[13] Ko JK, Um Y, Lee SM (2016) Effect of manganese ions on ethanol fermentation by xylose isomerase expressing *Saccharomyces cerevisiae* under acetic acid stress, Bioresour. Technol. 222: 422-430. https://doi.org/10.1016/j.biortech.2016.09.130.

[14] Wang FQ, Gao C.J, Yang CY, Xu P (2007) Optimization of an ethanol production medium in very high gravity fermentation, Biotechnol. Lett. 29: 233–236. https://doi.org/10.1007/s10529-006-9220-6.

[15] Zhao XQ, Bai Fw (2012) Zinc and yeast stress tolerance: Micronutrient plays a big role. J. Biotechnol. 158: 176–183. https://doi.org/10.1016/j.jbiotec.2011.06.038.

[16] Ku I, Ku S, Sakamoto T, Hasunuma T, Zhao X, Kondo A (2014) Zinc, magnesium, and calcium ion supplementation confers tolerance to acetic acid stress in industrial *Saccharomyces cerevisi*ae utilizing xylose, Biotechnol. J. 12: 1519-1525. https://doi.org/10.1002/biot.201300553.

[17] Shi H, Jiang Y, Yang Y, Peng Y, Li C (2020). Copper metabolism in Saccharomyces cerevisiae: An update. *Biometals*, pp.1-12.

[18] Farrés M, Piña B, Tauler R (2016) LC-MS based metabolomics and chemometrics study of the toxic effects of copper on *Saccharomyces cerevisiae*, Metallomics. 8: 790-798. https://doi.org/10.1039/C6MT00021E.

435 [19] Lihua W, Wang R, Zhan J, Huang W (2020) High levels of copper retard the growth of
436 *Saccharomyces cerevisiae* by altering cellular morphology and reducing its potential for ethanolic
437 fermentation, Int. J. Food Sci. Technol. 2720-2731. https://doi.org/10.1111/ijfs.14903.
438 [20] Teixeira DAR (2014) A Computational Platform for Gene Expression Analysis, MSc Thesis,
439 Informatics and Computing Engineering (MIEC), FEUP. https://hdl.handle.net/10216/75202.
440 [21] Abusamra H (2013) A comparative study of feature selection and classification methods for gene
441 expression data of glioma, Procedia Comput. Sci. 23 5-14. https://doi.org/10.1016/j.procs.2013.10.003.
442 [22] Zinati Z, Alemzadeh A, KayvanJoo AH (2016) Computational approaches for classification and
443 prediction of P-type ATPase substrate specificity in Arabidopsis, Physiol. Mol. Biol. Plants. 22: 163–
444 174. https://doi.org/10.1007/S12298-016-0351-5.
445 [23] Jack C, Liu H, Lin W, Tsai F (2020) Identification of contributing genes of Huntington's disease
446 by machine learning, BMC Medical Genom.13: 1-11. https://doi.org/10.1186/s12920-020-00822-w.
447 [24] Zinati Z, Sazegari S, Amin H, Tahmasebi A (2021) Mining transcriptome data to identify genes
448 and pathways related to lemon taste using supervised and unsupervised data learning methods, Hortic.
449 Environ. Biotechnol. 19. https://doi.org/10.1007/s13580-021-00337-y.
450 [25] RapidMiner Documentation. https://docs.rapidminer.com/. Accessed 10 Sep 2017.
451 [26] Cho S, Won H (2003) Machine learning in DNA microarray analysis for cancer classification, In
452 Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics. 189-198.
453 [27] Xing EP, Jordan MI, Karp RM (2001) Feature selection for high-dimensional genomic microarray
454 data, *Proc. 18th Int'l Conf. Machine Learning*. 601-608.
455 [28] Alagukumar S, Lawrance R (2015) A selective analysis of microarray data using association rule
456 mining, Procedia Comput. Sci, 47: 3-12. https://doi.org/10.1016/j.procs.2015.03.177.
457 [29] Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using
458 Principal Component Analysis and heatmap, Nucleic Acids Res. 43. W566-70.
459 https://doi.org/10.1093/NAR/GKV468.
460 [30] Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A (2013) Enrichr:
461 interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinformatics. 14
462 https://doi.org/10.1186/1471-2105-14-128.

463 [31] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL,
464 Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A (2019)
465 Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res. 44
466 w90-w97. https://doi.org/10.1093/nar/gkw377.
467 [32] Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, Galocha M, Godinho CP,
468 Martins LC, Bourbon N, Mota MN (2020) YEASTRACT+: a portal for cross-species comparative
469 genomics of transcription regulation in yeasts, Nucleic Acids Res. 48. D642-D649.
470 https://doi.org/10.1093/nar/gkz859.
471 [33] Hu Z, Killion PJ, Iyer VR (2007) Genetic reconstruction of a functional transcriptional regulatory
472 network. Nat. Genet. 39*:* 683-687. https://www.nature.com/articles/ng2012.
473 [34] Reimand J, Vaquerizas JM, Todd A.E, Vilo J, (2010) Luscombe, N.M. Comprehensive reanalysis
474 of transcription factor knockout expression data in *Saccharomyces cerevisiae* reveals many new targets,
475 Nucleic Acids Res. 38*:* 4768-4777. https://doi.org/10.1093/nar/gkq232.
476 [35] Davydenko S, Meledina T, Mittenberg A, Shabelnikov S, Vonsky M, Morozov A (2020)
477 Proteomics answers which yeast genes are specific for baking, brewing, and ethanol production,
478 Bioeng. 7*:* 147. https://doi.org/10.3390/bioengineering7040147.
479 [36] Somlo M, Reid RA, Krupa M (1997) An oligomycin-resistant adenosine triphosphatase and its
480 effects on cellular growth, mitochondrial oxidative phosphorylation and respiratory proton translocation
481 in *Saccharomyces cerevisiae*, Biochem j. 162: 51–59. https://doi.org/10.1042/bj1620051.
482 [37] Su R, Wang R, Guo S, Cao H, Pan J, Li C, Shi D, Tang Z (2011) In vitro effect of copper chloride
483 exposure on reactive oxygen species generation and respiratory chain complex activities of
484 mitochondria isolated from broiler liver, Biol. trace elem. Res. 144 668-677.
485 https://doi.org/10.1007/s12011-011-9039-4.
486 [38] Belyaeva EA, Sokolova TV, Emelyanova LV, Zakharova IO (2012) Mitochondrial electron
487 transport chain in heavy metal-induced neurotoxicity: effects of cadmium, mercury, and copper, Sci.
488 World J. 2012. 136063. https://doi.org/10.1100/2012/136063.
489 [39] Chen Y, Lu Z, Chen D, Wei Y, Chen X, Huang J, Guan N, Lu Q, Wu R, Huang R (2017)
490 Transcriptomic analysis and driver mutant prioritization for differentially expressed genes from a

Saccharomyces cerevisiae strain with high glucose tolerance generated by UV irradiation, RSC adv. 7: 38784-38797. https://doi.org/10.1039/C7RA06146C.

[40] Vijayraghavan S, Kozmin SG, Strope PK, Skelly DA, Lin Z, Kennell J, Magwene PM, Dietrich F.S, McCusker JH (2019) Mitochondrial genome variation affects multiple respiration and nonrespiration phenotypes in *Saccharomyces cerevisiae*, Genetics. 211: 773-786. https://doi.org/10.1534/genetics.118.301546.

[41] Steensma HY, Holterman L, Dekker I. van Sluis CA, Wenzel TJ (1990) Molecular cloning of the gene for the E1 alpha subunit of the pyruvate dehydrogenase complex from *Saccharomyces cerevisiae*, Eur. J. Biochem. 191: 769-74. https://doi.org/10.1111/j.1432-1033.1990.tb19186.x.

[42] Vest KE, Leary SC, Winge DR, Cobine PA (2013) Copper import into the mitochondrial matrix in *Saccharomyces cerevisiae* is mediated by Pic2, a mitochondrial carrier family protein, J. Biol. Chem. 288: 23884-23892. (2013). https://doi.org/10.1074/jbc.M113.470674.

[43] Shi NQ, Davis B, Sherman F, Cruz J, Jeffries TW (1999) Disruption of the cytochrome c gene in xylose-utilizing yeast *Pichia stipitis* leads to higher ethanol production, Yeast. 15: 1021-1030. https://doi.org/10.1002/(SICI)1097-0061(199908)15:11<1021::AID-YEA429>3.0.CO;2-V.

[44] Granados-Arvizu JA, Madrigal-Perez LA, Canizal-García M, González-Hernández JC, García-Almendárez BE, Regalado-González C (2019) Effect of cytochrome bc1 complex inhibition during fermentation and growth of *Scheffersomyces stipitis* using glucose, xylose or arabinose as carbon sources, FEMS yeast res. 19. https://doi.org/10.1093/femsyr/foy126.

[45] Meaden PG, Dickinson FM, Mifsud A, Tessier W, Westwater J, Bussey H, Midgley M (1997) The ALD6 gene of *Saccharomyces cerevisiae* encodes a cytosolic, Mg^{2+} activated acetaldehyde dehydrogenase, Yeast. 13 1319-27. https://doi.org/10.1002/(SICI)1097-0061(199711)13:14<1319::AID-YEA183>3.0.CO;2-T.

[46] Papapetridis I, van Dijk M, Dobbe AP, Metz B, Pronk JT, van Maris AJ (2016) Improving ethanol yield in acetate-reducing *Saccharomyces cerevisiae* by cofactor engineering of 6-phosphogluconate dehydrogenase and deletion of ALD6, Microb. Cell factories. 15: 67. https://doi.org/10.1186/s12934-016-0465-z.

518 [47] Yamanaka R, Tabata S, Shindo Y, Hotta K, Suzuki K, Soga T, Oka K (2016) Mitochondrial Mg^{2+}
519 homeostasis decides cellular energy metabolism and vulnerability to stress, Sci. rep. 6: 30027.
520 https://doi.org/10.1038/srep30027.
521 [48] Cipollina C, van den Brink J, Daran-Lapujade P, Pronk JT, Porro D, de Winde JH
522 (2008)*Saccharomyces cerevisiae* SFP1: at the crossroads of central metabolism and ribosome
523 biogenesis, Microbiology. 154: 1686-1699. https://doi.org/10.1099/mic.0.2008/017392-0.
524 [49] Zhou Q, Liu ZL, Ning K, Wang A, Zeng X, Xu J (2014) Genomic and transcriptome analyses
525 reveal that MAPK-and phosphatidylinositol-signaling pathways mediate tolerance to 5-hydroxymethyl-
526 2-furaldehyde for industrial yeast *Saccharomyces cerevisiae*, Sci. rep. 4: 6556.
527 https://doi.org/10.1038/srep06556.
528 [50] Adhikari H, Cullen PJ (2015)Role of phosphatidylinositol phosphate signaling in the regulation of
529 the filamentous-growth mitogen-activated protein kinase pathway, Eukaryot. cell. 14: 427-440.
530 https://doi.org/10.1128/EC.00013-15.
531 [51] Boorsma A, de Nobel H, ter Riet B, Bargmann B, Brul S, Hellingwerf KJ, Klis FM (2004)
532 Characterization of the transcriptional response to cell wall stress in *Saccharomyces cerevisiae*, Yeast.
533 21: 413-427. https://doi.org/10.1002/yea.1109. PMID: 15116342.
534 [52] Dziedzic SA, Caplan AB (2011) Identification of autophagy genes participating in zinc-induced
535 necrotic cell death in *Saccharomyces cerevisiae*, Autophagy. 7: 490-500.
536 https://doi.org/10.4161/auto.7.5.14872.
537 [53] Mennella TA, Klinkenberg LG Zitomer RS (2003) Recruitment of Tup1-Ssn6 by yeast hypoxic
538 genes and chromatin-independent exclusion of TATA binding protein, Eukaryot. cell. 2: 1288-1303.
539 https://doi.org/10.1128/EC.2.6.1288-1303.2003.
540 [54] Kurylenko O Ruchala J, Kruk B, Vasylyshyn R, Szczepaniak J., Dmytruk K, Sibirny A (2021) The
541 role of Mig1, Mig2, Tup1 and Hap4 transcription factors in regulation of xylose and glucose
542 fermentation in the thermotolerant yeast *Ogataea polymorpha*, FEMS Yeast Res. 21. foab029.
543 https://doi.org/10.1093/femsyr/foab029.

544 [55] Williams FE, Trumbly RJ (1990) Characterization of TUP1, a mediator of glucose repression in
545 *Saccharomyces cerevisiae*, Mol. cell. biol. 10: 6500-6511. https://doi.org/10.1128/mcb.10.12.6500-
546 6511.1990.
547 [56] Shi X, Zou Y, Chen Y, Ying H (2018) Overexpression of THI4 and HAP4 improves glucose
548 metabolism and ethanol production in *Saccharomyces cerevisiae*, Front. Microbiol. 9: 1444.
549 https://doi.org/10.3389/fmicb.2018.01444.

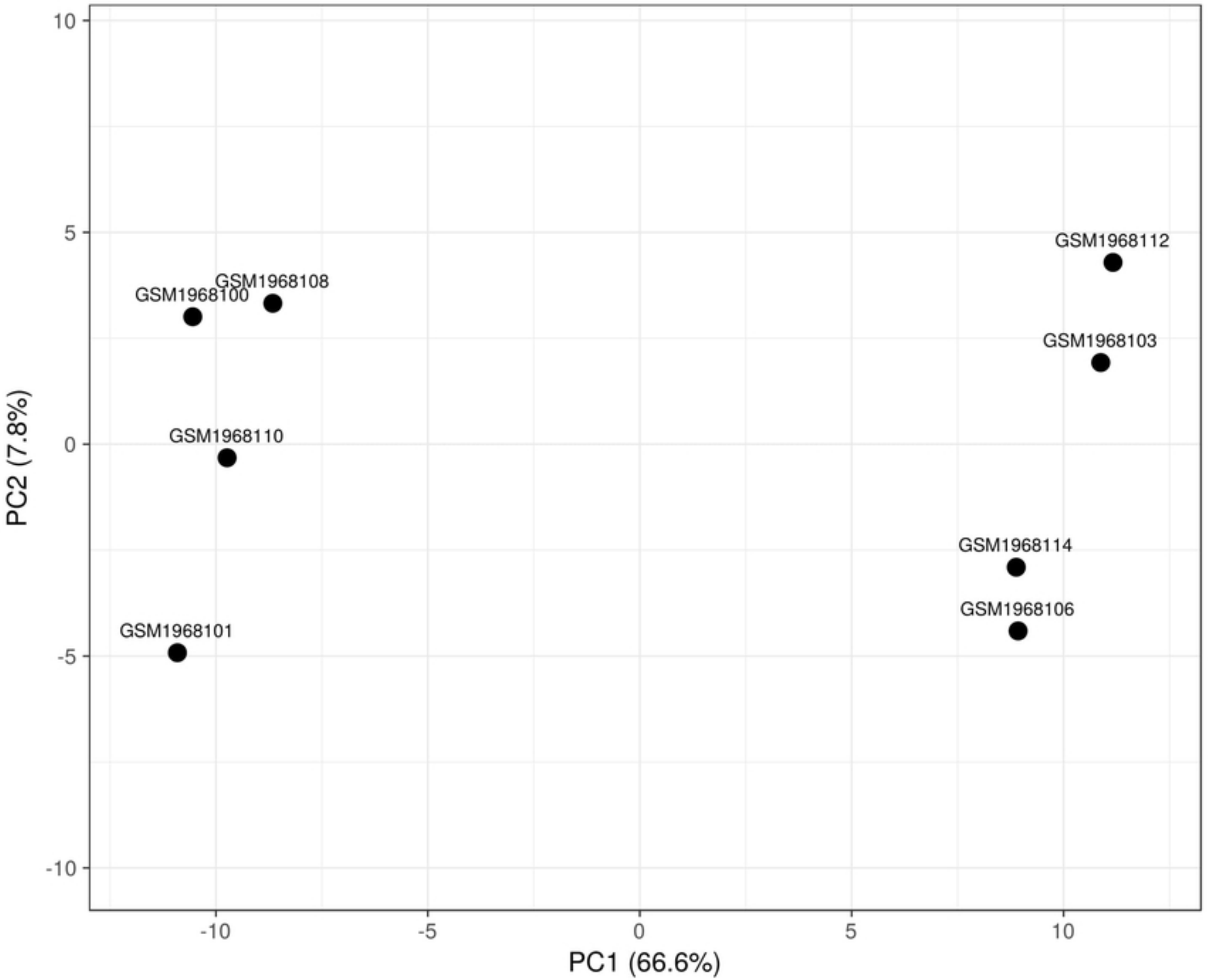
10
5
0
-5
-10
PC2 (7.8%)
GSM1968112
GSM1968108
GSM1968100
GSM1968103
GSM1968110
GSM1968114
GSM1968106
GSM1968101
-10
-5
0
5
10
PC1 (66.6%)

Fig 1

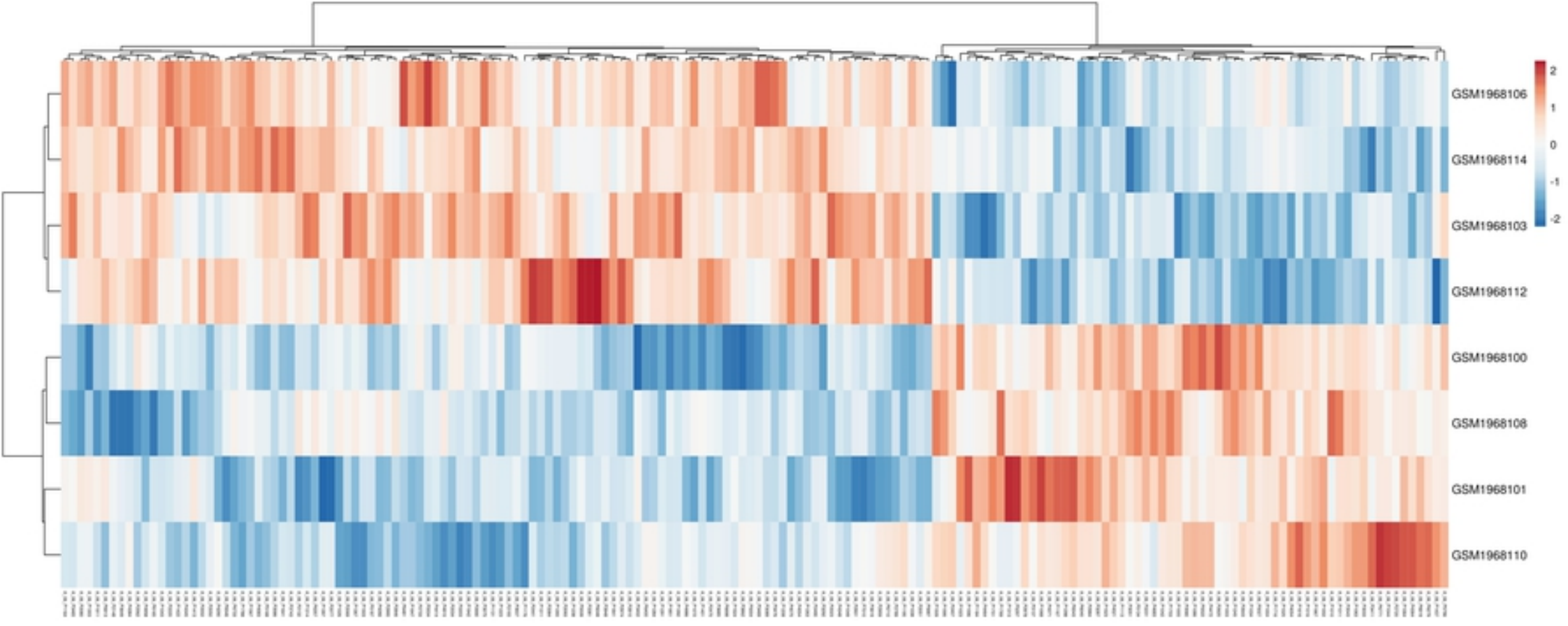
GSM1968106
GSM1968114
GSM1968103
GSM1968112
GSM1968100
GSM1968108
GSM1968101
GSM1968110
2
1
0
-1
-2

Fig 2

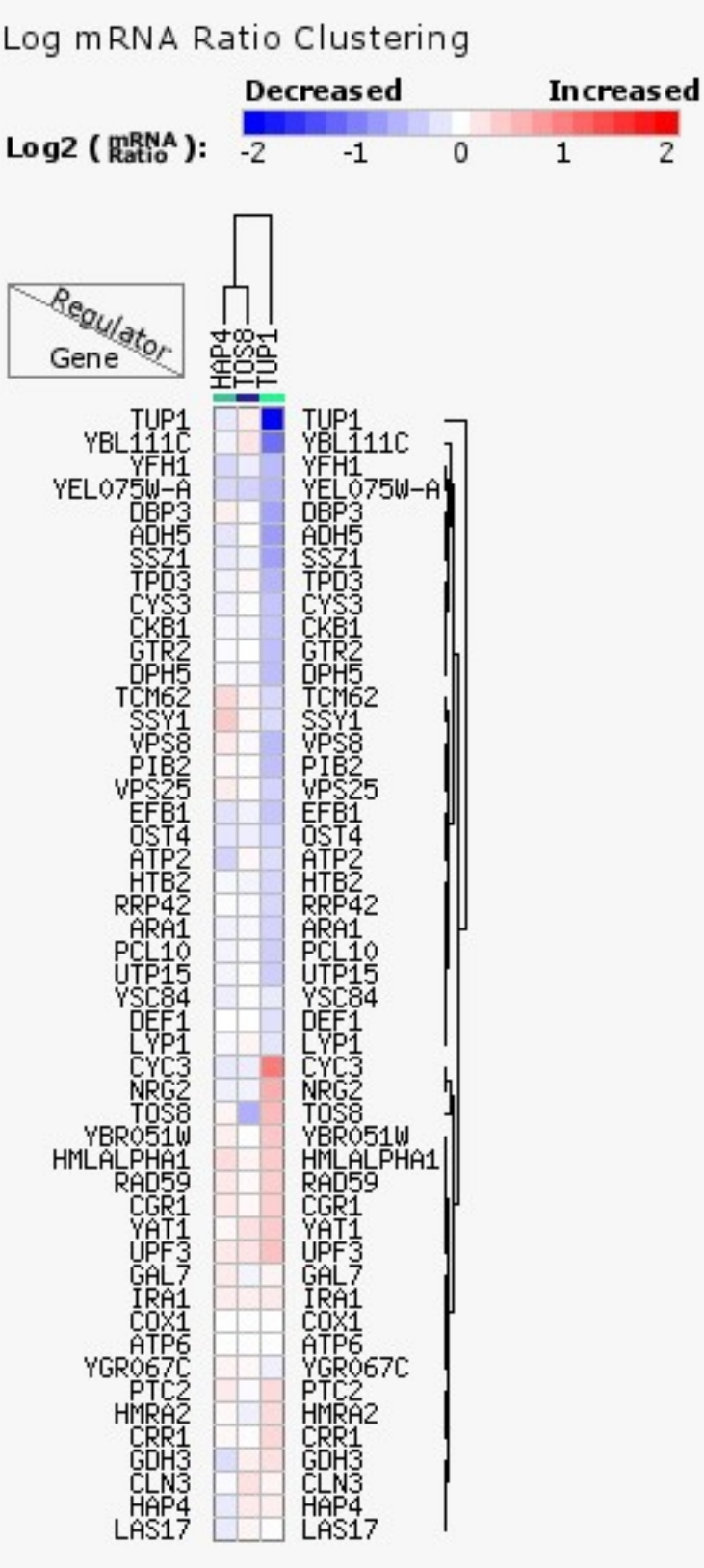
Log mRNA Ratio Clustering
Decreased
Increased
Log2 (mRNA Ratio):
-2
-1
0
1
2
Regulator
Gene
HAP4
TOS8
TUP1
TUP1
YBL111C
YFH1
YEL075W-A
DBP3
ADH5
SSZ1
TPD3
CYS3
CKB1
GTR2
DPH5
TCM62
SSY1
VPS8
PIB2
VPS25
EFB1
OST4
ATP2
HTB2
RRP42
ARA1
PCL10
UTP15
YSC84
DEF1
LYP1
CYC3
NRG2
TOS8
YBR051W
HMLALPHA1
RAD59
CGR1
YAT1
UPF3
GAL7
IRA1
COX1
ATP6
YGR067C
PTC2
HMRA2
CRR1
GDH3
CLN3
HAP4
LAS17

Fig 3

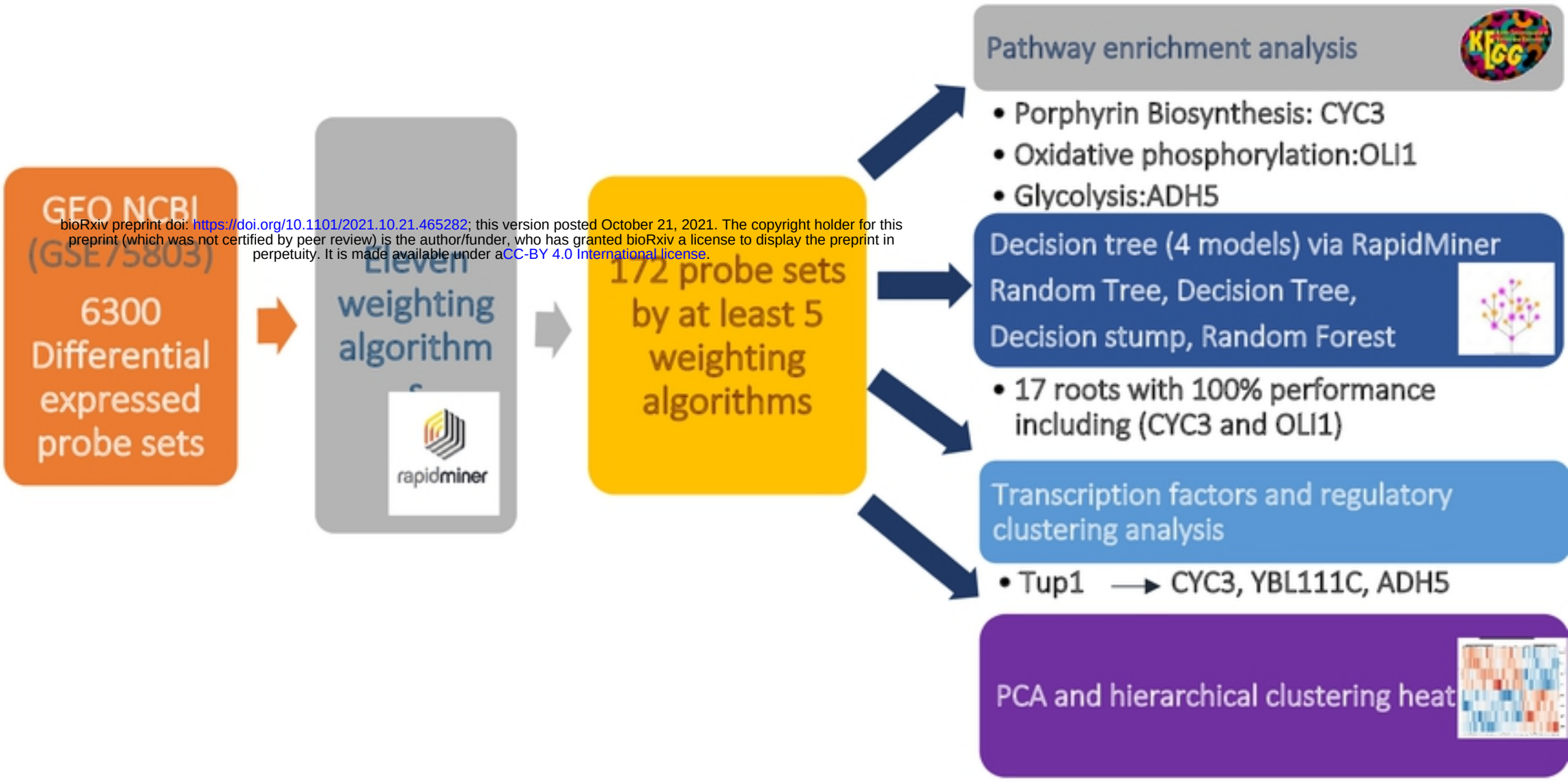

Fig 4