transport (*32*). The model was also used for tracer forecasts during the MINOS campaign.
40. E. Roeckner *et al*. "The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate," *MPI-Report 218* (Max Planck Institute for Meteorology, Hamburg, Germany, 1996).
41. E. Roeckner, L. Bengtsson, J. Feichter, J. Lelieveld, H. Rodhe, *J. Clim.* **12**, 3004 (1999).
42. N. A. Rayner, E. B. Horton, D. E. Parker, C. K. Folland, R. B. Hackett, Version 2.2 of the Global Sea-Ice and Sea Surface Temperature Data Set, 1903–1994, Climate Research Technical Note 74 (Hadley Centre, Bracknell, UK, 1996); available at www.met-office.gov.uk/research/hadleycentre/obsdata/GISST.html.
43. S. Nicholson, S. Some, B. Kone, *J. Clim.* **13**, 2628 (2000).
44. D. P. Rowell, C. K. Folland, K. Maskell, M. L. Ward, *Q. J. Meteorol. Soc.* **121**, 669 (1995).
45. A. Avila, F. Rodà, *Atmos. Environ.* **36**, 2881 (2002).
46. The Programme for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe (EMEP), at www.emep.int, offers estimates of emissions variable over time (since 1980) of European anthropogenic trace species.
47. D. Rosenfeld, *Science* **287**, 1793 (2000).
48. L. D. Rotstayn, U. Lohmann, *J. Clim.,* **15**, 2103 (2002).
49. A. S. Board *et al.*, *J. Geophys. Res.* **104**, 16181 (1999).
50. A. D. Clarke *et al.*, *J. Geophys. Res.* **106**, 32555 (2001).
51. H. Singh *et al.*, *Nature* **410**, 1078 (2001).
52. J. Sciare, E. Baboukas, N. Mihalopoulos, *J. Atmos. Chem.* **39**, 281 (2001).
53. S. Mylona, *Tellus* **48B**, 662 (1996).
54. J. van Aardenne, F. J. Dentener, J. G. J. Olivier, C. G. M. Klein Goldewijk, J. Lelieveld, *Global Biogeochem. Cycles* **15**, 909 (2001). The SO_2 emission estimates are inherently uncertain. The SO_2 decrease after 1980 is nevertheless consistent with sulfate concentration and deposition data (*27–31*, *45*, *46*).
55. Funding from the NSF contributed to this work. The King Air flights were supported by the Israeli Space Agency as part of the Mediterranean Israeli Dust Experiment (MEIDEX).

25 June 2002; accepted 17 September 2002

Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*

Tong Ihn Lee,[1]* Nicola J. Rinaldi,[1,2]* François Robert,[1]* Duncan T. Odom,[1] Ziv Bar-Joseph,[3] Georg K. Gerber,[3] Nancy M. Hannett,[1] Christopher T. Harbison,[1,2] Craig M. Thompson,[1]† Itamar Simon,[1] Julia Zeitlinger,[1] Ezra G. Jennings,[1,2] Heather L. Murray,[1] D. Benjamin Gordon,[1] Bing Ren,[1]‡ John J. Wyrick,[1]§ Jean-Bosco Tagne,[1] Thomas L. Volkert,[1] Ernest Fraenkel,[1] David K. Gifford,[3] Richard A. Young[1,2]‖

We have determined how most of the transcriptional regulators encoded in the eukaryote *Saccharomyces cerevisiae* associate with genes across the genome in living cells. Just as maps of metabolic networks describe the potential pathways that may be used by a cell to accomplish metabolic processes, this network of regulator-gene interactions describes potential pathways yeast cells can use to regulate global gene expression programs. We use this information to identify network motifs, the simplest units of network architecture, and demonstrate that an automated process can use motifs to assemble a transcriptional regulatory network structure. Our results reveal that eukaryotic cellular functions are highly connected through networks of transcriptional regulators that regulate other transcriptional regulators.

Genome sequences specify the gene expression programs that produce living cells, but how cells control global gene expression programs is far from understood. Each cell is the product of specific gene expression programs involving regulated transcription of thousands of genes. These transcriptional programs are modified as cells progress through the cell cycle, in response to changes in environment, and during organismal development (*1–5*).

Gene expression programs depend on recognition of specific promoter sequences by transcriptional regulatory proteins (*6–9*). Because these regulatory proteins recruit and regulate chromatin-modifying complexes and components of the transcription apparatus, knowledge of the sites bound by all the transcriptional regulators encoded in a genome can provide the information necessary to nucleate models for transcriptional regulatory networks. With the availability of complete genome sequences and development of a method for genome-wide binding analysis (also known as genome-wide location analysis), investigators can identify the set of target genes bound in vivo by each of the transcriptional regulators that are encoded in a cell's genome. This approach has been used to identify the genomic sites bound by nearly a dozen regulators of transcription (*10–13*) and several regulators of DNA synthesis (*14*) in yeast.

Experimental design. We used genome-wide location analysis to investigate how yeast transcriptional regulators bind to promoter sequences across the genome (Fig. 1A). All 141 transcription factors listed in the Yeast Proteome Database (*15*) and reported to have DNA binding and transcriptional activity were selected for study. Yeast strains were constructed so that each of the transcription factors contained a myc epitope tag. To increase the likelihood that tagged factors were expressed at physiologic levels, we introduced epitope tag coding sequences into the genomic sequences encoding the COOH terminus of each regulator, as described in (*16*). We confirmed appropriate insertion of the tag and expression of the tagged protein by polymerase chain reaction and immunoblot analysis. Introduction of an epitope tag might be expected to affect the function of some transcriptional regulators; for 17 of the 141 factors, we were not able to obtain viable tagged cells, despite three attempts to tag each regulator. Not all the transcriptional regulators were expected to be expressed at detectable levels when yeast cells were grown in rich medium, but immunoblot analysis showed that 106 of the 124 tagged regulator proteins could be detected under these conditions.

We performed a genome-wide location analysis experiment (*10*) for each of the 106 yeast strains that expressed epitope-tagged regulators (*17, 18*). Each tagged strain was grown in three independent cultures in rich medium (yeast extract, peptone, dextrose). Genome-wide location data were subjected to quality control filters and normalized, and the ratio of immunoprecipitated to control DNA was determined for each array spot. We calculated a confidence value (*P* value) for each spot from each array by using an error model (*19*). The data for each of the three samples in an experiment were combined by a weighted average method (*19*); each ratio was weighted by *P* value and then averaged. Final *P* values for these combined ratios were then calculated (*17, 18*).

Given the properties of the biological sys-

[1]Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]MIT Laboratory of Computer Science, 200 Technology Square, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.
†Present address: Akceli Inc., 1 Hampshire Street, Cambridge, MA 02139, USA.
‡Present address: Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093, USA.
§Present address: California Institute of Technology, Pasadena, CA 91125, USA.
‖To whom correspondence should be addressed. E-mail: young@wi.mit.edu

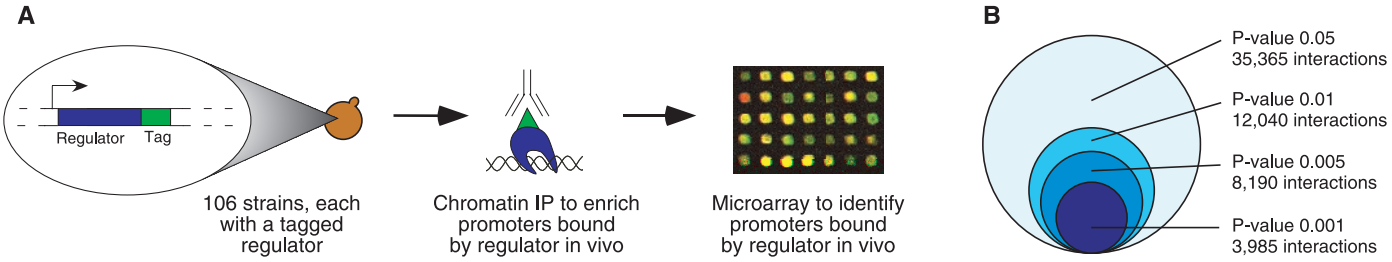


Fig. 1. Systematic genome-wide location analysis for yeast transcription regulators. (**A**) Methodology. Yeast transcriptional regulators were tagged by introducing the coding sequence for a c-*myc* epitope tag into the normal genomic locus for each regulator. Of the yeast strains constructed in this fashion, 106 contained a single epitope-tagged regulator whose expression could be detected in rich growth conditions. Chromatin immunoprecipitation (ChIP) was performed on each of these 106 strains. Promoter regions enriched through the ChIP procedure were identified by hybridization to microarrays containing a genome-wide set of yeast promoter regions. (**B**) Effect of *P* value threshold. The sum of all regulator-promoter region interactions is displayed as a function of varying *P* value thresholds applied to the entire location data set for the 106 regulators. More stringent *P* values reduce the number of interactions reported but decrease the likelihood of false-positive results.

tem studied here (cell populations, DNA binding factors capable of binding to both specific and nonspecific sequences) and the expectation of noise in microarray-based data, it was important to use error models to obtain a probabilistic assessment of regulator location data. The total number of protein-DNA interactions in the location analysis data set, using a range of *P* value thresholds, is shown in Fig. 1B. We selected specific *P* value thresholds to facilitate discussion of a subset of the data at a high confidence level, but note that this artificially imposes a "bound or not bound" binary decision for each protein-DNA interaction.

We generally describe results obtained at a *P* value threshold of 0.001 because our analysis indicates that this threshold maximizes inclusion of legitimate regulator-DNA interactions and minimizes false positives. Various experimental and analytical methods indicate that the frequency of false positives in the genome-wide location data at the 0.001 threshold is 6% to 10% (*17, 18*). For example, conventional, gene-specific chromatin immunoprecipitation experiments have confirmed 93 of 99 binding interactions (involving 29 different regulators) that were identified by location analysis data at a threshold *P* value of 0.001. Use of a high-confidence threshold should underestimate the regulator-DNA interactions that actually occur in these cells. We estimate that about one-third of the actual regulator-DNA interactions in cells are not reported at the 0.001 threshold (*17, 18*).

Regulator density. We observed nearly 4000 interactions between regulators and promoter regions at a *P* value threshold of 0.001. The promoter regions of 2343 of 6270 yeast genes (37%) were bound by one or more of the 106 transcriptional regulators in yeast cells grown in rich medium. Many yeast promoters were bound by multiple transcriptional regulators (Fig. 2A), a feature previously associated with gene regulation in higher eukaryotes (*20, 21*), suggesting that yeast genes are also frequently regulated through combinations of regulators. More than one-third of the promoter regions that are bound by regulators were bound by two or more regulators (*P* value threshold = 0.001), and, relative to the expected distribution from randomized data, a disproportionately high number of promoter regions were bound by four or more regulators. Because of the stringency of the *P* value threshold, we expect that this represents an underestimate of regulator density.

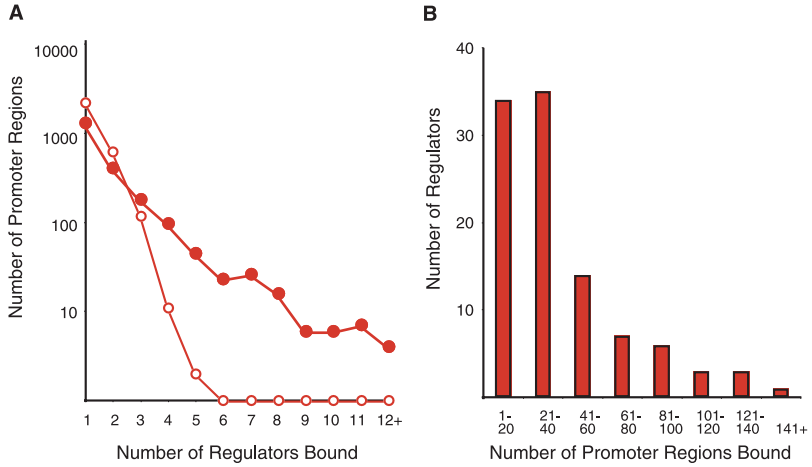


Fig. 2. Genome-wide distribution of transcriptional regulators. (**A**) Plot of the number of regulators bound per promoter region. The distribution for the actual location data (red circles) is shown alongside the distribution expected from the same set of *P* values randomly assigned among regulators and intergenic regions (white circles). At a *P* value threshold of 0.001, significantly more intergenic regions bind four or more regulators than expected by chance. (**B**) Distribution of the number of promoter regions bound per regulator.

The number of different promoter regions bound by each regulator in cells grown in rich medium ranged from 0 to 181 (*P* value threshold = 0.001), with an average of 38 promoter regions per regulator (Fig. 2B). The regulator Abf1 bound the largest number (181) of promoter regions. Regulators that should be active under growth conditions other than yeast extract, peptone, and dextrose were typically found, as expected, to bind the smallest number of promoter regions. For example, Thi2, which activates transcription of thiamine biosynthesis genes under conditions of thiamine starvation (*22, 23*), was among the regulators that bound the smallest number (3) of promoters. Identification of a set of promoter regions that are bound by specific regulators allowed us to predict sequence motifs that are bound by these regulators (*17, 18*).

Network motifs. The simplest units of commonly used transcriptional regulatory network architecture, or network motifs, provide specific regulatory capacities such as positive and negative feedback loops. We used the genome-wide location data to identify six regulatory network motifs: autoregulation, multicomponent loops, feedforward loops, single-input, multi-input, and regulator chain (Fig. 3). These motifs suggest models for regulatory mechanisms that can be tested. Descriptions of the algorithms used to identify motifs and a complete compilation of motifs can be obtained in (*18*).

An autoregulation motif consists of a regulator that binds to the promoter region of its own gene. We identified 10 autoregulation motifs with genome-wide location data for the 106 regulators (*P* value threshold = 0.001), which suggests that about 10% of yeast genes encod-

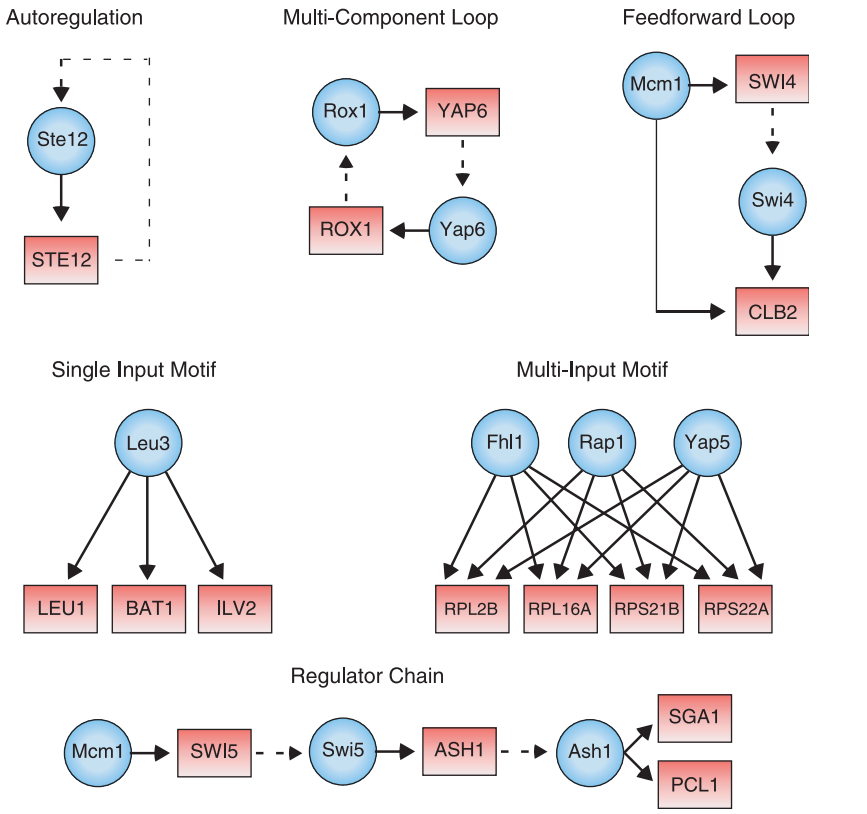


Fig. 3. Examples of network motifs in the yeast regulatory network. Regulators are represented by blue circles; gene promoters are represented by red rectangles. Binding of a regulator to a promoter is indicated by a solid arrow. Genes encoding regulators are linked to their respective regulators by dashed arrows. For example, in the autoregulation motif, the Ste12 protein binds to the *STE12* gene, which is transcribed and translated into Ste12 protein. These network motifs were uncovered by searching binding data with various algorithms. For details on the algorithms used and a full list of motifs found, see (*18*).

ing regulators are autoregulated. This percentage does not change substantially at less stringent *P* value thresholds. In contrast, studies of *Escherichia coli* genetic regulatory networks indicate that most (52% to 74%) prokaryotic genes encoding transcriptional regulators are autoregulated (*24, 25*).

Autoregulation is thought to provide several selective growth advantages, including reduced response time to environmental stimuli, decreased biosynthetic cost of regulation, and increased stability of gene expression (*24–28*). For example, upon exposure to mating pheromone, the concentrations of the pheromone-responsive Ste12 transcriptional regulator rapidly increase because Ste12 binds to and up-regulates its own gene (*10, 29*) (Fig. 3). The consequent increase in Ste12 protein leads to the binding of other genes required for the mating process (*10*).

A multicomponent loop motif consists of a regulatory circuit whose closure involves two or more factors (Fig. 3). We observed three multicomponent loop motifs in the location data for 106 regulators (*P* value threshold = 0.001). The closed-loop structure provides the capacity for feedback control and offers the potential to produce bistable systems that can switch between two alternative states (*30*). The multicomponent loop motif has yet to be identified in bacterial genetic networks (*24, 25*).

Feedforward loop motifs contain a regulator that controls a second regulator and have the additional feature that both regulators bind a common target gene (Fig. 3). The regulator location data reveal that feedforward loop architecture has been highly favored during the evolution of transcriptional regulatory networks in yeast. We found that 39 regulators are involved in 49 feedforward loops potentially controlling 240 genes in the yeast network (about 10% of genes that are bound in the genome-wide location data set).

In principle, a feedforward loop can provide several features to a regulatory circuit. The feedforward loop may act as a switch that is designed to be sensitive to sustained rather than transient inputs (*25*). Feedforward loops have the potential to provide temporal control of a process, because expression of the ultimate target gene may depend on the accumulation of adequate levels of the master and secondary regulators. Feedforward loops may provide a form of multistep ultrasensitivity (*31*), as small changes in the level or activity of the master regulator at the top of the loop might be amplified at the ultimate target gene because of the combined action of the master regulator and a second regulator that is under the control of the master regulator.

Single-input motifs contain a single regulator that binds a set of genes under a specific condition. Single-input motifs are potentially useful for coordinating a discrete unit of biological function, such as a set of genes that code for the subunits of a biosynthetic apparatus or enzymes of a metabolic pathway. For example, several genes of the leucine biosynthetic pathway are controlled by the Leu3 transcriptional regulator (Fig. 3).

Multi-input motifs consist of a set of regulators that bind together to a set of genes. We found 295 combinations of two or more regulators that could bind to a common set of promoter regions. This motif offers the potential for coordinating gene expression across a wide variety of growth conditions. For example, each of the regulators bound to a set of genes can be responsible for regulating those genes in response to a unique input. In this manner, two different regulators responding to two different inputs would allow coordinate expression of the set of genes under these two different conditions.

Regulator chain motifs consist of chains of three or more regulators in which one regulator binds the promoter for a second regulator, the second binds the promoter for a third regulator, and so forth (Fig. 3). This network motif is observed frequently in the location data for yeast regulators; we found 188 regulator chain motifs, which varied in size from 3 to 10 regulators. The chain represents the simplest circuit logic for ordering transcriptional events in a temporal sequence. The most straightforward form of this appears in the regulatory circuit of the cell cycle, where regulators functioning at one stage of the cell cycle regulate the expression of factors required for entry into the next stage of the cell cycle (*13*).

The regulatory motifs described above suggest models for gene regulatory mechanisms whose predictions can be tested with experimental data. One regulatory motif that caught our attention involved ribosomal protein genes; ribosomes are important protein biosynthetic machines, but transcriptional regulation of ribosomal protein genes is not well understood. Fhl1, a protein whose function was not previously known, forms a single-input regulatory motif consisting of essentially all ribosomal protein genes, but little else. No other regulator studied here exhibited this behavior. This predicts that loss of Fhl1 function should have a profound effect on ribosome biosynthesis if no other regulators are capable of taking its place. Indeed, a mutation in Fhl1 causes severe defects in ribosome biosynthesis (*32*), an observation that was difficult to interpret previously in the

absence of the genome-wide location data. Many ribosomal protein genes are also components of a multi-input motif involving Fhl1 and additional regulators (Fig. 3), which suggests that expression of these genes may be coordinated by multiple regulators under various growth conditions. This model and others suggested by regulatory motifs can be addressed with future experiments.

Assembling motifs into network structures. We assume that regulatory network motifs form building blocks that can be combined into larger network structures. An algorithm was developed that explores all the genome-wide location data together with the expression data from over 500 expression experiments to identify groups of genes that are both coordinately bound and coordinately expressed. In brief, the algorithm begins by defining a set of genes, *G*, that are bound by a set of regulators, *S*, with a *P* value threshold of 0.001. We find a large subset of genes in *G* that are similarly expressed over the entire set of expression data, and we use those genes to establish a core expression profile. Genes are then dropped from *G* if their expression profile is significantly different from this core profile. The remainder of the genome is scanned for genes with expression profiles that are similar to the core profile. Genes with a significant match in expression profiles are then examined to see if the set of regulators *S* are bound. At this step, the probability of a gene being bound by the set of regulators is used instead of the individual probabilities of that gene being bound by each of the individual regulators. Because we are assaying the combined probability of the set of regulators being bound and are relying on similarity of expression patterns, we can relax the *P* value for individual binding events and thus recapture information that is lost because of the use of an arbitrary *P* value threshold. The process is repeated until all combinations of genes bound by regulators have been considered. Additional details of the algorithm are available upon request. The resulting sets of regulators and genes are essentially multi-input motifs refined for common expression (MIM-CE). We expect these to be robust examples of coordinate binding and expression and therefore useful for nucleating network models.

We used the refined motifs to construct a network structure for the yeast cell cycle by an automatic process that requires no prior knowledge of the regulators that control transcription during the cell cycle. We selected the cell cycle regulatory network because of the importance of this biological process, the availability of extensive genome-wide expression data for the cell cycle (*2, 3*), and the extensive literature that can be used to explore features of a network model. Our goal was to determine whether the computational approach would construct the regulatory

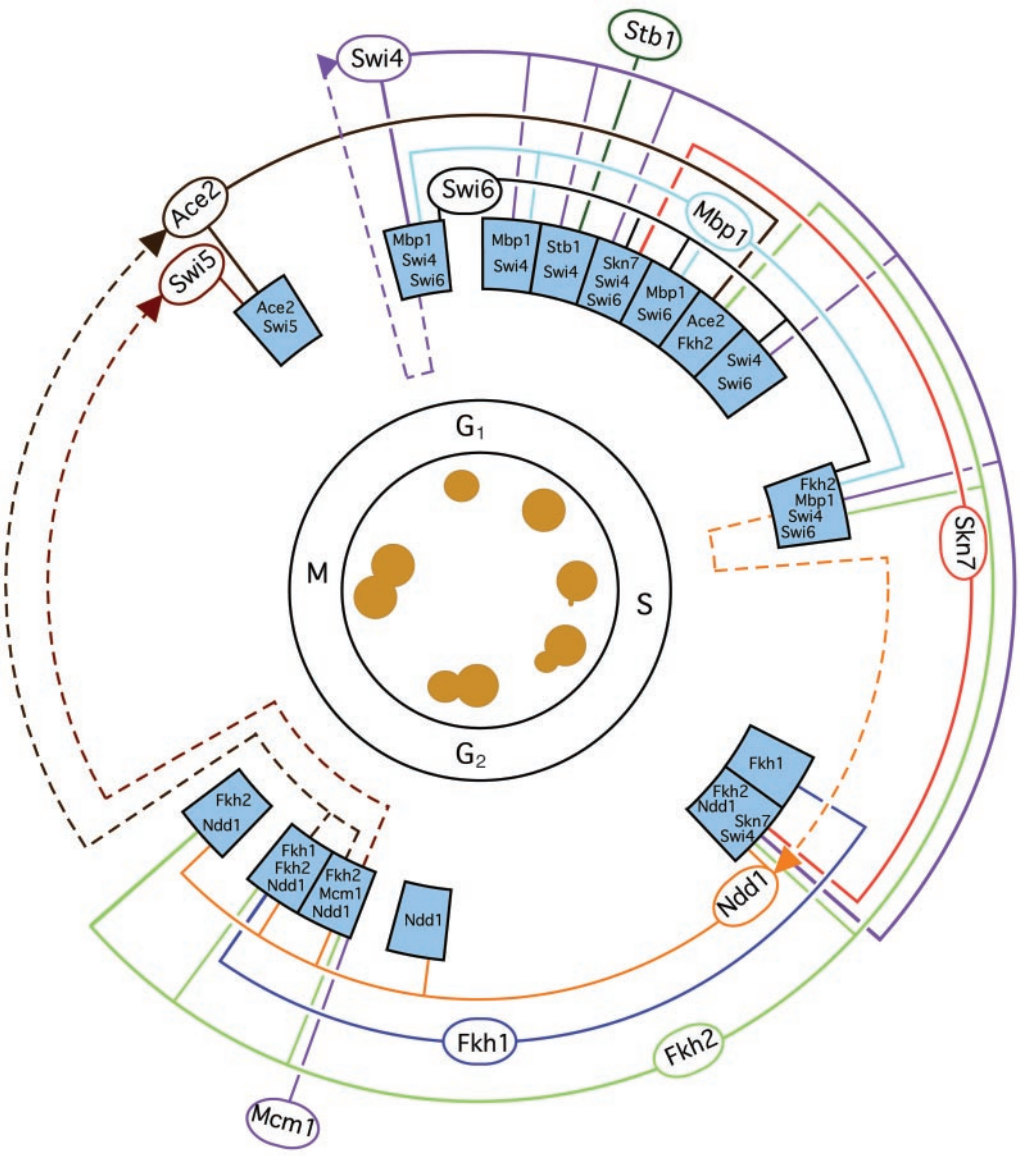
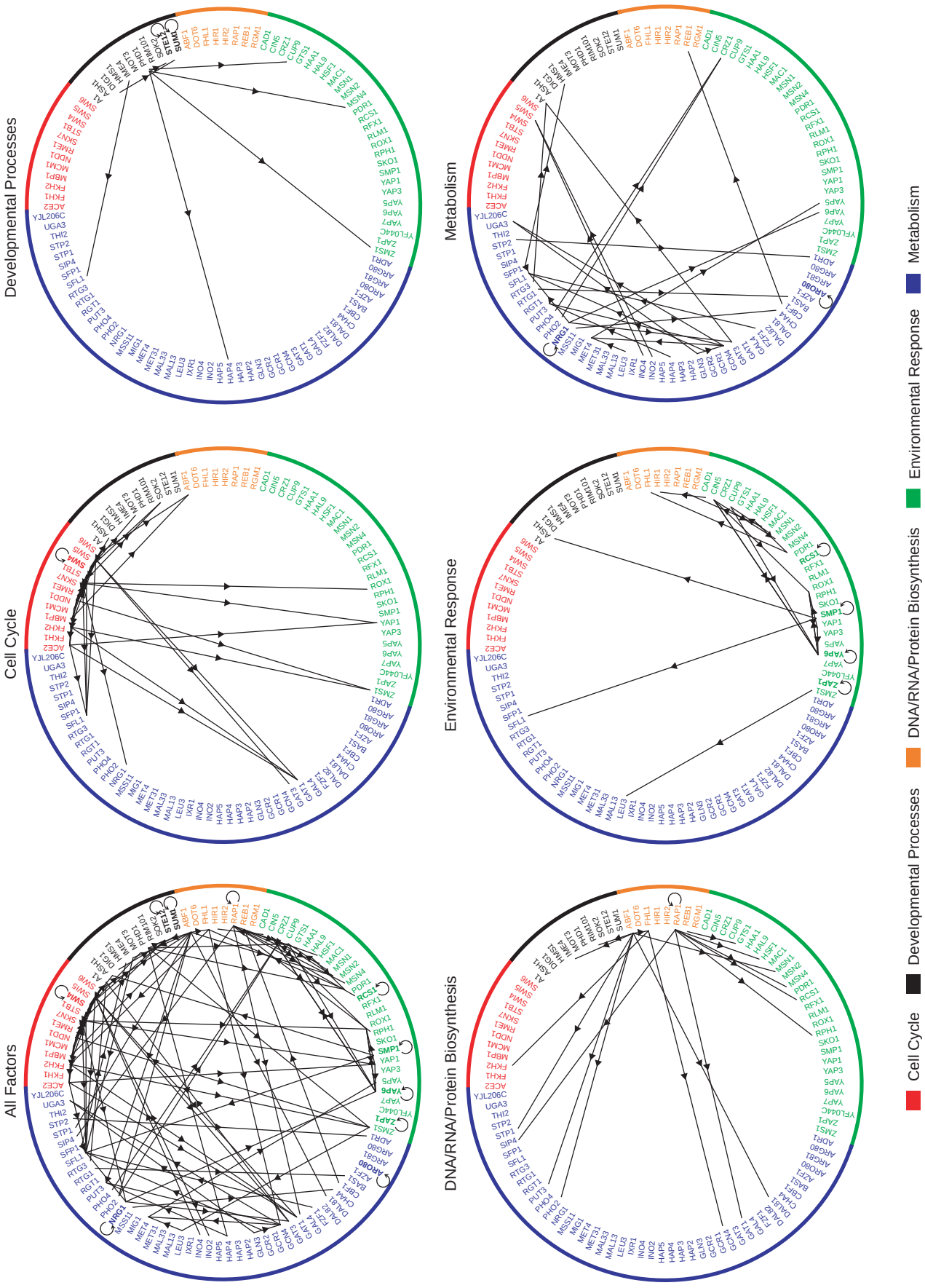


Fig. 4. Model for the yeast cell cycle transcriptional regulatory network. A transcriptional regulatory network for the yeast cell cycle was derived from a combination of binding and expression data (see text). Yeast cell morphologies are depicted during the various stages of the cell cycle. Each blue box represents a set of genes bound by a common set of regulators and coexpressed throughout the cell cycle. Text inside each blue box identifies the common set of regulators that bind to the set of genes represented by the box. Each box is positioned in the cell cycle according to the time of peak expression levels for the genes represented by the box. Regulators, represented by ovals, are connected to the sets of genes they regulate by solid lines. The arc associated with each regulator effectively defines the period of activity for the regulator. Dashed lines indicate that a gene in the box encodes a regulator found in the outer rings.

Fig. 5. (**Right**) Network of transcriptional regulators binding to genes encoding other transcriptional regulators. All 106 transcriptional regulators that were subjected to location analysis in rich medium are displayed in a circle and segregated into functional categories on the basis of the primary functions of their target genes, as indicated by the color key. Lines with arrows depict binding of a regulator (*P* value threshold = 0.001) to the gene encoding another regulator. Circles with arrows depict binding of a regulator to the promoter region of its own gene.


Developmental Processes
Metabolism
Cell Cycle
Environmental Response
All Factors
DNA/RNA/Protein Biosynthesis
Cell Cycle
Developmental Processes
DNA/RNA/Protein Biosynthesis
Environmental Response
Metabolism

logic of the cell cycle from the location and expression data without previous knowledge of the regulators involved. We reasoned that MIM-CEs that are significantly enriched in genes whose expression oscillates through the cell cycle (*3*) would identify the regulators that control these genes. We identified 11 regulators with this approach. To construct the cell cycle network, we generated a new set of MIM-CEs by using only the 11 regulators and the cell cycle expression data (*3*).

To produce a cell cycle transcriptional regulatory network model, we aligned the MIM-CEs around the cell cycle on the basis of peak expression of the genes in the group by means of an algorithm described in (*33*) (Fig. 4). Three features of the resulting network model are notable. First, the computational approach correctly assigned all the regulators to stages of the cell cycle, where they were shown to function in previous studies (*34*). Second, two regulators that have been implicated in cell cycle control but whose functions were ill-defined (*35–37*) could be assigned within the network on the basis of direct binding data. Third, and most important, reconstruction of the regulatory architecture was automatic and required no prior knowledge of the regulators that control transcription during the cell cycle. This approach should represent a general method for constructing other regulatory networks.

Coordination of cellular processes. Transcriptional regulators were often bound to genes encoding other transcriptional regulators (Fig. 5). For example, there were many instances in which transcriptional regulators within a functional category (for example, cell cycle) bound to genes encoding regulators within the same category. We have noted that cell cycle regulators bound to other cell cycle regulators (*13*), and this phenomenon was also apparent among transcriptional regulators that fall into the metabolism and environmental response categories. For example, the metabolic regulator Gcn4 bound to promoters for *PUT3* and *UGA3*, genes that encode transcriptional regulators for amino acid and other metabolic functions. The stress response activator Yap6 bound to the gene encoding the Rox1 repressor, and vice versa, which suggests positive and negative feedback loops.

We also found that multiple transcriptional regulators within each category were able to bind to genes encoding regulators that are responsible for control of other cellular processes. For example, the cell cycle activators bind to genes for transcriptional regulators that play key roles in metabolism (*GAT1*, *GAT3*, *NRG1*, and *SFL1*); environmental responses (*ROX1*, *YAP1*, and *ZMS1*); development (*ASH1*, *SOK2*, and *MOT3*); and DNA, RNA, and protein biosynthesis (*ABF1*). These observations are likely to explain, in part, how cells coordinate transcriptional regulation of the cell cycle with other cellular processes. These connections are generally consistent with previous experimental information about the relationships between cellular processes. For example, the developmental regulator Phd1 has been shown to regulate genes involved in pseudohyphal growth during certain nutrient stress conditions; we found that Phd1 also binds to genes that are key to regulation of general stress responses (*MSN4*, *CUP9*, and *ZMS1*) and metabolism (*HAP4*).

These observations have several important implications. The control of most, if not all, cellular processes is characterized by networks of transcriptional regulators that regulate other regulators. It is also evident that the effects of transcriptional regulator mutations on global gene expression, as measured by expression profiling (*1, 4, 5, 19, 38–48*), are as likely to reflect the effects of the network of regulators as they are to identify the direct targets of a single regulator.

Significance of regulatory network information. This study identified network motifs that provide specific regulatory capacities for yeast, revealing the regulatory strategies that were selected during evolution for this eukaryote. These motifs can be used as building blocks to construct large network structures through an automated approach that combines genome-wide location and expression data in the absence of prior knowledge of regulator functions. The network of transcriptional regulators that control other transcriptional regulators is highly connected, suggesting that the network substructures for cellular functions such as cell cycle and development are themselves coordinated at a transcriptional level.

It is possible to envision mapping the regulatory networks that control gene expression programs in considerable depth in yeast and in other living cells. More complete understanding of transcriptional regulatory networks in yeast will require knowledge of regulator binding sites under various growth conditions (*17, 18*) and experimental testing of models that emerge from computational analysis of regulator binding, gene expression, and other information. The approach described here can also be used to discover transcriptional regulatory networks in higher eukaryotes. Knowledge of these networks will be important for understanding human health and designing new strategies to combat disease.

References and Notes

1. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
2. R. J. Cho *et al*., *Mol. Cell* **2**, 65 (1998).
3. P. T. Spellman *et al*., *Mol. Biol. Cell* **9**, 3273 (1998).
4. A. P. Gasch *et al*., *Mol. Biol. Cell* **11**, 4241 (2000).
5. H. C. Causton *et al*., *Mol. Biol. Cell* **12**, 323 (2001).
6. T. I. Lee, R. A. Young, *Annu. Rev. Genet.* **34**, 77 (2000).
7. C. W. Garvie, C. Wolberger, *Mol. Cell* **8**, 937 (2001).
8. G. Orphanides, D. Reinberg, *Cell* **108**, 439 (2002).
9. M. Ptashne, A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002).
10. B. Ren *et al*., *Science* **290**, 2306 (2000).
11. V. R. Iyer *et al*., *Nature* **409**, 533 (2001).
12. J. D. Lieb, X. Liu, D. Botstein, P. O. Brown, *Nature Genet.* **28**, 327 (2001).
13. I. Simon *et al*., *Cell* **106**, 697 (2001).
14. J. J. Wyrick *et al*., *Science* **294**, 2357 (2001).
15. M. C. Costanzo *et al*., *Nucleic Acids Res.* **28**, 73 (2000).
16. M. Knop *et al*., *Yeast* **15**, 963 (1999).
17. See supporting data on *Science* Online.
18. Additional information is available at the authors' Web site: http://web.wi.mit.edu/young/regulator_network.
19. T. R. Hughes *et al*., *Cell* **102**, 109 (2000).
20. B. Lemon, R. Tjian, *Genes Dev.* **14**, 2551 (2000).
21. M. Merika, D. Thanos, *Curr. Opin. Genet. Dev.* **11**, 205 (2001).
22. Y. Kawasaki, K. Nosaka, Y. Kaneko, H. Nishimura, A. Iwashima, *J. Bacteriol.* **172**, 6145 (1990).
23. H. Nishimura, Y. Kawasaki, Y. Kaneko, K. Nosaka, A. Iwashima, *FEBS Lett.* **297**, 155 (1992).
24. D. Thieffry, A. M. Huerta, E. Perez-Rueda, J. Collado-Vides, *Bioessays* **20**, 433 (1998).
25. S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Nature Genet.* **31**, 64 (2002).
26. H. H. McAdams, A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).
27. A. Becskei, L. Serrano, *Nature* **405**, 590 (2000).
28. N. Guelzim, S. Bottani, P. Bourgine, F. Kepes, *Nature Genet.* **31**, 60 (2002).
29. J. W. Dolan, S. Fields, *Genes Dev.* **4**, 492 (1990).
30. J. E. Ferrell, *Curr. Opin. Cell Biol.* **14**, 140 (2002).
31. A. Goldbeter, D. E. Koshland Jr., *J. Biol. Chem.* **259**, 14441 (1984).
32. S. Hermann-Le Denmat, M. Werner, A. Sentenac, P. Thuriaux, *Mol. Cell. Biol.* **14**, 2905 (1994).
33. Z. Bar-Joseph, G. Gerber, D. K. Gifford, T. S. Jaakkola, I. Simon, Sixth Annual International Conference on Research in Computational Molecular Biology, Washington, DC, 18 to 21 April 2002 (ACM Press, New York, 2002), pp. 39–48.
34. I. Simon *et al*., *Cell* **106**, 697 (2001), and references therein.
35. B. A. Morgan, N. Bouquin, G. F. Merrill, L. H. Johnston, *EMBO J.* **14**, 5679 (1995).
36. N. Bouquin, A. L. Johnson, B. A. Morgan, L. H. Johnston, *Mol. Biol. Cell* **10**, 3389 (1999).
37. Y. Ho, M. Costanzo, L. Moore, R. Kobayashi, B. J. Andrews, *Mol. Cell. Biol.* **19**, 5267 (1999).
38. S. Chu *et al*., *Science* **282**, 699 (1998).
39. S. A. Jelinsky, L. D. Samson, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1486 (1999).
40. H. D. Madhani, T. Galitski, E. S. Lander, G. R. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12530 (1999).
41. M. C. Lopez, H. V. Baker, *J. Bacteriol.* **182**, 4970 (2000).
42. T. J. Lyons *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7957 (2000).
43. C. J. Roberts *et al*., *Science* **287**, 873 (2000).
44. A. F. Shamji, F. G. Kuruvilla, S. L. Schreiber, *Curr. Biol.* **10**, 1574 (2000).
45. K. J. Travers *et al*., *Cell* **101**, 249 (2000).
46. C. B. Epstein *et al*., *Mol. Biol. Cell* **12**, 297 (2001).
47. K. Natarajan *et al*., *Mol. Cell. Biol.* **21**, 4347 (2001).
48. F. Devaux, E. Carvajal, S. Moye-Rowley, C. Jacq, *FEBS Lett.* **515**, 25 (2002).

49. We thank J. Terragni, D. Pokholok, J. Lange, and E. Kanin for experimental support; J. Liang for assistance with the supporting Web site; and R. DeShaies, T. Ideker, and T. Jaakkola for helpful discussions. Supported by fellowships from the National Cancer Institute of Canada (F.R.), the Human Frontier Science Program (J.Z.), the Sloan/U.S. Department of Energy Program for Computational Molecular Biology (D.T.O.), the Program in Mathematics and Molecular Biology at Florida State University and Burroughs Wellcome Fund Interfaces Program (Z.B.-J.), and the Helen Hay Whitney Foundation (B.R.). Supported by grants from NIH and Corning Inc.

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5594/799/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S6

14 June 2002; accepted 16 August 2002