COMMUNICATION TO THE EDITOR

BIOTECHNOLOGY and BIOENGINEERING

Computational Identification of Altered Metabolism Using Gene Expression and Metabolic Pathways

Hojung Nam,[1] Jinwon Lee,[2] Doheon Lee[1]

[1]Department of Bio and Brain Engineering, KAIST, 373-1, Guseong-dong, Yuseong-gu, Deajeon 305-701, Republic of Korea; telephone: +82-42-350-4316; fax: +82-42-350-8680; e-mail: dhlee@biosoft.kaist.ac.kr

[2]Department of Chemical and Biomolecular Engineering, Sogang University, Seoul, Korea

Received 13 October 2008; revision received 9 February 2009; accepted 6 March 2009

Published online 20 March 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22320

ABSTRACT: Understanding altered metabolism is an important issue because altered metabolism is often revealed as a cause or an effect in pathogenesis. It has also been shown to be an important factor in the manipulation of an organism's metabolism in metabolic engineering. Unfortunately, it is not yet possible to measure the concentration levels of all metabolites in the genome-wide scale of a metabolic network; consequently, a method that infers the alteration of metabolism is beneficial. The present study proposes a computational method that identifies genome-wide altered metabolism by analyzing functional units of KEGG pathways. As control of a metabolic pathway is accomplished by altering the activity of at least one rate-determining step enzyme, not all gene expressions of enzymes in the pathway demonstrate significant changes even if the pathway is altered. Therefore, we measure the alteration levels of a metabolic pathway by selectively observing expression levels of significantly changed genes in a pathway. The proposed method was applied to two strains of *Saccharomyces cerevisiae* gene expression profiles measured in very high-gravity (VHG) fermentation. The method identified altered metabolic pathways whose properties are related to ethanol and osmotic stress responses which had been known to be observed in VHG fermentation because of the high sugar concentration in growth media and high ethanol concentration in fermentation products. With the identified altered pathways, the proposed method achieved best accuracy and sensitivity rates for the Red Star (RS) strain compared to other three related studies (gene-set enrichment analysis (GSEA), significance analysis of microarray to gene set (SAM-GS), reporter metabolite), and for the CEN.PK 113-7D (CEN) strain, the proposed method and the GSEA method showed comparably similar performances.

Correspondence to: D. Lee

Contract grant sponsor: Korean Systems Biology Program
Contract grant number: M10309020000-03B5002-00000
Contract grant sponsor: National Research Laboratory Program
Contract grant number: R0A-2005-000-10094-0
Contract grant sponsor: Ministry of Education, Science and Technology
Additional Supporting Information may be found in the online version of this article.

Biotechnol. Bioeng. 2009;103: 835–843.

KEYWORDS: metabolism; altered metabolism; metabolic pathway; gene expression

Introduction

Understanding altered metabolism is an important issue because altered metabolism is often revealed as a cause or an effect in pathogenesis (Dakubo et al., 2006; Eapen, 2007; Lohmander, 2004; Picton, 1998), and it is also a crucial factor when manipulating an organism's metabolism in metabolic engineering (Chavez-Bejar et al., 2008; Covert et al., 2004; Keasling and Chou, 2008; Kim et al., 2008). With the development of high-throughput technologies, deeper understanding of biological processes has arisen through the integration of data from genomics, proteomics, and metabolomics.

Previous studies have integrated transcriptional regulation and metabolism in an effort to understand the mechanisms of metabolic alteration. These studies can be grouped into three classes. The first approach identifies new relationships between metabolites and genes. For example, new associations between metabolites and transcription factors were identified by observing gene and protein expression profiles under different carbon sources (Ideker et al., 2001), and links between metabolites and transcription factors were inferred by using a joint model of genetic regulation and metabolic reactions (Yeang and Vingron, 2006). The second approach involves the development of new methodologies to analyze metabolic reactions. The effects of enzyme genetic regulation were combined with metabolic flux modeling to validate flux modes with biomass growth of cells (Covert et al., 2004; Covert and Palsson, 2002, 2003). The third approach identifies new

features of metabolic pathways by analyzing expression patterns of genes that encode enzymes. The transcriptional regulation of metabolic pathways were characterized by examining expression patterns of enzymes along the topology of metabolic pathways (Ihmels et al., 2002, 2004). The reporter reactions which are the reactions showing significant regulation changes are identified by integrating transcriptome, metabolome, and the topology of metabolic pathways (Cakir et al., 2006). The reporter metabolites whose neighboring enzymes show the most significant transcriptional changes in perturbations are identified (Patil and Nielsen, 2005).

The key contribution of Patil et al.'s work is that they computationally identified genome-scale altered metabolism by observing transcriptional changes of neighboring enzymes in a metabolic network, whereas other studies had identified altered metabolism with individual components of genes or proteins or flux control (Blankenhorn et al., 1999; Hayes et al., 2006; Perrenoud and Sauer, 2005; Stancik et al., 2002; Yohannes et al., 2004). However, changed transcription levels in neighboring enzymes do not guarantee changed concentration levels in metabolites. Several studies reported that even when the cells grow in a constant environment for prolonged periods of time, transcriptional levels, protein concentrations, and fluxes of a reaction correlated poorly (Daran-Lapujade et al., 2007; Greenbaum et al., 2003; Griffin et al., 2002; ter Kuile and Westerhoff, 2001; Yang et al., 2002). Hence, it is difficult to conclude that the concentrations of reporter metabolites change under perturbation conditions.

Lately, gene-set analysis methods for evaluating the expression of biological pathways are introduced and applied to metabolic pathways (Dinu et al., 2007; Subramanian et al., 2005). Gene-set enrichment analysis (GSEA) is one of the most well-established gene-set approaches that determines whether an a priori defined set of genes concordantly shows statistically significant differences between two phenotypes (Subramanian et al., 2005), significance analysis of microarray to gene set (SAM-GS) combines the regression-based test statistic for each individual gene in a pathway of interest into a pathway-level test statistic (Dinu et al., 2007). However, these methods do not provide metabolic pathways specific statistic because they are developed for analyzing universal biological pathways including protein interactions, signal transductions.

In this study, we propose a computational method that identifies genome-wide altered metabolism by analyzing predefined functional units of metabolic pathways in KEGG maps as opposed to analyzing the activities of each reaction (see Fig. 1). Until now, no high-throughput technology is available for the measurement of genome-wide metabolic activity. Therefore, we regard changes of metabolic genes' expression levels as significance indication for the possibility to the changes of associated reactions carrying metabolic flux. We then accumulate these indications to predict the altered activity of a metabolic pathway, and determine whether the activity of a pathway is altered or not based on the statistical tests. In the case of metabolic pathways, control of a metabolic pathway is accomplished by altering the activity of at least one rate-determining step enzyme, not all gene expressions of metabolic genes in the pathway demonstrate significant changes even if the pathway is altered (Hall et al., 2007; Nakamura et al., 2005). Therefore, we measure the alteration levels of a metabolic pathway by selectively observing significantly changed expression levels of genes in a pathway. The proposed method was applied to gene expression profiles of *Saccharomyces cerevisiae* cultivated in a very high-gravity (VHG) perturbation containing maltodextrin. The results of the proposed work were validated with concentrations of intracellular metabolites as measured by gas chromatography–mass spectrometry (GC–MS) profiling cultivated in conditions that were identical to those of a gene expression experiment (Devantier et al., 2005b). The method was found to achieve higher or comparable accuracy compared to other previous studies.

Materials and Methods

Figure 1 provides an overview of the proposed method. First, the significance of a change in the level of genes is extracted by applying *t*-tests to microarray data. This maps the scores of the altered gene levels to the enzymes and the transcription factors in a metabolic pathway. Second, an integrated metabolic network is generated by combining transcription factor information with metabolic pathways. Third, through the use of the significance levels of the changed expression of genes and the integrated metabolic pathways, the alteration scores of pathways are measured. The result of identified altered metabolism is validated with the concentration changes of intracellular metabolites.

Data Sets

In this work, three different sources of data were used to identify altered metabolism (see Fig. 1). First, gene expression profiles of two strains of *S. cerevisiae*, an industrial strain (Red Star (RS)) and a laboratory strain (CEN.PK 113-7D) were used. For both strains, the data were obtained from anaerobic batch cultures under two different cultivation conditions: growth in glucose-containing standard media and in maltodextrin-containing VHG mineral media (Devantier et al., 2005a). The second source was the transcriptional regulation information of *S. cerevisiae*. We first obtained 78 protein–DNA binding information with *P*-value less than 0.001 from (Lee et al., 2002), and assigned a regulation type (e.g., activator and inhibitor) to each binding information by referring Gene Ontology annotations at Saccharomyces Genome Database (Hong et al., 2008). The third was predefined functional units of 85 metabolic pathways from KEGG Pathway DB (Kanehisa et al., 2006, 2008; Kanehisa and Goto, 2000).

t-Test and *z*-Score

To calculate the alteration levels of a metabolic pathway, we first measure the significance of differential expression of

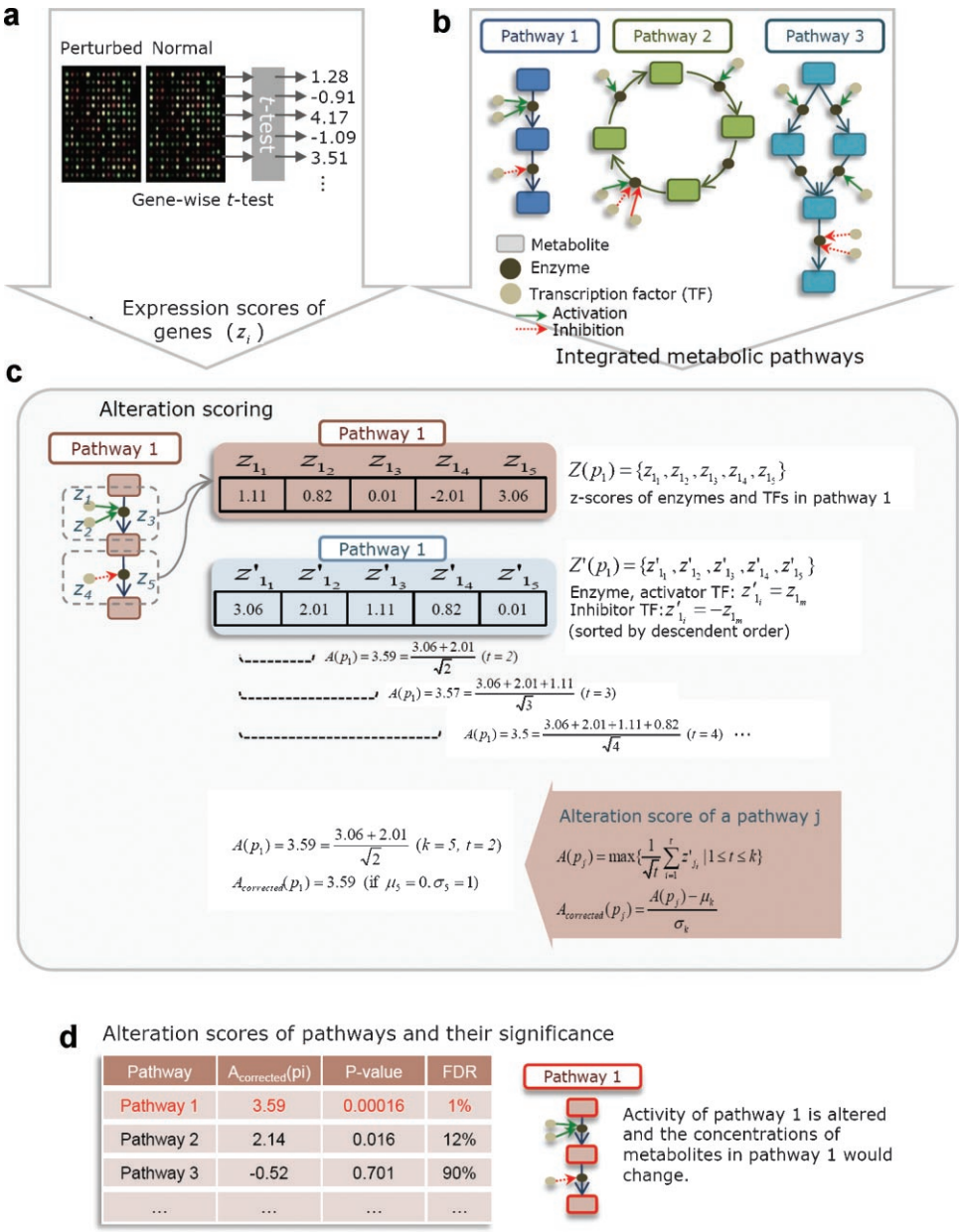

Figure 1. Method overview. **a**: Statistical test. Using microarray gene expression profiles, the standardized *z*-scores of genes, which represent significant changes in the levels of expression, are obtained by *t*-tests. **b**: Integrated metabolic pathways. For an integrated pathway, information from two biological processes was merged: (i) a metabolic reaction (metabolite, reaction, and enzyme) and (ii) a transcription (transcription factors and enzymes). **c**: Alteration scoring. The alteration score of a pathway is an aggregated form of the expression scores of genes in a pathway that maximize the aggregated score. **d**: Alteration scores of pathways and their significance. Once the alteration score of each pathway is calculated, significantly altered pathways are determined (FDR $\le$ 5%). [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

each gene under VHG stress. Here, we used the equal variance *t*-test to evaluate the significance of the change of a gene. The obtained *P*-value (p_i) of each gene ($gene_i$) was converted to a *z*-score (z_i) by using the inverse normal cumulative distribution function (θ^{-1}). A *z*-score follows a standard normal distribution, and a higher *z*-score indicates more significantly induced or reduced gene expression in a specific type of perturbation.

$$z_i = \theta^{-1}(1 - p_i)$$

Scoring Altered Metabolism

z-Scores (z_i) of enzymes and their transcription factors under VHG stress were observed to determine the alteration score of a pathway. Because there is typically at least one rate-determining step (or pacemarker) that determines the overall activity of a metabolic pathway, not all enzymes of a pathway consistently show significant transcriptional changes even if the pathway is altered (Hall et al., 2007; Nakamura et al., 2005). Therefore, expression levels of

significantly changed enzymes and their transcription factors in a pathway were selectively observed to measure the alteration levels of a metabolic pathway. First, a set of z-scores ($Z(p_j)$) for enzymes and transcriptional factors (TFs) in pathway j was assembled. In order to reflect the opposite effects of inhibitor transcription factors, the z-scores of the inhibitors were multiplied by -1. Then, a set of the z-scores in pathway j was sorted in descendent order ($Z'(p_j)$). By expanding from the maximum z-score, top t z-scores which maximize the alteration score were selected. An alteration score of a pathway ($A(p_j)$) is an aggregated form of the expression scores of genes (z_{j_i}) in a pathway j that maximize the aggregated score. In order to eliminate the size effectiveness of a pathway, normalization of each $A(p_j)$ score to $A_{\text{corrected}}(p_j)$ score was accomplished by correcting for background distributions based on the mean (μ_k) and standard deviation (σ_k) as derived from $A(p_j)$ scores of 10,000 random sets of k genes. The basic implication of the scoring function for identifying altered metabolism is a statistical test for the null hypothesis, "observed gene expression of enzymes and transcription factors of a metabolic pathway is purely random."

$$A(p_j) = \max\left\{\frac{1}{\sqrt{t}}\sum_{i=1}^{t} Z'_{j_i} | 1 \le t \le k\right\}$$

$$A_{\text{corrected}}(p_j) = \frac{A(p_j) - \mu_k}{\sigma_k}$$

Once we obtain corrected altered scores from methods described above, P-values were calculated using the inverse normal cumulative distribution function (θ^{-1}) and then we adjusted P-values to control for false discovery rate (FDR). A q-value of 0.05 for a pathway indicates that among all significant pathways selected at this threshold, 5 out 100 of them are expected to be false leads.

Results

Identified Altered Metabolic Pathways in VHG Fermentation

VHG fermentation technology is being used for industrial scale production of ethanol as a biofuel. The benefits of

Table I. Identified altered pathways (FDR $\le$ 5%).

Strain	Alteration type	KEGG id	Size 1 (k)	Size 2 (t)	Pathway name	Property	Score	P-value	FDR
RS	Activated	00190[a]	64	36	Oxidative phosphorylation	Energy metabolism	3.684	1.14E−04	0.00
	Activated	00020[a]	31	15	TCA cycle	Energy metabolism	3.546	1.95E−04	0.00
	Repressed	00230	89	62	Purine metabolism	Nucleotide metabolism	5.192	1.04E−07	0.00
	Repressed	00400	20	18	Phenylalanine, tyrosine and tryptophan biosynthesis	Amino acid metabolism	3.546	1.95E−04	0.00
	Repressed	00240	70	45	Pyrimidine metabolism	Nucleotide metabolism	3.426	3.06E−04	0.00
	Repressed	00300	15	13	Lysine biosynthesis	Amino acid metabolism	3.328	4.37E−04	0.00
	Repressed	00780[a]	7	5	Biotin metabolism	Metabolism of cofactors and vitamins	3.260	5.57E−04	0.00
	Repressed	00290	18	14	Valine, leucine, and isoleucine biosynthesis	Amino acid metabolism	3.197	6.94E−04	0.00
	Repressed	00220[a]	19	13	Urea cycle and metabolism of amino groups	Amino acid metabolism	3.007	0.001318	0.01
	Repressed	00260	43	27	Glycine, serine, and threonine metabolism	Amino acid metabolism	2.975	0.001465	0.01
	Repressed	00670	14	12	One carbon pool by folate	Metabolism of cofactors and vitamins	2.965	0.001512	0.01
	Repressed	00750	7	5	Vitamin B6 metabolism	Metabolism of cofactors and vitamins	2.871	0.002045	0.01
	Repressed	00310[a]	26	14	Lysine degradation	Amino acid metabolism	2.694	0.003527	0.01
	Repressed	00510[a]	31	24	*N*-Glycan biosynthesis	Glycan biosynthesis and metabolism	2.641	0.004135	0.02
	Repressed	00340	18	14	Histidine metabolism	Amino acid metabolism	2.414	0.007889	0.05
CEN	Activated	00500[b]	61	34	Starch and sucrose metabolism	Carbohydrate metabolism	4.386	5.77E−06	0.00
	Activated	00052[b]	29	19	Galactose metabolism	Carbohydrate metabolism	3.180	7.36E−04	0.03
	Repressed	00230	89	74	Purine metabolism	Nucleotide metabolism	6.207	2.70E−10	0.00
	Repressed	00240	70	55	Pyrimidine metabolism	Nucleotide metabolism	4.485	3.65E−06	0.00
	Repressed	00770[b]	10	8	Pantothenate and CoA biosynthesis	Metabolism of cofactors and vitamins	4.059	2.46E−05	0.00
	Repressed	00100[b]	21	19	Biosynthesis of steroids	Lipid metabolism	3.805	7.09E−05	0.00
	Repressed	00300	15	12	Lysine biosynthesis	Amino acid metabolism	3.715	1.01E−04	0.00
	Repressed	00340	18	15	Histidine metabolism	Amino acid metabolism	3.096	9.8E−04	0.01
	Repressed	00290	20	16	Valine, leucine, and isoleucine biosynthesis	Amino acid metabolism	2.881	0.001985	0.02
	Repressed	00350[b]	18	13	Tyrosine metabolism	Amino acid metabolism	2.761	0.002885	0.02
	Repressed	00400	19	15	Phenylalanine, tyrosine, and tryptophan biosyn	Amino acid metabolism	2.881	0.001982	0.02
	Repressed	00252[b]	7	6	Alanine and aspartate metabolism	Amino acid metabolism	2.568	0.005114	0.03
	Repressed	00750	31	22	Vitamin B6 metabolism	Metabolism of cofactors and vitamins	2.583	0.004897	0.03
	Repressed	00670	14	11	One carbon pool by folate	Metabolism of cofactors and vitamins	2.4931	0.006325	0.04
	Repressed	00260	43	26	Glycine, serine, and threonine metabolism	Amino acid metabolism	2.378	0.008715	0.05

Size 1 indicates the number of genes in the pathway.
Size 2 indicates the number of genes that are selected for score calculation.
[a]Altered pathways which are exclusively identified in RS.
[b]Altered pathways which are exclusively identified in CEN.

VHG fermentation include decreased water requirements and energy costs, increased productivity of ethanol. However, it was reported that ethanol and osmotic stresses decrease the fermentation efficiency (Devantier et al., 2005a,b; Wang et al., 2007). These ethanol and osmotic stresses are observed in VHG fermentation because of the high sugar concentration in growth media and high ethanol concentration in fermentation products. Hence there have been many efforts for understanding responses of ethanol and osmotic stresses to achieve more efficient fermentation. The responses caused by ethanol and osmotic stresses include decreased growth rate, lower viability, higher energy consumption, inhibition of general amino acid transport systems, and inhibition of the glucose transport system (Alexandre and Charpentier, 1998; Hu et al., 2007; Pham et al., 2006; Rautio et al., 2007).

In this study, we identified altered metabolic pathways for two strains of *S. cerevisiae* in VHG fermentation. Identified altered pathways of the RS industrial strain and CEN.PK 113-7D (CEN) laboratory strain are listed in Table I (FDR $\leq$ 5%). The genes and their corresponding reactions which maximized the alteration score in each pathway are fully listed in Supplementary Appendix 1. The proposed method identified altered metabolic pathways whose properties are related to ethanol and osmotic stress responses of carbohydrate metabolism, nucleotide metabolism, amino acid metabolism, and energy metabolism (see Table I). In the case of the RS strain, oxidative phosphorylation and TCA cycle which have the energy metabolism property were activated. Purine metabolism, pyrimidine metabolism, pantothenate and CoA biosynthesis, and other amino acid metabolism related pathways were repressed. In CEN strain, starch and sucrose metabolism as well as galactose metabolism were activated in VHG fermentation. Analogous to the results of RS strain, we identified amino acid metabolism related pathways which were repressed.

Because the RS strain was found to be more stress-tolerant and to be more efficient for fermentation (Devantier et al., 2005a), we focused on the differences between the identified pathways of the CEN and the RS strain. Exclusively identified altered pathways for each strain are marked in Table I. It should be noticeably that pathways related to energy metabolism (oxidative phosphorylation, TCA cycle) showed active-alteration in the RS strain and pathways related to carbohydrate metabolism (starch and sucrose metabolism, galactose metabolism) are activated in the CEN strain. These altered activities of metabolic pathways may guide the causes and effects for different fermentation efficiencies between the two trains.

Results Validation

To validate the performance of the proposed method, we compared the accuracies of our method, GSEA, SAM-GS, and reporter metabolite approach (Patil and Nielsen, 2005). First, we identified altered pathways in VHG fermentation by each method. Figure 2 shows detection of altered metabolic pathways by our proposed method, by the GSEA, and by the SAM-GS method using same gene expression data (FDR $\leq$ 5%). For the GSEA and the SAM-GS, no transcription factor information was used when constructing gene sets. Interestingly, no altered metabolic pathway was detected by the SAM-GS method while the proposed method and the GSEA method identified more than 10 altered metabolic pathways (Fig. 2a). As the

a

	Proposed method		GSEA		SAM-GS
	Activated	Repressed	Activated	Repressed	Altered
# of significantly altered pathways (RS strain)	2	13	0	11	0
# of significantly altered pathways (CEN strain)	2	13	1	15	0

b

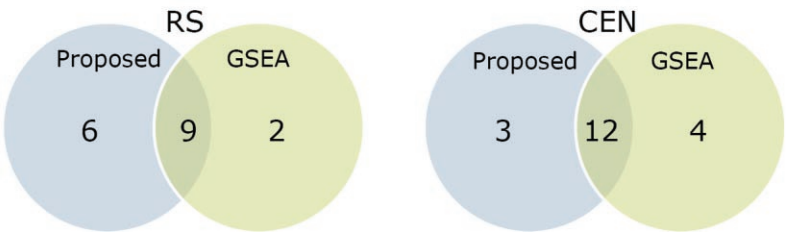

Figure 2. Results of the analysis by three methods (FDR $\leq$ 5%). **a**: The number of significantly altered metabolic pathways identified by the proposed method, the GSEA method, and the SAM-GS method (FDR $\leq$ 5%). **b**: Venn diagram of overlap between identified altered pathways. The figure shows the overlap between the altered pathways for both strains (RS, CEN) which are identified by our proposed method (blue) and the GSEA method (green). [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

Figure 2b shows, some of the altered pathways were commonly identified by the proposed method and the GSEA method, while, some altered metabolic pathways were identified uniquely. Full lists of identified altered pathways by the GSEA analysis are summarized in Supplementary Appendix 2.

We assume that the concentration of a metabolite would change (e.g., accumulate or be consumed) if the metabolite is a substrate or a product in an altered metabolic pathway. Thus, to validate the proposed method, changed concentrations of metabolites as measured by GC–MS analysis were compared to changes in metabolites as inferred by the proposed method. The concentration data of *S. cerevisiae* used here was cultivated in conditions that were identical to those of a gene expression experiment (Devantier et al., 2005b). (See Supplementary Appendix 3 for the detailed information for the concentration data.) Table II summarize the results and accuracy comparisons of changes (in the case of more than 1.2-fold changes for RS strain) of metabolites measured by GC–MS, inferred by the proposed method with or without TF information (FDR $\leq$ 5%), by the GSEA (FDR $\leq$ 5%), by the SAM-GS (FDR $\leq$ 5%), and by the reporter metabolites approach (FDR $\leq$ 5%). Taking into account various fold changes of the concentration of metabolites from 1.2- to 2.2-fold, the accuracy of altered metabolites from the proposed method were compared to results from the GSEA, the SAM-GS, and the reporter metabolites approach (See Supplementary Appendix 4 for the detailed information for the changed metabolites and their corresponding pathways). Note that a metabolite can belong to more than one metabolic pathway, so we determined concentration of a metabolite is changed if the metabolite is a substrate of at least one of the altered pathways.

Figure 3 depicts results of accuracy comparisons of the proposed method and other related studies from 1.2- to

Table II. Altered metabolites (RS strain, concentration change > 1.2-fold).

Name[a]	Concentration change (RS)[b]	Proposed method (with TF)[c]	Proposed method (without TF)[d]	GSEA[e]	SAM-GS[f]	Reporter metabolite[g]
Pyruvate	0	1	1	1	0	0
L-Glutamate	1	1	1	1	0	1
Oxaloacetate	1	1	1	0	0	0
Glycine	1	1	1	1	0	0
L-Alanine	0	0	0	0	0	0
Succinate	0	1	1	0	0	0
L-Lysine	1	1	1	1	0	0
L-Aspartate	1	1	1	1	0	0
L-Glutamine	1	1	1	1	0	0
L-Serine	1	1	1	0	0	0
L-Phenylalanine	1	1	1	1	0	0
L-Tyrosine	1	1	1	1	0	0
Fumarate	1	1	1	1	0	0
L-Leucine	1	1	1	1	0	0
L-Histidine	1	1	0	0	0	0
L-Proline	1	0	0	0	0	0
L-Asparagine	0	0	0	0	0	0
Citrate	1	1	1	0	0	0
L-Valine	1	1	1	1	0	0
L-Threonine	1	1	1	1	0	0
Lactate	0	0	0	0	0	0
L-Isoleucine	1	1	1	1	0	0
Citramalate	1	0	0	0	0	0
TP	—	16	15	12	0	1
FP	—	2	2	1	0	0
TN	—	3	3	4	5	5
FN	—	2	3	6	18	17
Accuracy[h]	—	82.6	78.3	69.6	21.7	26.1
Sensitivity[i]	—	88.9	83.33	66.7	0.0	5.6
Specificity[j]	—	60.0	60.0	80.0	100.0	100.0

See Supplementary Appendix 4 for the detailed information for the changed metabolites and their corresponding pathways.
[b]TP, true positive; FP, false positive (Type 1 error); TN, true negative; FN, false negative (Type 2 error).
[a]Names of metabolites.
[b]Concentration change measured by experiment. 1, more than 1.2-fold change; 0, no change.
[c]Concentration change inferred by proposed work with transcription factors. 1, changed (FDR $\leq$ 5%); 0, no change.
[d]Concentration change inferred by proposed work without transcription factors. 1, changed (FDR $\leq$ 5%); 0, no change.
[e]Concentration change inferred by GSEA. 1, changed (FDR $\leq$ 5%); 0, no change.
[f]Concentration change inferred by SAM-GS. 1, changed (FDR $\leq$ 5%); 0, no change.
[g]Concentration change inferred by reporter metabolite. 1, changed (FDR $\leq$ 5%); 0, no change.
[h]Accuracy: (TP + TN)/(TP + FP + TN + FN).
[i]Sensitivity: TP/(TP + FN).
[j]Specificity: TN/(FP + TN).

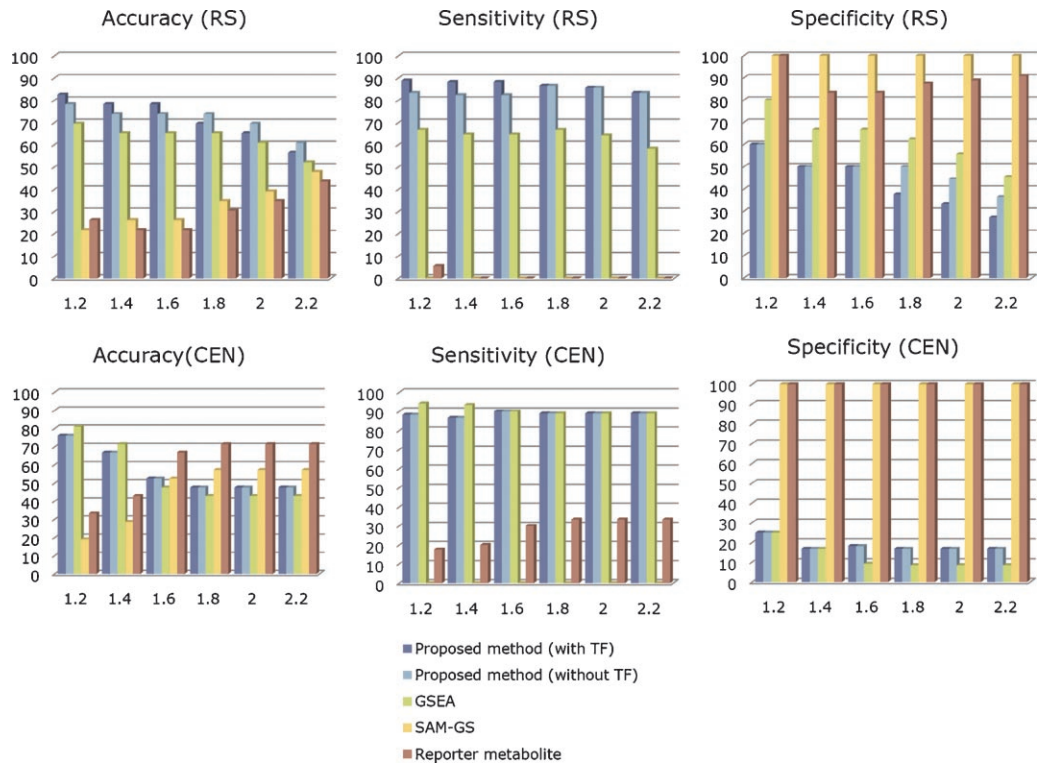

Figure 3. The results of accuracy tests (RS, CEN strain). The *X*-axis of the graphs illustrates the threshold of fold changes from 1.2 to 2.2. The *Y*-axis illustrates the percentage of accuracy, sensitivity, and specificity. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

2.2-fold changes for both two strains. As the graphs show, the proposed method achieved best accuracy and best sensitivity rates for the RS strain and we can see that the TF information tends to increase sensitivity rates. Due to small size of transcriptional regulation information, the results of accuracies with TF and without TF were not significantly different in the case of CEN strain. However, if the size of available transcriptional regulation information grows, we expect that the proposed scoring method could identify altered metabolic pathways more accurately. For the CEN strain, the proposed method and the GSEA method showed comparably similar performances. In the aspects of sensitivity and specificity tests for the two aforementioned strains, however, the proposed method showed less accurate specificity rates compare to the SAM-GS and the reporter metabolite method, but the proposed method clearly showed more accurate sensitivity results within the overall fold changes. This result shows that the proposed method can identify changed concentrations of metabolites with better or comparable accuracy compared to suggested methods earlier.

Discussion

Notably, it is generally accepted that the control of metabolic processes is hierarchical and originates at the level of transcription, moving on to translation and enzyme activity or usually a combination of them (Vemuri and Aristidou, 2005). Therefore, the proposed method as well as all other related studies which observe gene expression changes cannot detect alterations of post-transcriptional regulation because the gene expression data does not reveal the information of post-transcriptional effects. To account for effects of post-transcriptional regulation, we regard changed expression levels of metabolic genes not as the final determinants of changed levels of enzyme activity, but as significance indication for the possibility that their associated reactions carry metabolic flux. We then accumulate these indications to predict the altered activity of a metabolic pathway, and determine whether the activity of a pathway is altered or not based on the statistical tests. Although, in this work, we validated the results only with concentrations of intracellular metabolites due to the limitation of data availability, the proposed method can also be validated with changes of metabolic flux. To apply our method, the data set should satisfy the conditions: (1) the data set should have more than one gene expression sample per a perturbation to perform statistical tests, (2) the changes of metabolites or fluxes should had been observed in large scale (should had been observed in several pathways) to validate the changed activities of pathways. Therefore, if the data set which satisfies the aforementioned conditions is available, the proposed method can be applied and then

validated with concentrations of metabolites or metabolic flux rates.

Conclusion

This study proposes a computational method for identifying functional units of altered metabolism in metabolic pathways. Altered metabolisms were identified through statistical tests using gene expression and KEGG metabolic pathways. The proposed method was applied to *S. cerevisiae* gene expression and metabolic profile data. The method identified altered metabolic pathways whose properties are related to ethanol and osmotic stress responses like carbohydrate metabolism, nucleotide metabolism, amino acid metabolism, and energy metabolism which are observed in VHG fermentation because of the high sugar concentration in growth media and high ethanol concentration in fermentation products. With the identified altered pathways, the proposed method achieved best accuracy and sensitivity rates for the RS strain compared to other three previous studies (GSEA, SAM-GS, and reporter metabolite), and for the CEN strain, the proposed method and GSEA method showed comparably similar performances. In conclusion, the proposed work represents a method to identify altered metabolism through the use of gene expression along with the metabolic pathway information.

This work was supported by the Korean Systems Biology Program (no. M10309020000-03B5002-00000) and the National Research Laboratory Program (R0A-2005-000-10094-0) from the Ministry of Education, Science and Technology through the Korea Science and Engineering Foundation. We would like to thank Chung Moon Soul Center for BioInformation and BioElectronics for providing research facilities. Special thanks are also expressed to Dr. Rasmus Devantier for sharing the gene expression experimetal results.

References

Alexandre H, Charpentier C. 1998. Biochemical aspects of stuck and sluggish fermentation in grape must. J Ind Microbiol Biotechnol 20: 20–27.

Blankenhorn D, Phillips J, Slonczewski JL. 1999. Acid- and base-induced proteins during aerobic and anaerobic growth of *Escherichia coli* revealed by two-dimensional gel electrophoresis. J Bacteriol 181(7): 2209–2216.

Cakir T, Patil KR, Onsan Z, Ulgen KO, Kirdar B, Nielsen J. 2006. Integration of metabolome data with metabolic networks reveals reporter reactions. Mol Syst Biol 2:50.

Chavez-Bejar MI, Lara AR, Lopez H, Hernandez-Chavez G, Martinez A, Ramirez OT, Bolivar F, Gosset G. 2008. Metabolic engineering of *Escherichia coli* for L-tyrosine production by the expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from *Zymomonas mobilis*. Appl Environ Microbiol 74(10):3284–3290.

Covert MW, Palsson BO. 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. J Biol Chem 277(31):28058–28064.

Covert MW, Palsson BO. 2003. Constraints-based models: Regulation of gene expression reduces the steady-state solution space. J Theor Biol 221(3):309–325.

Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature 429(6987):92–96.

Dakubo GD, Parr RL, Costello LC, Franklin RB, Thayer RE. 2006. Altered metabolism and mitochondrial genome in prostate cancer. J Clin Pathol 59(1):10–16.

Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MA, de Groot MJ, Slijper M, Heck AJ, Daran JM, de Winde JH, Westerhoff HV, Pronk JT, Bakker BM. 2007. The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. Proc Natl Acad Sci USA 104(40):15753–15758.

Devantier R, Pedersen S, Olsson L. 2005a. Transcription analysis of *S. cerevisiae* in VHG fermentation: The genome-wide transcriptional response of *Saccharomyces cerevisiae* during very high gravity ethanol fermentations is highly affected by the stationary phase. Industrial Biotechnology 1(1):51–63.

Devantier R, Scheithauer B, Villas-Boas SG, Pedersen S, Olsson L. 2005b. Metabolite profiling for analysis of yeast stress response during very high gravity ethanol fermentations. Biotechnol Bioeng 90(6):703–714.

Dinu I, Potter JD, Mueller T, Liu Q, Adewale AJ, Jhangri GS, Einecke G, Famulski KS, Halloran P, Yasui Y. 2007. Improving gene set analysis of microarray data by SAM-GS. BMC Bioinformatics 8:242.

Eapen BR. 2007. Photosensitivity in Smith-Lemli-Opitz syndrome: A flux balance analysis of altered metabolism. Bioinformation 2(2):78–782.

Greenbaum D, Colangelo C, Williams K, Gerstein M. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol 4(9):117.

Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. 2002. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. Mol Cell Proteomics 1(4):323–333.

Hall RK, Wang XL, George L, Koch SR, Granner DK. 2007. Insulin represses phosphoenolpyruvate carboxykinase gene transcription by causing the rapid disruption of an active transcription complex: A potential epigenetic effect. Mol Endocrinol 21(2):550–563.

Hayes ET, Wilks JC, Sanfilippo P, Yohannes E, Tate DP, Jones BD, Radmacher MD, BonDurant SS, Slonczewski JL. 2006. Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in *Escherichia coli* K-12. BMC Microbiol 6:89.

Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G, Costanzo MC, Dwight SS, Engel SR, Fisk DG, et al. 2008. Gene Ontology annotations at SGD: New data sources and annotation methods. Nucleic Acids Res 36(Database issue):D577–D581.

Hu XH, Wang MH, Tan T, Li JR, Yang H, Leach L, Zhang RM, Luo ZW. 2007. Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*. Genetics 175(3):1479–1487.

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292(5518):929–934.

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. 2002. Revealing modular organization in the yeast transcriptional network. Nat Genet 31(4):370–377.

Ihmels J, Levy R, Barkai N. 2004. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. Nat Biotechnol 22(1): 86–92.

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30.

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: New developments in KEGG. Nucleic Acids Res 34(Database issue):D354–D357.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res 36(Database issue):D480–D484.

Keasling JD, Chou H. 2008. Metabolic engineering delivers next-generation biofuels. Nat Biotechnol 26(3):298–299.

Kim HU, Kim TY, Lee SY. 2008. Metabolic flux analysis and metabolic engineering of microorganisms. Mol Biosyst 4(2):113–120.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298(5594):799–804.

2004. Markers of altered metabolism in osteoarthritis. J Rheumatol Suppl 70:28–35.

Nakamura R, Okunuki H, Ishida S, Saito Y, Teshima R, Sawada J. 2005. Gene expression profiling of dexamethasone-treated RBL-2H3 cells: Induction of anti-inflammatory molecules. Immunol Lett 98(2):272–279.

Patil KR, Nielsen J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci USA 102(8):2685–2689.

Perrenoud A, Sauer U. 2005. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. J Bacteriol 187(9):3171–3179.

Pham TK, Chong PK, Gan CS, Wright PC. 2006. Proteomic analysis of *Saccharomyces cerevisiae* under high gravity fermentation conditions. J Proteome Res 5(12):3411–3419.

Picton SV. 1998. Aspects of altered metabolism in children with cancer. Int J Cancer Suppl 11:62–64.

Rautio JJ, Huuskonen A, Vuokko H, Vidgren V, Londesborough J. 2007. Monitoring yeast physiology during very high gravity wort fermentations by frequent analysis of gene expression. Yeast 24(9): 741–760.

Stancik LM, Stancik DM, Schmidt B, Barnhart DM, Yoncheva YN, Slonczewski JL. 2002. pH-dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. J Bacteriol 184(15):4246–4258.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102(43): 15545–15550.

ter Kuile BH, Westerhoff HV. 2001. Transcriptome meets metabolome: Hierarchical and metabolic regulation of the glycolytic pathway. FEBS Lett 500(3):169–171.

Vemuri GN, Aristidou AA. 2005. Metabolic engineering in the -omics era: Elucidating and modulating regulatory networks. Microbiol Mol Biol Rev 69(2):197–216.

Wang FQ, Gao CJ, Yang CY, Xu P. 2007. Optimization of an ethanol production medium in very high gravity fermentation. Biotechnol Lett 29(2):233–236.

Yang C, Hua Q, Shimizu K. 2002. Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in Synechocystis. Appl Microbiol Biotechnol 58(6):813–822.

Yeang CH, Vingron M. 2006. A joint model of regulatory and metabolic networks. BMC Bioinformatics 7:332.

Yohannes E, Barnhart DM, Slonczewski JL. 2004. pH-dependent catabolic protein expression during anaerobic growth of *Escherichia coli* K-12. J Bacteriol 186(1):192–199.