INVERSE METABOLIC ENGINEERING BY INTEGRATION OF MULTIPLE OMICS ANALYSES

Hiroshi Shimizu[1], Takashi Hirasawa[1], Katsunori Yoshikawa[1], Yuki Nakakura[2]
Keisuke Nagahisa[1], Chikara Furusawa[1], Yoshio Katakura[2], Suteaki Shioya[2]

1: Department of Bioinformatic Engineering, Graduate School of
Information Science and Technology,Osaka University,
2-1 Yamadaoka, Suita, Osaka 565-0871 Japan
2: Department of Biotechnology, Graduate School of Engineering,
Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871 Japan

An inverse metabolic engineering method with "multiple omics analyses" was applied to creation of a stress tolerant strain of *Saccharomyces cerevisiae*. DNA microarray data of laboratory and brewing strains under high ethanol concentration condition (transcriptomics data) were compared and analyzed by a clustering method. Sensitivity analysis of gene knockout mutant library was further performed (phenomics data). The selection of candidate genes for conferring stress tolerance was successfully performed.
Copyright © 2007 IFAC

Keywords: inverse metabolic engineering, DNA microarray, self-organizing map, hierarchical clustering, multiple omics analyses

1. INTRODUCTION

Microorganisms have multiple dimensional networks in the cells such as gene, protein, and metabolism networks. Recently comprehensive and analytical methodologies are available for studying the microorganisms at a post genome era (Shioya et al., 2006). Genome scale mathematical models for predicting the behaviour of the cells have been developed (Westerhoff and Palsson, 2000) to create desired phenotypes of the cells. However, there is limitation of bottom up approach, because all the functions of elements and their interactions are not predictable by forward methods only. Important factors and elements should be analyzed with omics data as "Inverse Problem". Establishment of inverse metabolic engineering with multiple omics data is highly desired.

The yeast *Saccharomyces cerevisiae* has been used in the brewing industries for production of alcoholic beverages. *S. cerevisiae* is also used in fuel ethanol production from biomass resources such as cellulose and starch. Yeast cells are usually exposed to some environmental changes during the course of fermentation, such as increase in osmotic pressure, accumulation of ethanol, and/or carbon dioxide (Attfield, 1997). Since ethanol is one of the main stress factors in fermentation processes using yeast, yeast strains that can grow well under high ethanol stress condition are necessary.

Determining important factors conferring desired phenotypes (ethanol stress tolerance in this case), based on comparative analyses between wild type strain and the strain with superior properties is called as inverse metabolic engineering (Bailey et al., 2002) Omics analysis has provided a great impact on inverse metabolic engineering (Bro and Nielsen, 2004). Systematic method to extract important factors from huge amount of genome-wide data is highly desired. As one successful example of genome

breeding, necessary mutations for high production of lysine were identified by comparison of genomes information of a wild-type strain with that of a high producing strain of *Corynebacterium glutamicum* (Ohnishi et al., 2002). Based on the DNA microarray analyses, identification of genes for improvement of galactose uptake of *S. cerevisiae*, (Bro et al., 2005), identification of gene for antibiotics resistance of *Escherichia coli* (Gill et al., 2002), molecular breeding for lysine production in *E. coli*, (Imaizumi et al., 2005), and creation of osmotic stress tolerant strains of *S. cerevisiae* (Hirasawa et al., 2006) were succeeded. However, when only one genome-wide comprehensive analysis is used for selection of candidate factors, many false positives are possible to be selected. Systematic design of experiments with multiple genome-wide analyses would provide true positives with high probability.

In this paper, we reported a method of the identification of target genes for construction of ethanol stress tolerant yeast strains, by integration of the information obtained from multiple comprehensive analyses. We first observed and compared the gene expression profiles of two yeast strains of *S. cerevisiae*, a laboratory strain and a brewing one, the latter of which is used for Japanese rice wine (sake) brewing and shows more tolerant to ethanol stress than the laboratory strain, under high ethanol concentration condition. After the DNA microarray analysis, clustering analysis based on a self-organizing map in combination with hierarchical clustering was performed to compare the expression pattern of each gene between the two yeast strains.

We also examined the sensitivity to ethanol in terms of the growth of the single-gene knockout mutants of the selected genes (phenomics data). Considering the results of above analyses, we selected the candidate genes for the construction of ethanol tolerant strains and actually constructed the yeast strains overexpressing each selected gene from the laboratory strain. The validity of our methodology was thus confirmed.

2. DNA MICROARRAY AND PHENOTYPIC ANALYSES UNDER ETHANOL STRESS CONDITION

Idea of inverse metabolic engineering is that important factors are extracted by comparing between a standard strain and one with superior characteristics (Bro and Nielsen, 2004). In this paper, we started comparison of gene expression profiles of two yeast strains of *S. cerevisiae*, a laboratory strain and a brewing one, the latter of which is used for sake brewing and it shows more tolerant to ethanol stress than the laboratory strain, under high ethanol concentration condition. Candidate genes were selected by a clustering analysis. Further selection was performed with phenotypic data of gene

knockout mutant library. High throughput cultivation of gene knockout mutants under high ethanol concentration condition was performed. The cultivation was done by using 96-well microtitre plates. Our concept of integration of DNA microarray
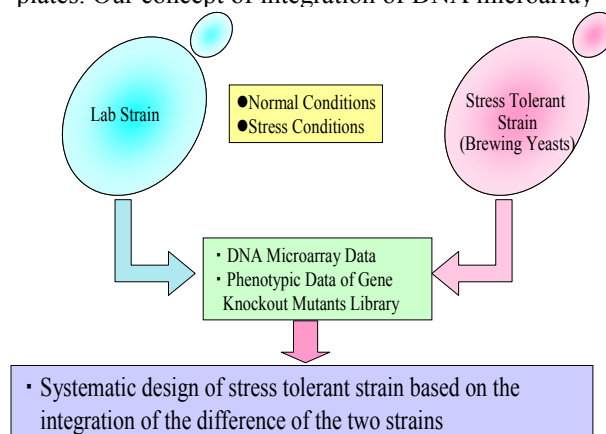

Fig.1 The concept of inverse metabolic engineering with multiple omics analyses.

2.1 Strains

The laboratory strains *S. cerevisiae* FY834 and the brewing strain *S. cerevisiae* IFO2347 (Kyokai No. 7, used for sake brewing) were used for DNA microarray analysis and the construction of recombinant yeast strains. For analysis of the ethanol sensitivity of single-gene knockout mutants, yeast MATalpha collection (glycerol stocks) and their parent strains *S. cerevisiae* BY4739 and BY4742, all of which were purchased from Open biosystems (USA), were used.

2.2 Cultivation Methods

Yeast cells were cultivated in yeast extract-peptone-dextrose (YPD) medium (1 % Bacto yeast extract, 2 % Bacto peptone, and 2 % glucose) or synthetic defined (SD) medium [0.67 % yeast nitrogen base without amino acid (Difco Laboratories, USA), 2 % glucose, 0.37 % leucine, 0.07 % lysine, 0.07 % histidine, 0.07 % tryptophan, and 0.07 % uracil] at 30 °C.

The growth of each strain cultivated in Sakaguchi flask was monitored by measuring the optical density of the culture medium at 660 nm (OD_{660}) with a spectrophotometer, UVmini-1240 (Shimadzu Corporation, Japan), and the growth of single-gene knockout mutants cultivated in 96-well microtiter plates (Corning Inc., USA) was monitored using Smart reader MPR-01 (Horiba Biotechnology Co., Ltd., Japan).

2.3 DNA Microarray Analysis

For DNA microarray analysis, yeast cells were cultivated in Sakaguchi flask containing 100 ml of

YPD medium at 30 °C with reciprocal shaking. After the cell growth reached the exponential growth phase (OD$_{660}$=about 1), 5 % (v/v, final concentration) ethanol was added to the culture broth. Before or 15, 30, 60 120 and 180 min after adding ethanol, cells were harvested by centrifugation, frozen in liquid nitrogen immediately, and then stored at -80 °C until the preparation of total RNA samples. Total RNA was extracted by the hot phenol method (Köhrer and Domdey, 1991) and purified with RNeasy mini kit (Qiagen K. K., Japan).

Cy3- or Cy5-labelled cDNA targets were prepared from 25 µg of total RNA samples and hybridized with probes on the slide glass-based microarray, Yeast gene chip ver. 2 (DNA Chip Research Inc. Japan). Methods for the preparation of cDNA targets, hybridization to and the washing of DNA microarray, spot detection, fluorescence intensity measurement and gene expression analysis were previously described by Hirasawa et al. (2006).

3. A CLUSTERING ANALYSIS METHOD FOR DNA MICROARRAY DATA

3.1 Clustering Analysis Method

To compare the expression patterns of each gene in the laboratory and brewing strains obtained by DNA microarray analysis, clustering analysis using a self-organizing map (SOM) was performed (Fig. 2). SOM is one of the neural network algorithms (Kohonen, 1998) and many research groups have reported its applicability to DNA microarray data analysis (Tamayo et al., 1999; Törönen et al., 1999; Huang et al., 2003). For our clustering analysis, the combined time course data sets of each gene in two strains were used.

Initially, there were given numbers of clusters and each cluster contained the vector according to the input data, which was called the weight vector. The weight vectors were updated during the process of SOM clustering. Finally, clusters, which are defined as the sets of genes whose similarities of expressions defined as Pearson's correlation coefficients were maximal with respect to the corresponding weight vector, were obtained. In the original SOM, clustering results have ambiguity because of their dependence on the order of data input. This ambiguity makes it difficult to compare and analyze the different clustering results. To avoid this problem, batch learning SOM (Kanaya et al., 2001), in which results do not depend on the order of data input, was introduced (hereafter, 'SOM' refers to batch learning SOM). One of the major problems of various clustering methods including SOM is to determine the optimal number of clusters. To solve this problem, Akaike's information criterion (Akaike, 1974), which represents the optimum balance between input data and the complexity of the model, was applied to estimate the optimum number of clusters.

Microarray data contain experimental errors; therefore, genes have the possibility of getting shifted from one cluster to another. Hence the degree of 'shift' of the cluster member caused by the addition of errors to each microarray data was investigated on the basis of the similarity of weight vectors. The range of experimental errors was defined as the expression ratios between 1/2 and 2, in which more than 99 % of data were contained, determined by multiple measurements using the same sample, as described in a previous paper (Yoshikawa et al., 2004).
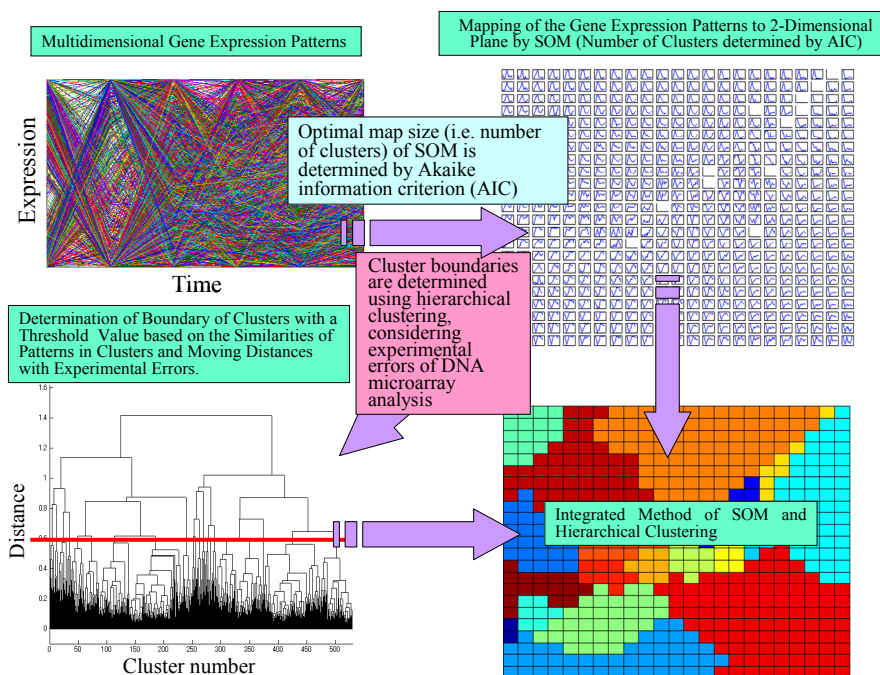

Fig.2 Integration method of clustering of DNA microarray data of SOM and hierarchical clustering.

On the basis of this evaluation of errors, the effect of errors on clustering results obtained using SOM was simulated as follows. First, synthetic errors, which follow the normal distribution and are within the same range of experimental errors, were generated and added to the original data. Then, the generated data were clustered again according to the SOM algorithm on the clustering results of original data. For all the input data, we calculated the similarity between the weight vector of a cluster to which the original data belonged and that to which data with artificial errors belonged. According to the set of similarities showing the change in clustering results caused by errors, we determined the threshold of similarity as that at which 80 % of obtained similarities exceed it. If the similarity between two given clusters is lower than the threshold then they should be further clustered into the same cluster, because experimental errors can cause the interchange of cluster members among these clusters.

For the further clustering of clusters obtained using the SOM, the hierarchical clustering method was adopted. The number of clusters was determined using the threshold mentioned above. These clusters are those that cannot be further subdivided considering the experimental errors of DNA microarray analysis.

3.2 Extraction of Candidate Genes based on DNA Microarray Data

To select candidate genes for constructing ethanol stress tolerant yeast on the basis of comparison of gene expression patterns between the laboratory and brewing strains without applying biological knowledge, clustering analysis was performed. We used the combined data of the expression profiles (time course of the ratio of gene expression after ethanol addition to that before addition) of each gene in the laboratory and brewing strains for clustering. As a result of clustering, 729 (=27 x 27) clusters were obtained using the SOM and the obtained clusters were further clustered into 29 clusters by the hierarchical clustering method as shown in Fig.3.

We selected clusters including candidates of target genes with specific expression patterns to construct ethanol stress-tolerant strains. Three clusters were selected, each of which included (i) genes significantly expressed only in the brewing strain, (ii) genes expressed in both laboratory and brewing strains following the addition of ethanol, but the expression ratios of these genes are higher in the brewing strain than in the laboratory strain and (iii) genes more significantly expressed in the laboratory strain than in the brewing strain. However, still we have more than four hundred genes as candidates, further selection is necessary.
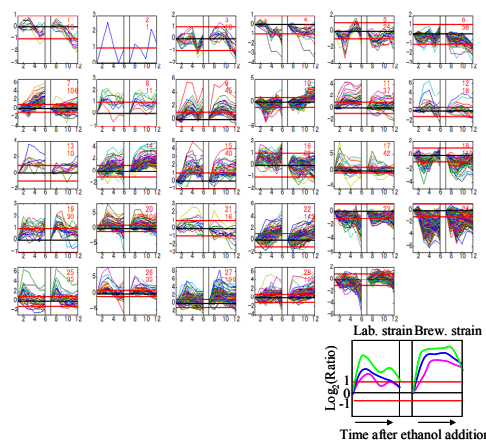

Fig. 3 Result of clustering analysis of DNA microarray.

4. PHENOTYPIC DATA OF GENE KNOCKOUT MUTANTS LIBRARY

If some genes that are essential for growth under the high ethanol concentration conditions are knocked out, yeast cells are not able to grow under such stress condition. Thus, we analyzed ethanol sensitivity of about 400 knockout mutants of the genes that might be related to ethanol stress tolerance by clustering analysis, in order to confirm whether each gene selected by the clustering analysis is necessary for growth under the high ethanol concentration condition. For ethanol sensitivity check, each knockout mutant was cultivated in 96-well microtiter plates at 30 ^{o}C in YPD medium with and without 5 % ethanol and the growth of the mutants was monitored. Cultivation experiments were performed more than twice, and when the ethanol sensitivity of the mutants was observed, the sensitivity was confirmed by flask culture experiments. The ethanol sensitivity of the yeast strains was evaluated by comparing specific growth rates during exponential growth with and without 5 % of ethanol.

All the mutants in which each gene categorized into category (i) was deleted did not show ethanol sensitivity. From category (ii), some single gene deletion mutants showed ethanol-sensitive growth in the experiments of microtiter plate cultivation, but significant sensitivity to ethanol was not observed in the flask culture experiments. Interestingly, five mutants of biosyntheses pathway genes, categorized into category (iii), showed ethanol-sensitive growth in microtiter cultivation experiments and this sensitivity was confirmed by flask culture experiments (Hirasawa et al., 2007). This result suggests that the expression of the genes related to this biosynthesis might be responsible for the ethanol stress tolerance of yeast cells.
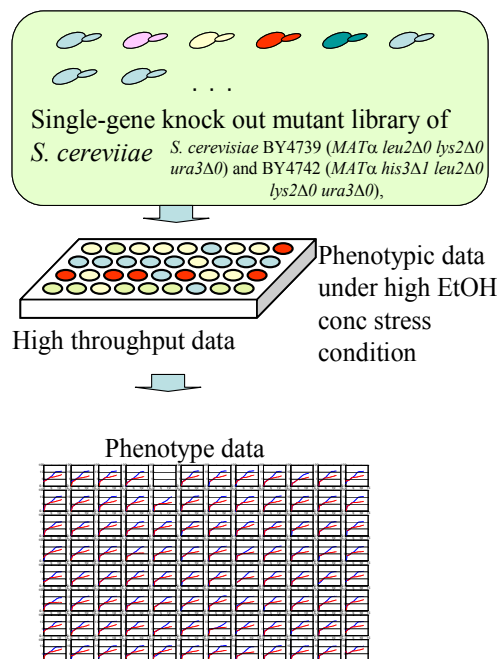
Fig.4 Phenotypic data analysis for selection of candidate genes conferring ethanol stress tolerance.

5. CREATION OF STRESS TOLERANT STRAIN BY OVEREXPRESSING CANDIDATES GENES

From the results of DNA microarray analysis and phenotypic data analysis of knockout mutants, it was speculated that yeast cells might acquire ethanol stress tolerance by the enhancement of the expression of a biosynthesis gene(s). We introduced plasmids carrying these biosynthesis gene(s) into the laboratory strain, and the specific growth rates of the constructed strains were compared with those of the parent and the control vector.

The specific growth rates of the laboratory strain transformed with plasmids carrying the biosynthetic gene(s), respectively, were higher than those of the laboratory strain under high ethanol concentration condition. The specific growth rates of the candidate genes-overexpressing strains were almost similar to that of the ethanol-tolerant strain IFO2347. These results indicate that the enhancement of the expression of the biosynthesis genes is effective for conferring ethanol stress tolerance (Hirasawa et al. 2007).

Bro and Nielsen (2004) described that these omics analyses have great impact on the inverse metabolic engineering. In our study, we could identify the novel target genes, concerning construction ethanol stress tolerant strains, by comparing the gene expression patterns between the laboratory and brewing strains by transcriptome analysis. Further selection of the candidate gene was performed based on phenotypic data analysis of gene knockout mutant library. We confirmed the validity of our methodology to identify the target genes for construction of stress tolerant strains.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* AC-19, 716-723.

Attfield, P.V. (1997). Stress tolerance: the key to effective strains of industrial baker's yeast. *Nature Biotechnol.* **15**, 1351-1357.

Bailey, J.E., A. Sburlati, V. Hatzimanikatis,, K. Lee, and W.A. Renner, and P.S. Tsai (2002). Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol. Bioeng.* 79, 568-579.

Bro, C., Nielsen, J. (2004). Impact of '-ome' analyses on inverse metabolic engineering. *Metab. Eng.* **6**, 204-211.

Bro, C., S. Knudsen, B. Regenberg, L. Olsson, and J. Nielsen (2005). Improvement of galactose uptake in *Saccharomyces cerevisiae* through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering. *Appl. Environ. Microbiol.* **71**, 6465-6472.

Gill, R.T., S. Wildt, S. Ziesman, and G. Stephanopoulos (2002). Genome-wide screening for trait conferring genes using DNA microarray. *Proc. Natl. Acad. Sci. USA* **99**, 7033-7038.

Hirasawa, T., Y. Nakakura, K. Yoshikawa, K. Ashitani, K. Nagahisa, C. Furusawa, Y. Katakura, H. Shimizu, and S. Shioya (2006). Comparative analysis of transcriptional responses to saline stress in the laboratory and brewing strains of *Saccharomyces cerevisiae* with DNA microarray. *Appl. Microbiol. Biotechnol.* **70,** 346-357.

Hirasawa, T., K. Yoshikawa, Y. Nakakura, K. Nagahisa, C. Furusawa, Y. Katakura, and S. Shioya, (2007). Identification of target genes

conferring ethanol stress-tolerance to yeast *Saccharomyces cerevisiae* based on DNA microarray data analysis. submitted

Huang, J., H. Shimizu, and S. Shioya (2003). Clustering gene expression pattern and extracting relationship in gene network base on artificial neural networks. *J. Biosci. Bioeng.* **96**, 421-428.

Imaizumi, A., R. Takikawa, C. Koseki, Y. Usuda, H. Yasueda, H. Kojima, K. Matsui, and S. Sugimoto (2005). Improved production of L-lysine by disruption of stationary phase-specific *rmf* gene in *Escherichia coli*. *J. Biotechnol.* **117**, 111-118.

Kanaya, S., M. Kinouchi, T. Abe, Y. Kudo, Y. Yamada, T. Nishi, H. Mori, Y. Ikemura (2001). Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene* **276**, 89-99.

Kohonen, T. (1998). The self-organizing map. *Neurocomputing* **21**, 1-6.

Köhrer, K. and H. Domdey (1991). Preparation of high molecular weight RNA. *Methods Enzymol.* **194**, 398-405.

Ohnishi, J., S. Mitsuhashi, M. Hayashi, S. Ando, H. Yokoi, K. Ochiai, and M. Ikeda (2002). A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L-lysine-producing mutant. *Appl. Microbiol. Biotechnol.* **58**, 217-223.

Shioya, S., H. Shimizu, T. Hirasawa, K. Nagahisa, C. Furusawa, G. Pandey, and Y. Katakura (2006). Metabolic pathway recruiting through genomic data analysis for industrial application of *Saccharomyces cerevisiae*. *Biochem Eng J.*, on-line appeared.

Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA.* **96**, 2907-2912.

Törönen, P., M. Kolehmainen, G. Wong, and E. Castrén (1999). Analysis of gene expression data using self-organizing maps. *FEBS Lett.* 451, 142-146.

Westerhoff, H. and B. Palsson (2004). The evolution of molecular biology to system biology, *Nature Biotechnol.* , **22**(10), 1249-1252.

Yoshikawa, K., G. Pandey, T. Hirasawa, Y. Katakura, K. Nagahisa, C. Furusawa, S. Shioya and H. Shimizu, Analysis of DNA microarray data using self-organizing map and hierarchical clustering, 3P-01-055, In: *Proceedings of 10th Asia Pacific Confederation of Chemical Engineering*, *Kitakyushu, Japan. (APCChE 2004)* Kataoka, K. (Ed.)