

## CHAPTER 4

---

# Transcription Regulatory Networks in Yeast Cell Cycle

Nilanjana Banerjee and Michael Q. Zhang\*

### Introduction

The functional genomics techniques for mapping transcription regulatory networks have evolved based on advances in experimental approaches and the kinds of data generated. Studies in yeast have emphasized powerful genetic approaches that are not available in other higher eukaryotic organisms. As a consequence, yeast is particularly amenable for analyzing transcriptional regulatory mechanisms *in vivo* under true physiological conditions. With its small genome (predicted to encode roughly about 6200 proteins) and its tractable genetics, *Saccharomyces cerevisiae* has played a prominent role in the development of many methodologies for functional genomics.<sup>1</sup> Various high throughput expression techniques, such as SAGE and microarrays, have been developed that exploit the huge body of transcription data and provide rapid, parallel surveys of gene-expression patterns for hundreds of thousands of genes in a single assay. Several computational algorithms have been developed and applied to uncover coregulated genes or causal relationships from the large-scale gene expression data. As transcription is mainly controlled and regulated by the binding of transcription factors (TFs) to the promoter DNA sequence, significant progress has also been made in identifying these *cis*-regulatory elements in the promoters, giving more insights to gene function and regulation pathways.<sup>2</sup> Recently, other high-throughput methods have been developed for measuring the interactions between DNA and TFs *in vivo*. Microarray-based chromatin immunoprecipitation assays (ChIP-chip), have enabled genome-wide location analysis of TF-binding *in vivo*, offering another powerful tool in dissecting the global regulatory networks. Also, sequencing of multiple yeast species have provided an opportunity to look for conserved functional modules. In this chapter, we discuss the functional genomics approaches to map regulatory networks from combinations of sequence data, genome-wide gene expression data and ChIP data in the context of the cell cycle regulation of the budding yeast, *Saccharomyces cerevisiae*. These approaches extract key aspects of regulatory mechanisms such as identifying target genes and *cis*-regulatory elements important for a TF or combination of TFs under a particular condition or perturbation. They also help mapping interactions between trans- and cis- transcription modules (defined by a TF and target genes) - giving a more systematic view of the mechanistic underpinnings of gene expression networks.

---

\*Michael Q. Zhang—Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, U.S.A. Email: mzhang@cshl.edu

## Identification of Target Genes and Their *cis*-Regulatory Elements

### *Gene Expression Analysis*

To map a transcription regulatory network it is essential to identify the transcription factor and its target genes or genes that are coregulated. Information about the transcript levels is fundamental in providing some of these connections. To understand how a genetic system is regulated, a typical approach is to monitor the system's responses to perturbations. After a perturbation, one of the first questions we can ask is which genes have been up-regulated (or down-regulated). If the perturbation consists of a TF knockout or over-expression, by sorting expression levels (relative to the control), one could in principle identify its target genes (activated or repressed).

The difficulty in the above approach lies in the fact that many of these target genes may not be the primary targets; they may contain secondary targets unless the mRNA samples were collected fast enough or translation was blocked. However, identifying patterns of gene expression and grouping genes into expression classes may provide much greater insight into their biological function, because many genes belonging to the same complex (e.g., ribosome) or to the same regulatory pathway tend to have similar or correlated expression profiles. For instance, if two or more genes have correlated (or anti-correlated) expression profiles in different experiments or at different time-points, these genes may be coregulated and possibly functionally related. Different metrics, like Euclidean distance, correlation coefficient, ranked correlation coefficient and mutual information based measure have all been used to quantify the similarity (or distance) between the expression patterns. After choosing the similarity measure in the expression profile space, supervised or unsupervised clustering methods may be used to study the gene expression matrix.<sup>3,4</sup>

### *Motif-Finding Algorithms*

A subsequent approach to understand the interaction between a TF and its target genes, is to further study the binding characteristics of a TF through its *cis*-regulatory elements in the promoters.<sup>5</sup> Motifs that are common to a set of apparently coexpressed genes are plausible candidates for binding sites implicated in transcriptional regulation. Several studies looked at groups of coregulated genes to find over-represented oligonucleotide sequences.<sup>6,7</sup> They detected new candidate regulatory sites, as well as sites that had already been characterized. Other approaches developed statistical techniques to predict short oligomers that may be involved in the expression of groups of coregulated genes.<sup>8,9</sup> Their strategy looked for pentamers and hexamers that are over-represented among the upstream regions of genes whose expression peaks at a particular phase of the cell cycle. Modified versions of the Gibbs motif sampler has also been successfully used to detect longer motifs in the yeast cell cycle clusters.<sup>10,11</sup>

The approach of clustering genes according to their expression profile across many experiments is well suited for genes that covary under most circumstances. However, no expression-based clustering can find genes in a cluster that do not have similar motifs or find motifs in genes that are not functional. So new methods were needed where the clustering took the DNA sequence into account. One method suggested that the two stages—clustering of expression profiles followed by Gibbs sampling of sequences—may be combined and viewed as operating on the marginal distributions of a joint probabilistic model for both sequence and expression data.<sup>12</sup> In this case, the presence or absence of a motif will have an influence on which cluster a gene may be assigned to. The hope was that using an integrated approach and a better-formulated optimization problem will result in significantly improved discriminative power for regulatory signal identification. Another method (REDUCE) uses unbiased statistics to identify oligonucleotide motifs whose occurrence in the regulatory region of a gene

correlates with the level of mRNA expression.<sup>13</sup> In this work, linear regression analysis is used to infer the activity of the transcriptional module associated with each motif. Using the cell-cycle and sporulation experiments as examples, the authors reconfirmed almost all motifs found by clustering methods, at least to the extent of finding a related sequence motif that captures the same experimental signal. Among new results, they found that Mcm1 and Fkh2 are antagonistic outside of their phase (M/G2). They have examples that point to combinatorial effects in transcription regulation or groups of genes that covary in one circumstance but vary differently in another, for which expression-based clustering would be poorly suited.

Currently, ChIP-chip assays have become a popular method for identifying TF binding sites *in vivo*. However, these assays can only map the probable protein-DNA interaction loci within a couple of hundred basepair (up to 1 kilobase) resolution. MDscan was developed to incorporate ChIP ranking information to swiftly discover relevant motifs.<sup>14</sup> To pinpoint interaction sites down to the base-pair level, MDscan examines the ChIP-array-selected sequences and searches for DNA sequence motifs representing the protein-DNA interaction sites. MDscan combines the advantages of two widely adopted motif search strategies, word enumeration and position-specific weight matrix updating and incorporates the ChIP array ranking information to accelerate searches and enhance their success rates. Because MDscan enumerates only existing *w*-mers in the top sequences, its search time increases only quadratically with respect to the total number of bases in the top sequences for all motif sizes. Other programs like BioProspector,<sup>15</sup> CONSENSUS<sup>16</sup> and AlignAce<sup>17</sup> failed to do as well as MDscan in finding many of the important motifs from the ChIP-enriched genes of cell-regulated targets. With some modifications, MDscan has also been used as part of another algorithm, MotifRegressor, which assumes that effect of TFBM is strongest among genes with a dramatic increase or decrease in gene expression level in response to a condition. The authors argue that the method combines the advantages of matrix-based motif finding and oligomer motif-expression regression analysis, resulting in high sensitivity and specificity. Using the alpha-factor cell cycle data they found 273 significant motifs. They studied the motif effects (coefficients) during the cell cycle and found that the known cellcycle-related motifs MCM1, SWI5, MCB, SCB, and SFF have coefficients that fluctuate with the cell cycle while some cell cycle motifs (STE12, STRE, and others) influence expression through the cell cycle, but to a lesser extent than the known cell cycle regulators.

### ***Full Genome Comparative Analysis and Motif-Finding***

Recent analyses of the genomic sequences of a number of related yeast species have helped to distinguish between real and misannotated ORFs and to find conserved motifs that may be functional targets of TFs. Yeast strains closely related to *S. cerevisiae* can be divided into three sub-groups: *Saccharomyces sensu stricto*, *Saccharomyces sensu lato* and petite negative (the last two subgroups have fewer chromosomes and are significantly different physiologically from *S. cerevisiae*). It is important to assess the evolutionary distance where nonfunctional sequences have diverged enough to allow many functional sequence signals to stand out above the noise, the sequences retain enough overall similarity to enable their alignment. Usually several species need to be compared to lend sufficient acuity to the phylogenetic footprints.<sup>18</sup> The genomes of three *sensu stricto* strains (*S. mikatae*, *S. kudriavzevii* and *S. bayanus*) and two more distantly related strains (*S. castellii* and *S. kluyveri*) have been sequenced.<sup>18</sup> Both four-way genome sequence alignments over just the *sensu stricto* strains and six-way alignments over all the sequenced strains, including *S. cerevisiae* revealed many characterized ungapped motifs. In addition, they found 79 unknown conserved motifs. To predict which of these unknown motifs are functional they further group all sequences that reside upstream of genes that are functionally related or those that reside upstream of genes that exhibit a similar expression. Several of these are cell-cycle related and would have to be validated experimentally.

In a similar study, four *sensu stricto* species: *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus* were compared.<sup>19</sup> The authors systematically discovered conserved nucleotide patterns (gapped and ungapped motifs) by some expert-rules, and constructed a list of 72 genome-wide motifs, 42 of which did not match previously characterized motifs. Functions were assigned to the majority of these by their enrichment in gene categories assembled from GO annotation, ChIP and RNA gene expression studies. In addition, they showed evidence of combinatorial control of gene regulation, where motif combinations change the functional specificity of downstream genes.

## Transcription Regulatory Network Reconstruction

### Combinatorial Interactions

While there has been substantial work on clustering algorithms and motif-discovery algorithms, a more ambitious goal for functional genomics is to understand the structure and dynamics of intracellular networks. The logical first steps have been to decompose the networks into functional modules. These modules aim to capture various aspects surrounding the regulator-target gene relationship, often under specific conditions or regulatory context. Studying the interaction between interacting regulators addresses the complex, cooperative interactions required by combinations of TFs to execute an exponentially larger number of regulatory decisions.<sup>20-23</sup> One approach has been to screen for cooperatively binding TFs by correlating pairs of computationally derived motif-combinations with gene expression data.<sup>21</sup> Motif synergy maps can be generated to give a global view of the intense cross-talk between TFs under different cellular conditions. Presence of computationally derived motif-combinations in the promoter however, does not automatically give direct evidence of TF binding. As a result such analysis can potentially suffer from a large number of false positives in predicting functional TF binding sites.

Genome-wide location data<sup>24-26</sup> elucidates the *in vivo* physical interactions of TFs with their chromosomal targets on the genome and as a result it can provide a more reliable view of functional TF-binding site interaction. Several studies have used genome-wide location analysis to explore the yeast cell cycle gene expression program and showed that TFs that function during one stage of the cell cycle regulate those that function during the next stage.<sup>24,25</sup>

The approach used by GRAM (gene regulatory module),<sup>24</sup> examines DNA-binding data and identifies sets of genes that are bound by common sets of transcriptional regulators (Fig. 1). It then uses expression data to identify a subset of the coexpressed genes. Finally, the algorithm searches the DNA-binding data again, using less stringent criteria, to find more genes with similar expression that are also bound by the same transcription factors. The algorithm helps compensate the technical limitations in each data. It presents a useful alternative to using a single p-value threshold for binding events, because their method allows the p-value threshold to be relaxed if there is sufficient supporting evidence from the expression data.<sup>27</sup> Yet another strategy exploited ChIP-chip data (with direct evidence of TF binding) and genome-wide gene expression data<sup>29</sup> to rigorously assess cooperativity among TFs in the yeast cell cycle.<sup>28</sup> Statistically significant cooperative TFs were generated by exploring the effect of cooperative binding vs. independent binding of the TFs on gene expression. The assumption is that if two TFs are cooperative then they should both bind (either directly or through another DNA binding protein) to the promoters of their target genes and the expression profiles of these target genes would be similar. If they are not cooperative, more than likely both TFs will not bind to the same promoters. Even if they do bind, the target genes will likely be regulated by different mechanisms and as a result the expression profiles will not be as coherent overall. The results confirmed most previously characterized cell-cycle related cooperative TFs, validating the use of this measure as a predictor of potential cooperativity. In addition, they propose

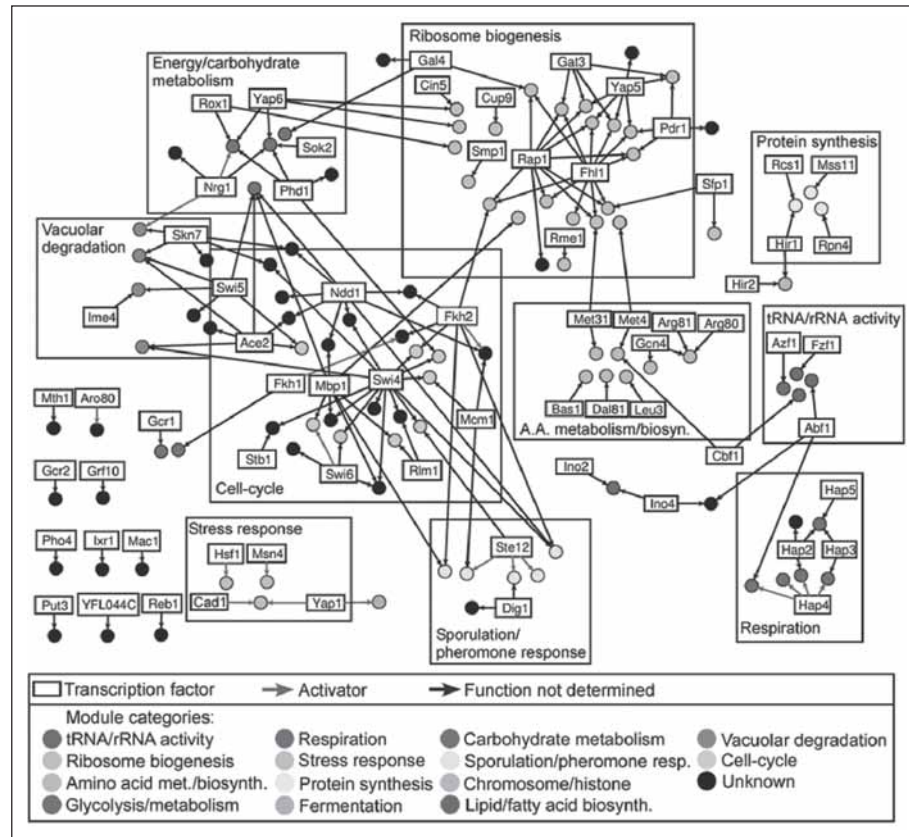


Figure 1. Gene modules network as discovered by the GRAM algorithm (Bar-Joseph et al. 2003). Many groups of connected genes and modules and regulators are involved in similar biological processes.

several novel cooperative TFs in cell cycle (e.g., Ndd1-Stb1, Ace2-Hsf1) and in other biological processes (e.g., Pdr1-Smp1 etc.). It is interesting that cell-cycle regulators interact with a strikingly large number of other protein classes. Many different processes in a cell during cell division have to be precisely coordinated with cell-cycle regulators. Such cooperativity suggests cross talk that is essential to coordinate different functions (Fig. 2).

In a related method, promoter sequence analysis was incorporated in order to infer not only the interacting TFs but also to assign their corresponding binding sites by iteratively and exhaustively searching for significant TF combinations and motif combinations up to the triplet level.<sup>30</sup> The authors were able to extend the previous chain of single regulators to an expanded chain of interacting regulators. These modules of interacting regulators at adjacent phases often share a common link that can bridge the continuity of the cycle. In addition, they identified similar modules that allow cell entry or exit of the cycle according to external signals at particular checkpoints (Fig. 3).

### Reconstructing Transcriptional Modules

Various mathematical techniques, such as differential equations, Bayesian and Boolean models and several statistical methods, have been applied to expression data in attempts to extract the underlying gene regulation networks.<sup>31</sup> Since the possible number of networks

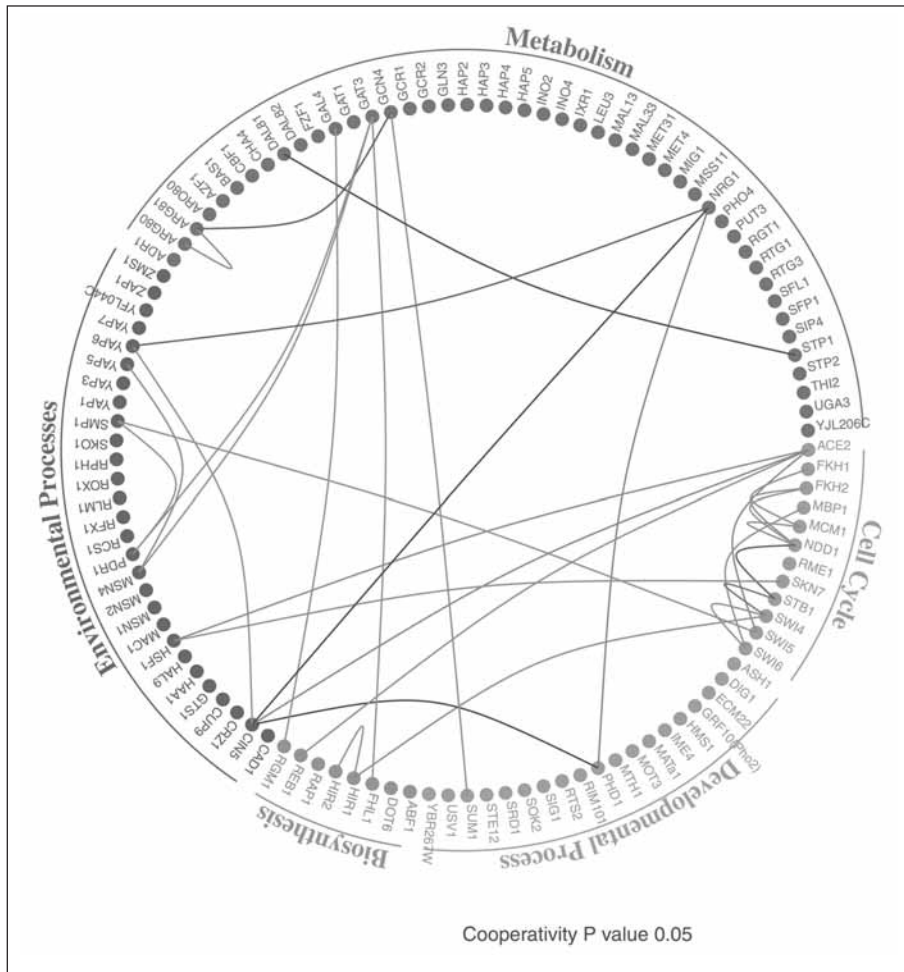


Figure 2. User interface to dynamically generate a TF cooperativity network with user-defined significance values. The intensity of a curved edge corresponds to the confidence in cooperativity based on cooperativity p-value ranging from 0.05 (black) to 0.0001 (gray). The TFs have been clustered according to their functional categories as defined by the MIPS database.

grows exponentially with the number of genes, it is not possible to derive a unique network with only limited data. To deal with the inherent complexity of network inference, Friedman and colleagues<sup>32</sup> examined local statistical properties of network components using Bayesian network approaches. With a large set of gene knockout expression data, they were able to extract a finer structure of interactions between genes, such as causality, mediation, activation and inhibition and uncovered some robust regulatory pathways. Recently, several studies have focused on computationally identifying condition-specific transcription modules (relating each module with regulators and target genes to the cellular conditions or perturbations that trigger it) and discovering interactions between such modules, by combinations of the DNA sequence, gene function and gene expression data.

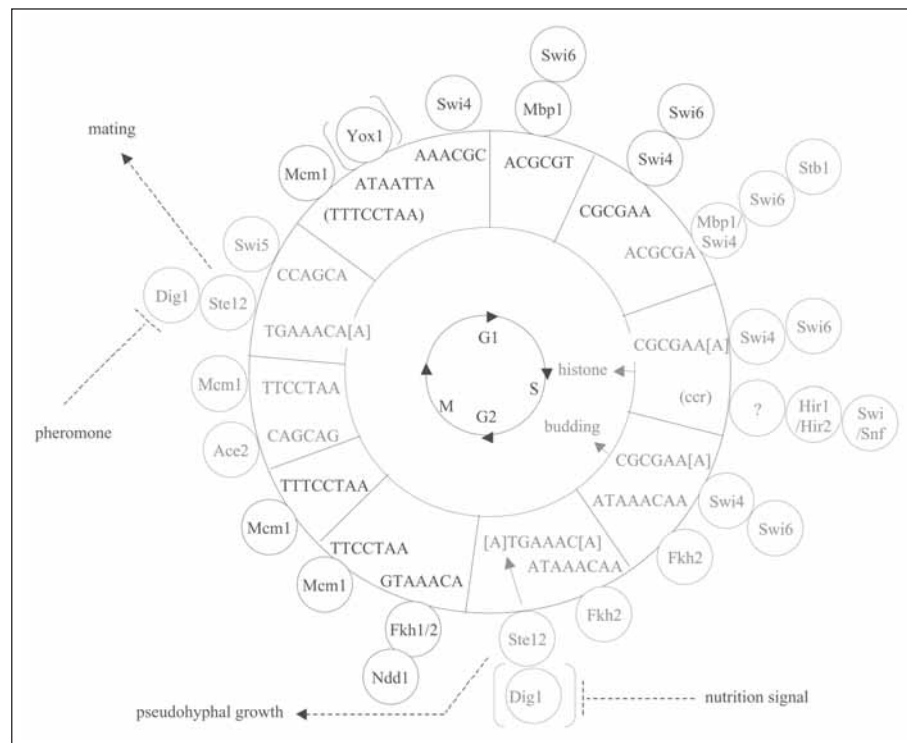


Figure 3. Motif-based cell cycle regulatory network.<sup>30</sup> Arched boxes indicate motif combinations and the circles indicate the binding of TFs to the motif combinations. The TFs and motif combinations in black were previously known and confirmed in this research. Those in red are putative novel combinations. TFs and motifs in parentheses mean those we could not detect but the presences are supported by literature. [A] in a motif indicates 'A' or null. Almost all TFs appear as elements of a combination, not as a single. {Swi5, Ste12, Dig1} shows a process joint combination between cell cycle and pheromone response/filamentous growth.

Another approach for inferring such regulatory modules integrates additional biological information, such as functional annotation or sequence information, with the analysis of gene expression data.<sup>33</sup> Here, genes may be assigned to several overlapping modules—a property that is essential for capturing the biologically relevant combinatorial regulation. The algorithm receives a set of genes as input and proceeds in two stages. In the first stage, the experimental conditions under which the input genes are coregulated most tightly are identified. They calculate the average change in the expression of the input genes for each condition and refer to them as the 'condition scores'. Only conditions with a large (absolute) score are selected. In the second stage, the algorithm selects from the whole genome those genes that show a significant and consistent change in expression under the conditions selected in the first stage. For each gene, the weighted average change in expression over these conditions is calculated, using the condition scores as weights. These average values are referred to as the 'gene scores'. Genes with large scores are selected to be part of the module. To assign a measure of reliability, the signature algorithm is applied to distinct input sets containing different subsets of the postulated transcription module. If the different input sets give rise to the same module, it is considered a reliable module.

The signature algorithm is a generalization of the standard Singular Value Decomposition method and can be used to extend and refine partial knowledge about a pathway using the available expression data. Specifically, by applying the signature algorithm to a given set of genes that are thought to participate in a particular cellular function, it is possible to (i) reject genes that are mistakenly included, (ii) retrieve additional genes that are also likely to be involved in the pathway and (iii) identify the experimental conditions under which these genes are coregulated. The algorithm has also been used to study the global structure of the transcription program. Applied to a diverse collection of input sets derived in three different ways (i) genes with a particular sequence in their upstream region, (ii) genes with related MIPS-functional annotation and (iii) cluster-related genes from the output of a hierarchical cluster algorithm. The reliable output sets led to the identification of 86 overlapping transcription modules where the genes of most modules participate in a module-specific cellular process.

Yet another approach for inferring regulatory modules utilizes a motif and information in its flanking region more explicitly.<sup>34</sup> In this work, the output of REDUCER is enhanced to more exactly identify both the target genes and the regulatory elements. A profile for each DNA motif and its flanking regions is built unlike the standard profile method, each gene's contribution to the profile is weighted by its mRNA expression in the corresponding experiment. The weighted profiles should favor true target genes of the TF. The authors identified conditions that activate a particular transcription module and if two transcription modules are both activated under a particular condition, it is possible that they may interact. Combinatorial interactions can be detected by examining genes shared by different modules.<sup>34</sup> A putative target gene of Mbp1, SPA2, was determined to interact with proteins in the signaling pathway upstream of other TFs. Among proteins that interact with Spa2, Ste20, Ste11, and Ste7 function in the upstream of TF Ste12 in the pheromone and filamentous growth pathways, and MKK1, Mkk2, and Slt2 are involved in the protein kinase pathway, which can activate TFs Swi4/6 complex and Rlm1. Therefore, it was argued that activation of one module such as the Mbp1 module may further tune the activity of other transcription modules such as the Ste12 module.

There have been several iterative learning procedures that search for the optimized model capturing gene interactions. One noteworthy approach for inferring regulatory networks utilizes a probabilistic graphical model method. This approach relies on the sometimes-violated assumption that the regulators are themselves transcriptionally regulated and that their expression profiles explain their activity level.<sup>35</sup> Their automated procedure takes as input a gene expression data set and a set of 466 candidate regulatory genes containing both known and putative transcription factors and signal transduction molecules. Given these inputs, the algorithm searches simultaneously for a partition of genes into modules and for a regulation program for each module that can explain the expression behavior of genes in the module. They define a space of possible models and use a Bayesian score to evaluate a model's fit to the data. The procedure uses the Expectation Maximization (EM) algorithm to search for the model with the highest score. Applying their method to gene expression data (in response to environmental changes) they inferred modules that mostly contained a functionally coherent set of genes. So they were able to identify groups of coregulated genes, their regulators, the behavior of the module as a function of the regulators' expression and the conditions under which the regulation takes place. A similar approach was also applied to infer regulatory modules from both gene expression data and promoter sequence data.<sup>36</sup>

### ***Constructing Multiple-Species Network***

Genome-wide comparative analysis has primarily been based on genomic sequence information. Recently two studies have attempted to measure evolutionarily conserved coexpression in a genome-wide scale and build 'multiple-species' networks. They argue that in experiments

limited to a single species, it would be difficult or even impossible to distinguish accidentally regulated genes from those that are physiologically important. The assumption is that coregulation of a pair of genes over large evolutionary distances implies that the coregulation confers a selective advantage, most likely because the genes are functionally related. DNA Microarray data for humans, flies, worms, and yeast have been used to identify gene interactions that are evolutionarily conserved.<sup>37</sup> The multiple-species network only maps those genes that have orthologs in other species and thus focuses on core, conserved biological processes; and interactions in the multiple-species network imply a functional relationship based on evolutionary conservation, whereas interactions using data from single species only indicate correlated gene expression. Most of the components were enriched for metagenes involved in similar biological processes, such as protein degradation, ribosomal function, cell cycle, metabolic pathways, and neuronal processes. Of the cell cycle metagenes, 30 are involved in regulating the cell cycle such as MEG2742 (encodes cyclin E) along with 80 that perform terminal cell cycle functions such as MEG1092 (encodes DNA polymerase-2). The remaining 131 genes were not previously known to be involved in the cell cycle, and so linking these genes to known cell cycle metagenes in the coexpression network suggests new cell cycle functions for these genes.

In a similar vein,<sup>38</sup> another comparative analysis of large datasets of expression profiles involved six evolutionarily distant organisms. All expression networks shared common topological properties, such as a scale-free connectivity distribution and a high degree of modularity. While these common global properties may reflect universal principles underlying the evolution or robustness of these networks, they do not imply similarity in the details of the regulatory programs. Rather, with a few exceptions, the modular components of each transcription program as well as their higher-order organization appear to vary significantly between organisms and are likely to reflect organism-specific requirements.

These studies suffer from several limitations. Expression profiles only cover a subset of all possible cellular conditions and thus provide only partial information about the underlying regulatory program. Moreover, this subset is typically very different for each organism, reflecting distinct physiologies as well as different research foci. One way to circumvent this problem is to restrict the data to a small subset of similar conditions, such as timepoints along the cell cycle.<sup>39</sup> Such an approach, however, drastically reduces the size of the dataset and limits the scope of comparison. The most serious problems may be the heterogeneity of the samples and conditions, expression profiles can be very different for different cell types within a single organism or even for different conditions/time-points for a single cell type, let alone the stochasticity of gene expression within a single cell.<sup>40</sup>

Even though there has been much progress in developing network models, it is important to note that the current experimental data from which networks are inferred is extremely noisy. The amount of samples, even in the largest experiments in the foreseeable future, does not provide enough information to construct a full detailed model with high statistical confidence. Compounded by these issues, there is a great need to integrate diverse data types and construct tools that will assimilate them into biological models.<sup>41</sup>

## Discussion

As the computational approaches to analyzing functional genomics data are further developed and refined, extracting and integrating orthogonal information will become increasingly important. Combination of sequence data, global expression profiling and binding site mapping has already produced a more complete picture of the genetic circuitry that is responsible for transcription regulation. Different types of large-scale data can be interrelated to reveal potentially important but not apparent relationships - for example, between gene expression and the position of genes on chromosomes,<sup>42</sup> or between gene expression and the subcellular localization of proteins,<sup>43</sup> or between gene expression and the protein interaction.<sup>44,45</sup> Ge-

omic and proteomic approaches have been integrated to build, test and refine a model of the galactose utilization pathway in *S. cerevisiae* by integrating both genomic and proteomic approaches.<sup>45,46</sup> Protein-protein interaction data and ChIP-chip data have yielded a statistically significant correlation between cooperatively acting TFs and their protein interaction profiles.<sup>46</sup> Emerging technologies, like metabolic footprinting, are beginning to distinguish between different physiological states of wild-type yeast and between yeast single-gene deletion mutants and lending valuable 'downstream' information.<sup>47</sup>

With the systematic combination of diverse data types and new functional genomics approaches a comprehensive understanding of complex transcription regulatory networks is beginning to emerge. But to efficiently dissect large amount of functional genomics data for transcription regulatory network studies, more promoter prediction tools,<sup>48</sup> more promoter extraction tools<sup>49</sup> and more specialized promoter databases, such as SCPD,<sup>50</sup> are clearly going to be urgently needed.

## References

1. Grunewald B, Winzler EA. Treasures and traps in genome-wide data sets: Case examples from yeast. *Nature Review Genetics* 2002; 3:653-661.
2. Zhang MQ. Large-scale gene expression data analysis: A new challenge to computational biologists. *Genome Research* 1999a; 9:681-688.
3. Brazma A, Vilo J. Gene expression data analysis. *FEBS Letters* 2000; 480:17-24.
4. Quackenbush J. Computational analysis of microarray data. *Nature Review Genetics* 2001; 2:418-427.
5. Bulyk M. Computational prediction of transcription-factor binding site locations. *Genome Biology* 2003; 5:201.
6. 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.
7. Brazma A, Jonassen I, Vilo J et al. Predicting gene regulatory elements in silico on a genomic scale. *Genome Research* 1998; 8:1202-1215.
8. Zhang MQ. Promoter analysis of coregulated genes in the yeast genome. *Computers and Chemistry* 1999b; 23:233-250.
9. Wolfsberg TG, Gabrielian AE, Campbell MJ et al. Candidate regulatory sequence elements for cell cycle-dependent transcription in *Saccharomyces cerevisiae*. *Genome Research* 1999; 8:775-792.
10. Spellman PT et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 1998; 9:3273-3297.
11. Tavazoie S, Hughes JD, Campbell MJ et al. Systematic determination of genetic network architecture. *Nat Genet* 1999; 22:281-285.
12. Holmes I, Bruno WJ. Finding regulatory elements using joint likelihood for sequence and expression profile data. *Proc Int Conf Intell Syst Mol Biol* 2000; 8:202-210.
13. Bussemaker HJ, Li H, Siggia ED. Regulatory element detection using correlation with expression. *Nature Genetics* 2001; 27:167-171.
14. Liu XS, Brutlag DL, Liu JS. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation Microarray experiments. *Nature Biotechnology* 2002; 20:835-839.
15. Liu X, Brutlag DL, Liu JS. BioProspector: Discovering conserved DNA motifs in upstream regulatory regions of coexpressed genes. *Pac Symp Biocomput* 2001; 127-138.
16. Hertz GZ, Hartzell III GW, Stormo GD. Identification of consensus patterns in unaligned DNA sequences known to be functionally related. *Comput Appl Biosci* 1990; 6:81-92.
17. Roth FP, Hughes JD, Estep PW et al. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotechnol* 1998; 16:939-945.
18. Cliften P, Sudarsanam P, Desikan A et al. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* 2003; 301:71-6.

19. Kellis M, Patterson N, Endrizzi M et al. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 2003; 423:241-454.
20. Wagner A. Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics* 1999; 15:776-784.
21. Pilpel Y, Sudarsanam P, Church G. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature Genetics* 2001; 29:153-159.
22. Hannenhalli S, Levy S. Predicting transcription factor synergism. *Nucleic Acids Research* 2002; 30:4278-4284.
23. GuhaThakurta D, Stormo GD. Identifying target sites for cooperatively binding factors. *Bioinformatics* 2001; 17:608-621.
24. Lee TI, Rinaldi NJ, Robert F et al. Transcriptional regulatory networks in *S. cerevisiae*. *Science* 2002; 298:799-804.
25. Simon I, Barnett J, Hannett N et al. Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* 2001; 106:697-708.
26. Horak CE, Luscombe NM, Qian J et al. Complex transcriptional circuitry at the G1/S transition in *S. cerevisiae*. *Genes & Development* 2002; 16:3017-3033.
27. Bar-Joseph Z, Gerber GK, Lee TI et al. Computational discovery of gene modules and regulatory networks. *Nature Biotechnology* 2003; 21:1337-1342.
28. Banerjee N, Zhang MQ. Identifying cooperativity among transcription factors controlling the cell cycle in yeast. *Nucleic Acids Research* 2003; 31:7024-7031.
29. Cho RJ et al. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* 1998; 2:65-73.
30. Kato M, Hata N, Banerjee N et al. Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biology* 2004; 5:R56.
31. Banerjee N, Zhang MQ. Functional genomics as applied to mapping transcription regulatory networks. *Curr Opin in Microbiol* 2002; 5:313-317.
32. Friedman N, Linial M, Nachman I et al. Using Bayesian networks to analyze expression data. *J Comput Biol* 2000; 7:601-620.
33. Ihmels J, Friedlander G, Bergmann S et al. Revealing modular organization in the yeast transcriptional network. *Nature Genetics* 2002; 31:370-377.
34. Wang W, Cherry JM, Botstein D et al. A systematic approach to reconstructing transcription networks in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 2002; 99:16893-16898.
35. Segal E, Shapira M, Regev A et al. Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics* 2003a; 34:166-76.
36. Segal E, Yelensky R, Koller D. Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics* 2003b; 19(Suppl 1):I273-I282.
37. Stuart JM, Segal E, Koller D et al. A gene-coexpression network for global discovery of conserved genetic modules. *Science* 2003; 302:249-55.
38. Bergmann S, Ihmels J, Barkai N. Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol* 2004; 2:E9.
39. Alter O, Brown PO, Botstein D. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proc Natl Acad Sci USA* 2003; 100:3351-3356.
40. Paulsson J. Summing up the noise in gene networks. *Nature* 2004; 427:415-418.
41. Hasty J, McMillen D, Isaacs F et al. Computational studies of gene regulatory networks: in numero molecular biology. *Nat Rev Genet* 2001; 2:268-279.
42. Cohen BA, Mitra RD, Hughes JD et al. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genetics* 2000; 26:183-186.
43. Drawid A, Jansen R, Gerstein M. Genome-wide analysis relating expression level with protein subcellular localization. *Trends Genet* 2000; 16:426-430.
44. Ge H, Liu Z, Church GM et al. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genetics* 2001; 29:482-486.
45. Ideker V, Ranish J, Christmas R et al. Integrated genomic and proteomic analysis of systematically perturbed metabolic network. *Science* 2001; 292:929-934.

46. Manke T, Bringas R, Vingron M. Correlating protein-DNA and protein-protein interaction networks. *J Mol Biol* 2003; 333:75-85.
47. Allen J, Davey HM, Broadhurst D et al. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnology* 2003; 21:692-696.
48. Davuluri R, Grosse I, Zhang MQ. Computational identification of promoters and first exons in the human genome. *Nature Genetics* 2001; 29:412-417.
49. Zhang T, Zhang MQ. Promoter extraction from GenBank (PEG): Automatic extraction of eukaryotic promoter sequences in large sets of genes. *Bioinformatics* 2001; 17:1232-1233.
50. Zhu J, Zhang MQ. SCPD: A promoter database of yeast *saccharomyces cerevisiae*. *Bioinformatics* 1999; 15:607-611.