A network of transcriptionally coordinated functional modules in Saccharomyces cerevisiae

Allegra A. Petti and George M. Church

Supplementary data	*"Supplemental Research Data"* **http://www.genome.org/cgi/content/full/gr.3847105/DC1**
References	This article cites 52 articles, 27 of which can be accessed free at: **http://www.genome.org/cgi/content/full/15/9/1298#References**
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Notes

To subscribe to *Genome Research* go to:
http://www.genome.org/subscriptions/

Methods

A network of transcriptionally coordinated functional modules in *Saccharomyces cerevisiae*

Allegra A. Petti[1] and George M. Church[1]

Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

Recent computational and experimental work suggests that functional modules underlie much of cellular physiology and are a useful unit of cellular organization from the perspective of systems biology. Because interactions among modules can give rise to higher-level properties that are essential to cellular function, a complete knowledge of these interactions is necessary for future work in systems biology, including in silico modeling and metabolic engineering. Here we present a computational method for the systematic identification and analysis of functional modules whose activity is coordinated at the level of transcription. We applied this method, Search for Pairwise Interactions (SPIN), to obtain a global view of functional module connectivity in *Saccharomyces cerevisiae* and to provide insight into the biological mechanisms underlying this coordination. We also examined this global network at higher resolution to obtain detailed information about the interactions of particular module pairs. For instance, our results reveal possible transcriptional coordination of glycolysis and lipid metabolism by the transcription factor Gcr1p, and further suggest that glycolysis and phosphoinositide signaling may regulate each other reciprocally.

[Supplemental material is available online at www.genome.org.]

The phenotype of a unicellular organism is determined by an integrated network of genes, proteins, and metabolites that participate in reciprocal regulatory relationships. Creating a quantitative description of this network—a goal that has recently engendered the dedicated discipline of systems biology—is essential to understanding, predicting, and manipulating cellular behavior. A first step toward this goal is deciphering the connectivity of the network, that is, the pattern of interactions among its components. Given the complexity of this undertaking, the integrated network is often treated as a group of superimposed subnetworks, including the gene regulatory, protein, and metabolic networks. A corollary task is to determine whether the network's topology reflects its organizational principles. Using biological networks with relatively well-characterized connectivity, quantitative analyses of network topology have revealed that these networks are modular—they can be clustered into nodes that are more densely connected to each other than to nodes in other clusters (Ravasz et al. 2002; Barabasi and Oltvai 2004; Yook et al. 2004).

In both natural and man-made systems, modularity is a fundamental design principle whereby components are partitioned according to common physical, regulatory, or functional properties. In biology, the exact meaning of modularity depends on the network under consideration. For example, modules in the protein network often have a straightforward physical interpretation as a static molecular complex (such as the ribosome) or a dynamic signaling pathway (such as a MAP kinase cascade) (Schwikowski et al. 2000; Spirin and Mirny 2003). Gene regulatory networks, in contrast, tend to display regulatory modules, in which every gene is controlled by the same transcription factors (TFs) under the same environmental conditions (Tavazoie et al. 1999; Segal et al. 2003). Both networks exhibit functional modu-

larity. Defined as groups of genes, proteins, and other molecules involved in a common subcellular process, functional modules transcend the heuristic subdivision of the integrated network into gene regulatory, protein, and metabolic networks. It has been proposed that the functional module is the most relevant organizational unit of a cell from the perspective of systems biology (Hartwell et al. 1999), and a growing body of work supports the idea that such modules underlie much of cellular physiology.

For natural and man-made systems, the relative independence of modules has significant implications for system engineering (by evolution or by man): A module can be selectively altered without perturbing the behavior of the rest of the system. However, higher-level properties that emerge from intermodule coordination are often critical to the behavior of the system as a whole (Hartwell et al. 1999). This need for module integration is balanced with the benefits of module independence by ensuring that only a few components in each module interface with other modules. With respect to (biological) functional modularity, modules are partially isolated from each other by the biochemical specificity of protein–protein and protein–DNA interactions. They contain a large number of internal components that do not interact with other modules, and a small number of input and/or output components that do (Alon 2003).

Most investigations of functional modularity have sought to define the modules' components (Rives and Galitski 2003; Spirin and Mirny 2003; Han et al. 2004; Ma et al. 2004; Pereira-Leal et al. 2004), and several databases use published literature to categorize every gene in *Saccharomyces cerevisiae* according to its cellular function (Ashburner et al. 2000; Mewes et al. 2002). However, several studies have found examples of physical and regulatory interactions between modules (Ihmels et al. 2002, 2004; Snel et al. 2002; Danial et al. 2003; Rives and Galitski 2003; Segal et al. 2003; Han et al. 2004; Segre et al. 2005). For instance, Gal4p, a well-characterized transcription factor traditionally associated with galactose metabolism, was also found to regulate genes associated with uracil metabolism, such as the uracil transporter gene *FUR4*, thereby facilitating the addition of uridine 5'-diphosphate to galactose (Ren et al. 2000).

[1]**Corresponding authors.**
E-mail allegra.conbrio@post.harvard.edu; fax (617) 432-7266.
E-mail http://arep.med.harvard.edu/gmc/email.html; fax (617) 432-7266.

A more global understanding of functional module interactions is essential to several endeavors of systems biology, such as predicting cellular-level phenotypes arising from, for example, genetic manipulation of microorganisms, and creating accurate mathematical models of cellular function (Kholodenko et al. 2002). This understanding will also help answer broader questions about how intermodule regulatory connections are distributed across the cell. One possibility is that most cellular functions are coordinated by a central organizing module; another, that the connections are uniformly distributed among modules.

These considerations motivated us to automate the identification of potentially coregulated functional module pairs in *S. cerevisiae*. Such coregulation can be achieved through a variety of mechanisms, including enzymatic reactions, shared metabolites, and coordinated transcription. We focus on transcription, which, by promoting the coexistence of biochemical components necessary for coordination by other mechanisms, may represent the most fundamental level of coregulation. Drawing on experimental evidence that a transcription factor can coordinate the activity of two pathways by regulating genes in both pathways (Ren et al. 2000; Iyer et al. 2001; Lieb et al. 2001), we developed an algorithm, SPIN (Search for Pairwise INteractions), that examines genes from a pair of functional modules, and uses expression data to assess the possibility that a given TF(s) transcriptionally coregulates genes in both modules. To obtain a comprehensive view of module coordination at a desired level of functional resolution, we applied SPIN to all pairwise combinations of 72 functional modules (Supplemental Fig. S1), which we defined according to the second-tier functional categories of the MIPS database (Mewes et al. 2002). For each module pair, the extent of transcriptional coregulation was determined with respect to four microarray expression data sets and each of 175 TFs. The results suggest extensive coregulation among functional modules, and afford a high-level view of interprocess coordination consistent with biological knowledge. SPIN also identifies previously unknown functional interactions and the TFs that mediate them, thereby providing novel insight into the overall functional and regulatory organization of *S. cerevisiae.*

Results

Single-TF analysis

SPIN accepts as input four data sets, including one set of transcription factor binding site (TFBS) data (for several TFs), one set of gene expression data, and two gene sets, each representing a predefined functional module. For each TF, SPIN determines whether the TF binds to genes in both modules. If so, it uses gene expression data to determine whether the TF confers upon these target genes an expression profile that is coherent and distinct from the expression profiles that characterize the functional categories (Fig. 1A). The output of SPIN is a list of statistically significant "triplets," that is, sets of two functional modules and one TF that appears to coordinate the modules' activity by coregulating genes in each. We consider only triplets in which each module contains at least four target genes.

To generate the results reported below, we used several combinations of TFBS and gene expression data. TFBS information for a total of 175 TFs was taken from two pre-existing data sets: One set, referred to as MSA (Multiple Sequence Alignment), was obtained using the sequence-alignment algorithm AlignACE (Roth et al. 1998) and the pattern-matching algorithm ScanACE to identify TF-binding motifs and their genomic locations, respec-
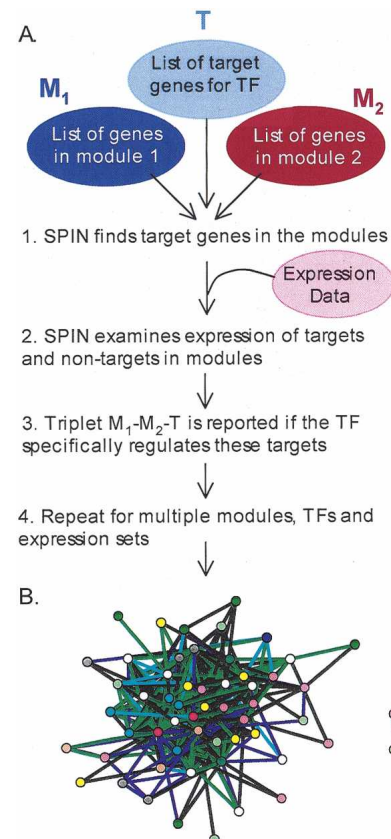
Figure 1. (*A*) General approach. (*B*) Global module interaction network (MSA TFs). Each node represents a functional module in the second-highest level of the MIPS hierarchy, and its color corresponds to the highest-level MIPS category of which the node is a member (see Fig. 3 legend). Each edge represents one or more MSA TFs that coordinate the transcription of genes in both modules in a statistically significant manner. Edge color represents the expression data set from which the interaction was inferred: (purple) cell cycle, (gray) diauxic shift, (turquoise) environmental stress, (green) MAPK. Supplemental Figure S4 is a more fully annotated version of this figure. This graph and others like it were drawn using Pajek (Batagelj and Mrvar 1998).

tively, as described in Pilpel et al. (2001) and Hughes et al. (2000); the second data set, referred to as Chip2, was obtained using ChIP-chip technology, and therefore constitutes direct physical evidence of in vivo TF binding (Lee et al. 2002). Because each DNA-binding motif in the MSA set is assumed to represent a binding site for a cognate TF, "TF" or "TFBS" will henceforth also refer to DNA motifs. Four microarray data sets, which measure mRNA expression profiles during the cell cycle, diauxic shift, response to environmental stress, and perturbations of the MAPK signaling pathways, were used. Functional modules were defined according to the second-tier functional categories in the MIPS database (in which published, largely experimental, evidence is manually curated in order to assign gene function). (Supplemental Fig. S1). However, SPIN can be used with any scheme of functional categorization. This is essential, because module definitions will inevitably coevolve with the annotation of the yeast genome.

To obtain a global view of module coordination at this level of functional resolution, we applied SPIN to all pairs of MIPS functional modules, using each TFBS data set and each expression data set. Table 1 summarizes these data for each data set

Table 1. Number of significant triplets identified by SPIN, using stringent criteria

	TFBS data set	
	MSA	Chip2
Total triplets passing size threshold	13,378	1154
Cell cycle	111	130
Diauxic shift	281	0
Environmental stress	29	0
MAPK	181	69
Unique category pairs across all expression sets	316	102

combination (complete results in Supplemental Figs. S2, S3). Results from the MSA TFs reveal extensive coordination among modules: 34 TFs were found to mediate 316 pairwise interactions involving 55 functional modules. Many module pairs are coordinated by more than one TF and/or in more than one set of expression data. Those coregulated by at least one TF under at least one set of expression data are shown in Figure 1B (annotated version in Supplemental Fig. S4). Results obtained using the Chip2 data set include 102 pairwise interactions among 36 categories, mediated by 25 TFs. A substantial portion (48%) of these interactions was also found using the MSA results (Supplemental Fig. S5). Overall, these results comprise a rich data set with which intermodule coregulation can be studied at different levels of detail, as we discuss below.

At low resolution, the results afford a global view of module connectivity that reveals the extent to which control of cellular activity is centralized, and the conditions under which each module is active. For both TFBS sets, the distribution of module connectivity, defined as the number of other modules with which a given module is coregulated, is not uniform (Fig. 2A). Those related to storage and transmission of genetic information, such as the *mRNA*, *Cell cycle*, *tRNA*, *Nucleotides*, *Differentiation*, and *DNA* modules, are among the most highly connected (see Table 2 for module abbreviations). Although there is no evidence that a single module acts as a cellular "central processing unit," these modules appear to have a particularly strong influence on the organization of cellular behavior. This is consistent with data on physical interaction modules in the protein–protein interaction network (Schwikowski et al. 2000; Tucker et al. 2001; Rives and Galitski 2003; Han et al. 2004). The functional module connectivity distributions show general (but imperfect) correlation between module size and number of partner modules (Pearson correlation coefficients = 0.87 [Chip2] and 0.64 [MSA]). A general correlation is expected, since central cellular processes may (1) involve many molecules, and (2) interface with numerous other processes. Connectivity distributions broken down by expression data set (Supplemental Fig. S6A–C) reveal that fewer interactions were deemed significant using the environmental stress expression data. The conditions in that data set elicit a broad transcriptional response known as the environmental stress response (ESR) (Gasch et al. 2000), but may affect few functional modules in a manner that is distinct from the ESR.

From a TF-centric perspective, the global view reveals a non-uniform distribution of TF connectivity (the number of pairwise interactions that a given TF mediates) analogous to the distribution of module connectivity (Fig. 2B). This supports the notion that some TFs (such as mRRPE, PAC, and SFF) are relatively general and are therefore involved in many processes, whereas others are more specific (Cliften et al. 2003). The module pairs coregu-

lated by each TF, and the conditions under which the coregulation was inferred, are largely consistent with known TF activities. For example, MCB, whose target genes are active during the G_1 phase of the cell cycle, was found to coregulate many modules related to cell cycle progression and transcription, all of which were identified using cell cycle expression data. Moreover, TFs that are known to act synergistically are often found to mediate the same pairwise interaction. For instance, many interactions mediated by mRRPE3 and PAC, which are thought to cooperatively regulate the expression of rRNA transcription and processing genes (Sudarsanam et al. 2002), involve one or more of the modules *rRNA*, *tRNA*, and *Amino-acyl tRNA synthetases*, and were inferred using diauxic shift expression data (Supplemental Fig. S6C). This may reflect the decrease in ribosome biogenesis that accompanies the diauxic shift (DeRisi et al. 1997).

Although the complete network is too large and highly connected to be informative about the details of transcriptional coordination, subnetworks of interest can be extracted and examined in increasing levels of detail. For instance, all modules that are coregulated with *Cell cycle* by MSA TFs are shown in Supplemental Figure S7; specific module pairs are discussed below.

Intersection of MSA and Chip2 results

To identify a small subset of the most biologically significant results, we selected statistically significant pairwise interactions that were identified using both binding site data sets (regardless of the identity of the motifs mediating the interaction). The resulting filtered data set contains 29 pairwise interactions among 22 modules, mediated by 39 motifs (Fig. 3; Supplemental Fig. S8). Many of these interactions and the TFs that mediate them are biologically well-grounded, but several less predictable relationships were also discovered. With respect to the former, there is a clique consisting of the *Cell cycle*, *DNA*, *mRNA*, and *Differentiation* modules. Every interaction within the clique is mediated by MBP1 (Chip2) and MCB (MSA), both of which correspond to the TF Mbfp. (MBP1 and SWI6 interact to form the Mbfp complex, which binds to the MCB DNA motif and regulates passage through START during the cell cycle [Horak et al. 2002].)

Other foreseeable relationships include the interactions of *Assembly of protein complexes* with *Ionic homeostasis* and *Mitochondrial transport*. The TFs HAP4 (Chip2) and HAP2/3/4 (MSA) (which represents a complex composed of HAP2, HAP3, and HAP4) mediate all of these interactions. They, along with HAP5, form a transcription factor complex that regulates genes encoding components of the TCA cycle and the electron transport chain (which are included in all four of these categories) under nonfermentative conditions. ABF1 (MSA) and HAP2/3/4 (MSA), which were found to coordinate *Assembly of protein complexes* with *Ionic homeostasis*, are another example of synergistic TFs that mediate the same interaction.

One of the more novel relationships is the coordination of *Cell cycle* with *Carbon*, which is mediated by several TFs including SWI5, CIN5, NDD1, NRG1 (all Chip2), and mMERE11 (MSA) (discussed further in Supplemental materials item S9).

TF pairs

A growing body of evidence shows that the activity and specificity of many TFs are increased by their acting in conjunction with other TFs (Pilpel et al. 2001; Sudarsanam et al. 2002). The single-TF approach to functional module coordination was therefore extended to identify pairs of TFs that act synergistically to coordinate the activity of functional modules. Analogous to the
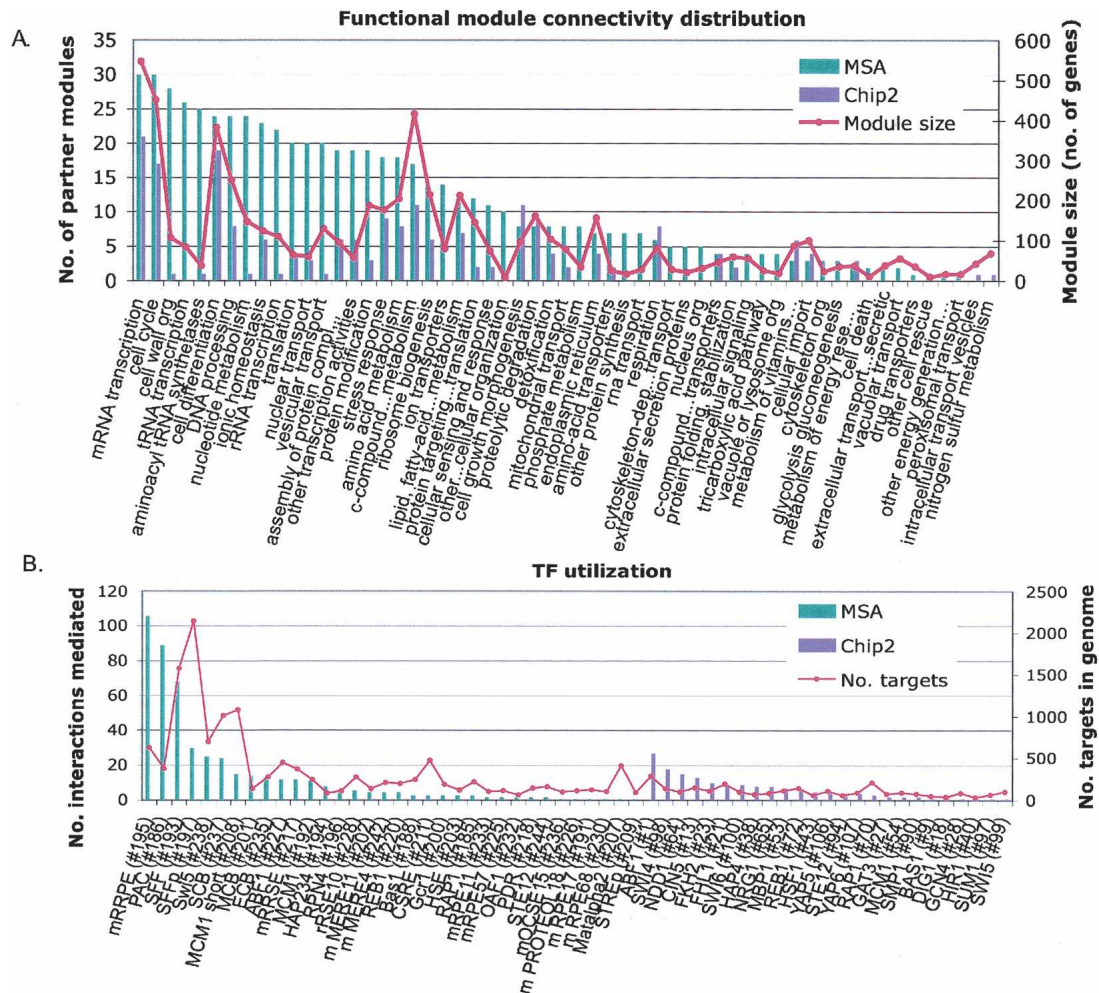
Figure 2. (*A*) Functional module connectivity histogram. For each module, the number of partner modules with which it is coordinated by at least one TF in at least one expression data set is shown. (Purple columns) MSA TFs, (green columns) Chip2 TFs, (pink line) module size. (*B*) TF utilization histogram. The number of pairwise module interactions mediated by each TF is shown on the *left*-hand axis. The *right*-hand axis shows the number of genome-wide promoter targets for each TF. (Purple column) MSA TFs, (pink column) Chip2 TFs, (pink line) promoter targets per TF.

single-TF analysis, coordination between all module pairs was analyzed by SPIN for all TF pairs in the MSA data set and all expression data sets. Using stringent filtering criteria, 26 TF pairs, comprised of 13 individual TFs, mediate 32 pairwise interactions among 23 different modules (Supplemental Fig. S10). This global TF-pair interaction network contains a few additional module interactions that were not identified by the single-TF analysis. The most highly connected modules (Supplemental Fig. S11) differ slightly from those identified using single TFs, and include *Carbon*, *mRNA*, *Cell cycle*, *Cell Sensing and Response*, and *Differentiation*. Of the 26 different TF pairs found to mediate the module interactions, 11 of them have been previously documented as synergistic (Pilpel et al. 2001). Most of the TF combinations (Supplemental Fig. S12) involve one of the TFs SFF, SFFp (an SFF variant), or MCM1-short (an MCM1 variant). Our requirement that each module contain a number of genes that are targets of both TFs (see Methods) may have biased our results toward these TFs, which have large numbers of targets. SFF pairs with the greatest number of additional motifs (Supplemental Fig. S13), which may explain why it was found to coordinate many module pairs in the single-TF analysis.

Analysis of module coregulation with respect to individual transcription factors

In addition to the global study, we investigated the regulatory relationships between several pairs of functional modules in detail. It is this level of analysis that yields the greatest number of practical biological insights: SPIN efficiently reproduces results from numerous previous studies, and provides new information that allows us to synthesize known and novel results into revised, more complete, biological hypotheses. Here, we discuss the coordination of "Carbon compound and carbohydrate metabolism" (*Carbon*) with "Lipid, fatty-acid, and isoprenoid metabolism" (*Lipids*).

The metabolic relationship between glucose and lipid metabolism—that each process generates intermediates of the other—is well-established (Berg et al. 2002; Kanehisa et al. 2004). However, recently identified physical and signaling interactions between glycolysis and lipid metabolism imply that these processes are further coordinated by more elaborate, multilevel regulatory mechanisms that are not fully understood. In order to study this coordination at the level of transcription, our algo-

Table 2. Commonly used abbreviations

Category name	Abbreviation
Carbon compound and carbohydrate metabolism	Carbon
Cell differentiation	Differentiation
DNA processing	DNA
Glycolysis and gluconeogenesis	Glycolysis
Lipid, fatty acid, and isoprenoid metabolism	Lipids
Metabolism of energy reserves	Reserves
mRNA transcription	mRNA
Nucleotide metabolism	Nucleotides
rRNA transcription	rRNA
Stress response	Stress
tRNA transcription	tRNA

rithm was applied to the functional module pairs *Carbon* and *Lipids*, and *Glycolysis* and *Lipids*.

Using stringent and lenient filtering criteria (see Methods), we found that the activity of *Carbon* and *Lipids* is coordinated by a variety of TFs using a variety of expression data sets. It appears that different TFs coordinate different subsets of genes in each process in a condition-specific manner. Using MAPK expression data, for instance, SWI4 (Chip2) was found to coordinate genes in each process that are associated with cell wall biosynthesis. Using cell cycle data, however, GCR1 (MSA) was found to coordinate the *Carbon–Lipids* interaction. We focus on this result because, compared to the other TFs that mediate this interaction, GCR1 is more specific to the glycolytic genes.

GCR1 is the primary transcriptional activator of the glycolytic enzymes (Lopez and Baker 2000), but has not been previously implicated in the metabolism of lipids, fatty acids, or isoprenoids. Here, however, GCR1 was associated with the expression of six genes associated with *Lipids* (Fig. 4). These genes are involved in biosynthesis of inositol and phosphoinositides, such as the signaling molecule $PI(4,5)P_2$ (*MSS4* and *INM1*), fatty acid biosynthesis (*ELO1*), ergosterol biosynthesis (*FPS1*), glycerol import (*ERG20*), and inositol-dependent transcriptional regulation of other phospholipid biosynthetic genes (*UME6*) (Elkhaimi et al. 2000; Christie et al. 2004).

These results draw together several recent studies, each of which reports on a different aspect of the relationship between glycolysis and lipid metabolism. Using a yeast proteome chip,

some glycolytic enzymes and glucose transporters were observed to bind phosphoinositides (Zhu et al. 2001), consistent with independent reports that most glycolytic enzymes are localized to the cell wall (in addition to the cytoplasm) (Hubbard et al. 1994; Lesage et al. 1994; Delgado et al. 2001; Honigberg and Purnapatre 2003; Willis et al. 2003; Young et al. 2003). Other work has shown that glucose stimulates cleavage and subsequent signal transduction by the phosphoinositide $PI(4,5)P_2$, as well as transcription of *INM1* [which encodes a $PI(4,5)P_2$ biosynthetic enzyme] (Murray and Greenberg 2000).

Taken together, these data suggest that phosphoinositides may couple glycolysis with $PI(4,5)P_2$ synthesis and glucose-induced $PI(4,5)P_2$ signaling. This raises the possibility that phosphoinositides or phosphoinositide-mediated signal transduction regulate the activity of glycolytic enzymes in a feedback loop. Our evidence for transcriptional coordination of glucose and lipid metabolism suggests that joint transcription by GCR1 enables components of both processes to coexist.

Discussion

Recent computational and experimental work has yielded increasing evidence that the cell is organized into functional modules—groups of genes, proteins, and other molecules that serve a particular cellular function. Although each module acts in relative isolation from the rest of the cell, higher-level properties that emerge from the coordination of and interactions among functional modules may be essential to cellular survival. Future work in systems biology may necessitate a more complete knowledge of these interactions than we currently possess. SPIN facilitates the identification of functional modules whose activity is transcriptionally coordinated, and yields qualitative information about the relationships between those modules. We analyzed several module interactions in detail, and discuss that between *Carbon* and *Lipids*.

Published experimental results point to various interactions between these functional modules, but their relationship has not been studied systematically. Our results indicate the TFs likely responsible for the coordination, the environmental conditions under which the TFs effect the coordination, and the genes that act at the interface of the modules, all of which suggest a mechanism and/or rationale for the interaction. For example, the analy-
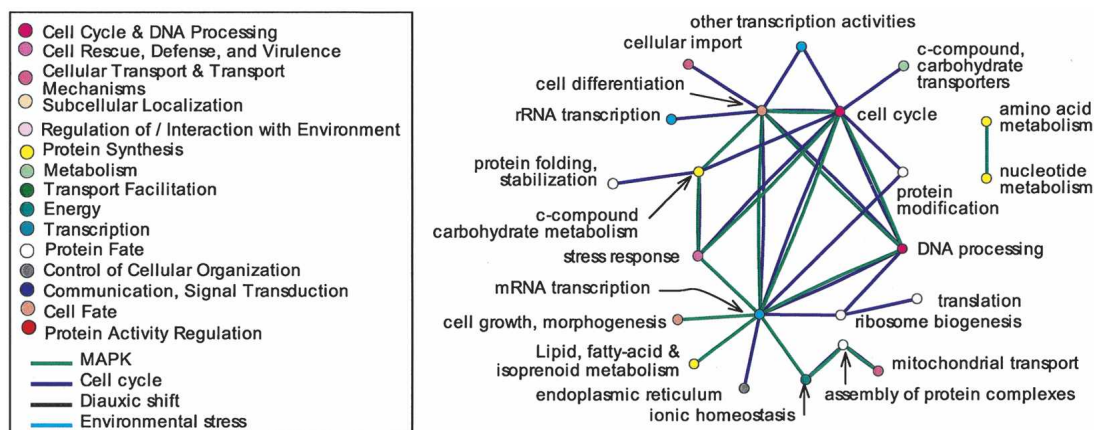
Figure 3. Duplicated interactions. Each pairwise interaction was identified using at least one MSA TF and at least one Chip2 TF. Each node represents a functional module in the second-highest level of the MIPS hierarchy, and the node color corresponds to the highest-level MIPS category of which the node is a member. In Supplemental Figure S8, each edge is labeled with the corresponding TFs.
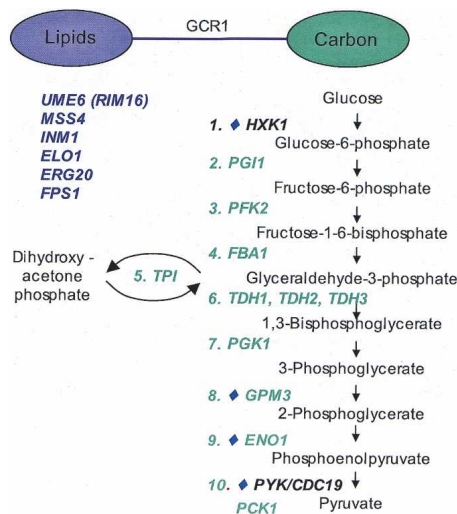
Figure 4. Interaction between *Lipid, fatty-acid, and isoprenoid metabolism* and *Carbon compound and carbohydrate metabolism*. The two modules (colored ovals) are coregulated by the MSA TF GCR1 (purple edge) using the cell cycle expression data set. GCR1 targets in each module are shown *below* the modules in purple or green type; those associated with glycolysis are organized according to the superimposed chart of glycolysis. Gene names preceded by a blue diamond were shown to bind phosphoinositides using proteome chips in Zhu et al. (2001).

sis of the interaction between carbon and lipid metabolism corroborates and integrates previous evidence that glycolysis can occur at the cell wall, and that glycolysis may regulate (or be regulated by) phosphoinositide signaling. Further experimental directions may be to determine whether phospholipids activate or inhibit glycolytic enzymes; if $PI(4,5)P_2$ signaling influences glycolysis; and whether glycolysis (or glucose alone) is necessary for the observed effect of glucose on $PI(4,5)P_2$ signaling.

In addition to this analysis, we performed a global analysis of module coordination in order to understand, at a particular functional resolution, how regulatory control is distributed across the cell. SPIN was applied to all pairwise combinations of 72 functional modules, and revealed extensive coordination among 55 of these modules, many of which are coordinated with more than one partner module, by more than one TF, and/or in more than one experimental condition. The distribution of module connectivity is not uniform. Although a single master regulator module was not identified, processes related to storage and transmission of genetic information, such as *mRNA*, *Cell cycle*, *tRNA*, *Nucleotides*, *Differentiation*, and *DNA* are most highly connected. (However, the observed distribution of connectivity may be different if different sets of expression data are used.) This suggests a control hierarchy in which these basic processes are central to the orchestration of cellular behavior. In related work, however, it has been shown that many functional modules that are coexpressed in *S. cerevisiae* are not coexpressed in other organisms (Ihmels et al. 2004). Extending the genome-wide analysis of module coordination to additional organisms might shed light on the evolution of interactions within complex cellular networks.

Methods

Overview

SPIN was motivated by experimental evidence that functional modules can be coordinately regulated via a TF that induces the

coexpression of a subset of genes associated with each module. Given two gene sets (each corresponding to a functional module), one TFBS data set, and one microarray expression data set, the algorithm loops through each TF to identify genes in each module that contain a corresponding TFBS. These "target genes" are scored using expression data to assess the evidence that the TF influences their expression. Two modules are said to be coordinately regulated by a particular TF if its target genes in each module (at least four) are coexpressed more coherently than, and in a pattern that is distinct from, non-target genes in those modules. The input data sets, scoring metrics, and methods of assessing statistical significance are discussed in detail below.

Functional module definitions

Each functional module is defined as the set of genes assigned to a particular functional category in the MIPS (Munich Information Center for Protein Sequences) *S. cerevisiae* database (Mewes et al. 2002; Supplemental Fig. S1). Using published information, MIPS assigns genes to functional categories in a hierarchical fashion, and each gene can be assigned to multiple functional categories. SPIN is designed to analyze pairs of functional categories that are from the same level of the hierarchy. Although any level can be used, we present results obtained using categories in the second-highest level of the hierarchy that correspond to defined physiological processes (e.g., cell cycle, ribosome biogenesis, the TCA cycle, and intracellular signal transduction). Category names are italicized, and Table 2 lists abbreviations for some commonly used ones.

Genome sequence data

Upstream region sequence data were downloaded from the AlignACE Web site (http://atlas.med.harvard.edu/download/index.html) (Roth et al. 1998; Hughes et al. 2000) and the *Saccharomyces* Genome Database (SGD) (Christie et al. 2004). These data contain 5018 upstream sequences, representing 6186 non-mitochondrial genes.

Transcription factor binding data

We used two different transcription factor (TF) binding data sets, each being a matrix in which the value of element (i, j) is 1 if the promoter of gene i contains a binding site for transcription factor j and 0 if not. The "MSA" set (Multiple Sequence Alignment) (Supplemental Figs. S14, S15) contains data for 62 TF DNA-binding motifs, represented as position-specific weight matrices (PSWMs), that were generated using the literature and the multiple sequence alignment program AlignACE (Roth et al. 1998) as described in Hughes et al. (2000), Pilpel et al. (2001), and Roth et al. (1998). The location of each motif in each promoter was determined using the pattern-matching program ScanACE (Hughes et al. 2000), which identifies and scores close matches to each consensus motif. A motif was considered present in a promoter if its match in the promoter scored equal to or better than one standard deviation below the mean score of the sequences constituting the PSWM (results obtained using a different score threshold are similar [data not shown]). For comparison, TF-binding data generated by an independent method were also used; the "Chip2" data set (Supplemental Figs. S16, S17) contains location data for 113 TFs, whose binding sites in 6270 promoters were determined using combined chromatin immunoprecipitation microarray hybridization (ChIP-chip) technology (Lee et al. 2002). Every pair of modules was analyzed twice—once with each TFBS data set. Results from the two sets are compared and considered together to create a composite view of intermodule coordination.

Expression data sets

For each TFBS data set, SPIN was run on four microarray expression data sets: a cell cycle time series across 15 time points (data points at 90 and 100 min were omitted because of experimental error) (Cho et al. 1998); a diauxic shift time series across seven time points (DeRisi et al. 1997); 42 environmental conditions pertaining to pH, oxidative, saline, and osmotic stress (Causton et al. 2001); and 56 conditions probing the MAPK signal transduction pathways (Roberts et al. 2000).

Scoring metric terminology and rationale

Given a pair of functional modules, a TFBS data set, and an expression data set, SPIN loops through each TF to identify genes in each functional category that contain a cognate binding site. These genes are referred to as "target genes," whereas genes that do not contain the TFBS are referred to as "non-target genes." Any combination of two functional categories and one TFBS is referred to as a "triplet." "Aggregate target (non-target) genes" refers to the union of target (non-target) genes from two functional categories, and "disjoint target (non-target) genes" refers to non-overlapping sets of target (non-target) genes from two functional categories.

Because the presence of a TFBS in a promoter does not guarantee that the TF regulates the downstream gene(s), triplets in which both modules contain target genes for a particular TF are scored to assess the potential biological significance of that TF's binding. We assume that if a TF influences the expression of its target genes, their expression profile will differ from that of non-target genes in the same functional module. This is consistent with biological reasoning that target genes, unlike non-target genes, may interact with other modules. We use two expression-based metrics to accommodate different possible types of coordination within a pair of modules. On one hand, there may be conditions in which the aggregate target genes (input/output nodes) are tightly coexpressed between the modules, but in which the aggregate non-target genes (internal nodes), participating in module-specific processes, are not significantly coexpressed between the modules (Fig. 5A); the "Expression Profile Convergence" score (C) determines whether the aggregate target genes are more highly coexpressed than the aggregate non-target genes. On the other hand, the aggregate non-target genes might also be coexpressed, but in a pattern that is different from that of the aggregate target genes (Fig. 5B); the "Expression Profile Distinctness" score (D) measures the extent to which the expression profile of the target genes differs from the expression profile of the non-target genes. (We have not observed any modules in which all genes are regulated by the same TF.)

Both scoring metrics are based on two simpler metrics that measure the coherence of a set of expression profiles: the "Expression Coherence Within" (E_w) and the "Expression Coherence Between" (E_b) scores. Given a group of genes N, E_w is the fraction of all pairwise expression profile correlation coefficients that exceed a data set-specific correlation threshold, and is designated by $E_w(N)$. It measures how tightly a group of genes is coexpressed. Given two groups of genes N and M, $E_b(N, M)$ is the fraction of $|N| * |M|$ pairwise expression profile correlation coefficients that exceed a data set-specific correlation threshold, where correlation coefficients are computed between genes from different groups only. It measures how tightly two groups of genes are coexpressed. For a particular expression data set, the correlation threshold is the value of the correlation coefficient at the 95th percentile (Pilpel et al. 2001; Sudarsanam et al. 2002; Zhu et al. 2002) of all pairwise correlation coefficients for that data set.

The Expression Profile Convergence score, C, determines
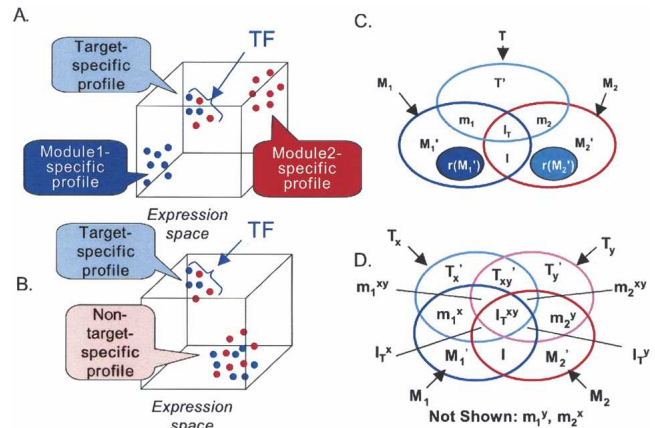
Figure 5. Coherence of target gene expression profiles. (*A*) Distinctness of target gene expression profiles. The cube represents a three-dimensional expression space. Each dot represents a gene from functional module 1 (blue) or functional module 2 (red), and the location of the gene on an axis represents the expression level of the gene at the time point or condition represented by the axis. The distance that separates two genes is inversely proportional to the "coherence" of their expression profiles. Target genes are shown as regulated by the TF (arrow). This distribution of genes represents a hypothetical scenario in which the intermodule coherence between target genes is higher than that between non-target genes. (*B*) Notation as in *A*. This distribution of genes represents a hypothetical scenario in which target genes and non-target genes have very different expression profiles, even though the intermodule coherence between target genes is similar to that of non-target genes. (*C*) Triplet scoring notation. M_1 = genes in module 1; T = genes in genome containing binding site for TF T; m_1 = genes in M_1 containing binding site for T ("target genes"); M_1' = genes in M_1 that do not contain a binding site for T ("non-target genes"); $r_{|x|}(M_1')$ = x randomly selected genes from M_1'; M_2, m_2, M_2', $r_{|x|}(M_2')$ defined analogously. Aggregate target genes = $(m_1 \cup m_2)$ or $(m_1 \cup m_2 \cup I_T)$; aggregate non-target genes = $(M_1' \cup M_2')$ or $(M_1' \cup I \cup M_2')$; I and I_T represent non-target and target genes, respectively, that are present in both modules. (*D*) Quartet scoring notation. Variables are analogous to those in *C*, but here there are two TFs, X and Y. T_x = genes in genome containing binding site for X; T_y = genes in genome containing binding site for Y. B, X, Y, and N represent module components that are targets of both X and Y, X only, Y only, or neither TF, respectively: $B = (m_1^{xy} \cup I_T^{xy} \cup m_2^{xy})$, $X = (m_1^{x} \cup I_T^{x} \cup m_2^{x})$, $Y = (m_1^{y} \cup I_T^{y} \cup m_2^{y})$, $N = (M_1' \cup I_T \cup M_2')$.

whether the coherence between the disjoint target genes (gene sets m_1 and m_2) is greater than that between the disjoint non-target genes (gene sets M_1 and M_2) (Fig. 5C; Supplemental Fig. S18):

$$C = E_b (m_1, m_2) \qquad (1)$$

The significance of the C score is ensured by controlling the False Discovery Rate (FDR); q-values were computed from p-values using the Q-value software (Storey and Tibshirani 2003). To obtain p-values for the C score, the E_b score of the disjoint target genes is compared to the E_b score of two randomly selected, disjoint gene sets (of sizes $|m_1|$ and $|m_2|$) from the same functional categories (Fig. 5C):

$$C^{rand} = E_b (r_{|m_1|} (M_1' \cup m_1), r_{|m_2|} (M_2' \cup m_2)) \qquad (2)$$

For a TFBS data set containing data for Q TFs, random partitioning is repeated $(3 \times Q/0.05) - 1$ times (Edgington 1995) to ensure that a p-value of 0.05, corrected for multiple hypotheses by the Bonferroni correction, can be obtained (although significant results are ultimately chosen by controlling the FDR, as men-

tioned above); the factor 3 is used to increase the power of the test.

The Expression Profile Distinctness score, D, compares the coherence within the target genes (the numerator) to that between the target genes and non-target genes (the denominator) (Fig. 5B,C; Supplemental Fig. S19). Note that it incorporates genes that are assigned to both functional categories (sets I and I_T for non-target and target genes, respectively, with dual assignments). Referring to Figure 5C, it is the ratio:

$$D = \frac{E_w(m_1 \cup I_T \cup m_2)}{E_b(m_1 \cup I_T \cup m_2, M_1' \cup I \cup M_2')} \quad (3)$$

As for the C score, the statistical significance of the D score is ensured by controlling the FDR. To obtain p-values for the D score, all genes in the two modules are randomly partitioned into sets of the same sizes as $m_1 \cup I_T \cup m_2$ and $M_1' \cup I \cup M_2'$ and D is recomputed, that is:

$$D^{rand} = \frac{E_w(r_{|m_1 \cup I_T \cup m_2|}(M_1 \cup M_2))}{E_b(r_{|m_1 \cup I_T \cup m_2|}(M_1 \cup M_2), r_{|M_1' \cup I \cup M_2'|}(M_1 \cup M_2))} \quad (4)$$

Given Q TFs, the randomization procedure is repeated $(3 \times Q/0.05) - 1$ times.

Motif combinations

A "quartet" is defined, in analogy to the triplet, as one pair of modules and one pair of motifs. For every quartet, the intersections among the gene sets were determined. As above, Convergence ($C_{quartet}$) and Distinctness ($D_{quartet}$) scores were defined and used to compare the expression coherence of double-target genes (that is, genes that are targets of both TFs, represented by gene set B) to that of non-target genes (set N) and single-target genes (set X or Y) in the same category (Fig. 5D):

$$C_{quartet} = E_b(m_1^{xy}, m_2^{xy}) \quad (5)$$

$$C_{quartet}^{rand} = E_b(r_{|m_1^{xy}|}(m_1^{xy} \cup m_1^x \cup m_1^y), r_{|m_2^{xy}|}(m_2^{xy} \cup m_2^x \cup m_2^y)) \quad (6)$$

$$D_{quartet} = \frac{E_w(B)}{E_b(B,(X \cup Y \cup N))} \quad (7)$$

$$D_{quartet}^{rand} = \frac{E_w(r_{|B|}(B \cup X \cup Y))}{E_b(r_{|B|}(B \cup X \cup Y), r_{|X \cup Y|}(B \cup X \cup Y) \cup N)} \quad (8)$$

A key distinction between these scores and their single-motif counterparts is that for motif pairs, randomizations are done only within the set of target genes, and exclude non-target genes. This modification is necessary in order to determine whether the expression profile conferred by the motif pair is significantly different from the expression profiles conferred by each motif individually.

Selecting significant coordinately regulated modules

Stringent criteria

A pair of modules is said to be coregulated by a particular TFBS or TFBS pair if the q-value (Storey and Tibshirani 2003) for both the D and C scores is at most 0.15. Because the significance of the two scores can be combined in a manner that accounts for their dependence, the overall q-value is lower than 0.15.

Lenient criteria

The above criteria are highly conservative in that (1) they require two different scores to be statistically significant, (2) the signifi-

cance of those scores was calculated using conservative randomization procedures, and (3) each triplet was required to have at least four target genes in each module. Because biological significance may not always manifest itself as statistical significance, it is possible that the high stringency of the conservative criteria results in false negatives. In several analyses (where noted) above, therefore, we also consider some triplets or quartets for which one score has a q-value ≤0.15, and the other score has a p-value ≤0.05 (not corrected for multiple hypotheses by controlling the FDR). Results obtained using these lenient criteria are labeled as such in the complete list of results in Supplemental Figures S2, S3 (triplets), and S10 (quartets).

References

Alon, U. 2003. Biological networks: The tinkerer as an engineer. *Science* **301:** 1866–1867.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25:** 25–29.

Barabasi, A.L. and Oltvai, Z.N. 2004. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5:** 101–113.

Batagelj, V. and Mrvar, A. 1998. Pajek—Program for large network analysis. *Connections* **21:** 47–57.

Berg, J., Tymoczko, J., and Stryer, L. 2002. *Biochemistry*. W.H. Freeman and Company, New York.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. 2001. Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12:** 323–337.

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., et al. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2:** 65–73.

Christie, K.R., Weng, S., Balakrishnan, R., Costanzo, M.C., Dolinski, K., Dwight, S.S., Engel, S.R., Feierbach, B., Fisk, D.G., Hirschman, J.E., et al. 2004. *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.* **32:** D311–D314.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. 2003. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301:** 71–76.

Danial, N.N., Gramm, C.F., Scorrano, L., Zhang, C.Y., Krauss, S., Ranger, A.M., Datta, S.R., Greenberg, M.E., Licklider, L.J., Lowell, B.B., et al. 2003. BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature* **424:** 952–956.

Delgado, M.L., O'Connor, J.E., Azorin, I., Renau-Piqueras, J., Gil, M.L., and Gozalbo, D. 2001. The glyceraldehyde-3-phosphate dehydrogenase polypeptides encoded by the *Saccharomyces cerevisiae* TDH1, TDH2 and TDH3 genes are also cell wall proteins. *Microbiology* **147:** 411–417.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278:** 680–686.

Edgington, E.S. 1995. *Randomization tests*. Marcel Dekker, New York.

Elkhaimi, M., Kaadige, M.R., Kamath, D., Jackson, J.C., Biliran Jr., H., and Lopes, J.M. 2000. Combinatorial regulation of phospholipid biosynthetic gene expression by the UME6, SIN3 and RPD3 genes. *Nucleic Acids Res.* **28:** 3160–3167.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. 2000. Genomic expression programs in the response of yeast cells to environmental changes.

Mol. Biol. Cell **11:** 4241–4257.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.R., et al. 2004. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430:** 88–93.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. 1999. From molecular to modular cell biology. *Nature* **402:** C47–C52.

Honigberg, S.M. and Purnapatre, K. 2003. Signal pathway integration in the switch from the mitotic cell cycle to meiosis in yeast. *J. Cell Sci.* **116:** 2137–2147.

Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., Piccirrillo, S., Gerstein, M., and Snyder, M. 2002. Complex transcriptional circuitry at the G$_1$/S transition in *Saccharomyces cerevisiae*. *Genes & Dev.* **16:** 3017–3033.

Hubbard, E.J., Jiang, R., and Carlson, M. 1994. Dosage-dependent modulation of glucose repression by MSN3 (STD1) in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **14:** 1972–1978.

Hughes, J.D., Estep, P.W., Tavazoie, S., and Church, G.M. 2000. Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **296:** 1205–1214.

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

Ihmels, J., Bergmann, S., and Barkai, N. 2004. Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20:** 1993–2003.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409:** 533–538.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32:** D277–D280.

Kholodenko, B.N., Kiyatkin, A., Bruggeman, F.J., Sontag, E., Westerhoff, H.V., and Hoek, J.B. 2002. Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci.* **99:** 12841–12846.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298:** 799–804.

Lesage, P., Yang, X., and Carlson, M. 1994. Analysis of the SIP3 protein identified in a two-hybrid screen for interaction with the SNF1 protein kinase. *Nucleic Acids Res.* **22:** 597–603.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nat. Genet.* **28:** 327–334.

Lopez, M.C. and Baker, H.V. 2000. Understanding the growth phenotype of the yeast gcr1 mutant in terms of global genomic expression patterns. *J. Bacteriol.* **182:** 4970–4978.

Ma, H.W., Zhao, X.M., Yuan, Y.J., and Zeng, A.P. 2004. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics* **20:** 1870–1876.

Mewes, H.W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S., and Weil, B. 2002. MIPS: A database for genomes and protein sequences. *Nucleic Acids Res.* **30:** 31–34.

Murray, M. and Greenberg, M.L. 2000. Expression of yeast INM1 encoding inositol monophosphatase is regulated by inositol, carbon source and growth stage and is decreased by lithium and valproate. *Mol. Microbiol.* **36:** 651–661.

Pereira-Leal, J.B., Enright, A.J., and Ouzounis, C.A. 2004. Detection of functional modules from protein interaction networks. *Proteins* **54:** 49–57.

Pilpel, Y., Sudarsanam, P., and Church, G.M. 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* **29:** 153–159.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297:** 1551–1555.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306–2309.

Rives, A.W. and Galitski, T. 2003. Modular organization of cellular networks. *Proc. Natl. Acad. Sci.* **100:** 1128–1133.

Roberts, C.J., Nelson, B., Marton, M.J., Stoughton, R., Meyer, M.R., Bennett, H.A., He, Y.D., Dai, H., Walker, W.L., Hughes, T.R., et al. 2000. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287:** 873–880.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* **16:** 939–945.

Schwikowski, B., Uetz, P., and Fields, S. 2000. A network of protein–protein interactions in yeast. *Nat. Biotechnol.* **18:** 1257–1261.

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. 2003. Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34:** 166–176.

Segre, D., Deluna, A., Church, G.M., and Kishony, R. 2005. Modular epistasis in yeast metabolism. *Nat. Genet.* **37:** 77–83.

Snel, B., Bork, P., and Huynen, M.A. 2002. The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci.* **99:** 5890–5895.

Spirin, V. and Mirny, L.A. 2003. Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci.* **100:** 12123–12128.

Storey, J.D. and Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100:** 9440–9445.

Sudarsanam, P., Pilpel, Y., and Church, G.M. 2002. Genome-wide co-occurrence of promoter elements reveals a *cis*-regulatory cassette of rRNA transcription motifs in *Saccharomyces cerevisiae*. *Genome Res.* **12:** 1723–1731.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. 1999. Systematic determination of genetic network architecture. *Nat. Genet.* **22:** 281–285.

Tucker, C.L., Gera, J.F., and Uetz, P. 2001. Towards an understanding of complex protein networks. *Trends Cell Biol.* **11:** 102–106.

Willis, K.A., Barbara, K.E., Menon, B.B., Moffat, J., Andrews, B., and Santangelo, G.M. 2003. The global transcriptional activator of *Saccharomyces cerevisiae*, Gcr1p, mediates the response to glucose by stimulating protein synthesis and CLN-dependent cell cycle progression. *Genetics* **165:** 1017–1029.

Yook, S.H., Oltvai, Z.N., and Barabasi, A.L. 2004. Functional and topological characterization of protein interaction networks. *Proteomics* **4:** 928–942.

Young, E.T., Dombek, K.M., Tachibana, C., and Ideker, T. 2003. Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *J. Biol. Chem.* **278:** 26146–26158.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., et al. 2001. Global analysis of protein activities using proteome chips. *Science* **293:** 2101–2105.

Zhu, Z., Pilpel, Y., and Church, G.M. 2002. Computational identification of transcription factor binding sites via a transcription-factor-centric clustering (TFCC) algorithm. *J. Mol. Biol.* **318:** 71–81.

Web site references

http://atlas.med.harvard.edu/download/index.html; AlignACE.