Identifying combinatorial regulation of transcription factors and binding motifs

Mamoru Kato*†, Naoya Hata†, Nilanjana Banerjee†‡, Bruce Futcher§ and Michael Q Zhang†

Addresses: *Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan. †Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. ‡George Mason University, School of Computational Sciences, 10900 University Boulevard, Manassas, VA 20110, USA. §Department of Molecular Genetics and Microbiology, University of Stony Brook, Stony Brook, NY 11794, USA.

Correspondence: Michael Q Zhang. E-mail: mzhang@cshl.edu

Abstract

Background: Combinatorial interaction of transcription factors (TFs) is important for gene regulation. Although various genomic datasets are relevant to this issue, each dataset provides relatively weak evidence on its own. Developing methods that can integrate different sequence, expression and localization data have become important.

Results: Here we use a novel method that integrates chromatin immunoprecipition (ChIP) data with microarray expression data and with combinatorial TF-motif analysis. We systematically identify combinations of transcription factors and of motifs. The various combinations of TFs involved multiple binding mechanisms. We reconstruct a new combinatorial regulatory map of the yeast cell cycle in which cell-cycle regulation can be drawn as a chain of extended TF modules. We find that the pairwise combination of a TF for an early cell-cycle phase and a TF for a later phase is often used to control gene expression at intermediate times. Thus the number of distinct times of gene expression is greater than the number of transcription factors. We also see that some TF modules control branch points (cell-cycle entry and exit), and in the presence of appropriate signals they can allow progress along alternative pathways.

Conclusions: Combining different data sources can increase statistical power as demonstrated by detecting TF interactions and composite TF-binding motifs. The original picture of a chain of simple cell-cycle regulators can be extended to a chain of composite regulatory modules: different modules may share a common TF component in the same pathway or a TF component cross-talking to other pathways.

Background

Gene expression is controlled by combinatorial interaction of transcription factors (TFs) and their binding motifs in DNA. Recent advances in genomic technology such as the DNA microarray have allowed systematic investigation of combinatorial control. However, the classic approach in microarray analysis is to cluster gene-expression patterns and to identify individual DNA sequence motifs specific to each expression

cluster [1-5]. The limitations of this approach are: it does not directly address combinatorial regulation by transcription factors; it does not identify the relevant transcription factor(s) even if an over-represented motif is found; and, because it uses a limited amount of information, the statistical significance of the results is limited, and so the approach will probably not be sufficiently powerful for a large genome.

More recently, more sophisticated methods [6-9] have been used to discover important motif combinations. However, motif discovery and manipulation methods cannot, on their own, determine which transcription factor binds to a particular motif or promoter. Recently, chromatin immunoprecipitation (ChIP) microarray data have become available which connect each of a large number of transcription factors to a large number of target genes. Lee *et al.* [10] have recently published ChIP microarray data for most of the transcription factors listed in the Yeast Proteome Database (YPD) [11]. The ChIP microarray technique is thought to provide strong *in vivo* evidence of direct binding of a specific protein complex to DNA [10,12-14], and motif-finding algorithms based on ChIP data have also been developed [15,16]. ChIP datasets and expression datasets are complementary kinds of data arising from different kinds of techniques, and in principle they can be profitably combined. When properly integrated, these datasets can identify not only target genes bound by multiple transcription factors, but also the corresponding regulatory motifs, with far greater statistical power than the non-integrated datasets. Banerjee and Zhang [17] have applied the method of Pilpel *et al.* [6] directly to ChIP microarray data to identify TF combinations. Similarly, an iterative approach was proposed [18] to improve expression clustering by identifying TF combinations using ChIP microarray data. But these methods fall short on identification of motif combinations. The integration of genome-wide ChIP data and expression data with combinatorial TF-motif analysis has therefore become an urgent issue.

Here we propose a novel method that further integrates these datasets and analyzes them to systematically identify combinations of both TFs and motifs. First, the method uses ChIP data for each transcription factor to identify over-represented motifs for that transcription factor. Second, for all possible combinations of over-represented motifs, it screens for those combinations found in genes that are transcribed at about the same time. Third, it further selects motif combinations found in genes that have strongly coherent expression patterns. Finally, it assigns particular TF combinations to the respective motif combinations by matching 'over-represented motifs' with 'over-represented TFs' (see Materials and methods). Taken together, the method outputs combinations of TFs and motifs that are specific to a functional gene set. We applied this method to yeast cell-cycle genes using both ChIP data and expression microarray data, and searched for up to three combinations of 6- to 9-mer motifs as well as combinations of TFs. In addition to previously known motifs, we found several new putative motif variations and combinations, and the corresponding TF combinations. We classified these over-represented motifs into three types of transcription-factor binding mechanisms and report novel combinations of TFs and motifs that are specific to particular cell cycle phases, and assign them to functional duties. Finally, combining all the results, we reconstructed a map of combinatorial regulation in the yeast cell cycle. This map highlights some important features in combinatorial regulation by TF modules. Furthermore, we have shown that by combining evidence from different, individually noisy, genomic resources, one can achieve much higher specificity, suggesting that this integrated approach will become essential when applied to large genomes.

Results
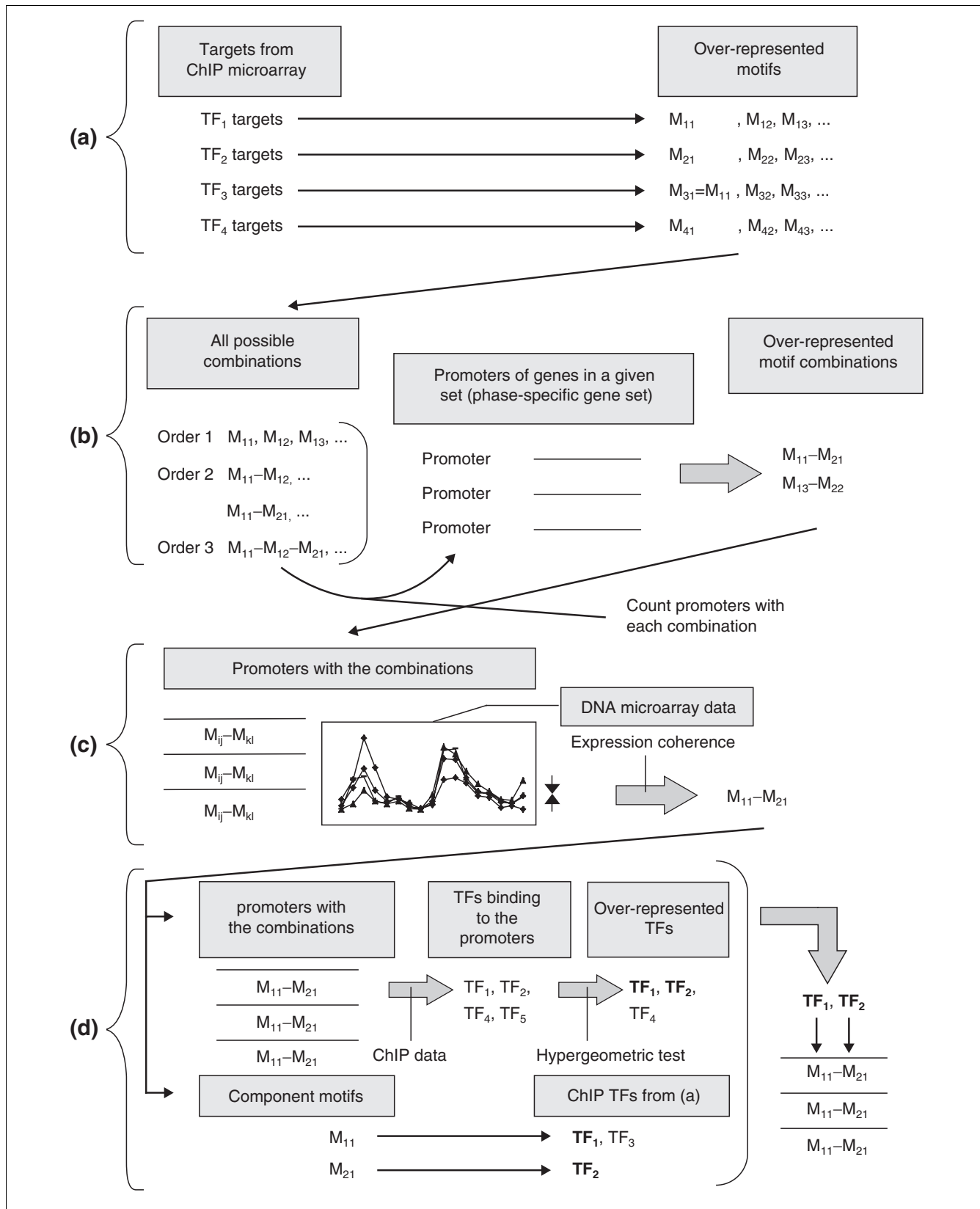### Binding motifs over-represented in ChIP data
We applied our method (Figure 1) to 113 transcription factors using the ChIP data of Lee *et al.* [10] and obtained over-represented motifs for each transcription factor (Table 1 and see also Table A at [19]). First we took the intersection of these target genes and cell-cycle genes (Figure 1). Next the method enumerated all possible 6-mer to 9-mer motifs and selected only motifs that were over-represented in the upstream regions of the intersection genes. Only 21 of the 113 TFs had over-represented motifs at our statistical criteria (see [19]); most of the other transcription factors do not have a cell-cycle specific role. However, our criteria were very stringent (see Materials and methods) to avoid false positives.

Figure 1 *(see following page)*
Overview of the method. **(a)** Finding over-represented single motifs from ChIP data. Target promoters of each TF are determined from ChIP data, and these promoters are searched for over-represented motifs. **(b)** Finding over-represented motif combinations specific to each cell-cycle phase. For all over-represented motifs found in (a), all possible first (single)-, second (double)- and third (triple)-order combinations of motifs are searched against the promoters of the G1, S, S/G2, G2/M, M/G1 or 'All' groups of cell-cycle genes. **(c)** When an over-represented motif combination is found, we require that the genes possessing that combination have more coherent patterns of gene expression than the patterns for the phase-specific gene group.**(d)** From ChIP data, we find the set of TFs (over-represented TFs, see Materials and methods) binding to the promoters with a motif combination (from (b) and (c)), and the set of TFs that can bind over-represented motifs (from (a)) constituting a motif combination. For each motif combination and its component motifs, we take the intersection of TFs from these two sets. Such intersection TFs (in bold in the (d)) are assigned to the component motifs of a combination (see Materials and methods).

(a)

Targets from ChIP microarray

Over-represented motifs

TF$_1$ targets → M$_{11}$, M$_{12}$, M$_{13}$, ...

TF$_2$ targets → M$_{21}$, M$_{22}$, M$_{23}$, ...

TF$_3$ targets → M$_{31}$=M$_{11}$, M$_{32}$, M$_{33}$, ...

TF$_4$ targets → M$_{41}$, M$_{42}$, M$_{43}$, ...

(b)

All possible combinations

Promoters of genes in a given set (phase-specific gene set)

Over-represented motif combinations

Order 1   M$_{11}$, M$_{12}$, M$_{13}$, ...

Order 2   M$_{11}$–M$_{12}$, ...

          M$_{11}$–M$_{21}$, ...

Order 3   M$_{11}$–M$_{12}$–M$_{21}$, ...

Promoter

Promoter

Promoter

M$_{11}$–M$_{21}$

M$_{13}$–M$_{22}$

Count promoters with each combination

(c)

Promoters with the combinations

M$_{ij}$–M$_{kl}$

M$_{ij}$–M$_{kl}$

M$_{ij}$–M$_{kl}$

DNA microarray data

Expression coherence

M$_{11}$–M$_{21}$

(d)

promoters with the combinations

TFs binding to the promoters

Over-represented TFs

M$_{11}$–M$_{21}$

M$_{11}$–M$_{21}$

M$_{11}$–M$_{21}$

ChIP data

TF$_1$, TF$_2$, TF$_4$, TF$_5$

TF$_1$, TF$_2$, TF$_4$

Hypergeometric test

Component motifs

ChIP TFs from (a)

M$_{11}$ → **TF$_1$**, TF$_3$

M$_{21}$ → **TF$_2$**

TF$_1$, TF$_2$

M$_{11}$–M$_{21}$

M$_{11}$–M$_{21}$

M$_{11}$–M$_{21}$

Figure 1 *(see legend on previous page)*

For many of the transcription factors in Table A at [19], there is existing knowledge about their mechanisms of DNA binding. From this knowledge, we can see that the relationship between a TF and its over-represented motifs fall into three categories: direct binding; piggy-back binding; or cross-binding (Figure 2). For example, it is known that Fkh2 (and to a lesser extent Fkh1), together with Mcm1, recruits Ndd1, and they control the transcription of G2/M genes [14,20] in the yeast cell cycle. GTAAACAA is known to be the direct-binding motif of both Fkh1 and Fkh2 [21]. Our results show that GTAAACA[A] ([A] is either A or N) is indeed over-represented in Fkh1 and Fkh2 ChIP data (Table 1). But GTAAACA is also over-represented in the ChIP data for Ndd1. Ndd1 does not directly bind to DNA but interacts with Fkh1 or Fkh2, both of which bind directly to DNA [20]. Thus, GTAAACA is a direct-binding motif for Fkh1 and Fkh2, but a piggy-back binding motif for Ndd1 (Figure 2a,b).

Mcm1 is involved in several biological processes including cell cycle control [14,20-22]. The transcription factor can form a homodimer and this is reflected in the dyad symmetry of its binding motif, TTACCNAATTNGGTAA [23], which is often referred to as the ECB [24]. Because we searched for 6-mers to 9-mers, we did not extract this 16-mer, but we did extract sub-sequences. For example, our motif TTTCCTAA (Table 1) is exactly half of the motif (TTTCCTAATTAGGAAA) found by Liu *et al.* [15] and is almost identical to half of the ECB [23]. Interestingly, the motif ATAATTA was associated with Mcm1 in M/G1 phase. This motif is likely to correspond to (T/C)AATTA, the binding site of the proteins Yox1 and Yhp1, which are recently characterized binding partners of Mcm1 in M/G1 [25]. Thus, ATAATTA may be a cross-binding motif of Mcm1 via Yox1 and Yhp1 (Figure 2c). Yox1 and Yhp1 were not among the 113 transcription factors assayed by ChIP [10], and so are not in the set of transcription factors for which we determined over-represented motifs. However, Horak *et al.* [26] did ChIP experiments for Yox1 and Yhp1. We searched for over-represented motifs for their ChIP targets, and in addition, for targets determined on the basis of mutagenesis experiments by Pramila *et al.* [25]. While we found the putative binding motif in the dataset of Pramila *et al.*, we could not find it, or any similar motifs, in the dataset of Horak *et al.* Although this is disappointing, it is consistent with the fact that Pramila *et al.* and Horak *et al.* largely disagreed on the genes regulated by Yox1 and Yhp1; Horak *et al.* defined 320 targets, whereas Pramila *et al.* defined 28 targets. Only two of these targets overlapped, whereas the expectation from picking random genes is an overlap of 1.5.

SBF, a complex containing Swi4 and Swi6, predominantly controls the expression of budding and cell-wall genes, and MBF, a related complex composed of Mbp1 and Swi6, functions in DNA replication [13]. The binding motifs of SBF and MBF (called SCB and MCB motifs) [27] are CRCGAAA and ACGCGT [3], respectively. In our results (Table 1), the most prominent motifs of the DNA-binding proteins Swi4 and
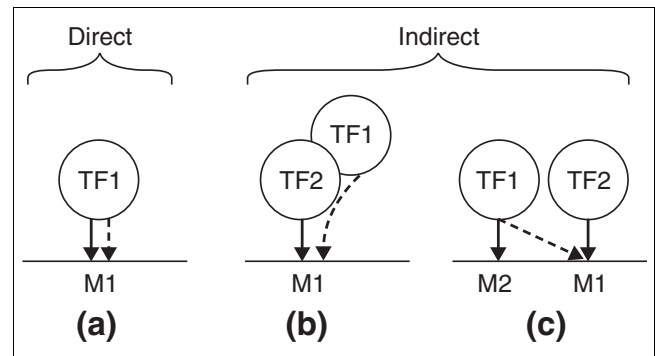
Figure 2
Over-represented motifs from ChIP data reflect three types of binding mechanisms. A solid arrow means that the TF binds directly to the motif. A dotted arrow means that a motif (for example, M1) is over-represented in TF1 ChIP data. This can occur **(a)** because of direct binding, or **(b, c)** because of two modes of indirect binding. (a) M1 is a direct-binding motif of TF1. (b) M1 is an indirect piggy-back binding motif of TF1 via TF2, where TF1 binds TF2, and TF2 binds the motif. For example, M1, TF1, and TF2 could correspond to GTAAACA (the Fkh2 motif), Ndd1, and Fkh2, respectively, as Ndd1 binds to Fkh2, which in turn binds to GTAAACA. (c) M1 is an indirect cross-binding motif of TF1 via TF2. For example, M1, TF1 and TF2 could correspond to GTAAACA (the Fkh2 motif), Mcm1, and Fkh2, respectively, since Mcm1 binds to its own motif (M2) on the same promoter as the Fkh2 motif.

Mbp1 [27] are CGCGAA and ACGCGT, consistent with the known motifs. We find that Swi6, which is a non DNA-binding cofactor of Swi4 and Mbp1 [27], has the same motifs, CGCGAA (see [19]) and ACGCGT. We interpret these as piggy-back binding motifs for Swi6 via Swi4 and Mbp1. We also find the novel variant CGCGTC, which is associated with as many as nine TFs, including SBF and MBF (see [19]). Conlon *et al.* [28] also pointed out this GCGTC motif in cell-cycle genes. We describe further studies of the CGCGTC motif in additional data provided at [19]. As expected, our integrated approach has substantially enhanced the signal-to-noise ratio. Such integration may not be necessary for the analysis of yeast, but will be crucial for analysis of higher eukaryotes. Among the nine TFs, Ash1 is thought to be a regulator of mating-type switching [29]. Hence this points to a possible connection between cell cycle and mating-type switching through the combination of SBF/MBF and Ash1. The binding motif of Ash1 is known to be YTGAT [29], but we found [A]GGCAC[C] and GCGGCA. Probably, these putative motifs are indirect binding motifs of Ash1, which suggests that Ash1 may cooperate with an unknown factor (or factors) through these motifs and in the end with SBF/MBF as well (see also [19]).

SCB- and MCB-like motifs were also found as over-represented motifs from the ChIP data for Stb1 and Ste12 (Table 1). Stb1 binds to Swi6 *in vitro* and is thought to interact with the Swi6 subunit of SBF and MBF to regulate transcription *in vivo* [30]. In our results, the most prominent motifs of Stb1 were CGCGAAAA and ACGCGA, which closely resemble the

Table 1

Over-represented motifs from ChIP data

ChIP TF	Phase	Over-represented motifs (exponential part of *p*-value without multiplicity correction)					Known motif	Known TF
FKH1	All	GTAAACA (-23)	GTAAACAA (-23)	AATAAACA (-10)			TAAACAAA	Fkh1 [15]
FKH2	SG2	TAAACAA (-11)					GTAAACAAA	Fkh1, Fkh2 [20,21]
FKH2	G2M	GTAAACA (-14)						
FKH2	All	GTAAACA (-25)	GTAAACAA (-22)	AAATAAACA (-12)	ATAAACAA (-09)	**CGCGTC (-09)**		
NDD1	G2M	GTAAACA (-09)						
NDD1	All	GTAAACA (-09)						
MCM1	G2M	*TTCCTAA (-11)*	*TTTCCTA (-11)*	*ATTTCC (-09)*	*ATTTCCT (-09)*		*TTACCNAATTNGGTAA*	Mcm1 [23]
MCM1	All	*ATTTCCT (-14)*	*TTCCTAA (-12)*	*TTTCCTAA (-12)*	*ATTTCCTA (-10)*	*AATTTCCT (-09)*	*TTTCCTAATTAGGAAA*	Mcm1 [15]
MCM1	MG1	*ATAATTA (-10)*					*YAATTA*	Yox1, Yhp1 [25]
MBP1	G1	**ACGCGT (-104)**	**CGCGTC (-100)**	**AACGCGT (-95)**	**ACGCGTC (-75)**	**AAACGCGT (-52)**	**ACGCGT**	Mbp1, Swi6 [3]
MBP1	All	**CGCGTC (-97)**	**ACGCGT (-85)**	**AACGCG (-81)**	**AACGCGT (-75)**	**ACGCGTC (-62)**	**CRCGAAA**	Swi4, Swi6 [3]
SWI4	G1	**CGCGAA (-56)**	**CGCGAAA (-55)**	**CGCGTC (-49)**	**ACGCGA (-44)**	**AACGCG (-43)**		
SWI4	All	**CGCGTC (-46)**	**CGCGAAA (-46)**	**ACGCGA (-40)**	**AACGCG (-34)**	**GACGCGA (-32)**		
SWI6	G1	**ACGCGT (-75)**	**AACGCGT (-69)**	**CGCGTC (-57)**	**ACGCGTC (-52)**	**AAACGCG (-41)**		
SWI6	All	**ACGCGT (-59)**	**CGCGTC (-59)**	**AACGCGT (-50)**	**ACGCGTC (-45)**	**CGACGCG (-37)**		
STB1	G1	**CGCGAAAA (-50)**	**ACGCGA (-44)**	**AACGCG (-38)**	**CGCGTC (-34)**	**AACGCGT (-33)**		
STB1	All	**CGCGAAAA (-49)**	**ACGCGA (-41)**	**CGCGTC (-32)**	**AACGCG (-32)**	**ACGCGAA (-30)**		
STE12	G1	**CGCGTC (-19)**	**CGCGAA (-13)**	**ACGCGA (-12)**			ATGAAAC	Ste12 [34]
STE12	All	TGAAACA (-20)	**GACGCGA (-17)**	ATGAAAC (-16)	**ACGCGAA (-10)**	TGAAACAA (-10)	TGAAACA	Ste12 [33]
DIG1	All	TGAAACA (-22)	ATGAAAC (-14)	**CGCGTC (-11)**				
ACE2	All	*ACCAGCA (-19)*					*CCAGCA*	Swi5 [39]
SWI5	G1	*ACCAGC (-11)*					*ACCAGC*	Ace2, Swi5 [37,38]
SWI5	MG1	*CCAGCA (-10)*					*RRCCAGCR*	Ace2, Swi5 [1]
SWI5	All	*ACCAGC (-13)*	*AACCAGC (-12)*	*CCAGCC (-12)*	*CCAGCA (-10)*	*ACCAGCA (-09)*		
MTH1	All	CAGCAG (-10)*						
MET4	All	***CACGTG (-29)***	*TCACGTG (-25)*	**ACTGTGG (-22)**	**GTGGCA (-12)**		*TCACGTG*	Cbf1, Met4, Met28 [40]
MET31	All	**TGTGGC (-14)**					**AAACTGTGG**	Met31, Met32 [1]
							AAANTGTGGC	Met31, Met32 [2]

Over-represented motifs of 6- to 9-mers were extracted from the intersection between ChIP targets of a TF and genes of a cell-cycle phase. The phases are G1, S, S/G2, G2/M, M/G1 or the union of 'All' the phases according to the classification of Spellman *et al.* [1]. The *p*-value indicates the degree of over-representation of a motif (without multiplicity correction). Motifs in the same typeface belong to the same group; * indicates motifs that do not belong to any particular group. R is A or G. This table shows ChIP TFs described in the main text but lists only the five highest-ranking motifs. Full results are listed in Table A at [19].

SCB and MCB, respectively. Together, these results imply the presence of the complexes Stb1+Swi6+Swi4, and Stb1+Swi6+Mbp1, which are different from the standard complexes SBF (Swi6+Swi4) and MBF (Swi6+Mbp1). These Stb1 motifs are piggy-back binding motifs via SBF and MBF. In the case of Ste12, in contrast, there is no evidence that Ste12 binds to SBF or MBF. Furthermore, many of the genes that have both the Ste12 direct-binding motif and the SCB- and MCB-like motif (CGCGTC) are bound by Ste12, SBF and MBF (see below). Thus, we suggest that the SCB- and MCB-like motif is a cross-binding motif of Ste12; that is, there is a group of genes where Ste12 binds to its direct binding site and SBF or MBF binds to its direct binding site in the same promoters.

Ste12 is involved in pheromone response and filamentous growth [22,31,32]. The known binding motifs of Ste12 are ATGAAAC and TGAAACA (called the PRE) [33,34]. Our results match these known motifs perfectly (Table 1). In addition, we find that Dig1 is associated with these motifs. Dig1 is an inhibitor of Ste12 [35]. Thus, the PRE motifs are presumably piggy-back binding motifs for Dig1 via Ste12. This result suggests that the Dig1+Ste12 complex binds DNA, despite the fact that Ste12 activity is inhibited, consistent with the results of Olson *et al.* [35].

Ace2 and Swi5 are transcription factors that function at the M/G1 boundary [14,36]. The direct-binding motifs of Ace2 and Swi5 are variously stated as ACCAGC [37,38] or RRC-CAGCR [1] or CCAGCA [39] (for Swi5). Our results are consistent with these motifs (Table 1).

Met4 and Met31 are involved in sulfur amino-acid metabolism [40] and may have a transcriptional role in cell-cycle control [1,2]. The transcriptional mechanisms differ between targets but two main mechanisms have been suggested [40]. First, the DNA-binding protein Cbf1 binds to its motif TCACGTC, and tethers a Met4+Met28 complex to this site. Second, the DNA-binding protein Met31 (or Met32) binds to its motif AAACTGTG, and likewise tethers a Met4+Met28 complex to this site. In our results, TCACGTG appears from the Met4 ChIP data, which matches the known binding motif of Cbf1, suggesting that TCACGTG is a piggy-back binding motif of Met4 via Cbf1. The motif ACTGTGG also appears in the Met4 ChIP data. This motif is similar to the known binding motif of Met31/Met32 and the motif AAACTGTGG of Spellman *et al.* [1]. Thus this motif is presumably a piggy-back binding motif of Met4 via Met31/Met32. Using the Met31 ChIP data, we find the over-represented motif TGT-GGC, which overlaps the motifs of Spellman *et al.* (AAACT-GTGG) and Tavazoie *et al.* [2] (AAANTGTGGC), and represents the direct-binding motif of Met31. Note that in the studies of Spellman *et al.* and Tavazoie *et al.*, which are based on expression clustering and motif searching, one has to try and identify the TF from other kinds of data after finding an over-represented motif. In our case, the ChIP data for the relevant TF gives this information directly.

Finally, we found that Mth1, which is involved in glucose signal transduction [41], binds to several cell-cycle genes, including some of the histone genes, and genes involved in budding and polarized growth. It was somewhat surprising to find Mth1 as a controller of cell-cycle genes. It is thought to act as a co-factor with other transcription factors involved in glucose signal transduction, and is not known to bind DNA directly. We found the motifs CAGCAG and CGCGTC over-represented in Mth1 ChIP data (see further investigation at [19]). We presume that Mth1 is a piggy-back or cross-binding transcription factor for these motifs. Possible candidates for the direct-binding factor are Swi5/Ace2 (for CAGCAG), or SBF or MBF (for CGCGTC), or Rgt1, as Rgt1 has a GC-rich

binding site [42] and is also involved in glucose signal transduction [41]. As glucose accelerates the growth of yeast cells, and therefore accelerates the cell cycle, it is possible that Mth1 is used to control a rapid cell cycle in response to the glucose growth signal. Indeed, Heideman and co-workers [43] have shown that expression of *CLN3*, a major activator of the cell-cycle program, is controlled in part by the availability of glucose.

Phase-specific combinations of TFs and motifs
The serial regulation of the yeast cell cycle is thought to occur as follows [14]: MBF (Mbp1+Swi6) and SBF (Swi4+Swi6) bind to the motifs ACGCGT [3] and CRCGAAA [3], respectively, to control the expression of late G1 genes [27]; Fkh1/Fkh2 and Mcm1 bind to GTAAACAAA [21] and TTACCNAAT-TNGGTAA [23], respectively, and recruit Ndd1 to control G2/M genes [20]; and Mcm1 and Ace2/Swi5 bind to the ECB motif and RRCCAGCR [1], respectively, to regulate M/G1 genes [24,25,36]. This general model is supported by many experiments [1,10,13,14,44]. Transcriptional control in S and S/G2 phases is less well characterized, but some studies suggest the involvement of SBF and Fkh1/Fkh2 [10,14,45].

Because we are interested in combinatorial control, we used the procedure shown in Figure 1**b**-d to search for combinations of TFs and motifs that are specific to each of the cell-cycle phases. For G1, we confirmed the regulatory role of MBF and SBF: Mbp1 and Swi6, and Swi4 and Swi6 are predicted to bind to ACGCGT and CGCGAA (and variants) respectively (Table 2). However, we further predict that Stb1, Mbp1 and/or Swi4 are associated with ACGCGA. Taken together with the results above, this suggests that putative complexes of Stb1+Swi6+Mbp1, and/or Stb1+Swi6+Swi4 bind to this motif and regulate some G1 genes. Lee *et al.* also found a significant number of targets bound by both Stb1 and Swi4 in G1 phase [10].

G1 phase also gave us the combinations {SBF, Ste12} and {MBF, Ste12} (see also data at [19]). Examples of genes in these categories include *PCL2*, *GIC2*, *MSB2*, *CRH1* and *SRL1*. At least some of these are genes involved in a normal G1 phase, but are also involved in mating and the pheromone response. For instance, *GIC2* is involved in polarized growth, which is needed for normal budding, but is also needed for mating. Perhaps surprisingly, some (for example, *PCL2*, *GIC2*) of these genes are strongly induced by alpha-factor (which acts via Ste12), while others (*MSB2*, *SRL1*) appear to be repressed.

In S phase, we find the novel combination {SBF, Fkh2, Hir1} (Table 2). A few genomic analyses [10,45] have indicated the involvement of SBF and Fkh1/Fkh2 in this phase, and one of them suggested that this combination may be associated with the regulation of histone genes [45]. To clarify our results, especially with regard to histone regulation, we divided the target genes of this combination into histone genes and other

Table 2

Phase-specific TF and motif combinations

Phase	Motif combinations*	Over-represented TFs†				Component motifs‡	§ChIP TFs			
G1	ACGCGT (-49)	Mbp1	Swi6			ACGCGT	Mbp1	Swi6		
	AAACGCG (-35)	Mbp1	Swi4	Swi6		AAACGCG	Mbp1	Swi4	Swi6	
	ACGCGA and CGCGAA (-29)	Mbp1	Stb1	Swi4		ACGCGA	Mbp1	Stb1	Swi4	
						CGCGAA			Swi4	Swi6¶
S	CGCGAA[A] and TAAACAA (-09)	Fkh2	Hir1¶	Swi4	Swi6	CGCGAA			Swi4	Swi6
						CGCGAAA			Swi4	
						TAAACAA	Fkh2			
S/G2	ATAAACAA (-06)	Fkh2				ATAAACAA	Fkh2			
G2/M	GTAAACA and TTAGGAA (-09)	Mcm1	Fkh1	Fkh2	Ndd1	GTAAACA		Fkh1	Fkh2	Ndd1
						TTAGGAA	Mcm1			
M/G1	TTAGGAAA (-07)	Mcm1				TTAGGAAA	Mcm1			
	TGAAACA[A] and CCAGCA (-09)	Swi5	Dig1	Ste12		TGAAACA		Dig1	Ste12	
						TGAAACAA			Ste12	
						CCAGCA	Swi5			
	AAACGC and ATAATTA (-05)	Mcm1	Swi4¶			AAACGC		Stb1¶		
						ATAATTA	Mcm1			

*Over-represented motif combinations for each cell-cycle phase. Numbers in parentheses represent the exponential part of the *p*-value for the over-representation. †Over-represented TFs in the promoters with a motif combination. ‡Motifs constituting the corresponding motif combination. §ChIP TFs, in the targets of which the 'component motif' to the left is over-represented. ¶TFs were derived only in either over-represented TFs or ChIP TFs but not in both (as described in Materials and methods).

genes (which are involved in budding, cell-wall synthesis, microtubules and the spindle-pole body). This division did not conflict with expression coherence, because scores (the average standard deviation scores) measuring expression coherence for both sets became lower as a consequence of this division. For the latter (that is, non-histone) genes, we found that the over-represented TFs (see Materials and methods) were Swi4, Mbp1 and Fkh2, but not Hir1 (*p*-value <10⁻³). Thus SBF and Fkh2 most probably bind to these motifs, which suggests that both these transcription factors probably regulate budding, cell-wall synthesis, and spindle-related genes in S phase.

The nine histone genes are the other major class of S-phase regulated genes. These are organized into five transcription units consisting of four divergently transcribed pairs (*HTA1-HTB1*, *HTA2-HTB2*, *HHT1-HHF1*, *HHT2-HHF2*) and *HHO1*. Histone mRNAs are regulated in at least three ways. First, there is cell-cycle regulated mRNA stability, such that the message is only stable during S-phase. Second, there is a negative transcriptional element that represses transcription at inappropriate times [46]; this repressive system involves the *HIR* genes - *HIR1*, *HIR2* and *HIR3*. Third, there is a positive transcriptional element that induces histone mRNA synthesis during S-phase [47]. Despite the fact that histones were the

first cell-cycle regulated genes discovered in yeast (and perhaps in any organism), the positive regulatory element and its transcription factor remain poorly characterized. The positive regulatory region includes repeats of the sequence GCGAAA [47], which closely resembles the SBF-binding site. However, histone mRNA abundance continues to oscillate through the cell cycle even in Swi4, Mbp1 and Swi6 single mutants [48,49], which argues that SBF is not essential for regulation, possibly because of the other two modes of regulation (mRNA stability, repression). Furthermore, as we suggest below, there might also be other redundant activators.

For instance, we identified Met4, in addition to Swi4, Hir1 and Hir2, as an over-represented TF for the histone promoters. The association of Met4 was completely unexpected because Met4 is thought to be a regulator of amino-acid metabolism [40], and the involvement of Met4 in histone gene regulation has never been reported. Furthermore, we found that the binding motifs (detected at the motif finding step) for Met4 also exist in some histone promoters. Thus Met4 may be a novel regulator of histone expression.

There have been three genome-wide ChIP experiments directed at the targets of MBF and SBF [10,13,14]. All five histone transcription units have been associated with SBF (and/

Figure 3
Reconstructed transcriptional regulation model of the yeast cell cycle. Segments of the cycle contain motif combinations. The TF and motif combinations in black were known previously and are also confirmed here. Those in red are new combinations (or previously poorly characterized combinations). TFs and motifs in parentheses are those we could not detect (for example, because the TF was not included in the ChIP experiments) but whose presence is indicated by some other experimental work [25,46]. Almost all TFs appear as elements of a combination. {Swi5, Ste12, Dig1} shows a joint-process combination between cell cycle and pheromone response or filamentous growth (see Discussion).

neously? For targets of {Fkh2, Mcm1, Ndd1}, it appears that activation is cooperative, not independent [20]. Furthermore, in the case of most joint-phase S genes and M/G1 genes, it appears that the peak of gene expression is sharp rather than broad [1] (that is, expression occurs only when both factors are simultaneously active, not over the whole time that either one or the other of the factors is active) again suggesting cooperativity rather than independence. Although a physical interaction between two factors is often the basis of cooperativity, other mechanisms might also play a part.

When these transcription factor combinations are connected, the resulting chain suggests that regulation is circularly relayed from an earlier TF to a later TF through their combination. Namely, it is relayed from Swi4 (SBF) to Fkh2 via {SBF, Fkh2}, Fkh2 to Mcm1 via {Fkh2, Mcm1, Ndd1}), Mcm1 to Swi4 via {Mcm1, Swi4}, and so forth (Figure 3). This feature is complementary to the finding of Simon *et al.* [14]. They found that transcription activators that function during one stage of the cell cycle regulate transcription activators that function during the next stage. Whereas their finding is pri-

marily focused on regulation between TFs, our chain connected by joint-phase combinations shows that the serial regulation of target genes is relayed through TF-motif combinations.

The apparent ability of cell-cycle factors such as SBF and Fkh2 to function cooperatively has some interesting consequences. It means that two factors can generate at least three peaks of expression: the SBF-only peak, the SBF plus Fkh2 peak and the Fkh2-only peak. But one can also imagine that some promoters might require both factors for gene expression, but have stronger motifs for one factor than the other. Thus a gene with a strong SBF motif and a weak Fkh2 motif might be expressed only in late S (when sufficient Fkh2 has accumulated to bind even the weak motif), while a gene with a weak SBF motif and a strong Fkh2 motif might be expressed only in early S (because later, SBF abundance might be too low to interact with the weak motif). Thus by using cooperativity, the cell could generate a continuum of peaks of expression over time using a small number of factors, and a large number of varied motifs, exactly as observed. Molecular experiments will be required to investigate this issue.

Finally, we note the existence of joint-process combinations, by which we mean combinations of TFs that allow genes to respond to two (or more) different transcriptional programs. Although cells undoubtedly have many such combinations, in our dataset the main examples involve Ste12, a regulator of the mating or pseudohyphal growth pathways. For instance, {SBF/MBF, Ste12} in G1 probably controls genes needed for G1 phase, but also independently needed for mating. Similarly {Swi5, Ste12, Dig1} in M/G1 may control genes needed for the M/G1 transition, but also important for either mating or for pseudohyphal growth.

In summary, we have extended the understanding of the yeast cell cycle by integrating ChIP-microarray analysis with expression analysis and motif-combinatorial analysis. Many of our findings from the integrated analysis confirm the results of previous analysis, hence validating our approach. However, we believe that the success of the non-integrated approaches was possible in part because *S. cerevisiae* has a small genome, its genes have very small regulatory regions, and the datasets are unusually good. As this type of genome-wide analysis moves to higher eukaryotes with larger genomes, we believe that non-integrated approaches will not have sufficient power to provide reliable results, whereas this integrated approach has overcome the limitations inherent in each individual approach. The added power of our integrated approach did allow us to find several interesting novel combinations of motifs and TFs, in particular those new joint-phase combinations. These new predictions lead directly to new hypothesis for new experiments. The computational integration of multiple approaches or datasets will be of increasing importance as more kinds of genomic resources, such as genome-wide protein-protein interaction data and

comparative genomics data, become available for more organisms. Indeed, for higher eukaryotes, where gene networks are more complex and regulatory regions are larger, we believe that integration of datasets will be absolutely essential.

Materials and methods
Finding over-represented single motifs
Full methods are described in [19]. This first step is represented briefly in Figure 1a. Among 4,339 nonredundant promoter sequences, we identified target genes (promoters) of a TF from each of 113 ChIP datasets [10] with *p*-value $<10^{-2}$. Next we took the intersection of these genes with each of six gene sets of Spellman *et al.* [1], that is, all the cell-cycle regulated genes and subclasses of G1, S, S/G2, G2/M and M/G1 phase genes. We defined each of these intersection sets as a foreground set. As a background (control) set, we chose the intersection of non-target genes of a given TF (with *p*-value in ChIP data >0.8), and non-cell cycle genes (genes other than all the cell-cycle regulated genes). From the 113 × 6 foreground-background pairs, we excluded pairs where the foreground set had fewer than 10 genes or the background set had fewer than 500.

For each foreground-background pair, we searched for over-represented motifs (6-9 bp) by a word-counting strategy [4,52,53]. The main differences of our approach from others are to use a set of non-target genes from ChIP data as a background set, and to use the contingency table test for manipulating the background set, whose size can be small. For all possible motifs, we calculated the statistic with Yates correction of the 2 × 2 contingency table test as follows:

$$T = \frac{N\left(\left|ad-bc\right|-N/2\right)^2}{\left(a+b\right)\left(c+d\right)\left(a+c\right)\left(b+d\right)}, \qquad (1)$$

	Foreground	Background
Motif	*a*	*c*
Non-motif	*b*	*d*

where *N* is the sum of *a*, *b*, *c*, and *d*; *a* and *b* are the occurrences of a given motif and other motifs than the given motif (non-motif) in a foreground set, respectively; *c* and *d* are the same in a background set. We counted these occurrences in both strands of the 600-bp upstream sequences. Then we calculated the *p*-value (without multiplicity correction) to see if each motif is significantly over-represented at the *p*-value threshold of 10^{-8}. We also required that: rank, according to *p*-value, of motifs in each data pair must be within the top 10; the number of upstream sequences with a motif must be greater than 25 in all the cell-cycle regulated genes; and *a* is 10 or more. We also excluded simple repeat motifs like AAAAAA, ATATATA and so on. Furthermore, for the list of

motifs obtained from each foreground-background pair, we merged similar motifs into extended ones (see [19]).

Finding over-represented motifs/motif combinations

For this second step (Figure 1b), we took each of the phase-specific gene sets (the G1, S, S/G2, G2/M and M/G1 genes as classified by Spellman *et al.* [1]) as a foreground set, and the intersection of the non-cell-cycle genes and genes with constant expression profiles (see [19]) as the background set. We then searched all possible order-1 combinations (single motifs), order-2 combinations (pairs), and order-3 combinations (triples) of motifs found. For each of the combinations, we calculated the statistic of Equation (1), where a and b are now the numbers of upstream sequences with and without a given combination in a foreground set, respectively; c and d are the same in a background set. Then we calculated the p-value (without multiplicity correction) to see if each combination is significantly over-represented at the p-value threshold of 2×10^{-15} and 2×10^{-5} for G1 and the other phases, respectively. We also required at least 60 occurrences of an upstream sequence for the G1 set, and at least eight occurrences for the other phase sets. Finally, for the list of obtained combinations, we merged associated combinations into extended combinations (for example merging (M1, M2) and (M2, M3) into (M1, M2, M3), see [19]). The p-value threshold for finding an associated combination pair was set to 2×10^{-5}.

Coherence of expression patterns

This procedure (Figure 1c) checks the coherence of expression profiles over time for genes that have a given motif combination in their upstream sequences. For measuring the coherence, the average standard deviation score is used:

$$Score = E_i(\sigma_g(X_{i,g})),$$

where $X_{i,g}$ is the normalized expression level of gene g at time i, σ_g is the standard deviation over genes, and E_i is the average over time. The lower the score, the closer the expression profiles are to the average. We calculated the score for genes having a motif combination in the upstream sequences for the data of Cho *et al.* [44]. We compared the score of genes having a motif combination with that of a phase-specific gene set. We selected only combinations whose genes had smaller scores than the phase-specific gene set.

Over-represented TFs in the promoters

This algorithm extracts TFs that are bound to the promoters with a motif combination (Figure 1d), based on the hypergeometric model:

$$P(t) = \frac{{}_T C_t \times {}_{K-T} C_{k-t}}{{}_K C_k}, \text{with} \quad {}_N C_n \equiv \binom{N}{n},$$

where K is the number of all promoters used and T is the number of promoters that are bound by TF_i (p-value in ChIP data $<10^{-2}$) among the K promoters, and k is the number of

promoters with a motif combination from a phase-specific gene set and t is the number of promoters that are bound by TF_i among the k promoters. For each of 113 TFs in ChIP data, this algorithm calculates the p-value

$$p = 1 - \sum_{i=0}^{t-1} P(i),$$

and then outputs TFs binding to a significant number of the promoters. We call such TFs over-represented TFs.

After checking the coherence of expression patterns, we used the above algorithm to find the set of over-represented TFs (p-value $<10^{-2}$; more stringent in G1, $<10^{-7}$) in the promoters with a motif combination comparing to binding TFs in all the promoters. From the procedure for finding over-represented single motifs, the set of TFs possible to bind the component motifs of a combination can be inferred. For each motif combination and its component motifs, we took the intersection of TFs from these two sets. We also kept those TFs for which there is additional experimental evidence even if they belonged to only one set. Thus we assigned each over-represented TF to each component motif. After arranging similar motif combinations (see [19]), we obtained the final results as shown in Table 2.

Full descriptions of the methods, detailed investigations for the transcription factor Mth1 and the binding motif CGCGTC, lists of putative histone regulators, lists of putative target genes of motif combinations, and lists of genes whose promoters have a complicated motif structure are all available at our website [19]. All the files are hierarchically structured and accessible on a web browser by tracking the hyperlinks.

References
1. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization.** *Mol Biol Cell* 1998, **9:**3273-3297.
2. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22:**281-285.
3. Zhang MQ: **Large-scale gene expression data analysis: a new challenge to computational biologists.** *Genome Res* 1999, **9:**681-688.
4. van Helden J, Andre B, Collado-Vides J: **Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies.** *J Mol Biol* 1998, **281:**827-842.
5. Roth FP, Hughes JD, Estep PW, Church GM: **Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation.** *Nat Biotechnol* 1998, **16:**939-945.
6. Pilpel Y, Sudarsanam P, Church GM: **Identifying regulatory networks by combinatorial analysis of promoter elements.** *Nat*

Genet 2001, **29:**153-159.

7. Liu X, Brutlag DL, Liu JS: **BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes.** *Pac Symp Biocomput* 2001:127-138.

8. GuhaThakurta D, Stormo GD: **Identifying target sites for cooperatively binding factors.** *Bioinformatics* 2001, **17:**608-621.

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

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

11. Costanzo MC, Crawford ME, Hirschman JE, Kranz JE, Olsen P, Robertson LS, Skrzypek MS, Braun BR, Hopkins KL, Kondu P, *et al.*: **YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information.** *Nucleic Acids Res* 2001, **29:**75-79.

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

13. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO: **Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF.** *Nature* 2001, **409:**533-538.

14. Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS, Young RA: **Serial regulation of transcriptional regulators in the yeast cell cycle.** *Cell* 2001, **106:**697-708.

15. Liu XS, Brutlag DL, Liu JS: **An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments.** *Nat Biotechnol* 2002, **20:**835-839.

16. van Steensel B, Delrow J, Bussemaker HJ: **Genomewide analysis of *Drosophila* GAGA factor target genes reveals context-dependent DNA binding.** *Proc Natl Acad Sci USA* 2003, **100:**2580-2585.

17. Banerjee N, Zhang MQ: **Identifying cooperativity among transcription factors controlling the cell cycle in yeast.** *Nucleic Acids Res* 2003, **31:**7024-7031.

18. Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK: **Computational discovery of gene modules and regulatory networks.** *Nat Biotechnol* 2003, **21:**1337-1342.

19. **Supplemental information** [http://rulai.cshl.org/kato/S00/index.htm]

20. Koranda M, Schleiffer A, Endler L, Ammerer G: **Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters.** *Nature* 2000, **406:**94-98.

21. Zhu G, Spellman PT, Volpe T, Brown PO, Botstein D, Davis TN, Futcher B: **Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth.** *Nature* 2000, **406:**90-94.

22. Shore P, Sharrocks AD: **The MADS-box family of transcription factors.** *Eur J Biochem* 1995, **229:**1-13.

23. Acton TB, Zhong H, Vershon AK: **DNA-binding specificity of Mcm1: operator mutations that alter DNA-bending and transcriptional activities by a MADS box protein.** *Mol Cell Biol* 1997, **17:**1881-1889.

24. McInerny CJ, Partridge JF, Mikesell GE, Creemer DP, Breeden LL: **A novel Mcm1-dependent element in the SWI4, CLN3, CDC6, and CDC47 promoters activates M/G1-specific transcription.** *Genes Dev* 1997, **11:**1277-1288.

25. Pramila T, Miles S, GuhaThakurta D, Jemiolo D, Breeden LL: **Conserved homeodomain proteins interact with MADS box protein Mcm1 to restrict ECB-dependent transcription to the M/G1 phase of the cell cycle.** *Genes Dev* 2002, **16:**3034-3045.

26. Horak CE, Luscombe NM, Qian J, Bertone P, Piccirrillo S, Gerstein M, Snyder M: **Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*.** *Genes Dev* 2002, **16:**3017-3033.

27. Koch C, Moll T, Neuberg M, Ahorn H, Nasmyth K: **A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase.** *Science* 1993, **261:**1551-1557.

28. Conlon EM, Liu XS, Lieb JD, Liu JS: **Integrating regulatory motif discovery and genome-wide expression analysis.** *Proc Natl Acad Sci USA* 2003, **100:**3339-3344.

29. Maxon ME, Herskowitz I: **Ash1p is a site-specific DNA-binding protein that actively represses transcription.** *Proc Natl Acad Sci USA* 2001, **98:**1495-1500.

30. Ho Y, Costanzo M, Moore L, Kobayashi R, Andrews BJ: **Regulation of transcription at the *Saccharomyces cerevisiae* start transition by Stb1, a Swi6-binding protein.** *Mol Cell Biol* 1999, **19:**5267-5278.

31. Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Walker WL, Hughes TR, *et al.*: **Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles.** *Science* 2000, **287:**873-880.

32. Lengeler KB, Davidson RC, D'Souza C, Harashima T, Shen WC, Wang P, Pan X, Waugh M, Heitman J: **Signal transduction cascades regulating fungal development and virulence.** *Microbiol Mol Biol Rev* 2000, **64:**746-785.

33. Dolan JW, Kirkman C, Fields S: **The yeast STE12 protein binds to the DNA sequence mediating pheromone induction.** *Proc Natl Acad Sci USA* 1989, **86:**5703-5707.

34. Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhauser R, *et al.*: **The TRANSFAC system on gene expression regulation.** *Nucleic Acids Res* 2001, **29:**281-283.

35. Olson KA, Nelson C, Tai G, Hung W, Yong C, Astell C, Sadowski I: **Two regulators of Ste12p inhibit pheromone-responsive transcription by separate mechanisms.** *Mol Cell Biol* 2000, **20:**4199-4209.

36. Doolin MT, Johnson AL, Johnston LH, Butler G: **Overlapping and distinct roles of the duplicated yeast transcription factors Ace2p and Swi5p.** *Mol Microbiol* 2001, **40:**422-432.

37. Dohrmann PR, Voth WP, Stillman DJ: **Role of negative regulation in promoter specificity of the homologous transcriptional activators Ace2p and Swi5p.** *Mol Cell Biol* 1996, **16:**1746-1758.

38. Knapp D, Bhoite L, Stillman DJ, Nasmyth K: **The transcription factor Swi5 regulates expression of the cyclin kinase inhibitor p40SIC1.** *Mol Cell Biol* 1996, **16:**5701-5707.

39. Zhu J, Zhang MQ: **SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*.** *Bioinformatics* 1999, **15:**607-611.

40. Blaiseau PL, Thomas D: **Multiple transcriptional activation complexes tether the yeast activator Met4 to DNA.** *EMBO J* 1998, **17:**6327-6336.

41. Mosley AL, Lakshmanan J, Aryal BK, Ozcan S: **Glucose-mediated phosphorylation converts the transcription factor Rgt1 from a repressor to an activator.** *J Biol Chem* 2003, **278:**10322-10327.

42. Hazbun TR, Fields S: **A genome-wide screen for site-specific DNA-binding proteins.** *Mol Cell Proteomics* 2002, **1:**538-543.

43. Newcomb LL, Hall DD, Heideman W: **AZF1 is a glucose-dependent positive regulator of CLN3 transcription in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 2002, **22:**1607-1614.

44. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, Davis RW: **A genome-wide transcriptional analysis of the mitotic cell cycle.** *Mol Cell* 1998, **2:**65-73.

45. Futcher B: **Transcriptional regulatory networks and the yeast cell cycle.** *Curr Opin Cell Biol* 2002, **14:**676-683.

46. Dimova D, Nackerdien Z, Furgeson S, Eguchi S, Osley MA: **A role for transcriptional repressors in targeting the yeast Swi/Snf complex.** *Mol Cell* 1999, **4:**75-83.

47. Osley MA, Gould J, Kim S, Kane MY, Hereford L: **Identification of sequences in a yeast histone promoter involved in periodic transcription.** *Cell* 1986, **45:**537-544.

48. Cross FR, Hoek M, McKinney JD, Tinkelenberg AH: **Role of Swi4 in cell cycle regulation of CLN2 expression.** *Mol Cell Biol* 1994, **14:**4779-4787.

49. Lowndes NF, Johnson AL, Breeden L, Johnston LH: **SWI6 protein is required for transcription of the periodically expressed DNA synthesis genes in budding yeast.** *Nature* 1992, **357:**505-508.

50. Cosma MP, Tanaka T, Nasmyth K: **Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter.** *Cell* 1999, **97:**299-311.

51. Wyrick JJ, Young RA: **Deciphering gene expression regulatory networks.** *Curr Opin Genet Dev* 2002, **12:**130-136.

52. Vilo J, Brazma A, Jonassen I, Robinson A, Ukkonen E: **Mining for putative regulatory elements in the yeast genome using gene expression data.** *Proc Int Conf Intell Syst Mol Biol* 2000, **8:**384-394.

53. Sinha S, Tompa M: **Discovery of novel transcription factor binding sites by statistical overrepresentation.** *Nucleic Acids Res* 2002, **30:**5549-5560.

comment

reviews

reports

deposited research

refereed research

interactions

information