Title
Reconstructing dynamic regulatory maps

Permalink
https://escholarship.org/uc/item/45g6b6vh

Journal
Molecular Systems Biology, 3(1)

ISSN
1744-4292

Authors
Ernst, Jason
Vainas, Oded
Harbison, Christopher T
et al.

Publication Date
2007

DOI
10.1038/msb4100115

Peer reviewed

Reconstructing dynamic regulatory maps

Jason Ernst[1], Oded Vainas[2], Christopher T Harbison[3,5], Itamar Simon[2] and Ziv Bar-Joseph[1,4,*]

[1] Machine Learning Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, [2] Department of Molecular Biology, Hebrew University Medical School, Jerusalem, Israel, [3] Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA, USA and [4] Department of Computer Science, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[5] Present address: Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ 08543, USA
* Corresponding author. Machine Learning Department, School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA.
Tel.: +1 412 268 8595; Fax: +1 412 268 3431; E-mail: zivbj@cs.cmu.edu

Even simple organisms have the ability to respond to internal and external stimuli. This response is carried out by a dynamic network of protein–DNA interactions that allows the specific regulation of genes needed for the response. We have developed a novel computational method that uses an input–output hidden Markov model to model these regulatory networks while taking into account their dynamic nature. Our method works by identifying bifurcation points, places in the time series where the expression of a subset of genes diverges from the rest of the genes. These points are annotated with the transcription factors regulating these transitions resulting in a unified temporal map. Applying our method to study yeast response to stress, we derive dynamic models that are able to recover many of the known aspects of these responses. Predictions made by our method have been experimentally validated leading to new roles for Ino4 and Gcn4 in controlling yeast response to stress. The temporal cascade of factors reveals common pathways and highlights differences between master and secondary factors in the utilization of network motifs and in condition-specific regulation.
Molecular Systems Biology 16 January 2007; doi:10.1038/msb4100115
Subject Categories: computational methods; chromatin & transcription
Keywords: dynamics; hidden Markov models; regulatory networks

Introduction

Understanding the dynamic programs that a cell utilizes in response to internal or external stimuli is an important challenge. These programs activate regulatory networks controlled by several transcription factors (TFs) (Harbison *et al*, 2004) and can involve a large number of genes (Natarajan *et al*, 2001). Direct information about this process has been obtained from genome-wide chromatin immunopre-cipitation (ChIP-chip) experiments and comparative motif studies that have been carried out to identify some of the regulators involved (Hahn *et al*, 2004; Harbison *et al*, 2004; Xie *et al*, 2005; Workman *et al*, 2006). Time-series microarray expression experiments are a complementary source of data, providing dynamic information about the expression of thousands of genes that are activated or repressed in response to stimuli such as environmental stress (Gasch *et al*, 2000).

An extensive literature has accumulated on methods to analyze and model time-series gene expression data. Some of these methods have focused on determining a continuous representation of the time-series expression data using splines (D'haeseleer *et al*, 1999; Bar-Joseph *et al*, 2003a). Other methods have focused on clustering genes while taking into account gene expression dynamics using methods such as auto-regressive equations (Ramoni *et al*, 2002), hidden Markov models (HMMs) (Schliep *et al*, 2003), and template-based methods (Ernst *et al*, 2005). Others have modeled time-series gene expression data using techniques such as differential equations (Chen *et al*, 1999), dynamic bayesian networks (Kim *et al*, 2003), and singular value decomposition (Holter *et al*, 2001). These methods, although useful, only provide a partial view of the transcriptional regulation process as they do not explicitly integrate information about TF–gene interactions such as ChIP-chip and motif data.

Most methods that integrate gene expression data with motif or ChIP-chip data do so without explicitly taking into account the dynamic nature of biological systems. A number of these methods combined a large number of expression data sets and motif data to infer transcription modules (Pilpel *et al*, 2001; Ihmels *et al*, 2002; Segal *et al*, 2003). Transcriptional modules are subsets of TFs and genes, such that genes in the same module tend to be similarly expressed and regulated by the same TFs across a number of experimental conditions. Bar-Joseph *et al* (2003b) inte-grated ChIP-chip data with expression data with a similar objective. Das *et al* (2006) presented a method that com-bined human expression data and motif information to identify active motifs, combinations of motifs and target genes under certain conditions. Although these prior methods provided important insights and often used time-series expression data sets, they did not take advantage of the sequential ordering of time points in an expression experi-ment, essentially treating time-series and static expression data in the same way.

A few recent methods have been proposed to integrate time-series expression data with ChIP-chip or motif data while taking into account the ordering of experiments in time-series data sets. For instance, time-series expression data were used to determine which genes were active at certain phases and then combined with ChIP-chip data using a trace-back algorithm to identify active TFs at these phases (Luscombe *et al*, 2004). This method in effect identified an ordered series of static regulatory graphs, but its direct connection with the dynamics of observed gene expression patterns is less clear. Other methods have relied more heavily on individual gene expression profile dynamics. For instance, Kundaje *et al* (2005) forms independent clusters of genes by using a joint probabilistic model for the dynamics of time-series expression profiles of genes and the motifs in their promoter regions. Others have integrated time-series expression data with ChIP-chip data to model the expression of individual genes (Lin *et al*, 2005) and interactions among TFs (Cokus *et al*, 2006) applying their techniques to model the cell cycle. Another method (Bonneau *et al*, 2006) used kinetic equations based on the time-series expression data to associate TFs with subsets of genes across a subset of experimental conditions.

Our objective is different from that of these prior works. We present a computational method that integrates the time-series expression data and ChIP-chip or motif information to infer an annotated *global* temporal map. This map describes the main transcriptional regulatory events leading to the observed time-series expression patterns and the factors controlling these events during a cell's response to stimuli. Our method focuses on bifurcation events. Bifurcation events occur when sets of genes that have roughly the same expression level up until some time point diverge (see Figure 1). Modeling expression patterns as results of a series of bifurcation events is consistent with a multilayer hierarchical model of gene regulation previously suggested for some organisms (Balázsi *et al*, 2005). Our method attempts to both detect these bifurcation events and explain them in terms of regulation by TFs. By focusing on detecting and explaining bifurcation events, we can determine the time when TFs are exerting their influence. The method also assigns genes to paths in the map based on their expression profiles and the TFs that control them. The model we use to learn these maps is based on an instance of an input–output hidden Markov model (IOHMM) (Bengio and Frasconi, 1995), where the ChIP-chip or motif data are the input and the observed expression data are the output.

We applied our method to study several stress responses in yeast. Our method was able to automatically infer many aspects of the temporal responses, some of which were previously known whereas others were new predictions. These new predictions range from low-level predictions
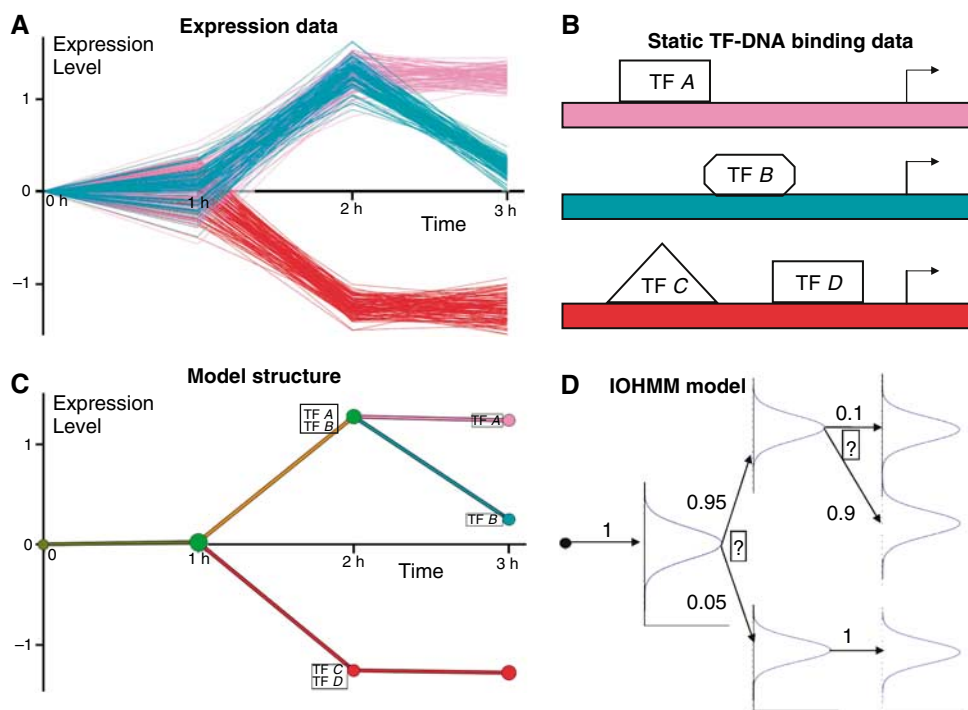

Figure 1 Model overview. (**A**) Plots of time series expression profiles generated to illustrate the model. (**B**) Static TF-DNA binding data—DREM integrates TF-gene regulatory relationships derived from ChIP-chip or motif data with the time series expression data. For this example a majority of the pink genes in (A) are regulated by TF *A*, the blue genes by TF *B* and the red genes by TF *C* and *D*. (**C**) The model structure inferred by DREM for the data in (A) and (B). After the model is derived genes are assigned to their most likely paths based on their expression profile as well as on the set of TFs that regulate them. TF labels appear on some of the paths out of splits. (**D**) IOHMM model—each state has a Gaussian emission distribution for the expression values and the transition probabilities for a gene depend on the set of TFs that regulates it. A logistic regression classifier (Krishnapuram *et al*, 2005) maps the set of regulating TFs to transition probabilities. The classifiers are denoted by question marks in the figure. Example transition probabilities are given for a gene which is regulated by TF *B*. These probabilities are greater for the states with distributions similar to those of TF *B* regulated genes. The TF information also affects the structure of the resulting IOHMM model. Based on this information some splits can be added and some splits are removed from the model.

regarding the timing of specific interactions to mechanistic predictions about the set of TFs controlling recovery from stress to predictions related to phenotypic changes. We have experimentally validated all types of these predictions leading to new roles for TFs in controlling yeast response to stress. We also used our temporal maps to compare different stress experiments and to identify a number of common control mechanisms. By using the time of activation that our method assigned to TFs, we were able to identify cascades of activators. Analysis of these cascades provides insights into the utilization of networks motifs and condition-specific regulation in response to stress.

Results

The Dynamic Regulatory Events Miner (DREM)

To combine time-series gene expression data and ChIP-chip or motif data, we extended an algorithm for learning IOHMMs (Bengio and Frasconi, 1995). IOHMMs are an extension of HMMs. HMMs have been used in the past to model sequential data including DNA and protein sequence data (Durbin *et al*, 1998) and to cluster gene expression data (Schliep *et al*, 2003). In HMMs as well as IOHMMs, hidden states are used to group genes by associating a cluster with each path through the hidden states over time. In our application, each hidden state is associated with one time point and represents a Gaussian distribution of the expression values for genes associated with it (Figure 1). IOHMMs extend HMMs in that they allow for an additional input set that does not necessarily change over time to control the transition probabilities of genes from state to state (Figure 1). In our application, we use static ChIP-chip experiments or motif data as this additional input information. Thus, our model encodes a mapping from ChIP-chip or motif data to the observed temporal expression values.

We constrain the transitions to enforce a tree structure among the hidden states. This allows us to model bifurcation events, places in the time series where subsets of genes that had similar expression values at prior time points diverge from each other. The identification of these events using an IOHMM is biased to those splits for which different sets of TFs regulate the divergent sets of genes. The algorithm searches over many possible treemap structures, training the parameters associated with the structures, and then scoring these structures using cross-validation to select the best one. The resulting set of hidden states and transitions between them leads to a global dynamic map. Each gene is then assigned to a specific path in the map based on its time-series expression data and ChIP-chip or motif data. Following this assignment, we compute association scores for TFs and splits using the hypergeometric distribution enrichment calculation. In the supplement, we discuss how to relax the tree constraints to allow paths to converge during recovery periods making the model more realistic (Supplementary Figure 14). For the results in the main text, we only used the sampled time points; however, as we show in Supplementary information, one can also use our model with interpolated values for time points that were not sampled (Supplementary Figure 15). Finally, we note that for the results in this paper, we limited the analysis to binary splits although the method also generalizes to higher order splits.

Complete details on the algorithm are given in the Methods section in Supplementary information.

A temporal map for amino-acid starvation response

To test the ability of the dynamic regulatory events miner (DREM) to learn dynamic maps from time-series and ChIP-chip data, we initially focused on the amino-acid (AA) starvation response pathway in yeast. As AAs are the basic structural components of proteins, yeast response to this stress by increasing AA synthesis and decreasing AA utilization is critical for its survival. For this condition, we have detailed time-series expression data (Gasch *et al*, 2000) as well as ChIP-chip experiments for 34 TFs (Harbison *et al*, 2004) (Supplementary Table 1). The time-series expression data were filtered to remove genes that did not change substantially at any time point (see Results section in Supplementary information). Figure 2A presents a temporal map derived by DREM using these two data sets. This map contained 11 unique paths controlled by a total of 15 TFs (using a TF split cutoff score of 0.001; see Materials and methods). In some cases, the same TF appears multiple times on the same path indicating multiple bifurcations with which the TF is significantly associated. As has been noted before, following AA starvation there is a massive transcription response involving both activation and repression. Our algorithm assigned to the first bifurcation two sets of TFs, Gcn4 and Cbf1 are associated with the genes that turn on upon starvation, whereas Fhl1, Sfp1, and Rap1 are associated with the repressed genes (Figure 2A and Supplementary Figure 1). These observations fit and expand current knowledge about this process. Gcn4 is indeed the master regulator of the AA starvation response (Hinnebusch and Fink, 1983; Natarajan *et al*, 2001), whereas Cbf1 has been previously associated with methionine biosynthesis (Baker and Masison, 1990). Our findings of Cbf1 association to the first bifurcation point suggest that it has a broader role in the cellular response to AA starvation. Indeed, genes regulated by Cbf1 and assigned to the higher path out of the first split were enriched not only for gene ontology (GO) (Ashburner *et al*, 2000) categories like sulfur AA metabolism (P-value $<10^{-9}$), which was previously noted (Thomas *et al*, 1992), but also for glutamate metabolism (P-value $<4 \times 10^{-6}$) and the more general AA metabolism category, with a lower P-value ($<7 \times 10^{-16}$). Gcn4 and Cbf1 are known to cooperate in the activation of the Met16 gene (O'Connell *et al*, 1995) and our finding of Gcn4 and Cbf1 as the two major activators in the AA response may suggest that such cooperativity is much more common. Indeed, we observe a significant overlap in genes bound by both Gcn4 and Cbf1 (P-value $<4 \times 10^{-8}$). Many of the TFs assigned to the second time point (secondary TFs) are known regulators of specific AA biosynthesis pathways, including Met4 as well as Arg81, Met32, and Dal82. Using GO, we determined that the set of genes assigned to the highest path after the first time point was enriched for AA metabolism (P-value $<9 \times 10^{-39}$) and subcategories such as glutamate biosynthesis (P-value $<10^{-13}$), sulfur AA metabolism (P-value $<6 \times 10^{-10}$), methionine metabolism (P-value $<2 \times 10^{-9}$), and arginine metabolism (P-value $<4 \times 10^{-9}$). Two other TFs assigned to this point,
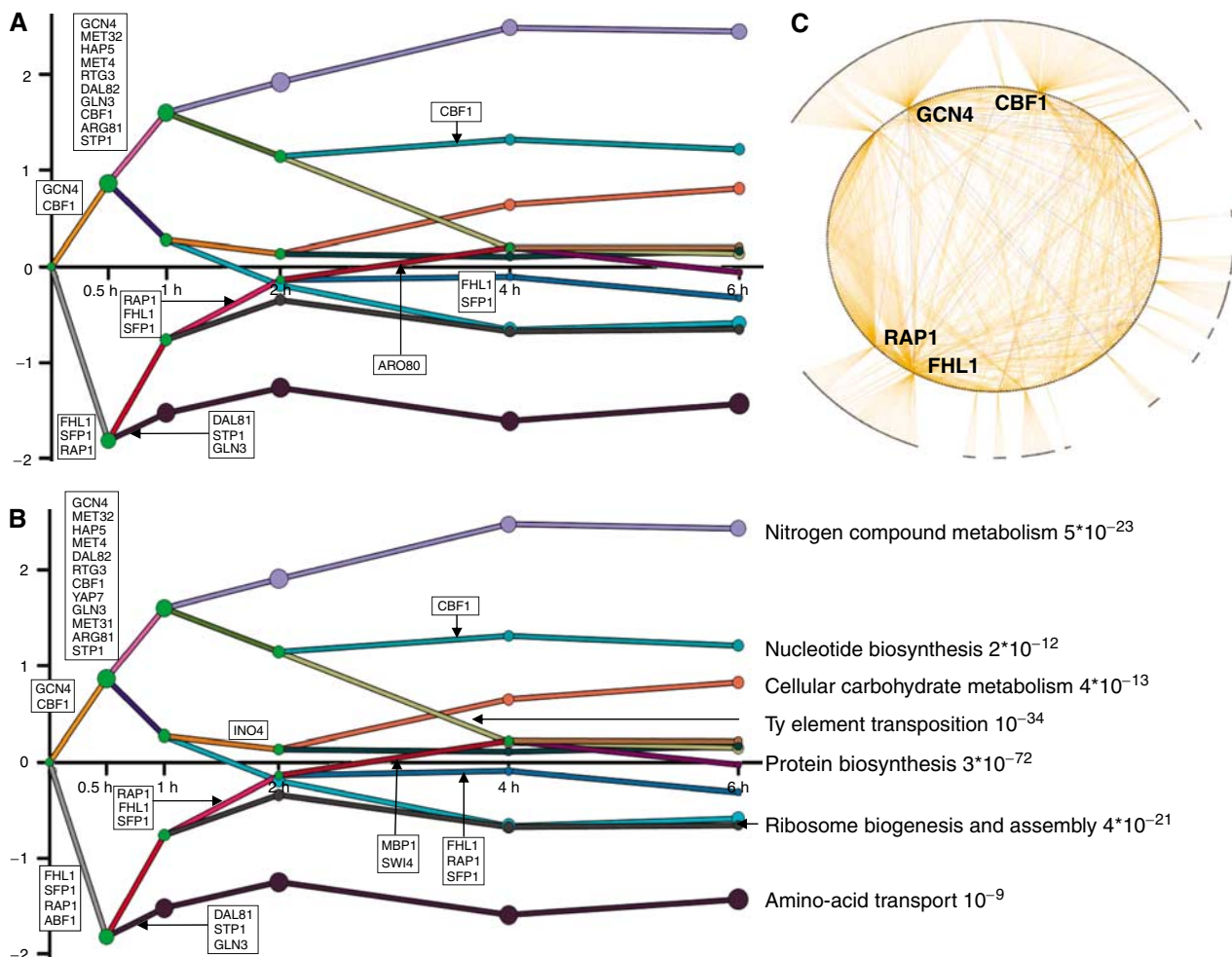
Figure 2 Dynamic regulatory map and static network for yeast response to AA starvation. (**A**) Dynamic map of yeast response to AA starvation using static input from condition-specific binding experiments and time-series expression data. TFs with split score below 0.001 appear next to the split they regulate, in ranked order of scores. Nodes in the graph represent hidden states. The area of a node is proportional to the s.d. of the expression of the genes assigned to that node. Green nodes represent split nodes. Many of the TFs were correctly assigned to the time points they are known to regulate. For example, Gcn4, which is a known master regulator of AA starvation response, is correctly assigned to the first split. Many of the TFs assigned to the second split regulate specific AA biosynthesis pathways. (**B**) Dynamic map of yeast response to AA starvation using input from both condition- and non-condition-specific ChIP-chip experiments. Several additional TFs not profiled with a condition-specific ChIP-chip experiment under the AA condition were determined to be participating in the response and recovery processes. These included Abf1, Swi4, Mbp1, and Ino4. In addition to identifying these TFs as potential participants in the response, DREM also identifies their time of influence. (**C**) Static regulatory graph for AA starvation. Nodes correspond to genes or TFs. An edge implies that the TF binds the gene with a *P*-value < 0.005 in an AA starvation ChIP-chip experiment. Blue edges represent interactions between TFs. Whereas some properties of the networks can be derived from the static representation, many of the dynamic aspects of the system are lost when not using the time-series data.

Rtg3 and Gln3, are activated upon starvation in a TOR pathway-related manner (Crespo *et al*, 2002).

The set of genes assigned to the repressed path out of the first split is highly enriched in categories such as ribosome biogenesis and assembly (*P*-value < 10^{-91}) and ribosome (*P*-value < 10^{-83}) consistent with what has been observed before for AA starvation response (Gasch *et al*, 2000). A majority of the ribosomal genes were determined to be bound by the ribosomal TFs Rap1, Fhl1, or Sfp1. It has been previously noted that Rap1 and Fhl1 remain bound to the promoter regions of ribosomal genes under environmental stress (Wade *et al*, 2004; Rudra *et al*, 2005). An additional TF, Ifh1, which has recently been implicated in having an important role in controlling the expression of ribosomal genes

under stress (Wade *et al*, 2004; Rudra *et al*, 2005) was not part of the set of TFs for which a ChIP-chip experiment was performed in AA starvation conditions.

Extending condition-specific dynamic maps using general binding data

Although the map derived by DREM provided explanations for several bifurcation points, others could not be explained using the limited set of TFs with ChIP-chip data under the AA starvation condition. We have hypothesized that some of these points could be explained by TFs, which were previously not known to be involved in this process and so were not originally profiled with ChIP-chip experiments in AA starvation condi-

tions. To test this, we have employed DREM again, this time augmenting the static input data with an additional 75 TFs profiled with ChIP-chip experiments in other conditions in yeast, primarily yeast complete growing media (YPD) (Harbison *et al*, 2004). To reduce false positives in this ChIP-chip data, we used a post-processed version of the data that also requires the presence of an evolutionarily conserved motif (see Materials and methods). The map derived from this additional data (Figure 2B) contained new TFs that explained a number of temporal events (Supplementary Tables 2 and 3). Some of these TFs were added to previously annotated time points. For instance, Yap7 was assigned to the set of secondary TFs on the top split. Abf1 was assigned to the first split and was determined to control genes that were downregulated early on. This agrees with previous studies that implicated Abf1 in being involved in regulating the ribosome (Della Seta *et al*, 1990; Planta 1997). Two G1 cell-cycle activators, Swi4 and Mbp1, now appear on a split starting at the 2-h point on the recovery path. This path was enriched for genes previously shown to be cell cycle regulated (Spellman *et al*, 1998) (*P*-value $<9 \times 10^{-6}$), where enrichment was based on all genes going into the split. Another factor, Ino4, which has been implicated in phospholipid biosynthesis (Klig *et al*, 1988) and more recently has been suggested to be a global regulator of gene expression (Santiago and Mamoun, 2003; Workman *et al*, 2006), was determined to be activating genes starting around the 2-h time point. Several of the genes in the path that DREM determined to be controlled by Ino4 were known lipid metabolism genes (GO *P*-value $<6 \times 10^{-5}$). The induction of Ino4-regulated lipid biosynthesis genes may be connected to the immediate need of membrane used for the autophagocytosis process. This process is utilized by yeast in order to regulate the equilibrium between proteins and the diminishing set of AAs due to the starvation condition (Petiot *et al*, 2002).

Validating interactions and mechanistic predictions

The temporal map derived by DREM suggests that Ino4 is activating genes as part of a recovery mechanism several hours after AA starvation. However, Ino4 had previously only been profiled with a ChIP-chip experiment in YPD media (Harbison *et al*, 2004). To validate the prediction of Ino4's role in AA starvation conditions, we first carried out ChIP–PCR experiments in which we checked the *in vivo* association of the Ino4 protein to several of its targets. We selected four of the genes bound by Ino4 in the ChIP-chip experiment in YPD media with a *P*-value <0.005 that were assigned to the path most significantly controlled by Ino4 (brown path in Figure 2B and genes in Figure 3A). We compared the occupancy of Ino4 in the promoter of these genes in synthetic complete + D-glucose (SCD) media and 4 h after AA starvation. As Figure 3B shows, for three of the four genes, the occupancy rate of Ino4 at 4 h after AA starvation is at least twice its rate in SCD media, suggesting that Ino4 indeed plays a role in regulating these genes during AA starvation, as predicted by DREM. To further characterize Ino4's increased activity in regulating genes in AA starvation, we carried out a genome-wide binding experiment for Ino4 in SCD conditions and at 4 h into AA starvation. As Figure 3C shows, Ino4 binds to many more genes in AA

starvation conditions as compared to its binding in SCD (125 more genes at a *P*-value <0.001 in at least one repeat and 66 more genes at a *P*-value <0.005; see also Table I). Of the 207 genes bound by Ino4 at a *P*-value <0.005, 34 were among the 422 genes assigned to the path that DREM identified as most associated with Ino4 (*P*-value $<2 \times 10^{-6}$). Furthermore, as the plots show (Figure 3D and Supplementary Figure 2), almost all of the genes assigned to this path that are bound by Ino4 in either AA starvation, SCD media, or both are more significantly bound in AA starvation conditions.

Temporal maps for the regulation of stress response in yeast

We next combined condition- and non-condition-specific binding data and used DREM to construct temporal regulatory maps for a number of other stress conditions (Figure 4A and Supplementary Figures 3, 5, 6, and 8). All of these had time-series expression data available (Gasch *et al*, 2000). The number of TFs for which we used condition-specific binding data varied from 28 for hydrogen peroxide (Supplementary Table 5), one for heat shock, and none for DTT and cold shock. All condition-specific binding data were obtained from Harbison *et al* (2004) except the heat-shock binding data, which were obtained from Hahn *et al* (2004). As in the AA starvation example above, we extended the input to include non-condition-specific binding data post-processed using motif data from Harbison *et al* (2004). Again, DREM was able to reconstruct a number of known cascades and suggest new regulatory roles.

For example, for heat shock, Hsf1 was identified as a 'master' regulator controlling the initial activation response, which is consistent with previous studies (Bonner *et al*, 1992). Msn2- and Msn4-regulated genes were also over-represented on the highest expressed paths of the heat-shock model. The set of genes assigned to the path that was still increasing at 10 min was overenriched for GO categories such as carbohydrate metabolism (*P*-value $<10^{-14}$), response to stress (*P*-value $<3 \times 10^{-9}$), and protein folding (*P*-value $<6 \times 10^{-7}$). All 15 of the protein folding genes assigned to this path were also bound by Hsf1.

For peroxide, Yap1 and Skn7 were two of the TFs correctly assigned to activated paths in the first bifurcation point, consistent with a previous report (Lee *et al*, 1999). Other TFs associated with regulating initially activated genes include Yap7, Rpn4, Msn2, Msn4, Gcn4, Aft2, and Put3. The genes on the initially activated path were enriched for GO categories such as aldehyde metabolism (*P*-value $<2 \times 10^{-9}$) and response to oxidative stress (*P*-value $<6 \times 10^{-9}$). Along the repressed paths are cell-cycle TFs Mcm1, Sum1, Swi5, and Swi6 and ribosomal TFs Fhl1 and Rap1, all of which were detected without hydrogen peroxide binding data (Supplementary Tables 6 and 7).

Using these temporal stress response maps, we looked for common mechanisms employed by yeast in response to stress.

Repressed pathways
While the identity of many of the activators varied depending on the stress condition, we observed two pathways that were generally repressed under the stress conditions we looked at.
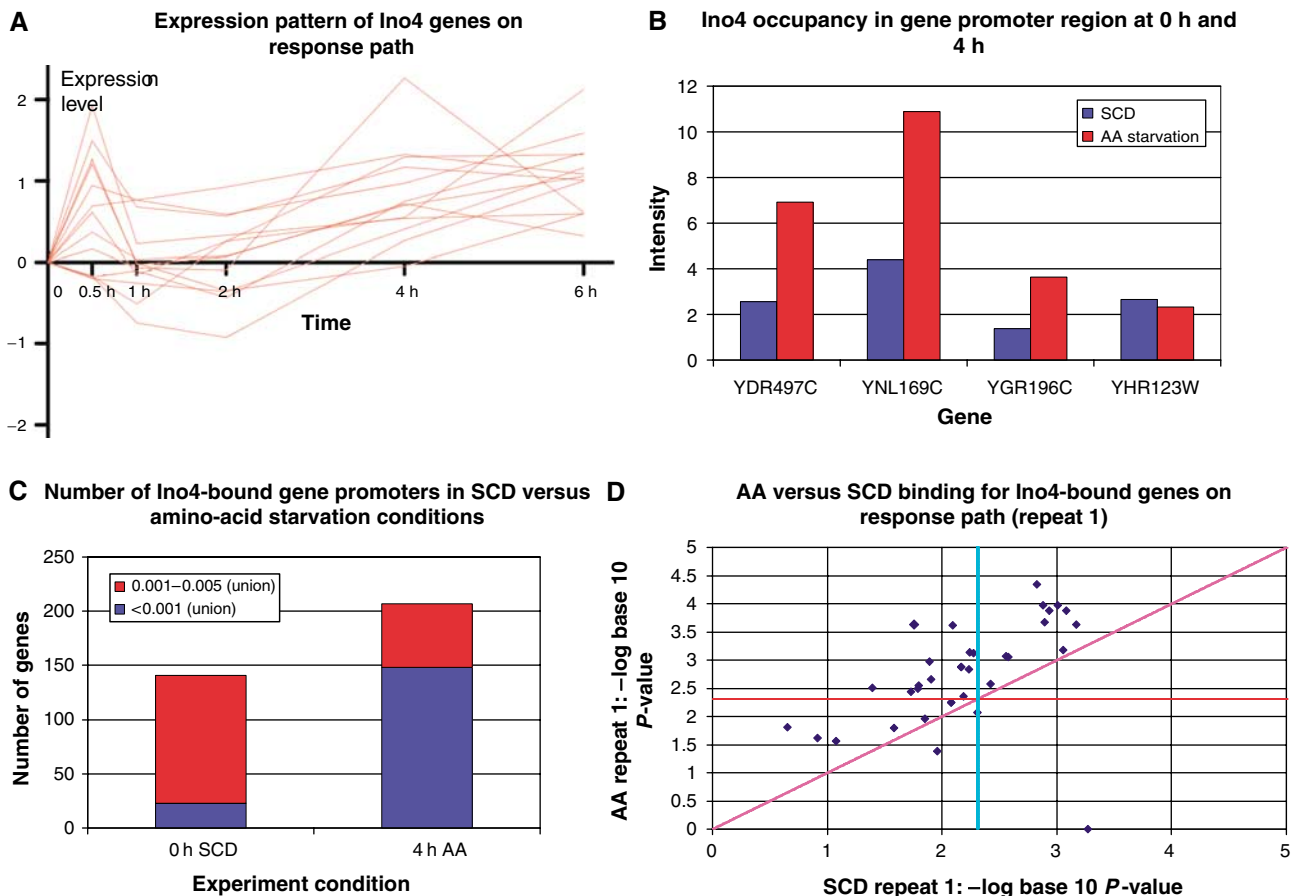
A Expression pattern of Ino4 genes on response path

B Ino4 occupancy in gene promoter region at 0 h and 4 h

C Number of Ino4-bound gene promoters in SCD versus amino-acid starvation conditions

D AA versus SCD binding for Ino4-bound genes on response path (repeat 1)

Figure 3 The role of Ino4 in regulating response to AA starvation. (**A**) Expression profiles of 13 genes in AA starvation that were assigned to the brown path in Figure 2B. These 13 genes were all bound by Ino4 in a ChIP-chip experiment in YPD media with a *P*-value < 0.005 and have an evolutionarily conserved Ino4 motif. It was predicted by DREM that Ino4 was activating these and other genes starting around 2 h (see also Supplementary Table 4). (**B**) Occupancy rates of Ino4 in the promoter region of four genes regulated by Ino4, before and at 4 h after AA starvation. For three of these four genes, the Ino4 promoter occupancy rates were at least two-fold higher following AA starvation than in synthetic complete + D-glucose (SCD) media before AA starvation. (**C**) Comparison of the number of genes bound by Ino4 before and 4 h after AA starvation using a whole-genome binding experiment. We compared the lists using two different *P*-value cutoffs (0.001 and 0.005). Genes were counted if they are bound at the appropriate *P*-value in at least one of the two repeats. At the 0.001 *P*-value cutoff, there is almost a six-fold enrichment for Ino4-bound genes 4 h after AA starvation. (**D**) Comparison of binding *P*-values for genes assigned to the main path determined by DREM to be regulated by Ino4 in one of the repeats (see also Supplementary Figure 2). The plots are the negative log base 10 of the binding *P*-value for genes that were bound with a *P*-value < 0.005 in one or more of the Ino4 binding experiments and are on the identified Ino4 response path. The horizontal and vertical lines represent a *P*-value significance of 0.005. Anything to the right of the vertical line is significant under normal growth conditions. Anything above the horizontal line is significant in the AA starvation experiment. Anything above the diagonal line is more significant in the AA starvation experiment. This plot indicates that these genes were bound more significantly in AA starvation conditions than SCD conditions.

The first pathway showed a similar pattern of repression and recovery in all of the reconstructed temporal regulatory maps except for the map for cold shock. This pathway included the ribosomal genes and their primary TFs (Rap1, Sfp1, and Fhl1). These genes are repressed steeply and quickly. However, these genes also recovered quickly approaching their pre-treatment levels. In cold shock, the ribosomal genes were assigned to an activated path consistent with a previously made observation (Gasch *et al*, 2000; Supplementary Figures 6 and 7).

Another common repressed pathway that was observed in AA starvation, heat shock, and cold shock was a pathway controlled by Swi4, Swi6, and Mbp1. This pathway primarily contained cell-cycle genes. For example, in the heat-shock pathway, there was a particularly strong enrichment for G1 cell-cycle genes (Spellman *et al*, 1998) (*P*-value < 4×10^{-20}). In comparison to the ribosomal genes, cell-cycle genes were

repressed at a slower rate and to a less significant level (Supplementary Figure 4). However, when they recover, they were expressed at a higher level than their initial (time point 0) value (Figure 4B). A possible explanation to the slow repression of the cell-cycle genes, which is followed by a strong activation, is that what we are actually seeing is the well-documented stress-related cell-cycle arrest. In an unsynchronized culture, the downregulation of G1 genes should be gradual as it takes time until all cells leave the G1 phase. On the other hand, in the recovery stage, the culture is relatively synchronized and therefore the G1 genes reach higher levels than in unsynchronized culture. To test the prediction that this pathway in heat shock indeed represents differences in the cell-cycle phase distribution, we counted the budding index of cells following heat shock. As Figure 4C shows, 20–40 min after heat shock, cells enter S-phase arrest (increase in the

Table I Experimental results for Ino4 binding in amino-acid starvation

Ino4-binding experiment	Intergenic region <0.001	Genes <0.001	Intergenic region <0.005	Genes <0.005
SCD repeat 1	16	21	88	114
AA repeat 1	68	89	122	153
Increase in AA repeat 1	325%	324%	39%	34%
SCD repeat 2	5	6	78	94
AA repeat 2	46	59	109	147
Increase in AA repeat 2	820%	883%	40%	56%

Summary of binding in Ino4 genome-wide ChIP-chip binding experiments. The table gives the number of intergenic regions bound and the number of associated genes at both the 0.001 and 0.005 *P*-value significance levels for two repeats. Only intergenic regions with assigned genes are included in the intergenic region counts. The table also gives the percentage increase in AA conditions compared to that in SCD media, which is the condition before starvation. These percentages indicate the increased activity of Ino4 at 4 h into AA starvation.
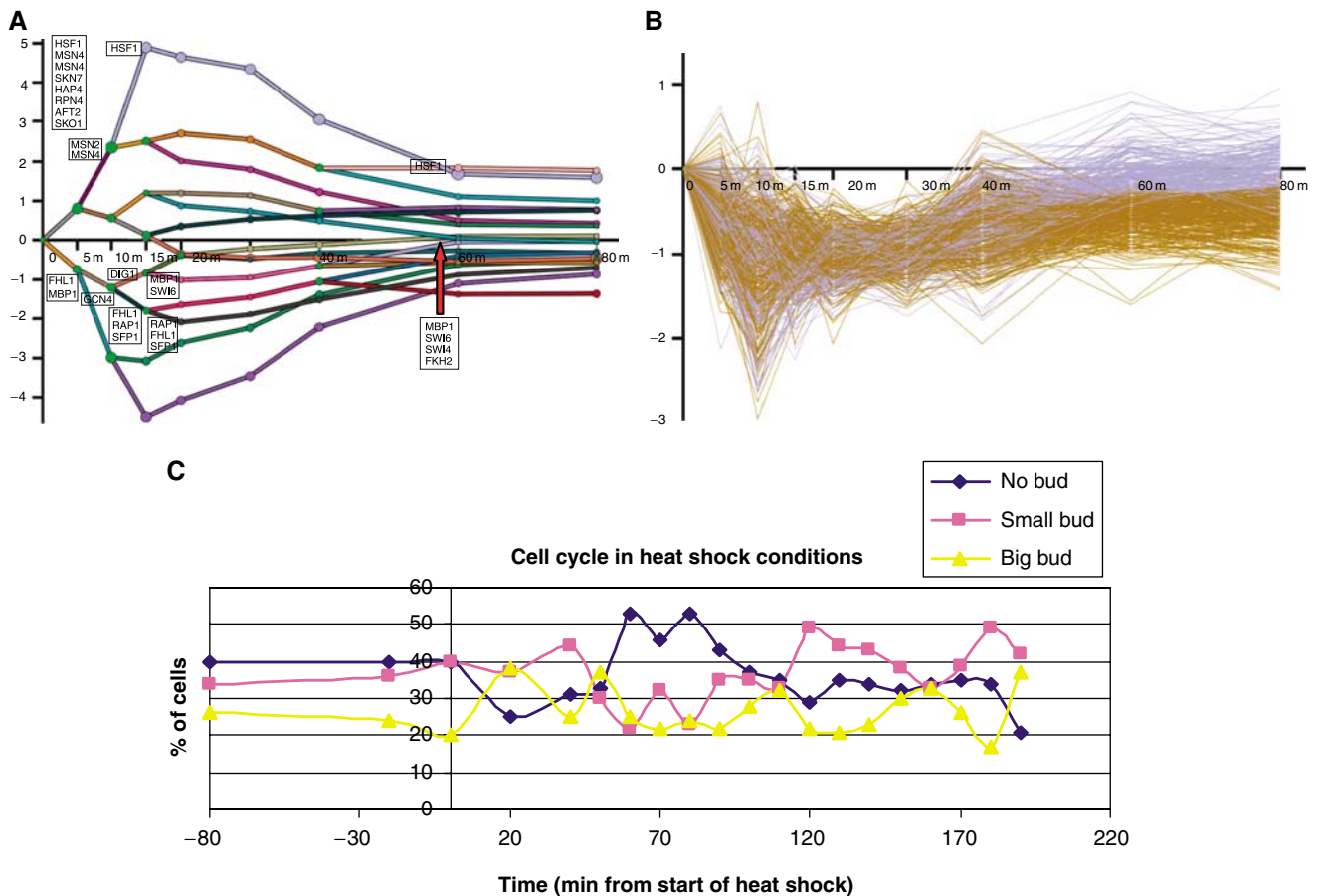

Figure 4 Cell-cycle recovery from stress. (**A**) Dynamic regulatory map derived for heat shock. As part of the recovery process, several genes regulated by Swi4, Swi6, Mbp1, and Fkh2 increase their expression level, reaching a level higher than their original (nontreated) values. Many of these genes are G1 genes (*P*-value $< 4 \times 10^{-20}$ based on a set of 300 G1 genes from Spellman *et al*, 1998). (**B**) Expression profiles of genes assigned to the bifurcation event at the 40 min time point. Blue profiles represent genes assigned to the upper path of this bifurcation node. This path was controlled by the above cell-cycle TFs. Note that the expression of many of these genes rises above their initial expression value starting at the 60 min time point. (**C**) Budding index before and after heat shock. Cells are initially arrested and as predicted by DREM the percentage of G1 cells peaks starting at the 60 min time point. Following that peak, cells resume their cell-cycle activity in a more synchronized manner.

fraction of cells with small buds). Later, at around 50 min, the cell cycle resumes. However, the stress causes partial synchronization of the cells (note the higher percentage of cells in G1 60 min after heat shock compared to unsynchronized cells). These results agree well with the predictions made by DREM. The increase in the expression values of the Swi4–Swi6 controlled path at the 60 min time point is the result of the increase of synchronization. Following

this, cells become partially synchronized, at least for the next 120 min.

Master regulators and condition-specific regulation in response to stress
Although the identity of the activators varied, a few activators were identified to have a much more significant control of the

initial change in expression levels in response to AA starvation (Gcn4, Cbf1, Rap1, Fhl1, and Sfp1) and heat shock (Hsf1, Msn2/4, and Skn7). This can be seen most clearly in the AA map where the majority of the activators in these conditions were activating genes in later time points. This type of response can result from cells that constitutively express a small number of master regulators so that they can react quickly to stress whereas the later TFs are only expressed as part of the response process. This is consistent with previous studies. For example, Msn2 and Msn4 are regulated at the level of nuclear exclusion (Gorner *et al*, 1998), Cbf1 also already exists in the cell before starvation and its transcript level is not affected by methionine starvation (Mellor *et al*, 1990), and

Gcn4 and Hsf1 are found in association to a subset of their target gene promoters even before stress (Hahn *et al*, 2004; Harbison *et al*, 2004). These differences are also apparent in the initial expression levels of the regulators. As Figure 5C shows, during the first two time points, where their impact on the genes they regulate is the highest, master regulators do not drastically change their own expression when compared with pretreatment levels. In contrast, the expression levels of most secondary TFs increase, indicating that many are transcriptionally regulated.

To further study this point, we looked at the condition-specific regulation activities of TFs in AA starvation and heat shock by dividing the TFs into two groups: the first contained
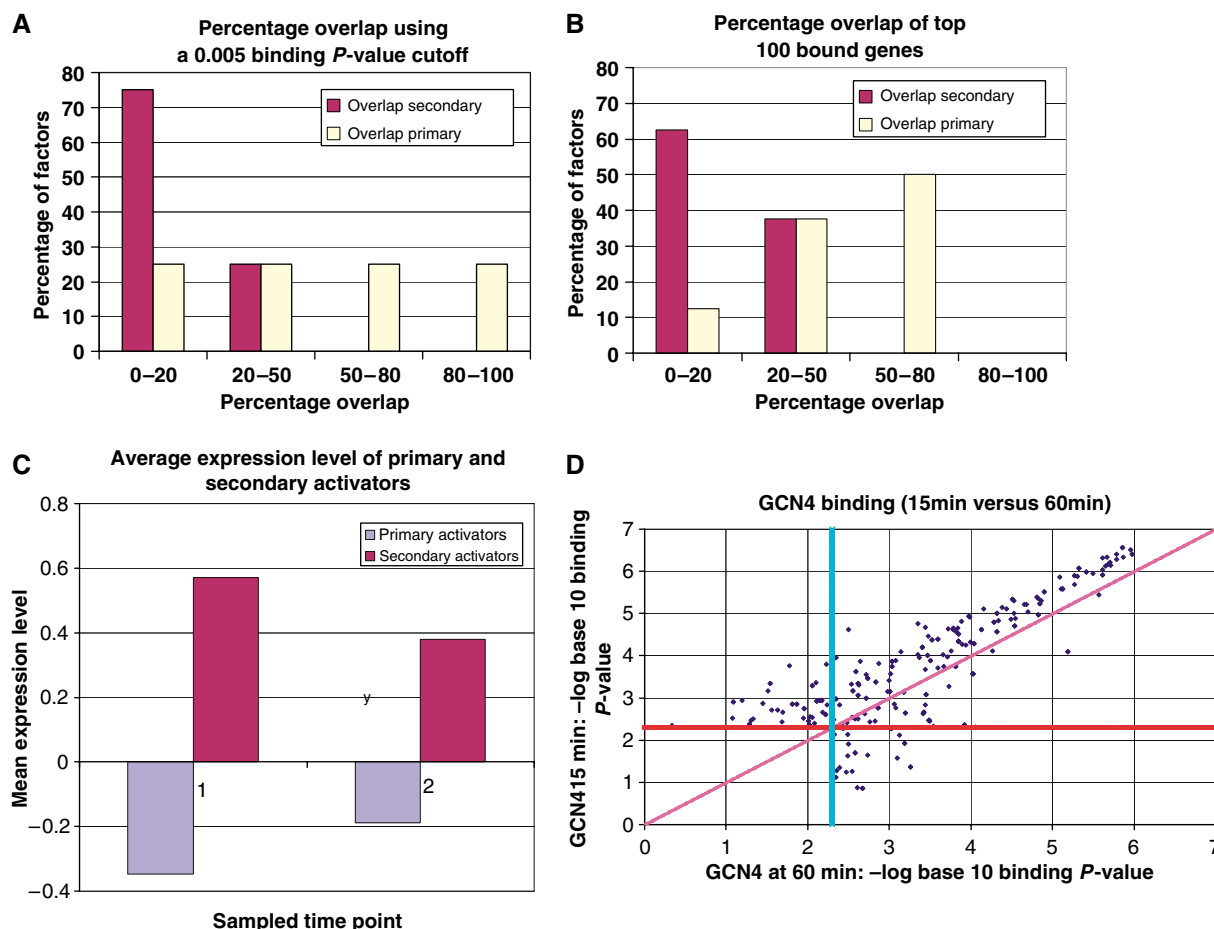

Figure 5 Distribution of binding overlap between YPD media and stress condition for different sets of TFs. We group TF into two sets. The first subset contains the TFs assigned to first split points in AA starvation or heat that were profiled with a ChIP-chip experiment in the condition (8, primary TFs Cbf1, Gcn4, Fhl1, Rap1, and Sfp1 from AA starvation in Harbison *et al* (2004) and for heat Msn2 and Skn7 from Harbison *et al* (2004) and Hsf1 from Hahn *et al* (2004)) and the secondary TFs assigned to second split point profiled with a ChIP-chip experiment in the condition (8, secondary TFs Arg81, Dal82, Gln3, Hap5, Met32, Met4, Rtg3, and Stp1 from AA starvation in Harbison *et al*, 2004). (**A**) Percent overlap for each of these two sets when binding (under both stress and YPD media conditions) is determined using a 0.005 *P*-value cutoff. Note the difference between the distribution of overlap for primary and other TFs. Whereas the majority of TFs display a big difference in the set bound genes under stress and YPD media, many primary TFs bind to a large percentage of stress-regulated genes in YPD media as well. This difference is even bigger in (**B**) where we plot the overlap for the top 100 genes (ordered by their binding *P*-values) in each condition. Whereas most TFs (and most secondary TFs) drastically alter the subset of genes they regulate under stress, half of the primary factors bind to more than 50% of the same genes in both conditions. Although the binding strength may be different under stress, these results indicate that many of the primary pathways are maintained, in low levels, under YPD media as well. (**C**) Average expression levels for primary and secondary factors for the first two time points in the AA starvation and heat-shock experiments. Whereas the average expression levels for the secondary factors are much higher when compared to their untreated levels, the levels for the primary factors do not change significantly between the two conditions. (**D**) Comparison of binding *P*-values for Gcn4 in the MMS condition at 15 and 60 min. Points above the diagonal correspond to genes that were bound more significantly at 15 min than at 60 min. As can be seen from the plot, a substantial majority of genes were bound more significantly at 15 min than at 60 min. Points to the right of the vertical had a *P*-value <0.005 in MMS at 60 min, whereas points above the horizontal line had a *P*-value <0.005 in MMS at 15 min.

TFs that were determined to control the initial response and the second contained the rest of the TFs. We compared the binding targets of these TFs under YPD media and in the condition that they regulate. As part of the response to stress, several yeast TFs begin to regulate genes that were not regulated by them in YPD media (Bar-Joseph *et al*, 2003b; Luscombe *et al*, 2004). As can be seen in Figure 5A, regulators controlling the first bifurcation point (master regulators) showed a much larger overlap in the set of genes bound in stress and YPD media compared to the TFs regulating genes at later time points (secondary regulators). Whereas six of the eight TFs assigned to a first point had more than 20% overlap between condition-specific and YPD media binding data, and four had more than 50% overlap, six of the eight TFs first appearing at a second split had less than a 20% overlap and none had more than 50% overlap (Supplementary Tables 9 and 10). These results hold even if we ignore the actual *P*-values and focus on the list of 100 top bound genes in each condition (Figure 5B). Again, these differences indicate that, at least for some types of stress, cells maintain the ability to quickly activate the initial response pathways.

In addition to differences in the condition-specific activity between master and secondary regulators, we have also observed differences in the utilization of different network motifs. As Supplementary Figures 9 and 10 show, genes bound by master regulators in a feed forward loop (FFL) displayed consistently higher expression levels when compared to genes regulated by the same TFs in a multiple input (MI) or single input (SI) motifs. In contrast, for many secondary TFs, we have not observed a large difference between the expression levels of FFL-controlled genes and genes controlled by the other two network motifs. The ability of master regulators to utilize FFLs by consistently expressing some of the genes at a higher level during a response may help cells fine-tune their response to various stresses. Indeed, whereas 45% of the genes bound by Gcn4 in an FFL are known AA biosynthesis genes (based on GO), only 21 and 13% of the genes bound in an MI or an SI, respectively, are assigned to this category. Thus, although many genes are activated initially, only a few of them will remain expressed at a high level in a later point as their expression requires the additional binding of a secondary factor. In this way, an FFL serves as a filtering motif, removing signals that were erroneously activated and maintaining those that are required for the actual response pathway (Mangan and Alon, 2003).

Determining the activation time of regulators

As the results above indicate, most TFs that are activated during stress either change or expand the set of genes they regulate. To identify these new sets, ChIP-chip experiments are often used to determine the roles of several TFs in various response pathways (Bar-Joseph *et al*, 2003b; Harbison *et al*, 2004; Workman *et al*, 2006). However, even when TFs are known, or suspected, to be involved in such pathway, the actual time in which they are activated may vary. Master regulators are activated early on, whereas secondary regulators are activated later. The ability of DREM to determine a time point for carrying out such experiment can help in accurately recovering the role a factor plays in a response

pathway. To study this, we have looked at the activation of Gcn4 as part of the response to methyl-methanesulfonate (MMS) stress in yeast. In a previous study, it was determined, using an experiment 60 min after the induction of stress, that Gcn4 did not expand the set of genes it regulates when compared to the set it regulates in YPD media (Workman *et al*, 2006). We used DREM to reconstruct a dynamic map for this system (Supplementary Figure 8). The DREM map inferred for this condition made two predictions about Gcn4: (1) that Gcn4 was expanding the set of genes it regulates at the 15 min time point when compared to YPD media and (2) that Gcn4 binding would likely be less intense at 60 min as the expression of the main Gcn4-controlled paths decreased at that time point. To test whether the temporal predictions of DREM were correct, we carried out genome-wide binding experiments at three time points: 0 (YPD media), 15, and 60 min. As predicted by DREM, we found a large expansion in the set of genes regulated by Gcn4 at the 15 min time point. Whereas 45 genes were bound by Gcn4 in YPD media (using a 0.005 *P*-value cutoff), 235 genes were bound at the 15 min time point and a smaller number (212 genes) at the 60 min time point (Supplementary Table 8). In addition, for the vast majority of Gcn4-bound genes, the binding *P*-value was more significant at the 15 min time point when compared to the 60 min point, indicating that Gcn4 is indeed more active earlier in the response as predicted by DREM (Figure 5D). We note that our results, indicating that Gcn4 is expanding the set it regulates even at the 60 min time point, differ from previously reported results (Workman *et al*, 2006). There could be several reasons for this discrepancy, the most likely source is the noise associated with any high-throughput experiment. However, the fact that our 60 min results agree well with the 15 min results, with the expression data, and with the findings of a previous study (Natarajan *et al*, 2001) indicates that Gcn4 is indeed activating genes as part of the MMS-response pathway.

Verifying the advantage of integrating time-series expression and ChIP-chip data

To verify the advantage gained from integrating time-series expression data and ChIP-chip data, we tested whether either data alone could have reproduced results similar to those obtained when combining them. First, we generated a randomized version of the AA ChIP-chip data by randomizing the genes each TF was bound to while holding the number of genes bound by each TF fixed. We then applied DREM to this randomized ChIP-chip data and the original AA gene expression data. This procedure resulted in maps that had no, or very few, TF labels (Supplementary Figure 11). Specifically, we found that activators that were determined to be 'master' regulators using the real binding data were not assigned to the first split using the randomized values and most of the AA biosynthesis regulators were not assigned to any of the splits. Second, we applied an HMM model to the time-series data without using ChIP-chip while still enforcing the same tree structure requirements on the hidden states. To compare the HMM model with the IOHMM model of Figure 2B, we did use the ChIP-chip data as a post-processing step to score the TFs at each split. We found that most of the relevant TFs were

substantially more significant with an IOHMM than with an HMM (Supplementary Figure 12). To further validate that the grouping of genes were more biologically meaningful when using the IOHMM compared to an HMM, we performed GO enrichment comparisons and found more GO categories to be enriched in an IOHMM model (Supplementary Figure 13). Combined, these results demonstrate the importance of the ChIP-chip data for determining an accurate regulation model. To demonstrate the importance of the time-series data for inferring regulatory models, as opposed to ChIP-chip data alone, we present a regulation graph in Figure 2C. This graph uses only the ChIP-chip AA-binding data. Although one can rank TFs based on the number of genes they regulate, it is clear from the figure that it would be very hard to infer the dynamics of the response process from the ChIP-chip data alone.

Discussion

The recent availability of expression and ChIP-chip and motif data has led to a number of efforts aimed at reconstructing regulatory networks. To date, these efforts primarily focused on determining a static graph representation of the underlying network. These efforts have led to many insights regarding the overall organization of networks (Barabasi and Oltvai, 2004), network motifs (Milo *et al*, 2002), and the set of interactions in various biological systems (Pilpel *et al*, 2001; Ihmels *et al*, 2002; Bar-Joseph *et al*, 2003b; Segal *et al*, 2003; Workman *et al*, 2006).

The computational method we presented in this paper, DREM, takes a different approach providing a global dynamic view of the gene regulation. This approach has a number of advantages when compared to methods that derive static graphs. First, biological systems are dynamic. TFs may bind different genes at different time points. Thus, the ability of DREM to derive dynamic maps that associate TFs with the genes they regulate and their activation time points may lead to better insights regarding the system being studied. These insights may include the identification of master regulators that control the initial response and secondary regulators that are responsible for more specific pathways. It may also help explain several aspects of the observed response including the condition-specific activity of factors and the activation of certain network motifs. As timing is available in these maps, some of the paths and the factors regulating them may be linked to predictions regarding the timing of specific phenotypes. Second, many TFs are post-transcriptionally regulated. For these TFs, it is hard to determine an activation time when using only their expression data. When studying biological systems using ChIP-chip methods, researchers rely on previous knowledge and other data sources to determine which factors to profile under the condition of interest (Bar-Joseph *et al*, 2003b; Hahn *et al*, 2004; Harbison *et al*, 2004; Workman *et al*, 2006). DREM's ability to use general motif data or ChIP-chip data from other experimental conditions for deriving temporal regulatory maps presents a useful complementary approach for determining TFs on which to study with a ChIP-chip experiment. As we have shown for Ino4, these predictions may lead to new regulatory roles for some of the factors. Importantly, for these factors, DREM also indicates

a time point at which these ChIP-chip experiment should be carried out. Determining the right point leads to a more accurate set of regulators as we have shown with Gcn4 in MMS. Finally, we note that although we presented DREM in the context of analyzing a large number of stress-response data sets, it can be applied equally well to study a single condition of interest.

The accuracy of the models generated by DREM and their predictive power is another indicator to the importance of data integration. As was noted in the past (Jansen *et al*, 2003), each data source provides only a partial view of the activity in the cell. By integrating diverse data sets, we can improve over the results obtained by each datum on its own. For example, in the context of clustering expression data, a key question is the number of clusters to use. Another problem relates to noise and the small number of time points measured. DREM addresses these problems by integrating ChIP-chip or motif data with the time-series data. This leads to more natural derivations of clusters based on bifurcation points and improves the resulting clusters as we showed using GO (Supplementary Figure 13).

Like any other computational method, DREM is highly dependent on the input data. DREM relies on the availability of high-quality time-series expression data. Here, the sampling rate may play an important role in the ability of DREM to derive accurate regulatory maps. For example, although we observed initial activation by a few master regulators in a number of different conditions, the initial response to peroxide was determined to be controlled by nine TFs. This may be the result of the sampling rate. If this rate is too low, regulatory effects may be aggregated in some of the time points, preventing DREM from associating TFs with their correct activation time. Determining the appropriate sampling rate is an important problem (Simon *et al*, 2005). In some cases, DREM can be used to identify a problem with the sampling rate that has been used and to suggest places in which more samples are needed. However, even in cases in which an experiment is well sampled, TF labels can aggregate at earlier time points, as assignments of genes to paths is based on all time points.

The models derived by DREM are currently limited to tree structures with the option to also model convergence of paths from a common split. Although this is motivated by previous biological observations (Balázsi *et al*, 2005), in some cases it may be more natural to allow other types of path merges and resplits. DREM also does not explicitly model regulation through other mechanisms, such as chromosome remodeling and mRNA degradation. Transition probabilities in DREM are computed using logistic regression, which does not capture all types of combinatorial interactions. Another limitation of DREM is that the output dynamic map model is sometimes chosen from a number of possible dynamic maps with similar scores. However, when this happens, these different maps usually share most of the important splits.

Although DREM was applied to learn networks in yeast, the growing availability and diversity of ChIP-chip (Cam *et al*, 2004), motif (Xie *et al*, 2005), and time-series expression data make it possible to use DREM for many different species such as human, mouse, and flies, among others. The ability to derive dynamic networks from such data may lead to new

Materials and methods

Pre-processing input data

We first generated a binary matrix of predictions of TF–gene regulatory interactions. For predictions based on condition-specific ChIP-chip data, a '1' was encoded if the binding *P*-value of the TF to the gene's promoter region was <0.005, otherwise a '0' was encoded. For TFs without condition-specific ChIP-chip data, we followed a version of the regulatory code of Harbison *et al* (2004). For these TFs, a TF–gene pair had a '1' encoded in the matrix if the TF bound the promoter of the gene with a *P*-value <0.005 in at least one ChIP-chip experiment and there was a motif in its promoter that was evolutionarily conserved in at least two other yeast species. If no ChIP-chip data were available for a gene, then a '0' was encoded for all entries. All time-series data were transformed to start at '0' so that the value at each time point represents the log ratio change from an unstressed control. Genes were filtered if there was more than one missing value or if the gene did not change sufficiently at any time point (see Resutls section in Supplementary information).

Dynamic regulatory events miner algorithm

Each state of the probabilistic model is associated with a Gaussian distribution. A tree structure was used among the states and their transitions. At time point 0, there is one state, which is the root of tree. Every state except those associated with the last time point has at least one child, and for the results in this paper we allowed not more than two children. Any state having more than one child has a logistic regression classifier with L_1 loss penalty (Krishnapuram *et al*, 2005) associated with it. This classifier maps the set of predicted TF interactions for a gene to a probability distribution of transitions to each of the child states. To learn a dynamic regulatory map, the DREM algorithm first performs a search over tree structures. A randomly selected subset of genes is used to train the Gaussian distribution parameters and the classifiers in the tree structure under consideration. The remaining genes are used to score the various tree structures considered. Training is carried out using a version of the Baum–Welch algorithm (Durbin *et al*, 1998). After the best scoring structure is found using the test set of genes, weakly supported splits are pruned to avoid overfitting the test set of genes. After a final model structure is selected, all genes are used to train the parameters of the final model. See Methods section in Supplementary information for full details.

DREM software: The DREM software is available for download at http://www.sb.cs.cmu.edu/drem.

Inferring gene assignments and TF scores

Genes are assigned to their most likely path through the model using the Viterbi algorithm (Durbin *et al*, 1998). The assignment of genes to paths through the models is used to determine if certain paths are overenriched for genes regulated by certain TFs. Overenrichment scores are used for the association of TFs with paths. These scores are obtained using the hypergeometric distribution, with a lower score meaning a stronger association. The base set of genes for the hypergeometric distribution can be just the genes going into the previous split giving a TF split association score, or all genes on the microarray giving an overall association score of a TF for a path.

GO *P*-values

GO *P*-values were computed in the DREM software based on the hyper-geometric distribution. All *P*-values reported are uncorrected, but are still significant at the 0.01 level when correcting for multiple hypothesis testing using a randomization procedure (Ernst and Bar-Joseph, 2006).

Saccharomyces cerevisiae strain list

For the immunoprecipitation experiments, we used Myc-tagged W303 yeast strain obtained as a gift from Rick Young. The genotype of the Ino4 strain is MATa:ade2-1:trp1-1:can1-100:leu2-3,112:his3-11,15: ura3:GAL+:psi+:INO4::9myc:TRP1 and of the Gcn4 strain is MATa:ade2-1:trp1-1:can1-100:leu2-3,112:his3-11,15:ura3:GAL+:psi +:GCN4::9myc:TRP1.

Growth condition

For the AA starvation experiments, cells were grown in complete minimal medium (SCD) to early-log phase. Cells were collected by centrifugation and resuspended in an equal volume of minimal medium lacking AAs and adenine (YNB−AA, 2% glucose, 20 mg/l uracil) and allowed to grow. Samples for location analysis were taken before resuspension in AA starvation conditions and 4 h afterwards.

For the MMS experiments, cells were grown in YPD media to early-log phase at 30°C until the culture reached an OD_{600} of 0.8–1.0. MMS (Sigma) was added to a final concentration of 0.03%, and the culture was grown for an additional hour. Samples for genome-wide location analysis were taken before adding MMS and 15 and 60 min after adding MMS.

Chromatin immunoprecipitation–PCR

Bound proteins were formaldehyde-crosslinked to DNA *in vivo*, followed by cell lysis and sonication to shear DNA. Crosslinked material was immunoprecipitated with an anti-myc antibody, followed by reversal of the crosslinks to separate DNA from protein (Aparicio 1999; Orlando, 2000). Enrichment for Ino4-binding site was measured by semiquantitative PCR using primers designed for the detection of upstream regions of the genes YDR497C, YNL169C, YGR196C, and YHR123W. Primer sequences are as follows: YDR497C: TAGCGCACC AAACTGAAAGA, AAGCGCATATACTTAGTTCTCTCCA; YNL169C: CG ACCAAGAAGGATTTGAGC, CCAGCACCTTTTTGGTGTTT; YGR196C: CGCTTTCCAGAAAAAGGGTA, CGTCGTTTGTTTGTTTGGTG; YHR 123W: TGGCAAAATACAGAACACAGG, TATGCTCAGTCCAGCCCTTT. As a negative control, primers for the upstream region of Cts1 were used (AGTGGTTGGTTGGTGGGAATA; TCTTTGACCAATGCCTATGAA). The quantization of the enrichment was performed by calculating the ratio of the IP signal and the input signal for the target gene divided by the IP ratio and the input ratio of the negative control gene (Cts1), utilizing the software TINA. TINA is software for quantification of band intensity. After PCR, the fragments were separated on agarose gel (1%) and monitored by a CCD camera. Bands intensities were quantified using TINA 2.09d quantification software (Raytest, http://www.raytest.de).

Chip on chip

Genome-wide location analysis was performed as described previously (Ren *et al*, 2000). Briefly, following purification of the DNA in the ChIP procedure, immunoprecipitated DNA and DNA from an unenriched sample were amplified and differentially fluorescently labeled by ligation-mediated PCR. These samples were hybridized to a microarray consisting of spotted PCR products representing the intergenic regions of the *S. cerevisiae* genome. The data have been deposited into ArrayExpress with the accession numbers E-MEXP-905 (Gcn4 experiments) and E-MEXP-906 (Ino4 experiments).

Budding index calculation

Cells grown continuously at 30°C were collected by centrifugation, resuspended in an equal volume of 37°C medium, and returned to 37°C for growth. Samples were collected at time points as described in Figure 4C. For each time point, cells were mildly sonicated in order to separate clumps, counted and divided into three groups: 'No Bud'—a single cell; 'Small Bud'—a cell with a small bud attached; and a 'Big Bud'—a cell with a bud that is more than half of the size of the cell it is attached to.

Supplementary information

Supplementary information is available at *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

References

Aparicio OM (1999) Characterization of proteins bound to chromatin by immuno-precipitation from whole-cell extracts. *Curr Protocols Mol Biol* **4:** 21.3.10–21.3.12

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* **25:** 25–29

Baker RE, Masison DC (1990) Isolation of the gene encoding the *Saccharomyces cerevisiae* centromere-binding protein *CP1*. *Mol Cell Biol* **10:** 2458–2467

Balázsi G, Barabási A, Oltvai ZN (2005) Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proc Natl Acad Sci USA* **102:** 7841–7846

Barabasi A, Oltvai Z (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5:** 101–113

Bar-Joseph Z, Gerber G, Gifford D, Jaakkola T, Simon I (2003a) Continuous representations of time series gene expression data. *J Comput Biol* **10:** 241–256

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

Bengio Y, Frasconi P (1995) An input output HMM architecture. *Adv Neural Inform Process* **7:** 427–434

Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V (2006) The Inferelator: an algorithm for learning parsimonious regulatory networks from systems biology data sets *de novo*. *Genome Biol* **7:** R36

Bonner JJ, Heyward S, Fackenthal DL (1992) Temperature-dependent regulation of a heterologous transcriptional activation domain fused to yeast heat shock transcription factor. *Mol Cell Biol* **12:** 1021–10230

Cam H, Balciunaite E, Blais A, Spektor A, Scarpulla RC, Young R, Kluger Y, Dynlacht BD (2004) A common set of gene regulatory networks links metabolism and growth inhibition. *Mol Cell* **16:** 399–411

Chen T, He HL, Church GM (1999) Modeling gene expression with differential equations. *Pac Symp Biocomput* **4:** 29–40

Cokus S, Rose S, Haynor D, Grønbech N, Pellegrini M (2006) Modelling the network of cell cycle transcription factors in the yeast *Saccharomyces cerevisiae*. *BMC Bioinform* **7:** 381

Crespo JL, Powers T, Fowler B, Hall MN (2002) The TOR-controlled transcription activators GLN3, RTG1, and RTG3 are regulated in response to intracellular levels of glutamine. *Proc Natl Acad Sci USA* **99:** 6784–6789

D'haeseleer P, Wen X, Fuhrman S, Somogyi R (1999) Linear modeling of mRNA expression levels during CNS development and injury. *Pac Symp Biocomput* **4:** 41–52

Das D, Nahlé Z, Zhang MQ (2006) Adaptively inferring human transcriptional subnetworks. *Mol Syst Biol* **2:** 29

Della Seta F, Ciafre SA, Marck C, Santoro B, Presutti C, Sentenac A, Bozzoni I (1990) The ABF1 factor is the transcriptional activator of the L2 ribosomal protein genes in *Saccharomyces cerevisiae*. *Mol Cell Biol* **10:** 2437–2441

Durbin R, Eddy S, Krogh A, Mitchison G (1998) *Biological Sequence Analysis*. Cambridge: Cambridge University Press

Ernst J, Bar-Joseph Z (2006) STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7:** 191

Ernst J, Nau G, Bar-Joseph Z (2005) Clustering short time series gene expression data. *Bioinformatics* **21** (Suppl 1)**:** S159–S168

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11:** 4241–4257

Gorner W, Durchschlag E, Martinez-Pastor MT, Estruch F, Ammerer G, Hamilton B, Ruis H, Schuller C (1998) Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Genes Dev* **12:** 586–597

Hahn JS, Hu Z, Thiele D, Iyer VR (2004) Genome-wide analysis of the biology of stress responses through heat. *Mol Cell Biol* **24:** 5249–5256

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104

Hinnebusch AG, Fink GR (1983) Positive regulation in the general amino acid control of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **80:** 5374–5378

Holter NS, Maritan A, Ciepak M, Fedoroff NV, Banavar J (2001) Dynamic modeling of gene expression data. *Proc Natl Acad Sci USA* **98:** 1693–1698

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

Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M (2003) A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* **302:** 449–453

Kim SY, Imoto S, Miyano S (2003) Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief Bioinform* **4:** 228–235

Klig LS, Hoshizaki DK, Henry SA (1988) Isolation of the yeast INO4 gene, a positive regulator of phospholipid biosynthesis. *Curr Genet* **13:** 7–14

Krishnapuram B, Figueiredo M, Carin L, Hartemink A (2005) Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Trans Pattern Anal Mach Intell* **27:** 957–968

Kundaje A, Middendorf M, Gao F, Wiggins C, Leslie C (2005) Combining sequence and time series expression data to learn transcriptional modules. *IEEE/ACM Trans Comput Biol Bioinform* **2:** 194–202

Lee J, Godon C, Lagniel G, Spector D, Garin J, Labarre J, Toledano MB (1999) Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. *J Biol Chem* **274:** 16040–16046

Lin L, Lee H, Li W, Chen B (2005) Dynamic modeling of *cis*-regulatory circuits and gene expression prediction via cross-gene identification. *BMC Bioinform* **6:** 258

Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M (2004) Genomic (2004) analysis of regulatory network dynamics reveals large topological changes. *Nature* **431:** 308–312

Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* **100:** 11980–11985

Mellor J, Jiang W, Funk M, Rathjen J, Barnes CA, Hinz T, Hegemann JH, Philippsen P (1990) CPF1, a yeast protein which functions in centromeres and promoters. *EMBO J* **9:** 4017–4026

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* **298:** 824–827

Natarajan K, Meyer MR, Jackson BM, Slade D, Roberts C, Hinnebusch AG, Marton MJ (2001) Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol Cell Biol* **21:** 4347–4368

O'Connell KF, Surdin-Kerjan Y, Baker RE (1995) Role of the *Saccharomyces cerevisiae* general regulatory factor CP1 in methionine biosynthetic gene transcription. *Mol Cell Biol* **15:** 1879–1888

Orlando V (2000) Mapping chromosomal proteins *in vivo* by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* **25:** 99–104

Petiot A, Pattingre S, Arico S, Meley D, Codogno P (2002) Diversity of signaling controls of macroautophagy in mammalian cells. *Cell Struct Funct* **27:** 431–441

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

Planta RJ (1997) Regulation of ribosome synthesis in yeast. *Yeast* **13:** 1505–1518

Ramoni MF, Sebastiani P, Kohane IS (2002) Cluster analysis of gene expression dynamics. *Proc Natl Acad Sci USA* **99:** 9121–9126

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306–2309

Rudra D, Zhao Y, Warner JR (2005) Central role of Ifh1p–Fhl1p interaction in the synthesis of yeast ribosomal proteins. *EMBO J* **24:** 533–542

Santiago TC, Mamoun CB (2003) Genome expression analysis in yeast reveals novel transcriptional regulation by inositol and choline and new regulatory functions for Opi1p, Ino2p, and Ino4p. *J Biol Chem* **278:** 38723–38730

Schliep A, Schonhuth A, Steinhoff C (2003) Using hidden Markov models to analyze gene expression time course data. *Bioinformatics* **19** (Suppl 1)**:** S255–S263

Segal E, Yelensky R, Koller D (2003) Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioninformatics* **19** (Suppl 1)**:** S273–S282

Simon I, Siegfried Z, Ernst J, Bar-Joseph Z (2005) Combined static and dynamic analysis for determining the quality of time-series expression profiles. *Nat Biotech* **23:** 1503–1508

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9:** 3273–3297

Thomas D, Jacquemin I, Surdin-Kerjan Y (1992) MET4, a leucine zipper protein, and centromere-binding factor 1 are both required for transcriptional activation of sulfur metabolism in *Saccharomyces cerevisiae. Mol Cell Biol* **12:** 1719–1727

Wade JT, Hall DB, Struhl K (2004) The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. *Nature* **432:** 1054–1058

Workman CT, Mak HC, McCuine S, Tagne JB, Agarwal M, Ozier O, Begley TJ, Samson LD, Ideker T (2006) A systems approach to mapping DNA damage response pathways. *Science* **312:** 1054–1059

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434:** 338–345