Deciphering gene expression regulatory networks

John J Wyrick* and Richard A Young†

In the past year, great strides have been made in our understanding of the regulatory networks that control gene expression in the model eukaryote *Saccharomyces cerevisiae*. The development and use of a number of genomic tools, including genome-wide location and expression analysis, has fueled this progress. In addition, a variety of computational algorithms have been devised to mine genomic sequence for conserved regulatory motifs in co-regulated genes. The recent description of the genetic network controlling the cell cycle illustrates the tremendous potential of these approaches for deciphering gene expression regulatory networks in eukaryotic cells.

Addresses
Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA; Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
*e-mail: jjwyrick@mit.edu
†e-mail: young@wi.mit.edu

Current Opinion in Genetics & Development 2002, **12**:130–136

0959-437X/02/$ – see front matter
© 2002 Elsevier Science Ltd. All rights reserved.

Abbreviations
GMEP genome-mean expression profile
SVD singular value decomposition

Introduction

Cells respond to environmental changes by reprogramming expression of specific genes throughout the genome. The transcription rate of a particular gene is determined by the interaction of diverse regulatory proteins — transcriptional activators and repressors — with specific DNA sequences in the gene's promoter. How a collection of regulatory proteins accomplishes the task of regulating a set of genes can be described as a regulatory network (Figure 1). Some investigators are beginning to map the regulatory networks that govern gene expression throughout living cells. For example, a mathematical model of the regulatory network that controls the lysis/lysogeny decision in bacteriophage-λ has been constructed [1,2]. Regulatory networks in eukaryotic cells are much more complex than networks in bacteriophage: even the simple eukaryote *Saccharomyces cerevisiae* has >200 proteins that regulate transcription of its ~6200 genes.

Our ability to map gene-regulatory networks in eukaryotic cells has been enhanced greatly by the sequencing of genomes and the development of new tools to study genome expression. The yeast *S. cerevisiae* was the first eukaryote to have its genome sequenced [3], and has proven to be a workhorse for functional genomics. The past few years have seen an explosion of genome-wide expression data from yeast cells exposed to dozens of different environmental stimuli (e.g. see [4,5]) or deleted for one of many hundred different genes (e.g. see [6]). More recently, DNA microarrays have been used to profile the genomic binding sites of transcription factors and other DNA-binding proteins [7••–9••,10]. Computational algorithms have been developed that identify potential regulatory sequences in promoter regions throughout the genome (reviewed in [11]). In this review, we discuss the contributions of expression analysis, genome-location studies, and computational analysis of promoter sequence elements to our understanding of gene-expression regulatory networks in *S. cerevisiae*.

Genome-wide expression analysis

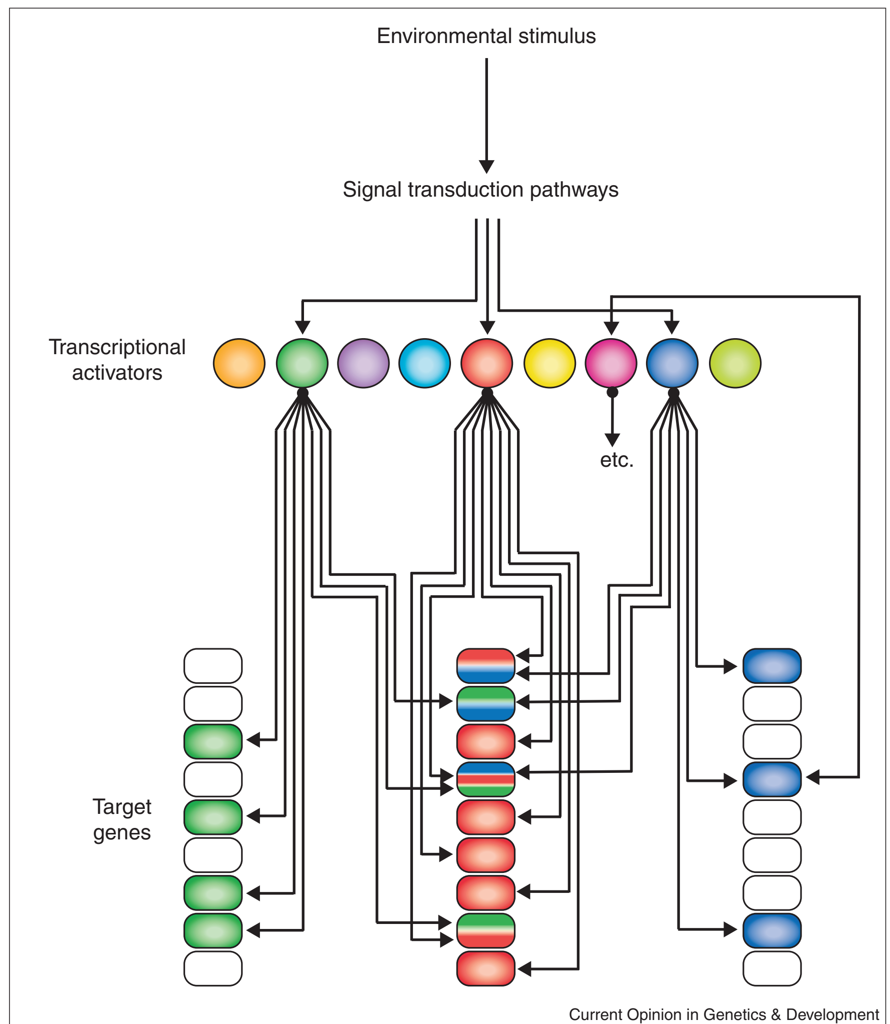
Genome-wide expression analysis involves the use of oligonucleotide or cDNA microarrays to measure, in a massively parallel fashion, the mRNA levels of many or all genes in a genome [12–15] (see Figure 2a). Genome-wide expression analysis has been used to investigate the regulatory networks controlling a variety of cellular processes in yeast, including the cell cycle [16–18], phosphate metabolism [19], galactose metabolism [20•], zinc metabolism [21], copper ion homeostasis [22], amino acid biosynthesis [23], sporulation [24,25], glucose repression [26], response to pheromone [27], and the general stress response [4,5]. This trove of data has been analyzed with a variety of clustering and pattern-finding algorithms to group together genes with similar patterns of expression [28–31].

As mRNA levels are the output of gene-expression regulatory networks, it is theoretically possible to use expression data to reverse engineer the architecture of the controlling regulatory networks. A number of groups have tackled this problem using singular value decomposition (SVD) analysis [32–35]. In these studies, SVD analysis was used to find underlying patterns or 'modes' in expression data, with the intention of linking these modes to the action of transcriptional regulators. An alternative approach is to use prior knowledge of the regulatory network's architecture to design competing models, and then use Bayesian belief networks to pick the model that best fits the expression data [36]. Gifford and co-workers have used this approach to distinguish between two competing models for galactose regulation [37]. Friedman and co-workers have used Bayesian networks to analyze genome-wide expression data in order to identify significant interactions between genes in a variety of metabolic and regulatory pathways [38,39].

Genome-wide location analysis

The information provided by expression analysis is the product of all the regulatory events that impinge on gene expression. To understand how genes are controlled by transcriptional regulatory proteins, an additional, more direct measure is needed. To this end, a number of techniques have been developed to identify the genomic

Figure 1



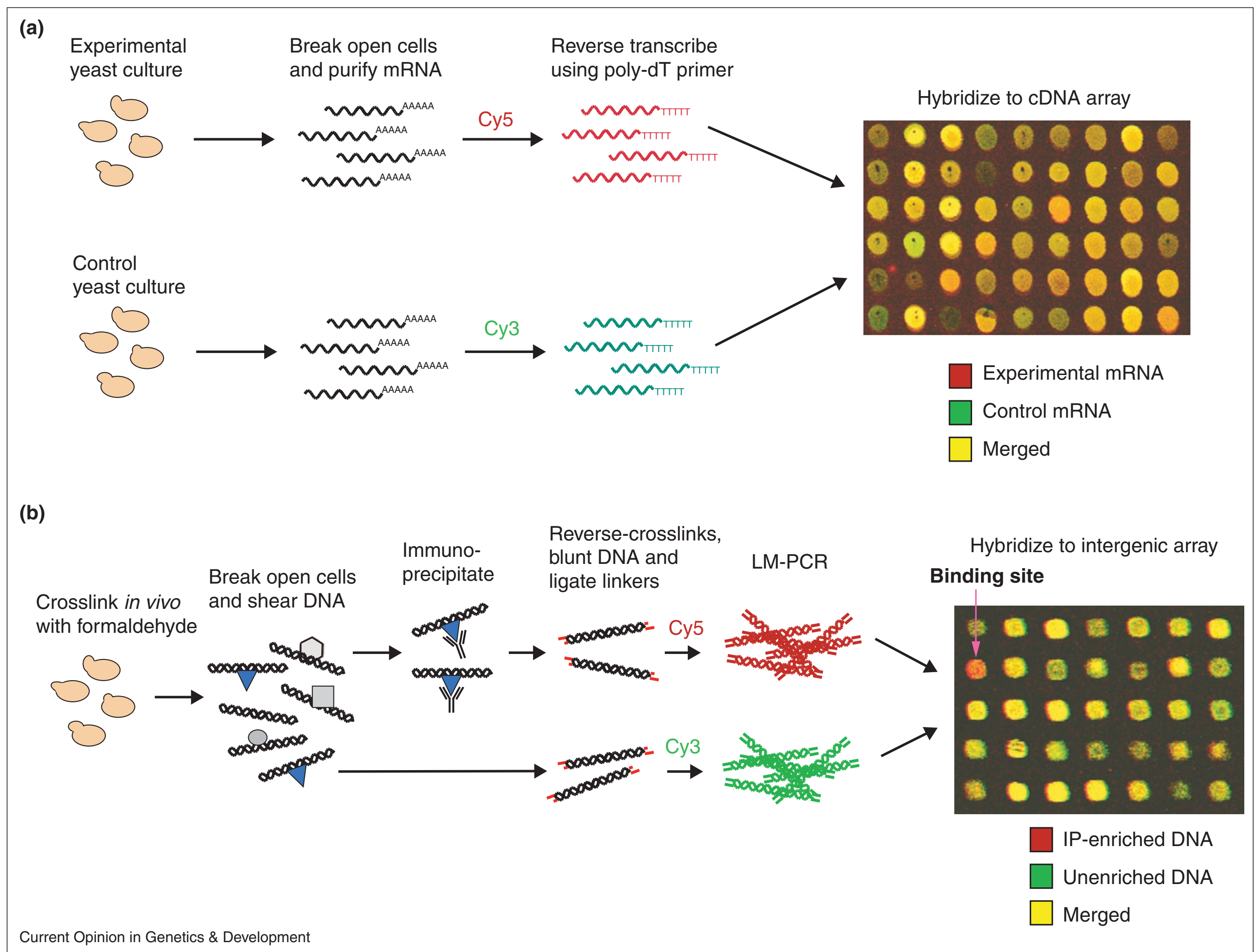
A model gene expression regulatory network. The colored circles represent distinct transcriptional activators. The rectangular ovals represent potential target genes in the genome. The color of the rectangular oval indicates which transcriptional activator is regulating its expression in response to the environmental stimulus; in addition, arrows point from each transcriptional activator to its regulated genes. Note that this model can be thought of as an individual regulatory network or as a collection of regulatory networks.

binding sites of transcriptional regulators [7••,8••,40–42]. We believe that the most powerful of these methods is genome-wide location analysis [7••,8••]. This approach combines a chromatin immunoprecipitation protocol, which has been previously used to study protein–DNA interactions at a small number of specific DNA sites [43], with DNA microarray analysis (Figure 2b). DNA is enriched by immunoprecipitation with an antibody against the transcription factor of interest. Enriched and un-enriched (control) DNA is labeled with Cy5 and Cy3 dyes, respectively, and then is hybridized to DNA microarrays containing the complete set of yeast intergenic regions. Intergenic regions enriched for Cy5 fluorescence relative to Cy3 correspond to transcription factor binding sites.

The binding of a transcriptional activator to the promoter region of a gene suggests that the activator has a regulatory effect on the gene, but it is also possible that the activator does not fully or even partially control the gene. Identifying the set of promoters where factor binding correlates with gene-expression increases the probability that the factor binding site is associated with adjacent gene expression [7••,9••]. This approach also helps resolve potential ambiguities that result from the fact that some of the intergenic regions contain the promoters for two divergently transcribed genes. Furthermore, the sets of promoters/genes found within the intersection of binding and expression datasets contains few false positives, because the noise that is inherent in each of the microarray datasets tends to be either reduced or eliminated within this intersection.

Initial studies using genome-wide location analysis focused on Gal4 [7••]. Gal4 is a well-characterized transcriptional activator that regulates genes required for galactose metabolism. Genome-wide location analysis of Gal4 identified a set of ten genes that were bound by Gal4 and whose expression was induced in galactose media. All seven of the known Gal4 target genes were identified, in addition to three novel target genes (*PCL10*, *MTH1* and *FUR4*), which are functionally linked to galactose metabolism. The fact that a single genome-wide location experiment

Figure 2



Genome-wide expression and location analyses **(a)** Schematic summary of the technical steps involved in genome-wide expression analysis using cDNA microarrays. **(b)** Schematic summary of the technical steps involved in genome-wide location analysis. The purple arrow points to a spot where the red intensity is over-represented, identifying a region bound *in vivo* by the protein under investigation.

correctly identified each of the Gal4 targets determined from much previous study attests to the power of the method.

Several interesting themes have emerged from the location analysis experiments that have been reported to date. One striking observation is that the sequence which is bound by a specific factor occurs at many more sites in genomic DNA than are actually bound by the factor. For example, the Gal4 consensus binding motif ($CGGN_{11}CCG$), determined from structural studies [44], is found in the promoters of >200 genes (*Saccharomyces cerevisiae* Promoter Database: http://cgsigma.cshl.org/jian/; [45]), but Gal4 binds to and regulates only 10 of these promoters *in vivo* [7••]. Similar observations have been made for other transcriptional activators [8••,46•], suggesting that selective binding to a subset of potential target sequences *in vivo* is a general feature of transcriptional activators. Lieb *et al.* [46•] have investigated this issue using the transcriptional activator Rap1 as a model. They surveyed the yeast genome for motifs that fit the Rap1 consensus binding sequence ACACCCRYACAYM. Although many of these matches were identified in intergenic sequences, a substantial number were located in open-reading frames (ORFs). However, genome-wide location analysis of Rap1 revealed that few of these ORFs (14%) were bound *in vivo*. Among intergenic regions, Rap1 showed a preference for binding to motifs found in promoters (46%) relative to motifs found in non-promoter intergenic regions (17%). This binding preference could not be accounted for by either the sequence or frequency of the motifs in promoter regions, indicating that the specificity of Rap1 binding is determined, in part, by some unknown genome-wide mechanism.

How do cells modify the binding specificity of transcriptional activators? Two general mechanisms [47] have been proposed: first, through cooperative binding of multiple

transcriptional activators with different binding specificities, or second by inhibiting binding at selective sites through the formation of repressive chromatin structure. The first mechanism parallels studies in mammalian systems in which multiple transcription factors bind cooperatively to a promoter as an enhanceosome [48]. The second mechanism is particularly attractive in light of the recently described histone code, in which it has been proposed that covalent modifications of histone amino-terminal tails serve as a fundamental mechanism regulating protein binding [49]. Whether either (or both) of these mechanisms is used in yeast cells to modify transcription factor binding remains an open question.

A second theme emerging from genome-wide location studies is that transcriptional activators bind to a subset of their target promoters even in non-inducing conditions. This has been demonstrated most clearly for the transcriptional activator Ste12, which regulates pheromone-responsive genes. Genome-wide location analysis of Ste12 following pheromone treatment identified 29 Ste12 target genes, which function in a variety of cellular processes involved in mating [7••]. Ste12 binds to the promoters of a subset of its target genes (e.g. *STE12*, *PCL2*, *FIG2* and *FUS1*) prior to pheromone treatment, and the expression of these genes is induced immediately following pheromone treatment [7••,27]. Gal4 [7••] and Rap1 [46•] are also pre-bound to a subset of their target promoters, indicating that this may be a general feature of transcriptional activators. The pre-bound target genes are not activated in non-inducing conditions because the negative regulators Dig1/Dig2 and Gal80 act to inhibit the function of Ste12 and Gal4, respectively, prior to induction. Interestingly, Ste12 binds to its own promoter and induces its transcription immediately following pheromone treatment. This may explain the increase in the number of sites bound by Ste12 following pheromone treatment, and is an interesting example of a positive feedback loop in gene-expression networks.

DNA motif-finding algorithms

A gene's expression pattern is largely determined by short promoter sequences (or motifs) that serve as transcription-factor binding sites. Hence, identifying and characterizing regulatory motifs is an important step in deciphering gene-expression regulatory networks. Two general approaches have been used to identify conserved regulatory motifs: first, comparing promoter DNA sequences of closely related species, and second, comparing promoter DNA sequences of co-regulated genes in the same species. The first of these methods takes advantage of the continuing flood of genomic sequence information to identify evolutionarily conserved promoter sequences, with the assumption that these sequences define potential transcription-factor binding sites. A study by Johnston and co-workers [50•] compared small segments of genomic sequence from several closely related Saccharomyces species. From this analysis, they identified a number of conserved DNA motifs, many of which had been previously identified as transcription factor binding sites. These encouraging results will hopefully spur efforts toward obtaining complete genome sequences for this family of *Saccharomyces* species, so this analysis can be extended genome-wide.

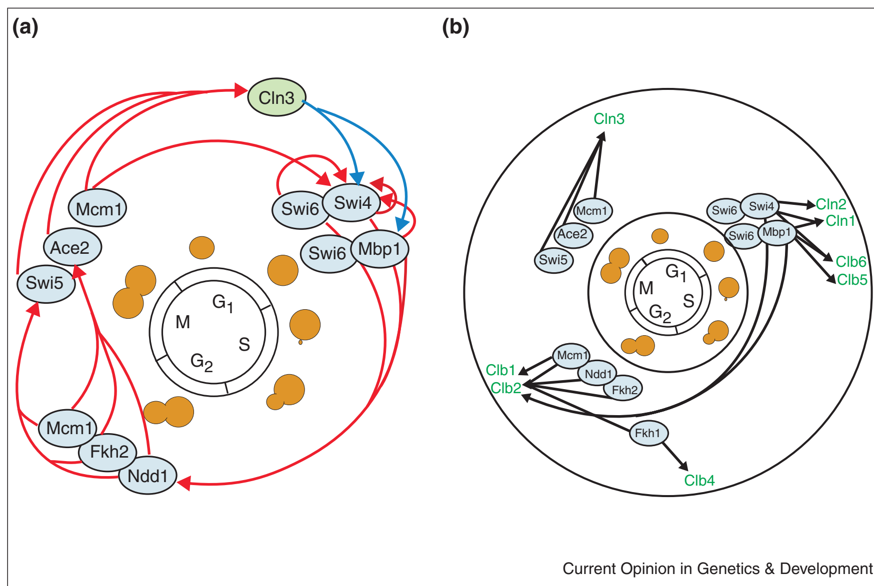
The second method takes advantage of a variety of computational algorithms that have been devised to identify conserved sequence motifs in the promoters of co-regulated genes (reviewed in [11]). In many of these studies, groups of co-regulated genes are defined by clustering analysis of genome-wide expression data sets, and then the promoter sequences of these gene clusters are analyzed for statistically over-represented DNA motifs [16,51–53]. Alternatively, potential DNA motifs can be classified by their genome-mean expression profile (GMEP), which is calculated by averaging the expression profiles of all genes that contain the motif in their promoters [54]. DNA motifs with coherent, non-random GMEPs are likely to represent functional regulatory motifs.

These motif-finding algorithms focus on detecting individual regulatory elements. Many eukaryotic genes, however, are bound by multiple transcription factors that act synergistically to regulate transcription. Two recent papers [55••,56•] describe computational algorithms designed to identify combinations of regulatory motifs. Bussemaker *et al.* [56•] used a linear model, in which regulatory motifs contribute additively to the expression level of a gene, to analyze the cell cycle and sporulation expression data sets. Pilpel *et al.* [55••] searched a database of putative regulatory motifs for motif pairs that had synergistic GMEPs. The identified motif pairs were used to build motif synergy maps and to analyze the causal relationship between individual motifs and expression patterns. Interestingly, they found that combinations of a small number of regulatory motifs could account for a complex set of expression patterns.

Deciphering the cell cycle regulatory network

A recent study of the yeast cell cycle illustrates the great potential of using genomic tools to decipher gene-expression regulatory networks [9••]. In this study, a combination of genome-wide location and expression analysis was used to investigate how the nine known cell cycle transcriptional activators (Mbp1, Swi4, Swi6, Mcm1, Fkh1, Fkh2, Ndd1, Swi5, and Ace2) regulate the expression of ~800 cell-cycle genes. The data reveal that distinct sets of these nine transcriptional activators regulate genes expressed in different stages of the cell cycle. Mbp1, Swi4, and Swi6 bind predominately to the promoters of late G_1 genes, Mcm1, Fkh2, and Ndd1 to G2/M genes, and Swi5 and Ace2 to M/G_1 genes. Fkh1, on the other hand, binds to genes expressed in the G_1, S, and G_2/M phases of the cell cycle. Strikingly, the data also revealed that each of these sets of stage-specific transcriptional activators also regulates the expression of one or more activators involved in the next stage of the cell cycle, forming a fully connected regulatory network (see Figure 3a). In addition, the

Figure 3



Serial regulation of transcription regulator and cyclin genes during the yeast cell cycle (based on Simon *et al.* [9••]). **(a)** Model of the closed circuit of transcriptional activators regulating cell-cycle progression. Blue-shaded ovals represent the transcriptional activators; the green-shaded oval represents the cyclin Cln3. The red arrows indicate binding of a transcriptional activator to the promoter of another regulatory factor. The blue arrows represent post-transcriptional regulation by Cln3/Cdc28. Each stage-specific set of activators (e.g. Mcm1, Fkh2, Ndd1) regulate the expression of one or more activators (e.g. Swi5, Ace2) involved in the next stage of the cell cycle. **(b)** Model of transcriptional regulation of cyclin genes during cell-cycle progression. Black arrows indicate binding of a transcriptional activator to the promoter of a cyclin gene. Each stage-specific set of transcriptional activators regulate key cyclin genes needed for progression through the cell cycle.

expression of a variety of protein regulators (e.g. cyclins) of the cell cycle is also regulated by these sets of stage-specific transcriptional activators (see Figure 3b). For example, the Mcm1/Fkh2/Ndd1 complex regulates the M/G_1 transcriptional activators *SWI5* and *ACE2*, and the cyclins *CLB1* and *CLB2*, which promote entry into mitosis [57]. Hence, a key insight from this study is that the cell cycle is regulated by a connected circular network of transcriptional activators.

An interesting feature of the yeast cell cycle transcriptional regulatory network is that pairs of activators exhibit partially redundant functions. These transcriptional activator pairs (e.g. Swi4 and Mbp1, Fkh1 and Fkh2, Ace2 and Swi5) have similar DNA-binding domains, similar binding sequences, but only partially overlapping target genes *in vivo* [9••]. Previous models of cell-cycle regulation suggested that each of these transcriptional activator pairs shares its function with its partner (i.e. they are functionally redundant), as each of these six transcriptional activator genes are not essential for cell-cycle progression. A striking example of this is seen for the *FKH1* and *FKH2* genes. Deletion of either gene individually has little effect, but deletion of both results in striking changes in the expression of a number of cell cycle regulated genes [18], suggesting that Fkh1 and Fkh2 regulate the same sets of target genes *in vivo*. However, genome-wide location data reveal that in wild-type cells, Fkh1 and Fkh2 share only 22% of their target genes, and hence are only partially redundant. Presumably, this partial redundancy allows Fkh1 and Fkh2 to substitute for each other when one has been deleted, yet be responsible for distinct functions in wild-type cells. An interesting possibility is that cells evolved pairs of regulators with overlapping target genes to ensure smooth transitions during the cell cycle. If a single activator regulated each stage of the cell cycle, there would be greater potential for disruptive changes in the gene-expression program.

Conclusions

It seems likely that, given the recent flurry of genome-wide location and expression papers this past year, the functional target genes of all yeast transcriptional activators might be identified in the very near future. Such data would provide a foundation for a complete map of the regulatory networks controlling gene expression in a eukaryotic cell. Analyzing these data, however, will be a considerable challenge, and new and more powerful computational tools are needed to build a gene-expression network using these data. Hence, the interface between mathematical modeling and DNA microarray experiments should prove to be fertile ground in the years ahead.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. McAdams HH, Shapiro L: **Circuit simulation of genetic networks.** *Science* 1995, **269**:650-656.

2. Arkin A, Ross J, McAdams HH: **Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells.** *Genetics* 1998, **149**:1633-1648.

3. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M *et al.*: **Life with 6000 genes.** *Science* 1996, **274**:563-567.

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

5. Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA: **Remodeling of yeast genome expression in response to environmental changes.** *Mol Biol Cell* 2001. **12**:323-337.

6. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD *et al.*: **Functional discovery via a compendium of expression profiles.** *Cell* 2000, **102**:109-126.

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

See annotation [8••].

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

These papers [7••,8••] provide the first descriptions of the genome-wide location analysis technique. Ren *et al.* [7••] use genome-wide location analysis to define the target genes of the yeast transcriptional activators Gal4 and Ste12. Iyer *et al.* [8••] use this technique to identify the genes regulated by the cell-cycle transcription factors SBF and MBF.

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

This paper illustrates the great potential of using genome-wide location and expression analysis to decipher gene-expression regulatory networks. The authors investigate how the nine yeast cell cycle transcriptional activators regulate the cell cycle gene expression program. These data reveal that the cell cycle is regulated by a connected circular network of transcriptional activators.

10. Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM: **Genome-wide location analysis of ORC and MCM proteins: high-resolution mapping of replication origins in *S. cerevisiae*.** *Science* 2001, **294**:2357-2360.

11. Ohler U, Niemann H: **Identification and analysis of eukaryotic promoters: recent computational approaches.** *Trends Genet* 2001, **17**:56-60.

12. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-470.

13. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW: **Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.** *Proc Natl Acad Sci USA* 1996, **93**:10614-10619.

14. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H *et al.*: **Expression monitoring by hybridization to high-density oligonucleotide arrays.** *Nat Biotechnol* 1996, **14**:1675-1680.

15. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.

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

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

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

19. Ogawa N, DeRisi J, Brown PO: **New components of a system for phosphate accumulation and polyphosphate metabolism in *Saccharomyces cerevisiae* revealed by genomic expression analysis.** *Mol Biol Cell* 2000, **11**:4309-4321.

20. • Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science* 2001, **292**:929-934.

The authors of this paper describe an innovative approach to model cellular pathways. They use genome-wide expression analysis, proteomics, and a database of two-hybrid protein interactions to model the galactose metabolic pathway.

21. Lyons TJ, Gasch AP, Gaither LA, Botstein D, Brown PO, Eide DJ: **Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast.** *Proc Natl Acad Sci USA* 2000, **97**:7957-7962.

22. Gross C, Kelleher M, Iyer VR, Brown PO, Winge DR: **Identification of the copper regulon in *Saccharomyces cerevisiae* by DNA microarrays.** *J Biol Chem* 2000, **275**:32310-32316.

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

24. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I: **The transcriptional program of sporulation in budding yeast.** *Science* 1998, **282**:699-705.

25. Primig M, Williams RM, Winzeler EA, Tevzadze GG, Conway AR, Hwang SY, Davis RW, Esposito RE: **The core meiotic transcriptome in budding yeasts.** *Nat Genet* 2000, **26**:415-423.

26. Lutfiyya LL, Iyer VR, DeRisi J, DeVit MJ, Brown PO, Johnston M: **Characterization of three related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*.** *Genetics* 1998, **150**:1377-1391.

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

28. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998. **95**:14863-14868.

29. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: **Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation.** *Proc Natl Acad Sci USA* 1999, **96**:2907-2912.

30. Roth FP, Hughes JD, Estep PW, Church GM: **Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation.** *Nat Biotechnol* 1998, **16**:939-945.

31. Niehrs C, Pollet N: **Synexpression groups in eukaryotes.** *Nature* 1999, **402**:483-487.

32. Alter O, Brown PO, Botstein D: **Singular value decomposition for genome-wide expression data processing and modeling.** *Proc Natl Acad Sci USA* 2000, **97**:10101-10106.

33. Holter NS, Maritan A, Cieplak M, Fedoroff NV, Banavar JR: **Dynamic modeling of gene expression data.** *Proc Natl Acad Sci USA* 2001, **98**:1693-1698.

34. Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR, Fedoroff NV: **Fundamental patterns underlying gene expression profiles: simplicity from complexity.** *Proc Natl Acad Sci USA* 2000, **97**:8409-8414.

35. Raychaudhuri S, Stuart JM, Altman RB: **Principal components analysis to summarize microarray experiments: application to sporulation time series.** *Pac Symp Biocomput* 2000, 455-466.

36. Gifford DK: **Blazing pathways through genetic mountains.** *Science* 2001, **293**:2049-2051.

37. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA: **Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks.** *Pac Symp Biocomput* 2001, 422-433.

38. Pe'er D, Regev A, Elidan G, Friedman N: **Inferring subnetworks from perturbed expression profiles.** *Bioinformatics* 2001, **17(Suppl 1)**:S215-S224.

39. Friedman N, Linial M, Nachman I, Pe'er D: **Using Bayesian networks to analyze expression data.** *J Comput Biol* 2000. **7**:601-620.

40. Bulyk ML, Gentalen E, Lockhart DJ, Church GM: **Quantifying DNA–protein interactions by double-stranded DNA arrays.** *Nat Biotechnol* 1999, **17**:573-577.

41. Bulyk ML, Huang X, Choo Y, Church GM: **Exploring the DNA-binding specificities of zinc fingers with DNA microarrays.** *Proc Natl Acad Sci USA* 2001, **98**:7158-7163.

42. van Steensel B, Delrow J, Henikoff S: **Chromatin profiling using targeted DNA adenine methyltransferase.** *Nat Genet* 2001, **27**:304-308.

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

44. Marmorstein R, Carey M, Ptashne M, Harrison SC: **DNA recognition by GAL4: structure of a protein-DNA complex.** *Nature* 1992, **356**:408-414.

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

46. Lieb JD, Liu X, Botstein D, Brown PO: **Promoter-specific binding of**
• **Rap1 revealed by genome-wide maps of protein–DNA association.** *Nat Genet* 2001, **28**:327-334.

The authors use genome-wide location analysis to identify the binding sites of the repressor–activator protein Rap1, and the silencing proteins Sir2, Sir3, Sir4. They demonstrate that Rap1 binds preferentially to sequence motifs located in promoter-containing intergenic regions.

47. Biggin MD: **To bind or not to bind.** *Nat Genet* 2001, **28**:303-304.

48. Merika M, Thanos D: **Enhanceosomes.** *Curr Opin Genet Dev* 2001, **11**:205-208.

49. Jenuwein T, Allis CD: **Translating the histone code.** *Science* 2001, **293**:1074-1080.

50. Cliften PF, Hillier LW, Fulton L, Graves T, Miner T, Gish WR,
• Waterston RH, Johnston M: **Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis.** *Genome Res* 2001, **11**:1175-1186.

Comparative sequence analysis of genomic regions from a number of related *Saccharomyces* yeast species reveals conserved promoter elements that in many cases have regulatory function.

51. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22**:281-285.

52. Wolfsberg TG, Gabrielian AE, Campbell MJ, Cho RJ, Spouge JL, Landsman D: **Candidate regulatory sequence elements for cell cycle-dependent transcription in *Saccharomyces cerevisiae*.** *Genome Res* 1999, **9**:775-792.

53. Jelinsky SA, Estep P, Church GM, Samson LD: **Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes.** *Mol Cell Biol* 2000, **20**:8157-8167.

54. Chiang DY, Brown PO, Eisen MB: **Visualizing associations between genome sequences and gene expression data using genome-mean expression profiles.** *Bioinformatics* 2001, **17(Suppl 1)**:S49-S55.

55. Pilpel Y, Sudarsanam P, Church GM: **Identifying regulatory**
•• **networks by combinatorial analysis of promoter elements.** *Nat Genet* 2001, **29**:153-159.

This paper outlines a novel computational algorithm to find synergistic motif combinations in the promoters of co-regulated genes. The authors apply this algorithm to yeast expression data sets from the cell cycle, sporulation, and stress response, in order to map the connections between the transcriptional regulators controlling these biological processes. They find that these maps are highly interconnected, indicating that combinatorial transcriptional regulation is a feature of yeast regulatory networks.

56. Bussemaker HJ, Li H, Siggia ED: **Regulatory element detection**
• **using correlation with expression.** *Nat Genet* 2001, **27**:167-171.

The authors describe a computational method for identifying new regulatory motifs. In this method, genome-wide expression data is fit with a linear model in which multiple DNA motifs can contribute additively to a gene's expression level. They test this method on the diauxic shift, cell cycle, and sporulation expression data sets, and identify several novel regulatory motifs.

57. Surana U, Robitsch H, Price C, Schuster T, Fitch I, Futcher AB, Nasmyth K: **The role of CDC28 and cyclins during mitosis in the budding yeast *S. cerevisiae*.** *Cell* 1991, **65**:145-161.