REVIEW

Constructing transcriptional regulatory networks

Alexandre Blais and Brian David Dynlacht¹

Department of Pathology, New York University Cancer Institute, New York University School of Medicine, New York, New York 10016, USA

Biological networks are the representation of multiple interactions within a cell, a global view intended to help understand how relationships between molecules dictate cellular behavior. Recent advances in molecular and computational biology have made possible the study of intricate transcriptional regulatory networks that describe gene expression as a function of regulatory inputs specified by interactions between proteins and DNA. Here we review the properties of transcriptional regulatory networks and the rapidly evolving approaches that will enable the elucidation of their structure and dynamic behavior. Several recent studies illustrate how complementary approaches combine chromatin immunoprecipitation (ChIP)-on-chip, gene expression profiling, and computational methods to construct blueprints for the initiation and maintenance of complex cellular processes, including cell cycle progression, growth arrest, and differentiation. These approaches should allow us to elucidate complete transcriptional regulatory codes for yeast as well as mammalian cells.

If you come to a fork in the road, take it.
—Yogi Berra

Cells must continually adapt to changing conditions by altering their gene expression patterns. One of the central effectors involves transcriptional regulatory interactions. The recent development of high-throughput methods and computational approaches has made it possible to survey these complex molecular interactions by modeling them as networks (for examples, see [Jeong et al. 2000, 2001; Newman 2003; Barabasi and Oltvai 2004]). Because transcription is controlled at many different levels (e.g., post-translational modification of factors, specific interactions with coactivators, thermodynamics of protein–protein, and protein–DNA interactions), it is obvious that any gene regulation network fits into a network of networks (or global network) that represents not

only transcription factor (TF)–DNA interactions but also the factors that modulate these interactions biochemically.

We focus here on transcriptional regulatory networks for two reasons. First, this area has received much attention in the past decade, due in large part to the development of high-throughput genomic approaches and an array of computational tools. In addition, the process of gene expression is often the *primum mobile*, the origin and effector of a response, wherein the information contained within a genome is interpreted and then ultimately used to produce the building blocks (proteins) required for a given response. In this review, we illustrate recent developments in the area of genomics and computational biology that have allowed several laboratories to elucidate regulatory networks in organisms as diverse as yeast and mammals. We explore how recent innovations have provided new insights into control of the mammalian cell cycle and differentiation, and we highlight both the caveats and future prospects of these approaches.

Properties of biological networks

Delineating the topology and dynamics of biological networks tells us a great deal about how these networks originate and how they enable the cell to respond to its environment and perform complex biological functions. For an extensive discussion of the principles underlying biological networks, we refer the reader to a recent review [Barabasi and Oltvai 2004]. Biological networks are usually depicted as nodes connected by edges. Nodes represent proteins, genes, or enzymatic substrates that translate extracellular signals from the environment. Edges often represent direct molecular interactions, regulatory interactions (such as the binding of a TF to the promoter of its target genes), or the sharing of functional properties. One important characteristic of biological networks is their scale-free structure: The number of nodes that make a large number of connections with other nodes (referred to as “hubs”) is much lower than the number of nodes with few connections. This is thought to confer a hierarchical structure, whereby hubs play a central role in directing the cellular response to a given stimulus. The fact that most nodes make a small

[**Keywords:** Bioinformatics; ChIP-on-chip; expression profiling; transcriptional regulatory networks]

¹**Corresponding author.**

E-MAIL brian.dynlacht@med.nyu.edu; FAX (212) 263-6157.

Article and publication are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1325605>.

number of connections renders a biological network more robust (less sensitive to random perturbations), although at the same time making it very sensitive to directed inactivation of a critical hub. Another aspect of scale-free biological networks is that they constitute “ultra-small worlds,” because only a few steps are necessary to join any two nodes (fewer than in randomly organized networks). This presumably facilitates the efficient propagation and integration of signals. One other notable characteristic of biological networks is the relative paucity of hubs that connect directly to one another. This propensity of biological networks distinguishes them from other real-world networks (such as social interaction), where hubs tend to interconnect. We can envision two possible explanations for this observation. One is that there might be a size limit beyond which a hub, while still being functional, renders the whole network too sensitive to directed inactivation. A more pragmatic explanation is that, since a cell carries out many distinct biological processes, it may need a certain level of compartmentalization that cannot be achieved if everything is directly connected.

Network motifs

Although a network may be modeled to describe all possible regulatory interactions occurring under any condition, it is more practical to study in great detail smaller portions of the network that can be considered autonomous. Such a subnetwork unit is referred to as a module, where nodes are connected functionally or physically. The nodes may represent the set of genes that share a common regulatory TF or that are expressed under the same specific set of conditions. Studies of the regulatory networks governing cell cycle progression and myogenesis provide examples (see below). We can reduce the complexity of a network further by considering its motifs (Fig. 1; Odom et al. 2004; Yeager-Lotem et al. 2004).

Figure 1. Transcriptional regulatory network motifs. Depiction of the most common motifs in transcriptional networks. Similar motifs were described previously (Lee et al. 2002).

Network motifs describe how single nodes connect with their neighbors. Examples include the single-input motif, which describes the connection between a target gene and its sole transcriptional regulator; the multiple-input motif, in which a target gene is regulated by a group of factors; and the feed-forward loop, in which the product of one TF regulates the expression of a second TF, and both factors together regulate the expression of a third gene. Network motifs, by their intrinsic behavior, help us understand how networks oversee different tasks, and different motifs predominate depending on the type of network or module (Yeager-Lotem et al. 2004). For example, a transcriptional regulatory module dominated by single-input motifs has a simple structure and is expected to have an “all-or-none” response, whereas a module or subnetwork in which multiple-input motifs predominate will be expected to have a more subtle and graded response. Networks characterized by multiple feed-forward loops tend to be stable rather than transient (Yeager-Lotem et al. 2004).

Networks undergo condition-specific rewiring

It is necessary to understand both the topology of a network (interconnectivity of nodes) and how this topology changes with time or environmental conditions, since not all nodes are active at any given time. The dynamics of a global network have recently been examined computationally in yeast, where a majority of TF hubs were identified as active in more than one specific physiological setting, although few were active in all settings (Luscombe et al. 2004). The terms “endogenous” and “exogenous” have been introduced to describe network components that regulate processes in very different ways. Endogenous subnetworks are defined as regulatory structures controlling processes that are temporally complex and intrinsic to the cell (examples include cell cycle and sporulation). They are characterized by a multistage architecture of their regulatory network. The TF hubs that regulate them have a relatively small number of targets, which often tend to be other TFs, and this tendency generates high local interconnectivity. These hubs are generally somewhat distant from the “terminal effectors” of these processes, being separated by several nodes. All of these properties suggest that these processes are regulated in a complex manner and over a relatively long period of time. On the other hand, exogenous subnetworks are established to allow the cell to respond more quickly to a variety of stimuli, such as drastic environmental changes. These regulatory networks generally involve relatively few TFs. However, these factors have a large number of targets, which are often the “terminal effectors” that coordinate the cell’s response to stimuli.

The work by Harbison et al. (2004) experimentally confirms these concepts. By conducting a large number of location analysis experiments on yeast TFs in a number of different experimental conditions, it is possible to analyze how the compendium of target genes changes under various circumstances. TFs can be classified into four categories based on their ability to recognize their

targets: Condition-invariant factors bind the same set of targets under any condition, while condition-enabled factors bind targets only under certain circumstances, condition-expanded factors bind additional targets in specific circumstances, and condition-altered factors bind different targets under distinct circumstances. This study underscores the dynamic behavior of transcriptional regulatory networks and provides a basis for understanding how regulatory networks can be rewired in a condition-dependent manner.

Constructing transcriptional regulatory networks

In order to understand the topology and dynamics of transcriptional regulatory networks governing biological processes such as the cell cycle or differentiation, approaches are devised to evaluate (1) the identity and expression level of interacting nodes, (2) how interactions change with time (e.g., through a cell cycle or during differentiation), and (3) the phenotypic impact of disrupting key nodes. The complexity of the eukaryotic transcriptional regulation machinery reflects the multitude of responses that it controls and makes elucidation of transcriptional regulatory networks a very difficult task. This leads to obvious questions regarding the mechanisms by which a specific transcriptional response is elicited, including how a given signaling pathway activates a particular TF, how temporal specificity is generated, and the origins of target specificity. It is thus presently difficult, if not impossible, to accurately account for all levels of regulation, and therefore, some assumptions are made. For example, it is often assumed that the steady-state level of an mRNA (measured in an expression profiling experiment using DNA microarrays) is indicative of the rate of transcription or of the level of protein translated from that mRNA. Further, it is often assumed that if a TF is expressed, it is active, although it is clear that dimerization, post-translational modifications, subcellular localization, and other factors must also be considered.

Recently, much progress has been made toward the development of methods that take into account some of the considerations described above. The most important contributions come from genomics, and two approaches have contributed substantially to the elucidation of regulatory networks: genome-wide expression profiling and the combination of chromatin immunoprecipitation (ChIP) with promoter DNA microarrays (known as ChIP-on-chip, ChIP chip, or location analysis), which identifies direct target genes under a given set of conditions. The use of expression profiling to construct gene regulatory networks has been reviewed previously (Banerjee and Zhang 2002; Ihmels et al. 2002, 2004; Bergmann et al. 2004; Siggia 2005). A third approach, genome-wide RNA interference (RNAi) screens, will also substantially contribute to our ability to construct global transcriptional regulatory networks (for review, see Baum and Craig 2004). Other technical innovations include the indirect assessment of transcription rate by measurement of mRNA decay rates (Holstege et al. 1998; Wang et al.

2002; Nachman et al. 2004), and the evaluation of promoter co-occupancy by pairs of TFs (Geisberg and Struhl 2004).

One powerful approach aimed at studying regulatory networks governed by a TF of interest is depicted in Figure 2. ChIP-on-chip is performed with cells grown under a variety of conditions (e.g., different cell cycle phases, developmental stages). Expression profiling is conducted on identical populations to determine the expression levels of each node in the network and to infer the effects of TF binding on the expression of its targets. Expression profiling is subsequently performed on cells from knock-out animals or using cells that have been treated with siRNAs, since factor occupancy alone does not provide definitive functional information. Computational methods are then used to extract correlations between binding and gene expression and to generate testable predictions based on the new observations. This last point is depicted in Figure 2 (upward arrow), wherein reiterative ChIP-on-chip is used to verify predictions of combinatorial regulation. The strength of this multifaceted approach is its ability to provide complementary information that, when taken together, overcomes the weak-

Figure 2. Complementary approaches to decipher transcriptional regulatory networks. Blue boxes indicate experimental approaches, and yellow boxes indicate the knowledge obtained. Directionality of arrows suggests the order in which the experiments can be performed. (PWMs) Position-weight matrices representing TF-binding sites.

nesses inherent in each individual approach. We describe and discuss these complementary approaches below and then provide a few examples that illustrate how they can be used to elucidate the organization of complex transcriptional regulatory networks in eukaryotes.

Recent approaches aimed at elucidating transcriptional regulatory networks

ChIP-on-chip

Several approaches have recently been developed to identify genomic TF-binding sites (Table 1). ChIP-on-chip was developed first in yeast and subsequently applied to mammalian cells and flies. Here, cells are grown under various conditions and fixed with a reversible cross-linker. Chromatin is sonicated and enriched with antibodies against a specific TF, DNA is purified and labeled in parallel with DNA derived from input chromatin or chromatin “enriched” with a negative control antibody, and both samples are hybridized to a single array containing segments of genomic DNA. ChIP-on-chip analysis has several features that make it invaluable for deciphering gene regulatory networks. First, in its simplest form, living cells that express native levels of protein are used, supplanting the need for overexpression and thereby avoiding potential loss of specificity. Second, the method focuses on *direct* interactions between regulator and target. This feature is significant because it allows us to define the number of intermediates (intervening nodes) between a TF and its target, in contrast with gene expression profiling or genetic experiments that ascribe a role for regulators in a process but that cannot distinguish direct from secondary effects. In addition, since multiple TFs, particularly those belonging to a family, may recognize the same DNA sequence, purely computational or *in vitro* approaches are prone to failure, whereas location analyses are only restricted by antibody specificity and are able to distinguish targets of highly related TFs (Cam et al. 2004; Odom et al. 2004; Blais et al. 2005; E. Balciunaite and B. Dynlacht, unpubl.). Each

of these advantages is a prerequisite for constructing an accurate regulatory network. On the other hand, ChIP-on-chip is limited by factors such as antibody accessibility to its epitope and by the fact that negative results are generally not interpretable. Moreover, knowledge of the precise location of a TF on a target promoter provides no information regarding its function.

Several innovations have been introduced to improve the accuracy of identification of TF-binding sites. Initially, ChIP-on-chip was performed using microarrays consisting of printed PCR products representing the proximal promoters of yeast (Ren et al. 2000; Iyer et al. 2001), human (Ren et al. 2002; Cam et al. 2004; Odom et al. 2004), or mouse genes (Blais et al. 2005). A few reports also made use of microarrays representing CpG islands, based on the premise that these sequences are more likely to overlap with regulatory elements or promoters (Weinmann et al. 2002; Wells et al. 2003; Kondo et al. 2004). The main caveat of using CpG island microarrays is that these loci are poorly annotated and often do not correspond to regulatory regions. On the other hand, proximal promoter arrays are also biased specifically for regions surrounding the transcription start site, preventing the identification of distal TF-binding sites. While this may not be a problem for the identification of targets of factors (such as E2F) that are known to bind close to the start site, it is problematic for proteins that recognize distant enhancers or downstream elements (such as p53, which is also known to recognize intronic sequences). Indeed, a number of studies have identified TF-binding sites located far (several or many kilobases) from 5' transcription start sites (Martone et al. 2003; Cawley et al. 2004; Euskirchen et al. 2004), further emphasizing the caveats inherent in using proximal promoter arrays. A second problem encountered with arrays of printed PCR products is their relatively low resolution, which is no greater than 1 kb if the promoter is uniquely represented by a PCR product of this size. This problem can be circumvented by two complementary approaches: the use of smaller probes (such as short PCR products or long oligonucleotides) or the use of locus tiling, whereby several probes are used to span a locus, usually with short

Table 1. Comparison of different methods aimed at identifying genomic targets

Method	Benefits	Caveats	References
ChIP on CpG island arrays	Microarrays are easy to produce	Targets often ill-defined; poor annotation; limited by antibody performance/availability; low resolution; somewhat biased	(Weinmann et al. 2002; Wells et al. 2003; Kondo et al. 2004)
ChIP on proximal promoter arrays	Straightforward	Limited by antibody performance/availability; low resolution; somewhat biased	(Ren et al. 2000, 2002; Iyer et al. 2001; Cam et al. 2004; Odom et al. 2004)
ChIP on genome tiling arrays	High resolution; less biased	High cost; requires more material or more amplification	(Cawley et al. 2004)
DamID and biotinylation methods	Circumvents the need for an antibody	Exogenously expressed factor may not behave exactly like the endogenous factor	(van Steensel and Henikoff 2000; Orian et al. 2003; Bianchi-Frias et al. 2004; Viens et al. 2004)
STAGE or SAGE	Circumvents the need for a microarray; unbiased	Subtractive hybridization may be necessary; high-throughput sequencing needed	(Kim et al. 2004; Chen and Sadowski 2005)

intervals between probes. When combined with the use of a scoring algorithm that considers the signal generated by a probe as well as its neighbors, these approaches allow a higher degree of resolution in the identification of TF-binding sites (Cawley et al. 2004).

Alternative approaches to ChIP-on-chip

In ChIP-on-chip studies, chromatin bound by a given TF (or marked by modified histone residues) is enriched with a specific antibody. Thus, antibody specificity and availability become important considerations. In yeast, this problem can be circumvented by recombining an epitope tag into any gene encoding a chromatin-associated protein. In flies, two methods that bypass this limitation involve exogenous expression of a chimeric DNA-binding protein. The first one, called DamID, entails the expression of a fusion protein linking bacterial DNA adenine methyltransferase (Dam) with the TF of interest, resulting in the methylation of DNA adjacent to its targets (van Steensel and Henikoff 2000; Orian et al. 2003; Bianchi-Frias et al. 2004). DNA is then sequentially isolated, digested with a restriction enzyme cutting only Dam-methylated DNA, size-fractionated, labeled, and hybridized to a DNA microarray for the identification and quantitation of methylated loci. One additional characteristic of the DamID method is that methylation marks are likely to have a much longer half-life than is the protein–DNA complex that generated them. While this is advantageous for signal enrichment, this property becomes a potential drawback when evaluating how recruitment of the protein to DNA is modulated with time or according to different environmental conditions.

The second approach, which thus far has been used only in gene-specific ChIP assays, involves coexpression of an *Escherichia coli* protein, biotinylating enzyme BirA, with the DNA-binding protein bearing a biotin-acceptor sequence (Viens et al. 2004). This generates an *in vivo* biotinylated DNA-bound protein that can be efficiently purified through streptavidin affinity. The major drawback of this and the DamID method is that the factors are ectopically expressed. Although a given TF may be expressed at or near physiological levels, its expression and activity profile may not perfectly reflect that of the endogenous protein. This is important if one is seeking to reveal not only the complete repertoire of targets but also the dynamic regulatory properties of that factor.

Recently, another approach adapted from serial analysis of gene expression (SAGE) was developed and applied to both yeast (Kim et al. 2004) and mammalian cells (Chen and Sadowski 2005; Labhart et al. 2005). Termed STAGE (for sequence tag analysis of genomic enrichment) or SABE (serial analysis of binding sites), it circumvents the need for microarrays to identify immunoprecipitated loci. Immunoprecipitated DNA sequence tags are concatamerized, cloned, and sequenced. Each tag represents a genomic locus, and provided genomic sequence data are available, the sequence tags can be assigned to a genomic location. This method is potentially

very important because it does not rely on microarrays, which makes it truly unbiased: It allows for the detection of protein–DNA interactions anywhere in the genome, whereas microarray-based methods are limited by the number and coverage of represented loci. However, both techniques are dependent on high-throughput sequencing, which must be extensive to ensure complete coverage and which makes it less convenient to perform time-course experiments where several samples are analyzed in parallel. Moreover, in some cases it is not possible to unambiguously assign short tags to a single genomic location. In addition, another potential drawback is the need for a subtractive hybridization step, at least in mammalian cells, where it is essential to reduce the intrinsic noise resulting from isolation of repetitive sequences during ChIP.

Although location analysis has been performed on two human tissues, liver and pancreatic islets (Odom et al. 2004), an important current limitation of ChIP-on-chip is the need for relatively large amounts of homogeneous material, preventing the use of rare cell populations, obtained through tissue dissociation, microdissection, or fluorescence-activated cell sorting (FACs). These are often the most interesting populations, since they are distinguished from other cells spatially or temporally through distinct genetic regulatory programs. Increasing the sensitivity of ChIP-on-chip, by improving immunoprecipitation and amplification efficiencies, will overcome this limitation. In addition, microarray coverage is expanding enormously, and whole-genome arrays spanning mammalian genomes should be available in the foreseeable future, necessitating the development of computational tools to analyze the abundant data.

One final point must be emphasized regarding the use of ChIP-on-chip: Knowledge of the location of a given TF does not provide information about whether the factor actually regulates a nearby gene under the prevailing conditions. There are many examples in which the recruitment of a TF does not correlate with transcriptional status (i.e., induction or repression) of its target genes (Martone et al. 2003; Blais et al. 2005). Such observations could be explained by combinatorial regulation by additional TFs or by recruitment of coactivator or corepressor proteins. Therefore, additional functional analyses are always required to complement ChIP-on-chip data and thereby achieve a more accurate depiction of regulatory networks.

The use of gene knock-outs and RNAi to identify functional regulatory interactions

It is necessary to employ approaches that complement factor location analyses to demonstrate functional interactions between a factor and its target. One approach is to alter the binding site for the factor within a target promoter, instead of ablating the factor itself. In higher eukaryotes, this is generally accomplished through the stable integration of an exogenous reporter construct into chromatin. The advantage here is that all other targets of the factor are left intact, and it is less likely to

cause undesirable secondary effects arising from genetic ablation. Alternatively, the effect of genetic ablation, suppression through RNAi, or overexpression of the TF of interest can be measured. These studies are often performed in conjunction with genome-wide expression profiling. This approach has been used widely, particularly before the advent of genomic arrays and location analysis (Muller et al. 2001; Bergstrom et al. 2002; Huang et al. 2003). The obvious drawback of this method—that it is often impossible to discern direct and secondary effects—becomes less problematic once direct targets of a factor have been identified, and the role of a given factor as activator or repressor can be deduced based on whether target gene expression is enhanced or reduced. In addition, the systematic perturbation of TFs within a pathway could, with the help of computational methods, allow construction of a regulatory network, because under these conditions it may not be necessary to distinguish primary from secondary effects as long as all putative TFs in a pathway are in turn ablated.

The use of RNAi has several advantages over a conventional knock-out approach. In addition to its adaptability to high-throughput screens necessary for elucidation of extensive regulatory networks (Friedman and Perimon 2004; Sonnichsen et al. 2005), acute ablation by RNAi could bypass compensatory mechanisms, especially relevant for examining functionally redundant TF families. Thus, RNAi was used in conjunction with expression profiling, to dissect cell cycle regulatory pathway in *Drosophila*, wherein two groups of highly related transcriptional regulators (dE2F and RBF) control gene expression (Dimova et al. 2003).

Computational approaches

Given the importance of regulatory sequences in dictating genetic programs, it is not surprising that an increasing number of studies have focused on the content of these sequences to decipher transcriptional networks. There are two DNA sequence-based approaches to the elucidation of regulatory networks. The first one relies on the prior knowledge of TF-binding site preferences, whereas the other discovers new binding sites without prior consideration of the identity of the binding factor. They are both statistical approaches that harness the power of analyzing a large number of sequences.

Predicting targets of a given TF based on promoter sequence and binding site preferences (or position-weight matrices, PWMs) involves scanning a unique sequence, a group of sequences, or a whole genome and identifying regions in which the local sequence conforms to a consensus sequence or PWM. This approach relies heavily on a number of assumptions and largely ignores redundancies in recognition by related factors, chromatin structure, and the synergistic or antagonistic contributions of other proximal and distal factors. Some studies have taken a few of these factors into account while neglecting others, and thus far, no method has been shown to predict physiological binding sites with a high degree of accuracy (Tronche et al. 1997; Wasserman and

Fickett 1998; Kel et al. 2001; Elkon et al. 2003; Fernandez et al. 2003). For example, Fernandez et al. (2003) predicted Myc targets based on primary sequence (the presence of an E-box sequence motif) then verified the accuracy of their predictions by performing large-scale, gene-specific ChIP assays. These authors found that the accuracy of predicting binding sites located near the transcription start sites of genes was considerably greater than predictions regarding distal sites. This confirmed the notion that binding site context is important in making accurate predictions. Approximately 58% of the predicted promoter E-boxes surveyed were bound by Myc under normal conditions, indicating that additional variables must be considered to make accurate predictions.

The factor-binding site discovery approach attempts to identify short sequences occurring in a group of promoters more often than by chance alone. They are termed *de novo* motif finding algorithms as they do not rely on prior knowledge of preferred TF-binding site sequences. Several algorithms exist (e.g., AlignACE, MDScan, MEME, REDUCE), and they have been reviewed recently (Tomba et al. 2005). *De novo* approaches have been used in combination with data gathered by expression profiling, ChIP-on-chip analysis, or gene function annotation. The underlying assumption is that related promoters (sharing the same expression profile or the same biological function) are more likely to be regulated by the same TF(s) and to contain a similar binding site for this factor(s). The 5' regulatory regions of yeast genes clustered by expression profiles (regulons) were scanned with a motif-discovery algorithm (AlignACE) to identify *cis*-regulatory elements involved in generating their expression profiles (Roth et al. 1998; Tavazoie et al. 1999). Further, Segal et al. (2003) used gene expression data to group genes into coexpression modules and assumed that the regulators of those groups are also transcriptionally regulated. This analysis identified modules that were highly enriched for genes involved in similar or complementary cellular processes. Sequence motifs, or groups of motifs, enriched among the promoter regions of each module were then used to deduce regulatory programs. Segal et al. (2003) found that the yeast transcriptional regulatory network is highly modular and relies on combinatorial regulation, since functionally related modules share some, but not all, of their regulatory elements. More recent methods allow the identification of *cis*-regulatory elements that, when considered together, can predict the expression profiles of regulons with impressive accuracy (Beer and Tavazoie 2004). This method clusters genes from large amounts of expression profiling data into regulons, finds overrepresented sequences, and uses Bayesian networks to deduce the relationships between expression profiles and sequence motifs. This probabilistic approach bypasses the assignment of targets to a TF based on the expression of the factors and allows one to factor in additional parameters, such as location of factors determined by ChIP-on-chip, if they are available. The approaches mentioned here use large data sets obtained under many experimental conditions and identify, by reverse engineering, the regulatory ele-

ments dictating their expression pattern. However, a more useful algorithm able to predict the expression of a gene in a given condition based on its promoter sequence has not yet been described.

Integrated approaches

A number of factors can improve these two sequence-based computational approaches. First, the accuracy of their predictions is often increased by evaluating the presence of “strings” of TF motifs, or modules. This is based on the notion that specific gene expression patterns often result from the combined action of several TFs (Pilpel et al. 2001; Sharan et al. 2003; Kato et al. 2004). Second, consideration of the phylogenetic conservation of binding sites can enhance the accuracy of sequence-based predictions. Several methods have been described that allow the evaluation of interspecies conservation of promoter elements (Harbison et al. 2004; Sinha et al. 2004; Dieterich et al. 2005; Elemento and Tavazoie 2005). This approach relies on the premise that biologically important TF-binding sites are more likely to be retained during evolution than nonfunctional sequences. Third, combining sequence-based approaches with ChIP-on-chip analysis can substantially improve accuracy because ChIP-on-chip provides direct evidence of physical binding to a genomic location. Thus, by comparing a large number of sequences bound by a protein, *de novo* motif finding algorithms can help determine what is the preferred DNA-binding sequence of a poorly characterized TF for which location data are available. Moreover, detailed analysis of the promoters bound by a given factor may reveal the presence of binding sites for additional TFs, thereby suggesting combinatorial regulation partners. In Figure 3, we depict the result of ChIP-on-chip analysis of a TF (red diamonds, left) coupled with expression profiling (right). In the simplest case (Fig. 3A), there is a direct correlation between the recruitment of a TF and the induction of its target genes (Regulon A), because both binding and induction occur in condition #2. Thus, it is deduced that binding of this factor causes the induction of Regulon A. In cases where there is no correlation (Fig. 3B, where the TF is bound in both conditions, but Regulon B is induced in condition #2), binding of additional factors and combinatorial regulation could explain the change in expression of Regulon B. Here, analysis of promoter sequence and expression profiles might suggest the presence of an additional activating factor at the promoters of Regulon B (Fig. 3C). Alternatively, there could be a binding site for a transcriptional repressor in the promoters of Regulon B that would bind only in condition #1 (Fig. 3D), resulting in condition-specific repression. These two possibilities can easily be tested by performing ChIP-on-chip with antibodies against the newly implicated regulatory proteins (blue or green diamonds in Fig. 3C,D). Depleting these factors using RNAi and RT-PCR analysis (Fig. 2) would then confirm the model.

Recently, Harbison et al. (2004) combined an extensive amount of ChIP-on-chip data, six sequence motif finding

Figure 3. A ChIP-on-chip experiment (left) was performed in parallel with expression profiling (right) in two conditions. (A) Location analysis for a transcription factor (red diamond) identified a group of target genes (Regulon A), whose expression levels (black dots) are induced from condition #1 to #2. Since the factor binds only in condition #2, it is concluded that the factor is responsible for the induction of those genes. (B) The target genes are induced from conditions #1 to #2 (Regulon B), but the red factor binds in both conditions so that its recruitment to the target promoters cannot alone explain the induction of target genes. (C) Examination of the target promoters reveals the presence of a binding site (blue rectangle) for an additional transcription factor whose expression (right, blue dots) is itself induced from conditions #1 to #2. It is thus possible that the red and blue factors collaborate in regulating the expression of target genes in Regulon B. This hypothesis is confirmed when ChIP-on-chip is performed with an antibody against this additional transcription factor (blue diamond). (D) Examination of the target promoters in Regulon B reveals the presence of a binding site (green rectangle) for a transcriptional repressor whose expression (right, green dots) is reduced from condition #1 to #2. It is thus possible that the green factor antagonizes the effect of the red factor. ChIP-on-chip is performed with an antibody against this repressor (green diamond) to confirm this hypothesis.

algorithms, and phylogenetic conservation to construct a yeast transcriptional regulatory map. Phylogenetic comparison of sequences enriched in *de novo* motif searches across a spectrum of *Saccharomyces* species greatly improved the reliability of their results by filtering out spurious matches to preferred TF-binding site sequences. Importantly, however, this does not appear to be fail-proof, since most of the phylogenetically conserved binding site sequences were not bound by the factors in ChIP-

on-chip experiments. Since this is the most extensive factor location analysis performed to date, it strongly reinforces the notion that sequence alone cannot predict binding and that additional factors (specific conditions prevailing in the cell and recruitment of other proteins) heavily influence factor binding. When combined with computational methods, ChIP-on-chip also permitted a thorough evaluation of preferred binding site sequences and allowed Harbison et al. (2004) to evaluate target promoter architecture and draw important conclusions regarding the wiring of the yeast regulatory network. Thus, some promoters appear to have a single binding site for a unique factor (single input motifs), while others have multiple binding sites for the same or different factors. Instances in which two or more factors bind the same promoter are indicative of cooperativity or combinatorial regulation.

Another important lesson here is that the most powerful approaches toward understanding transcriptional regulatory networks are the ones that combine several strategies, merging published evidence with ChIP-on-chip and expression profiling under various conditions, genetic ablation or RNAi, and computational approaches.

Deconvoluting transcriptional regulatory networks

Genome size and the wealth of genetic data make yeast an attractive system for understanding transcriptional regulatory networks, and indeed, transcriptional control of its cell cycle was one of the first regulatory networks studied in great detail (Simon et al. 2001; Horak et al. 2002; Lee et al. 2002; Harbison et al. 2004). One approach combined location analysis data with expression of genes in each cell cycle phase to generate a dynamic picture of transcriptional regulation (Simon et al. 2001; Lee et al. 2002). The authors of these two studies assumed that transcriptional control of the cell cycle is governed by the concerted action of multiple factors and used a probabilistic algorithm that groups genes based on their expression pattern and on their coordinate binding by sets of factors. This allowed the identification of multi-input motifs refined for coexpression (MIM-CE). These motifs constitute small regulatory units that are functionally linked because (1) some TFs regulate several targets expressed at different cell cycle phases, and (2) some TFs are regulated by other factors expressed earlier in the cell cycle, thereby imposing a cyclical structure on the transcriptional network that mirrors the cell cycle. The approach of relying on MIM-CEs constitutes the basis of the GRAM algorithm and has also been used to discover gene modules within the yeast genome and to build a much larger gene regulatory network (Bar-Joseph et al. 2003).

A major drawback of several experimental approaches outlined here (genome-wide expression profiling and location analysis) is that they cannot be performed on multicellular organisms because of cell type complexity and the inability to perform genome-wide location analyses on limiting amounts of material. Nevertheless, the study

of developmental processes characterized by sequential gene activation in organisms where the fate of single cells can be traced in time and space has allowed the mapping of complex regulatory networks. One elegant example is that of endo-mesodermal specification in the sea urchin (Davidson et al. 2002, 2003). The sea urchin is an excellent model to study genetic regulation of development because its larva has a simple structure, and it is generated from the zygote after a small number of regulatory steps. Its development proceeds through spatially defined stages of gene expression modulated by extracellular cues that regulate the activity of TFs. Combinatorial regulation defines spatial territories and overlapping expression of TFs sets boundaries of gene expression. It is the existence of those boundaries that allows identical pluripotent cells to assume different developmental fates.

The approach used by Davidson and coworkers relies on *cis*-regulatory analysis: identifying the TFs and their target binding sites and assessing the significance of this binding. They used multiple tools to achieve this. Large-scale perturbation analyses were performed, such as injection of antisense oligonucleotides that prevent expression of specific genes or overexpression of genes that block specific functions or pathways. Direct and indirect effects were distinguished using phylogenetically conserved TF-binding site predictions and by perturbation rescue experiments (e.g., knock-down of a gene rescued by the exogenous expression of its downstream target). Gene expression levels were evaluated by quantitative RT-PCR or by subtractive hybridization coupled to cDNA macroarrays (a membrane-based type of array where individual clones from large cDNA libraries are printed). This enabled a genomic view of the endo-mesoderm specification network that indicates the time and location of gene expression as well as the impact on the network and the phenotypes resulting from a given perturbation. Furthermore, it also incorporates prior information and explains the impact of signaling interactions on *cis*-regulatory control mechanisms. Related approaches have been used by others to elucidate regulatory networks involved in worm, fly, and frog development (Maduro and Rothman 2002; Inoue et al. 2005; Koide et al. 2005; Levine and Davidson 2005). The ability to generate transgenic flies and worms, as well as the availability of completely sequenced genomes, are clear advantages for generating regulatory networks in these model systems.

Applying complementary approaches: regulatory networks in mammalian cells

Transcriptional control of cell cycle progression

In mammalian cells, transcriptional controls enforced by the E2F and retinoblastoma protein (pRB and the related p107 and p130 polypeptides, collectively termed pocket proteins) families of TFs play a major role in cell cycle progression (for review, see Stevaux and Dyson 2002; Cam and Dynlacht 2003; Blais and Dynlacht 2004; Bracken et al. 2004). However, a comprehensive under-

standing of the gene regulatory mechanisms that involve pRB and E2F and govern cell cycle arrest is lacking. In particular, a central question concerns the identity of E2F target genes and the precise pathways that enforce cell cycle arrest in response to growth-limiting cues. Moreover, if E2F plays a central role in cell cycle arrest, does it promote cell cycle exit via common growth regulatory networks? Our laboratory and others have begun addressing this problem systematically using ChIP-on-chip (Ren et al. 2002; Weinmann et al. 2002; Cam et al. 2004). This approach has been combined with gene expression profiling and computational approaches to (1) understand the functional relationship between E2F, pocket proteins, and their targets on a genome scale and (2) further elucidate the networks controlled by E2F and pocket proteins during the cell cycle.

Chromatin from growth arrested cells (serum starved, contact inhibited, or arrested by p16^{INK4a} overexpression) was immunoprecipitated using antibodies specific to either E2F4 or p130, and the resulting DNA was hybridized to proximal promoter microarrays containing 13,000 human genes. Parallel gene expression profiles were obtained. Results of these experiments indicated that both E2F4 and p130 are directed to the same set of targets under distinct growth arrest conditions (Cam et al. 2004). Notably, these targets were invariably repressed under each growth arrest condition. These studies suggest that E2F4 and p130 form stable, obligate repressor complexes. Since a majority of E2F4-p130 targets share a particular pattern of expression during cell cycle arrest (decreased expression), we conclude that they form a “cell cycle arrest module,” an essential element of the transcriptional regulatory network that is engaged to promote and/or sustain cell cycle exit (Fig. 4). This module is defined by both similarity of target gene expression profiles and regulation by common TFs. Use of a “centralized command structure” in which a repressor complex shuts down a large number of diverse cellular functions permits widespread, simultaneous propagation of the regulatory signal. Because the targets of E2F4 are repressed in conditions of cell cycle arrest, they cannot propagate a regulatory signal and appear as terminal nodes, even if they have TF activity. However, in proliferating cells, these same targets may be induced by activator E2Fs, allowing them to relay the regulatory signal. This suggests that this cell cycle arrest module may be relatively isolated from others at the transcriptional level in growth-arrested cells. This is in contrast to the myogenesis regulatory network (Blais et al. 2005; see below), in which many TFs are connected through multiple feed-forward loops and serial regulator chains.

The cell cycle arrest module governed by E2F4 does not constitute the smallest unit in our network (Fig. 4). For example, if we superimpose a protein–protein interaction map upon the network of E2F4 targets, it is clear that the larger module can be broken down into smaller modules (or submodules) composed of proteins that bind to one another or form a higher-order structure (e.g., nucleosome). This agrees with the observation that genes with similar functions or that participate in a com-

mon process are often transcriptionally coregulated. Thus, E2F4 targets appear to form submodules that carry out specific tasks (e.g., chromatin assembly, mitochondrial protein synthesis, DNA replication) whose suppression is necessary for cell cycle arrest. One prediction is that cell cycle arrest would not be complete, or efficient, if the connection between E2F4 and one of these submodules was inactivated. Thus, in terms of network dynamics, the cell cycle arrest module supervised by E2F4 serves to coordinate the repression of many separate cellular functions to prevent further proliferation. Careful examination of Figure 4 indicates that not all genes bound by E2F4 were repressed under conditions of cell cycle arrest. However, most submodules contain at least one repressed gene. If we assume that all components of a complex are essential for its function (i.e., that any one of them may be rate-limiting), then most submodules would be inactivated during cell cycle arrest.

Transcriptional control of myogenesis

The first steps of myogenic differentiation involve a cascade of transcriptional activation, initiated by the induction by MyoD and Myf5 of two other TFs, myogenin and MRF4. Collectively, these four basic helix–loop–helix (bHLH) factors are known as muscle regulatory factors (MRFs). MRFs then induce the expression of a large number of muscle function genes. These initial events ultimately lead to the formation of mature muscle (for reviews, see Buckingham 2001; McKinsey et al. 2002).

An *in vitro* model of skeletal myogenesis, the C2C12 murine myoblast cell line, was used to deconvolute transcriptional regulatory networks controlling muscle differentiation. Using an approach and methods similar to the ones described above, transcriptional targets of the MRFs, MyoD and myogenin, as well as those of MEF2, a factor that collaborates with MRFs to regulate myogenesis, were identified using ChIP-on-chip (Blais et al. 2005). Expression profiling of MyoD^{+/+} and MyoD^{-/-} primary myoblasts was used to evaluate the impact of MyoD binding on target gene expression. Examination of the expression of MyoD targets during differentiation indicated that only a subset of these genes is induced when myoblasts differentiate, suggesting either that MyoD is responsible for maintaining a steady level of expression of these targets regardless of differentiation state or that MyoD is inactive at the target gene promoter but is in a poised state, awaiting additional cues to activate transcription. Analysis of promoters bound by MyoD indicated that the binding sites for a number of TFs are specifically enriched among these sequences and that recruitment of these factors may be necessary for induction of their expression. ChIP-on-chip data also indicated that MyoD and myogenin bind overlapping but also distinct sets of targets. Analysis of sequence motifs enriched in each set of target promoters suggests a number of potential partners that could impart combinatorial control, thereby explaining their target specificity.

The MyoD–myogenin–MEF2 axis appears to represent a major hub in the network, governing many processes

Figure 4. The cell cycle arrest network module. Cell cycle arrest module centered around E2F4. We depict a subset of the transcriptional and protein–protein interactions that control the mammalian cell cycle by merging ChIP-on-chip with protein interaction data. Orange edges, based on ChIP-on-chip data, indicate transcriptional regulation and are observed during quiescence and contact inhibition. Green edges represent protein–protein interactions and may occur in other cell cycle phases; they were mined from the BIND and DIP online databases of interacting proteins (Xenarios et al. 2002; Alfaro et al. 2005). A large number of square-shaped nodes (nonregulatory factors) are targets of the E2F4 transcription factor, forming a (diamond-shaped) hub. E2F3, an E2F4 target with transcription factor activity, is also represented as a diamond. Nodes colored in red represent genes whose expression is repressed under conditions of mitogen deprivation and contact inhibition. Yellow circles (nodes) connected with black edges represent multiprotein cellular entities regulated by E2F4.

involved in myogenesis (synaptogenesis, muscle contractility). This hub also connects a striking array of TFs that oversee specialized functions, such as the response to stress, which could function as “terminal hubs.” The organization of the network suggests a hierarchical structure, whereby signals initiated by the master regulators (MRFs) are propagated through the network, amplifying and diversifying the initial inputs. Furthermore, the presence of complex, multi-input, and feed-forward motifs likely lends stability and robustness to the network. This organization is logical, given that myogenesis drastically (and permanently) changes the way a cell functions (determination of cell fate). It stands in con-

trast to more rapid responses to stress or cell cycle arrest regulated by E2F4, specialized events that do not have a long lasting impact on cell function. Thus, examining the global architecture of a given network, in particular its connectivity and its serial regulation by transcriptional factors, provides new biological insights and leads to specific predictions regarding the consequences of network perturbation.

Future prospects

Improvements in experimental and computational approaches outlined here and the availability of databases

that compile genome-wide location analysis results will certainly increase the pace with which transcriptional regulatory networks are revealed. Once we have established the “rules” through which occupancy by a set of promoter-selective factors dictates all histone and DNA modifications and regulation of large numbers of genes, it may be possible to eventually predict the behavior of a gene under any condition based on the linear sequence of elements contained within the gene. Methodological improvements will also provide opportunities to elucidate regulatory networks of increasing complexity, such as those involved in tissues, which can themselves be regarded as cellular networks. Many biological networks are marked by robustness, which stems from their extensive connectivity. One great promise of the study of biological networks is thus a better understanding of how organisms respond to their environment and how breakdown in the network results in disease.

In the future, regulatory networks will also need to account for temporal changes in gene expression, protein-protein interactions, and cellular compartmentalization. These three dimensional reconstructions will no doubt appear very complex. Such cellular networks are necessarily complex, and the more complex they become, the closer they are to mirroring the dynamic changes that occur in a living cell. It is human nature to abhor complexity; clearly, we must ultimately overcome this fear in order to reveal the complex networks that govern the life of a cell.

Acknowledgments

We apologize to the many colleagues whose work could not be cited due to space constraints. We thank Y. Kluger, N. Tanese, and I. Sanchez for productive discussions. This work was supported in part by a post-doctoral fellowship from the Fonds de la Recherche en Santé du Québec to A.B., and by grants from the NIH to B.D.D. (CA077245-8 and GM067132-03).

References

- Alfarano, C., Andrade, C.E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., Betel, D., Bobechko, B., Boutilier, K., Burgess, E., et al. 2005. The biomolecular interaction network database and related tools 2005 update. *Nucleic Acids Res.* **33**: D418–D424.
- Banerjee, N. and Zhang, M.Q. 2002. Functional genomics as applied to mapping transcription regulatory networks. *Curr. Opin. Microbiol.* **5**: 313–317.
- Bar-Joseph, Z., Gerber, G.K., Lee, T.I., Rinaldi, N.J., Yoo, J.Y., Robert, F., Gordon, D.B., Fraenkel, E., Jaakkola, T.S., Young, R.A., et al. 2003. Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.* **21**: 1337–1342.
- Barabasi, A.L. and Oltvai, Z.N. 2004. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**: 101–113.
- Baum, B. and Craig, G. 2004. RNAi in a postmodern, postgenomic era. *Oncogene* **23**: 8336–8339.
- Beer, M.A. and Tavazoie, S. 2004. Predicting gene expression from sequence. *Cell* **117**: 185–198.
- Bergmann, S., Ihmels, J., and Barkai, N. 2004. Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol.* **2**: E9.
- Bergstrom, D.A., Penn, B.H., Strand, A., Perry, R.L., Rudnicki, M.A., and Tapscott, S.J. 2002. Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. *Mol. Cell* **9**: 587–600.
- Bianchi-Frias, D., Orian, A., Delrow, J.J., Vazquez, J., Rosales-Nieves, A.E., and Parkhurst, S.M. 2004. Hairly transcriptional repression targets and cofactor recruitment in *Drosophila*. *PLoS Biol.* **2**: E178.
- Blais, A. and Dynlacht, B.D. 2004. Hitting their targets: An emerging picture of E2F and cell cycle control. *Curr. Opin. Genet. Dev.* **14**: 527–532.
- Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B.D. 2005. An initial blueprint for myogenic differentiation. *Genes & Dev.* **19**: 553–569.
- Bracken, A.P., Ciro, M., Cocito, A., and Helin, K. 2004. E2F target genes: Unraveling the biology. *Trends Biochem. Sci.* **29**: 409–417.
- Buckingham, M. 2001. Skeletal muscle formation in vertebrates. *Curr. Opin. Genet. Dev.* **11**: 440–448.
- Cam, H. and Dynlacht, B.D. 2003. Emerging roles for E2F: Beyond the G1/S transition and DNA replication. *Cancer Cell* **3**: 311–316.
- Cam, H., Balciunaite, E., Blais, A., Spektor, A., Scarpulla, R.C., Young, R., Kluger, Y., and Dynlacht, B.D. 2004. A common set of gene regulatory networks links metabolism and growth inhibition. *Mol. Cell* **16**: 399–411.
- Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**: 499–509.
- Chen, J. and Sadowski, I. 2005. Identification of the mismatch repair genes PMS2 and MLH1 as p53 target genes by using serial analysis of binding elements. *Proc. Natl. Acad. Sci.* **102**: 4813–4818.
- Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. 2002. A genomic regulatory network for development. *Science* **295**: 1669–1678.
- Davidson, E.H., McClay, D.R., and Hood, L. 2003. Regulatory gene networks and the properties of the developmental process. *Proc. Natl. Acad. Sci.* **100**: 1475–1480.
- Dieterich, C., Grossmann, S., Tanzer, A., Ropcke, S., Arndt, P.F., Stadler, P.F., and Vingron, M. 2005. Comparative promoter region analysis powered by CORG. *BMC Genomics* **6**: 24.
- Dimova, D.K., Stevaux, O., Frolov, M.V., and Dyson, N.J. 2003. Cell cycle-dependent and cell cycle-independent control of transcription by the *Drosophila* E2F/RB pathway. *Genes & Dev.* **17**: 2308–2320.
- Elemento, O. and Tavazoie, S. 2005. Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. *Genome Biol.* **6**: R18.
- Elkon, R., Linhart, C., Sharan, R., Shamir, R., and Shiloh, Y. 2003. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res.* **13**: 773–780.
- Euskirchen, G., Royce, T.E., Bertone, P., Martone, R., Rinn, J.L., Nelson, F.K., Sayward, F., Luscombe, N.M., Miller, P., Gerstein, M., et al. 2004. CREB binds to multiple loci on human chromosome 22. *Mol. Cell. Biol.* **24**: 3804–3814.
- Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. 2003. Genomic targets of the human c-Myc protein. *Genes & Dev.* **17**: 1115–1129.

- Friedman, A. and Perrimon, N. 2004. Genome-wide high-throughput screens in functional genomics. *Curr. Opin. Genet. Dev.* **14**: 470–476.
- Geisberg, J.V. and Struhl, K. 2004. Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. *Nucleic Acids Res.* **32**: e151.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J., et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99–104.
- Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**: 717–728.
- Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., Piccirillo, S., Gerstein, M., and Snyder, M. 2002. Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes & Dev.* **16**: 3017–3033.
- Huang, E., Ishida, S., Pittman, J., Dressman, H., Bild, A., Kloos, M., D'Amico, M., Pestell, R.G., West, M., and Nevins, J.R. 2003. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.* **34**: 226–230.
- Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., and Barkai, N. 2002. Revealing modular organization in the yeast transcriptional network. *Nat. Genet.* **31**: 370–377.
- Ihmels, J., Bergmann, S., and Barkai, N. 2004. Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20**: 1993–2003.
- Inoue, T., Wang, M., Ririe, T.O., Fernandes, J.S., and Sternberg, P.W. 2005. Transcriptional network underlying *Caenorhabditis elegans* vulval development. *Proc. Natl. Acad. Sci.* **102**: 4972–4977.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**: 533–538.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. 2000. The large-scale organization of metabolic networks. *Nature* **407**: 651–654.
- Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. 2001. Lethality and centrality in protein networks. *Nature* **411**: 41–42.
- Kato, M., Hata, N., Banerjee, N., Futcher, B., and Zhang, M.Q. 2004. Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol.* **5**: R56.
- Kel, A.E., Kel-Margoulis, O.V., Farnham, P.J., Bartley, S.M., Wingender, E., and Zhang, M.Q. 2001. Computer-assisted identification of cell cycle-related genes: New targets for E2F transcription factors. *J. Mol. Biol.* **309**: 99–120.
- Kim, J., Bhinge, A.A., Morgan, X.C., and Iyer, V.R. 2004. Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat. Methods* **2**: 47–53.
- Koide, T., Hayata, T., and Cho, K.W. 2005. *Xenopus* as a model system to study transcriptional regulatory networks. *Proc. Natl. Acad. Sci.* **102**: 4943–4948.
- Kondo, Y., Shen, L., Yan, P.S., Huang, T.H., and Issa, J.P. 2004. Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. *Proc. Natl. Acad. Sci.* **101**: 7398–7403.
- Labhart, P., Karmakar, S., Salicru, E.M., Egan, B.S., Alexiadis, V., O'Malley, B.W., and Smith, C.L. 2005. Identification of target genes in breast cancer cells directly regulated by the SRC-3/AIB1 coactivator. *Proc. Natl. Acad. Sci.* **102**: 1339–1344.
- Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**: 799–804.
- Levine, M. and Davidson, E.H. 2005. Gene regulatory networks for development. *Proc. Natl. Acad. Sci.* **102**: 4936–4942.
- Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., and Gerstein, M. 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**: 308–312.
- Maduro, M.F. and Rothman, J.H. 2002. Making worm guts: The gene regulatory network of the *Caenorhabditis elegans* endoderm. *Dev. Biol.* **246**: 68–85.
- Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., et al. 2003. Distribution of NF- κ B-binding sites across human chromosome 22. *Proc. Natl. Acad. Sci.* **100**: 12247–12252.
- McKinsey, T.A., Zhang, C.L., and Olson, E.N. 2002. Signaling chromatin to make muscle. *Curr. Opin. Cell. Biol.* **14**: 763–772.
- Muller, H., Bracken, A.P., Vernell, R., Moroni, M.C., Christians, F., Grassilli, E., Prosperini, E., Vigo, E., Oliner, J.D., and Hehl, K. 2001. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes & Dev.* **15**: 267–285.
- Nachman, I., Regev, A., and Friedman, N. 2004. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics* **20** (Suppl. 1): I248–I256.
- Newman, M.E.J. 2003. The structure and function of complex networks. *SIAM Rev.* **45**: 167–256.
- Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. 2004. Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**: 1378–1381.
- Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C., et al. 2003. Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network. *Genes & Dev.* **17**: 1101–1114.
- Pilpel, Y., Sudarsanam, P., and Church, G.M. 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* **29**: 153–159.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306–2309.
- Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes & Dev.* **16**: 245–256.
- Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* **16**: 939–945.
- Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. 2003. Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34**: 166–176.
- Sharan, R., Ovcharenko, I., Ben-Hur, A., and Karp, R.M. 2003. CREME: A framework for identifying cis-regulatory modules in human-mouse conserved segments. *Bioinformatics* **19** (Suppl. 1): i283–i291.
- Siggia, E.D. 2005. Computational methods for transcriptional regulation. *Curr. Opin. Genet. Dev.* **15**: 214–221.

- Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakkola, T.S., et al. 2001. Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**: 697–708.
- Sinha, S., Blanchette, M., and Tompa, M. 2004. PhyME: A probabilistic algorithm for finding motifs in sets of orthologous sequences. *BMC Bioinformatics* **5**: 170.
- Sonnichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E., et al. 2005. Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**: 462–469.
- Stevaux, O. and Dyson, N.J. 2002. A revised picture of the E2F transcriptional network and RB function. *Curr. Opin. Cell Biol.* **14**: 684–691.
- Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. 1999. Systematic determination of genetic network architecture. *Nat. Genet.* **22**: 281–285.
- Tompa, M., Li, N., Bailey, T.L., Church, G.M., De Moor, B., Eskin, E., Favorov, A.V., Frith, M.C., Fu, Y., Kent, W.J., et al. 2005. Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.* **23**: 137–144.
- Tronche, F., Ringeisen, F., Blumenfeld, M., Yaniv, M., and Pontoglio, M. 1997. Analysis of the distribution of binding sites for a tissue-specific transcription factor in the vertebrate genome. *J. Mol. Biol.* **266**: 231–245.
- van Steensel, B. and Henikoff, S. 2000. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**: 424–428.
- Viens, A., Mechold, U., Lehrmann, H., Harel-Bellan, A., and Ogryzko, V. 2004. Use of protein biotinylation in vivo for chromatin immunoprecipitation. *Anal. Biochem.* **325**: 68–76.
- Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. 2002. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci.* **99**: 5860–5865.
- Wasserman, W.W. and Fickett, J.W. 1998. Identification of regulatory regions which confer muscle-specific gene expression. *J. Mol. Biol.* **278**: 167–181.
- Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes & Dev.* **16**: 235–244.
- Wells, J., Yan, P.S., Cechvala, M., Huang, T., and Farnham, P.J. 2003. Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene* **22**: 1445–1460.
- Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M., and Eisenberg, D. 2002. DIP, the database of interacting proteins: A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**: 303–305.
- Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.Y., Alon, U., and Margalit, H. 2004. Network motifs in integrated cellular networks of transcription-regulation and protein–protein interaction. *Proc. Natl. Acad. Sci.* **101**: 5934–5939.

Constructing transcriptional regulatory networks

Alexandre Blais and Brian David Dynlacht

Genes Dev. 2005, **19**:

Access the most recent version at doi:[10.1101/gad.1325605](https://doi.org/10.1101/gad.1325605)

References

This article cites 74 articles, 24 of which can be accessed free at:
<http://genesdev.cshlp.org/content/19/13/1499.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Focused on your science.

USC
SCIENTIFIC
SIDE BY SCIENCE