DNA Research **10**, 1–8 (2003)

Discovery of Novel Transcription Control Relationships with Gene Regulatory Networks Generated from Multiple-disruption Full Genome Expression Libraries

Sachiyo Aburatani,[1] Kosuke Tashiro,[1,2] Christopher J. Savoie,[2] Mayumi Nishizawa,[1] Kaori Hayashi,[1] Yuzuru Ito,[1] Shigeru Muta,[1] Kiyoshi Yamamoto,[1] Miki Ogawa,[1] Akiko Enomoto,[1] Momoka Masaki,[1] Shouji Watanabe,[3] Yukihiro Maki,[3] Yoriko Takahashi,[3] Yukihiro Eguchi,[3] Yoshiyuki Sakaki,[4] and Satoru Kuhara[1,2,*]

Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan,[1] *Gene Networks Inc., 650 S. Winchester Blvd., San Jose, CA 95128, USA,*[2] *Research Institute, Mitsui Knowledge Industry Co., Ltd, Harmony tower 21st Floor 32-2 Honcho 1-chome, Nakano-ku, Tokyo 164-8721, Japan,*[3] *and RIKEN Genomic Science Center, 1-7-22, Suehiro-cho, Tsurumi, Yokohama, Kanagawa, Japan*[4]

(Received 18 January 2003)

Abstract

Gene regulatory networks elucidated from strategic, genome-wide experimental data can aid in the discovery of novel gene function information and expression regulation events from observation of transcriptional regulation among genes of known and unknown biological function. To create a reliable and comprehensive data set for the elucidation of transcription regulation networks, we conducted systematic genome-wide disruption expression experiments of yeast on 118 genes with known involvement in transcription regulation. We report several novel regulatory relationships between known transcription factors and other genes with previously unknown biological function discovered with this expression library. Here we report the downstream regulatory subnetworks for *UME6* and *MET28*. The elucidated network topology among these genes demonstrates *MET28*'s role as a nodal point between genes involved in cell division and those involved in DNA repair mechanisms.

Key words: expression data; disruptant; gene regulatory network; *MET28* regulation

1. Introduction

Methods designed to elucidate gene regulation pathways have been reported previously.[1–3] However, the inferred networks reported in these studies were derived from gene expression data sets derived from time course, cell cycle and environmental perturbation.[4,5] Control relationships inferred from such data sets are suspect since they are not based on comprehensive experimental data designed specifically to elucidate transcription-related regulatory control functions and can therefore only produce information on co-regulation of expression, not hierarchical dependency. To rigorously and precisely identify novel and complex gene regulatory networks from de novo expression data sets, a systematic and integrated strategy of expression experiments on genomic deletion mutants combined with suitable computational methods is necessary.[6–9]

We have implemented for the yeast genome a systematic, iterative approach that combines full-genome biological expression experiments with gene regulatory inference. First, we constructed a gene expression data library using full-genome yeast c-DNA microarrays. The library was comprised of expression experiments on 118 yeast strains, each with one gene disrupted by homologous recombination. From this data we use computational techniques to infer gene expression regulatory relationships. In this paper, we introduce a static Boolean network model based on a multi-level digraph approach which can handle large sets of expression data and can analyze large-scale gene regulatory networks at high speed. With the model we then examine the biological relevance of putative regulatory relationships using computerized visualization and simulation software and finally validate our findings on novel or biologically interesting subnetworks through their presence in other databases as well as through further biological experimentation, including

Communicated by Satoshi Tabata

* To whom correspondence should be addressed. Kyushu University, Tel. +81-92-642-3041, Fax. +81-92-642-3043, E-mail: kuhara@grt.kyushu-u.ac.jp

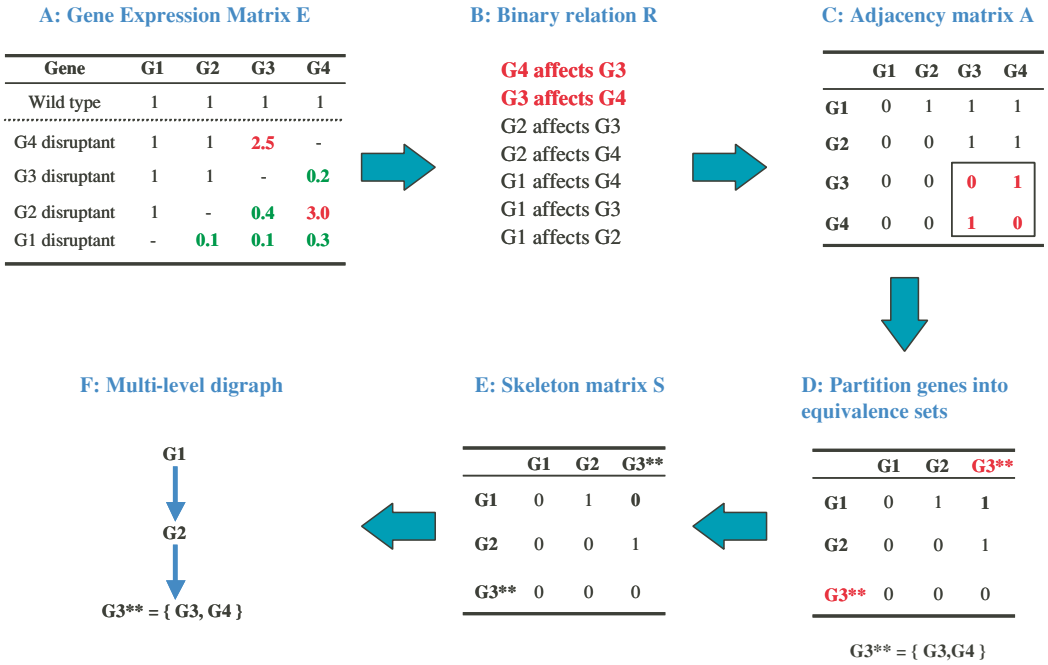

Figure 1. Construction of a gene regulatory subnetwork model using the Boolean method. A; Numerous profiles are integrated into an expression matrix. Each element of the matrix indicates the ratio of gene expression. B; evaluate binary relationships between genes. If gene 'G1' is deleted and the intensity of gene 'G2' is significantly altered as a result then gene 'G1' affects gene 'G2'. C; Identify looped regulation genes. If gene 'G3' and gene 'G4' affect each other mutually, they form a loop (strongly connected component) regulation. D; An equivalence set is introduced which treats a loop logically as a gene. An equivalence set is a group of genes that affect each other or as a group effect one discrete gene. E; A skeleton matrix is reconstructed. The shortest path relationships should be selected to build hierarchical connections. F; The regulation pathway can be formed from the skeleton matrix.

combinatorial disruption experiments.

2. Materials and Methods

2.1. Boolean network inference algorithms (Fig. 1)

A static Boolean network model based on the multi-level digraph approach infers gene regulatory networks by using a set of binary relations between genes. In the model, a network is expressed by a directed graph using symbols of a gene and a relation between two-paired genes represented as a node and an arc, respectively. A gene expression matrix E is created from a set of gene disruption experiments. The value of matrix element $E(a, b)$ indicates the expression ratio of gene 'b' to the normal condition, in the experiment of gene 'a' deletion. The inference procedures are as follows:

(A) Obtain the gene expression matrix E using several sets of the gene expression patterns resulting from disruption of one gene.

(B) Binary relation R; Using the gene expression matrix E, if the intensity of gene 'b' is changed higher than a given threshold value θ, or is changed lower than a given threshold $1/\theta$ resulting from the disruption of gene 'a,' it is defined that gene 'a' affects gene 'b' directly or indirectly. Thus the binary relation R is created by cutting the value of each element in the gene expression matrix E at the threshold (θ or $1/\theta$).

(C) Adjacency matrix A; The adjacency matrix A is derived directly from the binary relation R. If there is a relation that gene 'a' affects gene 'b', then the value of element (a, b) in the adjacency matrix A is set to 1; $A(a, b) = 1$.

(D) Equivalence set; If there is the relation, such that gene 'a' and 'b' affect each other, that is $A(a, b) = A(b, a) = 1$, we cannot decide which gene is located upstream or downstream in a gene regulatory network. This is the limitation and disadvantage of this method, however, we introduce an "*equivalence set*", which comprises a group of genes affecting each other and the group is treated as a one-gene node for determining influence on other genes. In Fig. 1, genes G3 and G4 are partitioned into a equivalence set G3**, or a new artificial gene G3**, and the adjacency matrix A is contracted to a 3×3 matrix. After this process, any two genes, gene 'a' and gene 'b', stand in only one relation, that is "Gene a affects gene b," "Gene b affects gene a" or "Gene a and b are independent." This shows that genes are in a semi-ordered (topologically sorted) matrix.

(E) Skeleton matrix S; Semi-ordered matrix A($n \times n$ matrix) includes indirect effects between genes. In order to remove them, we process as follows: The value of line i and column j in semi-ordered matrix A and skeleton matrix S are represented as $A(i, j)$ and $S(i, j)$, respectively. If $A(i, j)$ equals 1, $S(i, k)$ is set to $\max\{A(i, k) - A(j, k), 0\}(k = 1, \ldots, n)$. Thus, all indirect effects are removed from the semi-ordered matrix. In Fig. 1, the relation between gene G1 and gene G3** are removed, and the skeleton matrix S is constructed.

(F) Draw multi-level digraph; Arrows are drawn between nodes based on the value of each element in the skeleton matrix. The genes with parentheses indicate an equivalence set of genes. The multi-level digraph method shown in Fig. 1 is consistent with the experiment data E and the binary relation R.

2.2. Microarray experiments

We collected gene expression data using full-genome yeast c-DNA microarrays.[10] BY4741 (*MATa*, *HIS3D1*, *LEU2D0*, *MET15D0*, *URA3D0*) served as the wild type strain. Gene disruptions for strain BY4741 were purchased from Research Genetics, Inc. Cells were inoculated and grown in YPD medium (1% yeast extract, 1% bacto-peptone, 2% glucose) at 30°C until OD_{600} reached 1.0 in the logarithmic growth phase and then harvested to isolate mRNA for assay of gene expression. The respective parental strain was the control used for each disruptant strain.

2.3. Data Normalization

We measured the quantities of 5871 cDNA in microarray assays of 155 yeast disruptant species. A difference in fluorescent strength between Cy3, Cy5 causes bias of the expression quantity ratio. We normalized the expression quantity ratios of each expression profile. The ratio bias had a fixed trend in each spotted block, thus we calculated a linear regression to normalize the mean value ratio of each block to 1.0.

The logarithm value of the ratio was used to indicate the standard expression level; therefore, we found the logarithm value of the ratio and calculated the average and standard deviation (SD) of these log values (Table 1). The SD of expression levels of all spotted genes from the *UME6* (*YDR207C*) disruptant expression array for which UME6 is defined as a "Global Regulator" in Yeast Protein Database (YPD (http://www.incyte.com/sequence/proteome/databases/YPD.shtml)) was 0.4932. We thus recognized that an unacceptable number of errors exist in any array data whose overall SD was larger than 0.5 and we eliminated 37 expression experiments from this analysis.

2.4. Selection of genes for modeling

In YPD, 314 genes were defined as "Transcription factors," and 98 of these have previously been studied for control mechanism. The expression profile data of 552 genes including the genes controlled by these 98 "Transcription factors" were selected from 5871 profiles. Thus, we constructed the gene regulatory network from the expression profile data set based on the values of these 552 genes in 118 separate gene disruption experiments.

3. Results and Discussion

3.1. 552 gene member regulatory networks

Each of the genes in this library was selected for experimentation because it was reported in the Yeast Proteome Database (YPD) to be a factor involved in transcription regulation. Previously reported gene regulatory networks show that genes can interact with themselves as well as with other regulatory genes. To reconstruct hierarchical regulatory relationships from the expression library, we applied a novel Boolean algorithm that accommodates common looped regulatory relationships to our data.[11] As shown in Fig. 1, the gene regulatory relationships modeled by this method can be represented as a directed graph of upregulation or downregulation of gene expression between 2 given genes of the 5871 genes measured in each expression experiment. We constructed a 552-gene member model of the regulatory control relationships evident in the library. For further discussion, we constrained the data to a subnetwork model comprised of 98 well-known transcription factors. The resultant model contains a total of 552 nodes representing the included genes and 2953 putative regulatory links among these genes.

The relations describe the effects of one gene on the expression level of the other genes and, in this study, are obtained from the data of gene disruption experiments. Systematic analysis enables us to reconstruct a network that is consistent with all of the relation data.

3.2. Relationships among functional categories

We classified transcription factors in the network model according to their cellular functional roles (CFR) as defined in YPD. Figure 2 shows the control relationships among classified transcription factors in the network. We identified several control lines emanating from "carbohydrate metabolism" genes to all other functional gene groups. This finding is consistent with the energy-dependent nature of many cellular processes and metabolic pathways. As shown in Fig. 2, a distinct feature is that expression levels of lipid fatty-acid metabolism transcription factors were exclusively under control of carbohydrate metabolism transcription factors.

Table 1. List of disruptants.

Disrupted Gene name	Standard Deviation logarithm of ratio (in base 2)	selction	Dirupted Gene name	Standard Deviation logarithm of ratio (in base 2)	selection
UME6	0.4932	selected	CWP1	0.2575	selected
OSH1	0.3677	selected	TPK3	0.8840	not selected
HTB2	0.4381	selected	ASH1	0.2968	selected
HTA2	0.4062	selected	CNB1	0.4961	selected
ACH1	0.3297	selected	BAS1	0.6352	not selected
HAP3	0.3480	selected	YLL055C	0.3423	selected
YBL037C	0.4966	selected	GAT3	0.3905	selected
YBL055W	0.2859	selected	RIC1	0.3550	selected
YBL056C	0.6857	not selected	SPT8	0.3739	selected
BOI1	0.3176	selected	ACE2	0.3524	selected
SMP1	0.3208	selected	SWI6	0.3873	selected
HPC2	0.7516	not selected	CPR6	0.3532	selected
THI2	0.3780	selected	ECM22	0.3998	selected
MAK31	0.3824	selected	RCK2	0.3205	selected
BDF2	0.7060	not selected	YLR266C	0.3673	selected
GCS1	0.3845	selected	YLR278C	0.2928	selected
SOK1	0.7753	not selected	BUD6	0.3136	selected
YDR027C	0.7210	not selected	NIT3	0.5772	not selected
LYS14	0.4600	selected	BUD8	0.3264	selected
NRG1	0.3298	selected	BDF1	0.5327	not selected
INO2	0.4785	selected	LEU3	0.3664	selected
SWI5	0.4021	selected	GAL80	0.4073	selected
STB3	0.3364	selected	CMP2	0.6240	not selected
HST4	0.3405	selected	YML077C	0.4298	selected
YAR003W	0.3398	selected	YML082W	0.3183	selected
YDR214W	0.7977	not selected	ARGR2	0.3965	selected
ADR1	0.2955	selected	TUB3	0.2506	selected
HTA1	0.5430	not selected	SOK2	0.6848	not selected
BTT1	0.4487	selected	STB4	0.2767	selected
RMS1	0.4631	selected	TSP1	0.3109	selected
YAP6	0.7667	not selected	ARGR1	0.5233	not selected
YDR295C	0.8278	not selected	YMR100C	0.6441	not selected
SUM1	0.4828	selected	SAS2	0.5227	not selected
SWR1	0.4592	selected	RGM1	0.2902	selected
YDR341W	0.3903	selected	CAT8	0.4778	selected
CAD1	0.3674	selected	SCW10	0.2882	selected
GCN4	0.4791	selected	SPS18	0.3440	selected
HAT2	0.3774	selected	STB1	0.3536	selected
HPA3	0.4060	selected	DAL82	0.3750	selected
YER028C	0.2712	selected	KRE1	0.3071	selected
FIR1	0.3754	selected	RPD3	0.2987	selected
DOT6	0.3726	selected	YNR064W	0.3063	selected
YER131C	0.4844	selected	PHO80	0.8956	not selected
HAC1	0.3698	selected	SIN3	0.6336	not selected
RPO41	0.7767	not selected	YAP7	0.3854	selected
YFL053W	0.4188	selected	YOL042W	1.0904	not selected
YFR039W	0.5023	not selected	YOL054W	0.5047	not selected
SCW11	0.2715	selected	THI20	0.3857	selected
DST1	0.2939	selected	RTG1	0.3816	selected
SIP2	0.3850	selected	HAL9	0.6720	not selected
MIG2	0.2886	selected	INO4	0.3310	selected
SKI8	0.3201	selected	CIN5	0.2901	selected
HAP2	0.3306	selected	STD1	0.3062	selected
FZF1	0.3184	selected	SFL1	0.4158	selected
BUD9	0.3171	selected	YOR173W	0.5837	not selected
RME1	0.3380	selected	SAS5	0.7989	not selected
RSC1	0.4529	selected	TEA1	0.3619	selected
SCW4	0.2966	selected	PIP2	0.3702	selected
BGL2	0.3176	selected	HST2	0.3836	selected
OPI1	0.3983	selected	EGD1	0.3485	selected
EGD2	0.4532	selected	YPL061W	0.3019	selected
SKN7	0.4098	selected	RLM1	0.3104	selected
BAR1	0.3554	selected	YPL134C	0.7037	not selected
SDS3	0.4920	selected	UME1	0.3773	selected
MET28	0.3963	selected	BEM4	0.3766	selected
MUC1	0.3001	selected	YPL166C	0.6183	not selected
MRS1	0.3937	selected	CUP9	0.3730	selected
GZF3	0.2798	selected	PKA3	0.5210	not selected
LSM1	0.4317	selected	YPL230W	0.5565	not selected
NIT2	0.5389	not selected	GAL4	0.4492	selected
TPK1	0.4648	selected	YPR013C	0.6422	not selected
ASG7	0.2709	selected	YPR022C	0.7080	not selected
BUD4	0.3086	selected	YPR115W	1.0641	not selected
HMS2	0.3071	selected	YPR125W	0.3931	selected
SPT23	0.4320	selected	YPR145C	0.8397	not selected
RGT1	0.4011	selected	HPA2	0.3506	selected
STB6	0.3717	selected	YPR197W	0.4002	selected
MBR1	0.4121	selected			

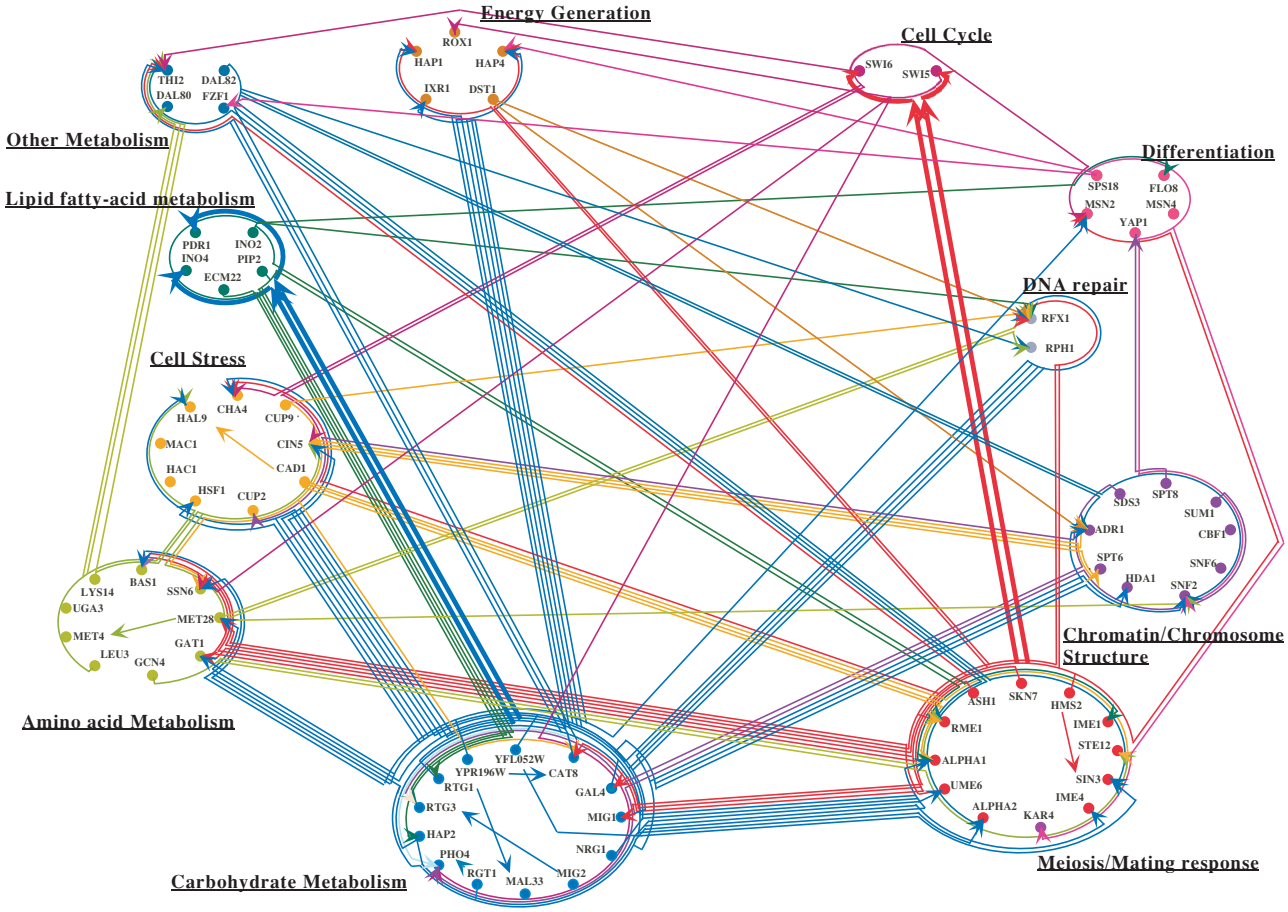

Figure 2. Transcription factor regulatory network model classified by cellular functional roles (CFR). In this model, there were 98 transcription factors grouped by cellular functional roles according to information provided in the Yeast Proteome Database. Genes inside the circles are grouped into a given cellular functional category. Regulatory control relationships are depicted with colored lines. Colors indicate the category of genes from which the control relationships emanate. Blue: Carbohydrate metabolism, Bluish purple: Chromatin/Chromosome structure, Brown: Energy generation, Dark Green: Other metabolism, Gray: DNA repair, Green: Lipid, fatty acid metabolism, Light Green: Amino acid metabolism, Orange: Cell stress, Red: Meiosis/Mating response, Pink: Differentiation, Purple: Cell cycle. The bold red lines indicate regulatory control of transcription factors related to cell cycle originating from genes in the meiosis/mating response group. The bold blue lines indicate control relationships emanating from the carbohydrate metabolism group exerting influence over genes in the lipid fatty-acid metabolism group.

In our CFR model, only 2 individual genes, *SKN7*, which can activate G1 cyclins through the cell cycle box (CCB) element and Mlu1 cell cycle box (MCB) element,[12] and *HMS2*, which codes a transcription factor with a probable role in pseudohyphal growth,[13] directly influenced cell cycle-related gene expression. These genes and their related cellular functions would necessitate some, albeit limited, interaction with the cell cycle process, whereas general control over cell cycle events by many genes would decrease the biological stability of cell division processes, which would not be evolutionarily beneficial. It is known, for example, that eukaryotic organisms use discreet and highly conserved cell-cycle checkpoints to ensure that nuclear division is restrained while DNA is undergoing replication or repair.[14,15] Likewise, it is apparent that expression levels of lipid fatty-acid metabolism transcription factors were affected by the expression of carbohydrate metabolism transcription factors. Also, *PDR1*, which recognizes the Pdr1p/Pdr3p response element (PDRE) to activate genes involved in multi-drug resistance of yeast,[16] is influenced by the expression of *CAT8*, which controls key enzymes of gluconeogenesis in yeast.[17] Interactions among proteins involved in phospholipid synthesis pathway with genes of the glucose response pathway, the lipid signaling pathway and other lipid synthesis pathways have been documented.[17] Genes encoding many enzymes of phospholipid biosynthesis contain variants of UAS_{INO} in their promoters and these are regulated in response to growth phase and nutrient starvation.[18] For example, Snf1p/Snf4p and Ire1p are reported to be key molecules that bind to the UAS_{INO} sequence, which are associated with the glucose response,[19] and are required for de-repression of UAS_{INO}-containing genes.

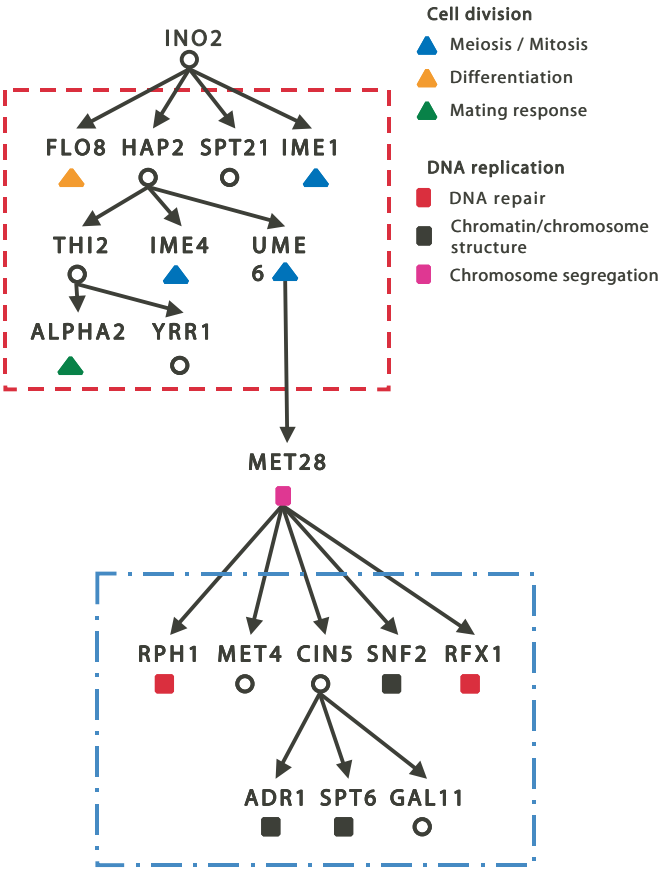

Figure 3. A detailed view of a gene regulatory subnetwork of transcription factors reconstructed from Boolean analysis of gene expression experiments on disruptant mutant yeast strains. Black lines indicate a regulatory relationship, with the arrows showing the direction of expression influence. The colors and shapes of the nodes denote the general categories of cellular function of the gene product according to their descriptions in the YPD. Genes related to cell division mechanisms are indicated with triangular nodes and genes related to DNA repair and chromosome structure are depicted with squares. The elucidated network shows novel topological control relationships among genes related to meiosis, the mating response and DNA structure and repair mechanisms via *UME6* and *MET28*. Genes related to meiosis and mating response genes are downstream of a cascade regulated by *INO2* and a further sub-grouping of genes related to DNA repair and structure appears hierarchically downstream of *MET28*.

3.3. MET28 *regulation*

The models were employed to further explore detailed relationships between expression regulatory genes, such as transcription factors with regulatory and non-regulatory genes, from all of the gene expression experimental data. We were able to characterize the regulatory roles of genes with unreported biological function by virtue of their expression control by and/or over genes with known function. Figure 3 shows the largest hierarchical transcription factor subnetwork elucidated by our network calculations. Novel control relationships among transcription factors involved in cell division regulation and DNA replication/repair regulation were found in this largest subnetwork. Two discrete functional branches in the subnetwork that correspond to cell division regulation and DNA replication/repair are linked by *UME6* and *MET28*, indicating the important role of these two transcription factors in coordinating the expression regulation of these interdependent regulatory pathways. *MET28*, as its name suggests, was previously characterized as a transcription factor related to methionine metabolism.[20] The novel putative role for Met28p in regulation of chromosome segregation is supported by its reported interaction with known chromosomal segregation component Smc1p, as part of a larger nexus of chromosomal segregation proteins, in mating-type two-hybrid assays.[21]

3.4. Detection of transcription factor binding sites

Through sequence analysis of coding sequences and upstream regions of genes in the above-mentioned subnetwork, we validated the sequence level control mechanisms between transcription factors and their target genes and DNA binding sequences. In the case of *UME6*, which is known as a global transcriptional regulation of many meiotic genes[22,23] and its control system, we performed Multiple Expectation-Maximization for motif Elicitation (MEME) analysis[24] of regions upstream of the 34 genes controlled by *UME6* in our model. We found two consensus sequences, TAGCCGCCGA and TGGGCGGCTA, that were present upstream of 14.7% and 32.4% of the 34 genes respectively, and that had significant P values according to the MEME search (Table 2 http://www.grt.kyushu-u.ac.jp/data/table2.html). According to the TRANSFAC database, TAGCCGCCGA is defined as the binding site of Ume6p[22] and TCGGCGGCA is reported to be the binding site of a repressor of *CAR1*[25] which were repressed by a three-component complex containing Ume6p (Ume6p, Sin3p, and Rpd3p).[26]

We identified 103 genes in the yeast genome in which these Ume6p-associated binding motifs can be found within 500 bp upstream of the coding sequence and divided those genes among 9 functional categories based on their descriptions in YPD (Table 3 http:// www.grt.kyushu-u.ac.jp/data/table3.html). *UME6* is a known global transcriptional regulator involved in pathway-specific repression and induction of many meiotic genes.[22,23] Our results support this literature and add new biological insight into the specific regulatory mechanisms of action in *UME6*-mediated meiotic transcription control. From the genes with putative *UME6*-related sequences according to expression profiles in *UME6* disruptant experiments, the expression of almost all genes related to meiosis containing the Ume6p binding motif were affected by the absence of *UME6* whereas the other functional categories showed no such association.

Aside from the Ume6p-related binding motifs, no other MEME consensus sequence was present upstream of the 11 meiosis-related genes, a finding which suggests that these 11 genes are regulated exclusively by *UME6* and that Ume6p directly influences their expression. Conversely, only two other genes possessed the putative binding sequence but did not show expression influence on the count of *UME6* in our experiments.

Experimentally driven discovery of network models of expression control allows for specific biological insights relevant to gene regulatory pathways that are not readily reconstructed from the available biological literature or present in pre-compiled pathway databases. We have shown here that such a system is useful in discovering novel gene function information as well as novel regulatory mechanisms. The use of this and similar strategies to elucidate hierarchical regulatory pathways from full genome expression libraries will allow for rapid insight into transcription regulation that can be applied to fields such as rational drug discovery and agrochemical targeting.

Acknowledgements: We thank Dr. Osamu Higuchi for providing the cDNAs for microarray production. This work was supported by JSPS research for the future program, the Genome Frontier Project, and a Grant-in-Aid for Scientific Research on Priority Areas and Genomefrontier from the MEXT. The authors contributed equally to this work.

References

1. Pilpel, Y., Sudarsanam, P., and Church, G. M. 2001, Identifying regulatory networks by combinatorial analysis of promoter elements, *Nature Genetics*, **29**, 153–159.
2. Friedman, N., Linial, M., Nachman, I., and Pe'er, D. 2000, Using Bayesian Networks to Analyze Expression Data, *J. Comput Biol.*, **7**, 601–620.
3. Somogyi, R. and Sniegoski, C. A. 1996, Modeling the complexity of genetic networks: Understanding multigene and pleiotropic regulation, *Complexity*, **1**, 45–63.
4. Spellman, P. T., Sherlock, G., Zhang, M. Q. et al. 1998, Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast *Saccharomyces cerevisiae* by Microarray Hybridization, *Mol. Cell Biol.*, **9**, 3273–3297.
5. Gasch, A. P., Spellman, P. T., Kao, C. M. et al. 1997, Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes, *Mol. Cell Biol.*, **17**, 2099–2106.
6. Akutsu, T., Miyano, S., and Kuhara, S. 2000, Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function, *J. Comp. Biol.*, **7**, 331–343.
7. Akutsu, T., Miyano, S., and Kuhara, S. 2000, Inferring qualitative relations in genetic networks and metabolic pathways, *Bioinformatics*, **16**, 727–734.
8. Brown, M. P., Grundy, W. N., Lin, D. et al. 2000, Knowledge-based analysis of microarray gene expression data by using support vector machines, *Proc. Natl. Acad. Sci. USA*, **97**, 262–267.
9. Ideker, T., Thorsson, V., Ranish, J. A. et al. 2001, Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network, *Science*, **292**, 929–934.
10. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. 1995, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science*, **270**, 467–470.
11. Shoudan, L. 1998, REVEAL, a general reverse engineering algorithm for inference of genetic network architectures, *Pac. Sym. Bio.*, **3**, 18–29.
12. Morgan, B. A., Bouquin, N., Merrill, G. F., and Johnston, L. H. 1995, A yeast transcription factor bypassing the requirement for SBF and DSC1/MBF in budding yeast has homology to bacterial signal transduction proteins, *EMBO J.*, **14**, 5679–5689.
13. Lorenz, M. C. and Heitman, J. 1998, Regulators of pseudohyphal differentiation in *Saccharomyces cerevisiae* identified through multicopy suppressor analysis in ammonium permease mutant strains, *Genetics*, **150**, 1443–1457.
14. Aguilera, A., Chavez, S., and Malagon, F. 2000, Mitotic recombination in yeast: elements controlling its incidence, *Yeast*, **16**, 731–754.
15. Dasika, G. K., Lin, S. C., Zhao, S., Sung, P., Tomkinson, A., and Lee, E. Y. 1999, DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis, *Oncogene*, **18**, 7883–7899.
16. Miura, F., Yada, T., Nakai, K., Sakaki, Y., and Ito, T. 2001, Differential display analysis of mutants for the transcription factor Pdr1p regulating multidrug resistance in the budding yeast, *FEBS Lett.*, **505**, 103–108.
17. Hiesinger, M., Roth, S., Meissner, E., and Schuller, H. J. 2001, Contribution of Cat8 and Sip4 to the transcriptional activation of yeast gluconeogenic genes by carbon source-responsive elements, *Curr. Genet.*, **39**, 68–76.
18. Carman, G. M. and Henry, S. A. 1999, Phospholipid biosynthesis in the yeast *Saccharomyces cerevisiae* and interrelationship with other metabolic processes, *Prog. Lipid Res.*, **38**, 361–399.
19. Jiang, R. and Carlson, M. 1997, The Snf1 protein kinase and its activating subunit, Snf4, interact with distinct domains of the Sip1/Sip2/Gal83 component in the kinase complex, *Mol. Cell Biol.*, **17**, 2099–2106.
20. Kuras, L., Cherest, H., Surdin-Kerjan, Y., and Thomas, D. 1996, A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediate the transcription activation of yeast sulfur metabolism, *EMBO J.*, **15**, 2519–2529.
21. Newman, J. R. S., Wolf, E., and Kim, P. S. 2000, A computationally directed screen identifying interacting coiled coils from *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci USA.*, **97**, 13203–13208.
22. Bowdish, K. S., Yuan, H. E., and Mitchell, A. P. 1995, Positive control of yeast meiotic genes by the negative regulator UME6, *Mol. Cell Biol.*, **15**, 2955–2961.
23. Rubin-Bejerano, I., Mandel, S., Robzyk, K., and Kassir, Y. 1996, Induction of meiosis in *Saccharomyces cerevisiae* depends on conversion of the transcriptional repressor

Ume6 to a positive regulator by its regulated association with the transcriptional activator Ime1, *Mol. Cell Biol.*, **16**, 2518–2526.

24. Bailey, T. L. and Elkan, C. 1994, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, AAAI Press Menlo Park, California, 28–36.
25. Luche, R. M., Sumrada, R., and Cooper, T. G. 1990, A cis-acting element present in multiple genes serves as a repressor protein binding site for the yeast CAR1 gene, *Mol. Cell Biol.*, **10**, 3884–3895.
26. Messenguy, F., Vierendeels, F., Scherens, B., and Dubois, E. 2000, In *Saccharomyces cerevisiae*, expression of arginine catabolic genes CAR1 and CAR2 in response to exogenous nitrogen availability is mediated by the Ume6 (CargRI)-Sin3 (CargRII)-Rpd3 (CargRIII) complex, *J. Bacteriol.*, **182**, 3158–3164.