10-21-2006

Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping

Albertha J. M. Walhout
University of Massachusetts Medical School

Unraveling transcription regulatory networks by protein–DNA and protein –protein interaction mapping

Albertha J.M. Walhout

Genome Res. 2006 16: 1445-1454 originally published online October 19, 2006
Access the most recent version at doi:10.1101/gr.5321506

Review

Unraveling transcription regulatory networks by protein–DNA and protein–protein interaction mapping

Albertha J.M. Walhout

Program in Gene Function and Expression and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

Metazoan genomes contain thousands of protein-coding and noncoding RNA genes, most of which are differentially expressed, i.e., at different locations or at different times during development, function, or pathology of the organism. Differential gene expression is achieved in part by the action of regulatory transcription factors (TFs) that bind to *cis*-regulatory elements that are often located in or near their target genes. Each TF likely regulates many targets in the context of intricate transcription regulatory networks. Up to 10% of a genome may encode TFs, but only a handful of these have been studied in detail. Here, I will discuss the different steps involved in the mapping and analysis of transcription regulatory networks, including the identification of network nodes (TFs and their target sequences) and edges (TF–TF dimers and TF–DNA target interactions), integration with other data types, and network properties and emerging principles that provide insights into differential gene expression.

Metazoan genomes contain thousands of protein- and RNA-encoding genes. Some genes are ubiquitously expressed, whereas others are expressed in a tightly controlled manner in only part of the organism, or under particular conditions during development or disease. In order to understand how differential gene expression is controlled at a genome-wide or systems level, it is important to identify all the *cis*-acting regulatory sequences and *trans*-acting factors involved, and how and when they interact to affect gene expression.

Differential gene expression can be regulated at the transcriptional and at the post-transcriptional level by three types of *trans*-acting factors (Fig. 1). Regulatory transcription factors (TFs) can activate or repress transcription by physically interacting with genomic *cis*-regulatory DNA elements that can be located in gene promoters, or at a greater genomic distance in enhancers, or in introns (Fig. 1A).

RNA binding proteins can interact with specific *cis*-regulatory RNA elements, for instance, that are located in the 3′ untranslated region of an mRNA molecule (Fig. 1B). The binding of RNA binding proteins regulates differential gene expression at the post-transcriptional level, by affecting transcript localization, translation, or degradation (for reviews, see Hieronymus and Silver 2004; Keene and Lager 2005).

microRNAs exclusively repress gene expression by physically interacting, through hybridization, with *cis*-regulatory elements located in the 3′ untranslated region of their target mRNAs (Fig. 1B). This hybridization results in the inhibition of translation and/or decreased mRNA stability (Ambros 2004; Du and Zamore 2005). Thus, TFs, RNA binding proteins, and microRNAs physically interact with their target genes, either at the DNA or at the mRNA level. Such regulator-target interactions are now being systematically mapped and modeled into regulatory networks. Because information about many genes and TFs is assembled into a single network model, transcription regulatory networks provide insight into the principles and properties that control differential gene expression at a systems level, rather than at the level of individual genes.

E-mail marian.walhout@umassmed.edu; fax (508) 856-5460.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.5321506.

What is a regulatory network?

Network models are composed of nodes and edges that describe relationships between nodes. In biological networks, the nodes are bioactive macromolecules such as proteins, DNA, RNA, and metabolites (Barabasi and Oltvai 2004). Two types of regulatory networks can be distinguished: transcription regulatory networks and post-transcription regulatory networks (Fig. 2). Each of these types of networks can be subdivided into physical and functional networks. Physical networks contain protein–protein, protein–DNA, protein–RNA, and/or RNA–RNA interactions (Fig. 2A,C). Functional networks incorporate the consequences of these physical interactions, e.g., activation or repression of gene expression (Fig. 2B,D). Ultimately, transcription and post-transcription regulatory networks need to be combined to obtain a comprehensive picture of all aspects of the regulation of differential gene expression in complex metazoan systems (Fig. 2E).

In this review, I will focus on the mapping of transcription regulatory networks. I will discuss the identification of predicted TFs and *cis*-regulatory sequences, i.e., network nodes, and the protein–protein and protein–DNA interaction mapping approaches that are being used to identify physical interactions between these nodes, i.e., network edges. I will discuss several emerging insights and hypotheses that can be derived from such networks, and the future challenges that lie ahead in this rapidly evolving field.

Identifying network nodes

Transcription regulatory networks contain two types of nodes: regulatory TFs and their target DNA sequences. Many different strategies have been employed to identify both types of nodes, including computational and experimental methods.

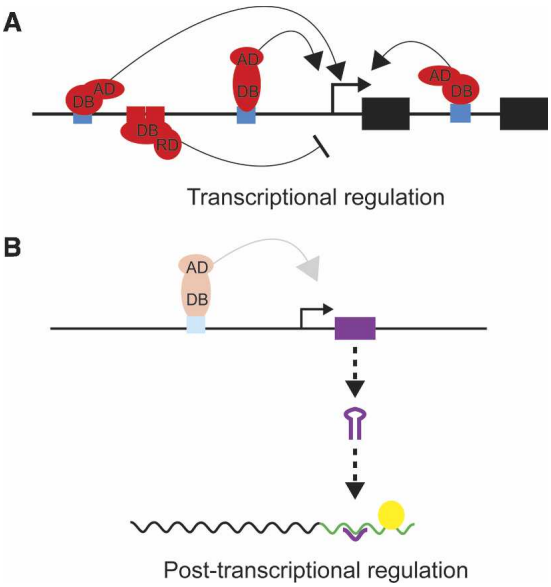

Figure 1. Regulators of gene expression physically interact with their targets. (*A*) Transcriptional regulation. Regulatory TFs function by binding to proteins and to DNA. Black boxes, exons; blue squares, *cis*-regulatory DNA elements; red ellipses, TFs; arrow, transcription start site. Curved arrows indicate activation of gene expression and blunt "arrow" indicates transcriptional repression. AD, transcription activation domain; RD, transcription repression domain; DB, DNA binding domain. (*B*) Post-transcriptional regulation. RNA binding proteins and microRNAs function by directly interacting with their target mRNAs. Arrow, transcription start site; purple box, microRNA gene; green line, 3′ UTR of target microRNA (purple line); yellow circle, RNA binding protein. Upstream regulation of miRNA expression (by the pink TF binding to the light blue element) is indicated and connects transcriptional and post-transcriptional gene regulation.

Regulatory transcription factors

Regulatory TFs are composed of at least two types of domains: a DNA binding domain, which serves to interact with its cognate DNA target sequence, and a transcription regulation domain, which serves to activate or repress transcription (Fig. 1A). TFs are grouped into families based on their predicted DNA binding domains. To date, more than 100 different DNA binding domains have been found (Kummerfeld and Teichmann 2006). These domains have been used to computationally predict which genes in a genome of interest encode regulatory TFs. However, computational prediction alone is insufficient to obtain comprehensive and high-quality TF predictions. For instance, we recently obtained a high-quality compendium of *Caenorhabditis elegans* TFs by a combination of computational prediction and extensive manual curation (Reece-Hoyes et al. 2005). By doing so, the number of false positive and false negative predictions was drastically reduced. It should be feasible to obtain such comprehensive predictions for other organisms, including human, as well. However, even manually curated collections are likely incomplete as not all DNA binding domains have yet been uncovered. For example, both yeast and *C. elegans* proteins that bind DNA but that do not possess a known DNA binding domain have recently been retrieved (Hall et al. 2004; Deplancke et al. 2006).

TF predictions have led to the observation that, in increasingly complex metazoan organisms, a larger proportion of the genome encodes TFs, compared with relatively simple, unicellular eukaryotes. For instance, the genome of the unicellular yeast *Saccharomyces cerevisiae* encodes ~200 predicted TFs (Harbison et al. 2004), which is ~3% of all protein-coding genes; the relatively simple metazoan nematode *C. elegans* contains 934 predicted TFs, which is ~5% of all protein-coding genes (Reece-Hoyes et al. 2005); and more complex eukaryotes such as humans may devote up to 10% of their coding power to regulatory TFs (Levine and Tjian 2003).

TFs interact with different types of DNA sequences, including promoters and *cis*-regulatory modules, and, within such larger elements, bind to specific *cis*-regulatory elements or TF binding sites. Considerable efforts are underway to identify each of these elements in order to decipher the "regulatory code" that controls differential gene expression. For instance, the ENCODE (ENCyclopedia of DNA elements) Consortium aims to identify all functional elements in the human genome (ENCODE Project Consortium 2004). So far the efforts of this consortium have focused on 1% of the genome, or 30 Mb of sequence, which contains ~600 predicted protein-coding genes. In order to gain

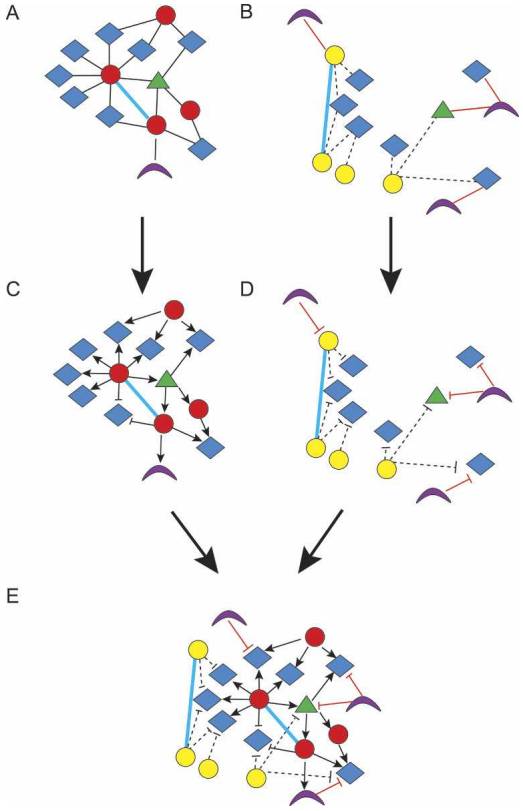

Figure 2. Regulatory networks. (*A*) Protein–DNA and protein–protein interaction network involving regulatory TFs and their target genes. (*B*) Protein–RNA, protein–protein, and microRNA–RNA interaction network involving RNA binding proteins and their target mRNAs and microRNAs and their target mRNAs. (*C*) Transcription regulatory networks. The transcriptional consequences of the protein interactions shown in *A* are included. (*D*) Post-transcription regulatory networks. The effects of the protein interactions shown in *B* on target gene expression are included. (*E*) Combined transcription and post-transcription regulatory networks. Red nodes, TFs; blue nodes, target genes; green node, a gene can be a target gene and encode a TF; yellow nodes, RNA binding proteins; purple nodes, microRNAs. Black edges, protein–DNA interactions; dashed edges, protein–RNA interactions; red edges, microRNA–RNA interactions; blue edges, protein–protein interactions. Arrows, activation; blunt "arrow," repression.

insight into the regulatory code of a genome, one of the first steps is to comprehensively identify all gene promoters.

Promoters

A gene promoter is defined as the regulatory sequence (a few hundred base pairs) that is located immediately upstream of the transcription start site (for review, see Maston et al. 2006). Eukaryotic protein and microRNA-encoding gene promoters are composed of two parts: a proximal promoter that serves as a recognition sequence for the pre-initiation complex and RNA polymerase II, and a distal promoter that performs a regulatory function by interacting with regulatory TFs. The identification of promoters is relatively straightforward in unicellular eukaryotes such as *S. cerevisiae* as its genome is compact (Goffeau et al. 1996): It contains very few introns and short intergenic regions (median length shorter than 400 bp). Hence, the intergenic regions contain most *cis*-regulatory elements and can be used as a proxy for gene promoters when transcription start sites have not been precisely mapped. Since the genomes of higher eukaryotes contain longer intergenic regions with many repeat sequences, and because transcription start sites are often poorly defined, it is more difficult to accurately pinpoint metazoan gene promoters. Moreover, higher eukaryotic genes may frequently be regulated from multiple, alternative promoters.

Several experimental approaches have been developed for transcription start site and, thus, promoter annotation. First, full-length cDNA sequencing has led to the annotation of thousands of transcripts for both the murine and human genome (Imanishi et al. 2004; Carninci et al. 2005). Second, the use of genome-wide tiling arrays has enabled the identification of 5′ and 3′ boundaries of transcripts (Carninci et al. 2005). Third, cap analysis of gene expression (CAGE) has been used to more precisely define transcription start sites in mammalian promoters (Carninci et al. 2006). Fourth, chromatin-immunoprecipitations (see below) with anti-TFIID and anti-RNA polymerase II antibodies have been used to identify many active human promoters (Kim et al. 2005a,b). Finally, by transient transfection assays, 387 gene promoters from the ENCODE regions that drive gene expression in at least one of 16 different cell-lines were identified (Cooper et al. 2006). Although a lot of progress has been made, it is likely that highly sensitive experimental methods need to be developed to identify promoters that are rarely active.

Cis–*regulatory modules*

Many gene promoters have been identified to date. However, the genome-wide identification of enhancers and silencers in higher eukaryotes has been relatively slow. This is because they can be located at a great genomic distance from the target's transcription start site(s) and can be found upstream, downstream, or within introns (for review, see Maston et al. 2006).

It has been postulated that functional TF binding sites often occur in clusters and form *cis*-regulatory modules (Davidson 2001). Recently, this hypothesis has been utilized by several groups for the computational prediction of *cis*-regulatory modules that may constitute enhancers or silencers (Aerts et al. 2003; Sharan et al. 2003; Gupta and Liu 2005; Blanchette et al. 2006; Hallikas et al. 2006). The methods used by these groups provide powerful tools to search for *cis*-regulatory modules containing consensus binding motifs for TFs for which the recognition sequence has been mapped. However, to date such information is only available for a limited number of TFs.

In addition to using computational tools to predict regulatory sequences, experimental methods can be used for the discovery of *cis*-regulatory modules (for review, see Elnitski et al. 2006). For instance, the observation that the genome is more accessible to DNaseI when TFs are bound, leading to DNaseI hypersensitive sites, can be used to identify *cis*-regulatory modules. Until recently, the unbiased, genome-wide mapping of such sites has been hampered by a lack of high-throughput "readout" methods that can be used to map such sites onto genome sequences. Several groups have already made significant progress toward this goal, for instance by combining DNaseI treatment with microarrays or massive parallel sequencing (Dorschner et al. 2004; Crawford et al. 2006a,b; Sabo et al. 2006). Integration with other types of data will be necessary to delineate the function of each DNaseI hypersensitive site and to find the transacting factors that bind to these sites.

TF binding sites and cis*–regulatory elements*

For a thorough understanding of transcription regulatory networks, it is not only important to find promoters and *cis*-regulatory modules, but also to precisely map the *cis*-regulatory elements located within these longer sequences. Individual *cis*-regulatory elements are short (usually <20 bp) DNA sequences that interact directly with regulatory TFs. Such TF binding sites have traditionally been mapped using a combination of deletion analyses and reporter gene expression (see, e.g., Davidson et al. 2002). However, such methods are not readily adaptable to high-throughput settings.

Recently, several methods have been employed to computationally identify putative *cis*-regulatory elements (Fig. 3) (for review, see Elnitski et al. 2006). The first method is based on the hypothesis that genes that are coexpressed under a particular condition are subject to control by the same TF(s). The advent of gene expression analysis by microarrays greatly facilitated the identification of coexpressed genes (DeRisi et al. 1997). Using a variety of computational algorithms, the regulatory regions of coexpressed genes can be interrogated for the occurrence of over-represented DNA sequences that may constitute binding sites for the TF responsible for the coexpression (for information and performance on such algorithms, see Tompa et al. 2005).

The second method, referred to as phylogenetic footprint-

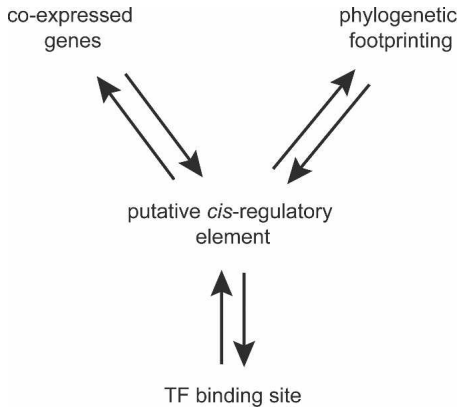

Figure 3. Different approaches that can be used for the identification of *cis*-regulatory DNA elements are highly complementary and interconnected. *Cis*-regulatory elements can be identified by interrogating the regulatory regions of coexpressed genes, by phylogenetic footprinting, or by experimentally identifying TF binding sites.

ing, is based on the conservation of functional *cis*-regulatory elements in closely related organisms. This method has been used to identify putative elements in yeast (Cliften et al. 2001, 2003; Kellis et al. 2003), *Drosophila melanogaster* (Glazov et al. 2005), and human genomes (Bejerano et al. 2004; Siepel et al. 2005; Woolfe et al. 2005; Xie et al. 2005). In addition to using computational tools to find putative *cis*-regulatory elements in complete genome sequences, experimentally mapped consensus TF binding motifs (see below) can also be used to interrogate a genome sequence of interest. However, depending on the length of the motif, many functional and nonfunctional sequences will be identified. Phylogenetic footprinting and coexpression can then be used to determine which motifs have a higher likelihood of being functional in vivo (Fig. 3; Elnitski et al. 2006).

Only a small portion of all TF binding sites that occur in a genome of interest have been identified to date. For instance, by comparative genomics, Xie and colleagues found 174 candidate DNA motifs that likely correspond to numerous TF binding sites in human promoters (Xie et al. 2005). However, these different elements may only represent ~10% of all TF binding motifs as the human genome may encode more than 2000 TFs (Levine and Tjian 2003), each of which likely binds DNA with different specificity and affinity. On the other hand, it is likely that some TFs from one family may have overlapping binding specificities, and that therefore the number of different TF binding motifs may be considerably less than 2000. In addition, certain TFs may exclusively bind to regulatory elements that are located in transcriptional enhancers or silencers. These TF binding motifs will be missed in studies that focus solely on promoter sequences.

The computational prediction of *cis*-regulatory modules has relied on the observation that TF binding sites are often clustered. However, the generality of this phenomenon has not been investigated and, thus, it is not clear how many functional, nonclustered TF binding sites occur in the genome. In addition, most researchers have focused on elements that are conserved between related species. Such phylogenetic footprinting likely increases the specificity of motif finding. However, the sensitivity will suffer from only interrogating conserved sequences because many important, species-specific elements are not conserved. The success of phylogenetic profiling for the identification of functional regulatory elements also depends on the evolutionary distance between the organisms used in the analysis: The use of closely related species may result in relatively low specificity and the use of distantly related species may result in high specificity, but relatively low sensitivity (Ruvinsky and Ruvkun 2003).

Identifying network edges

TF–TF dimers

Many TFs bind their target genes as dimers. For instance, bZIP, bHLH, and nuclear hormone receptor TFs all dimerize. The comprehensive identification of TF dimers requires the use of protein–protein interaction detection methods that can be used in (semi) high-throughput settings. One assay that is particularly suited to identify binary protein–protein interactions is the yeast two-hybrid system (Fields and Song 1989), and multiple putative TF homo- and heterodimers have already been found using this system (Li et al. 2004; Reece-Hoyes et al. 2005; Rual et al. 2005; Stelzl et al. 2005). Putative TF dimers have also been identified by protein arrays. For instance, Newman and Keating tested >2400 combinations of human bZIP protein–protein interactions and found multiple putative dimers (Newman and Keating 2003). Finally, 15 putative yeast TF–TF heterodimers have been identified by large-scale TAP-TAG purification methods (Gavin et al. 2006; Krogan et al. 2006).

Only a small portion of all TF dimers has been identified to date. For instance, <10% of all predicted TFs are present in the current *C. elegans* protein–protein interaction network (Li et al. 2004), even though at least 30% of all TFs belong to the bZIP, bHLH, or nuclear hormone receptor families (Reece-Hoyes et al. 2005). Since TF dimerization may be condition-dependent, many TF dimers have also likely been missed by TAP-TAG assays in yeast. In the future, it will be important to comprehensively map all dimerization interactions between TFs and to incorporate this information into network models (Fig. 2).

Interactions between TFs and their target genes/sequences

Protein–DNA interactions between TFs and their target DNA sequences can be mapped using two conceptually different strategies. First, one can identify for a TF or set of TFs of interest, the target genes, and/or *cis*-regulatory elements these TFs bind to. Alternatively, one can take a DNA sequence as a starting point and aim to identify the TFs that can interact with this sequence. We refer to these strategies as "TF-centered" and "gene-centered" methods, respectively (Fig. 4A; Deplancke et al. 2006).

TF–centered protein–DNA interaction mapping

The most widely used protein–DNA interaction mapping methods are TF-centered, and most are based on chromatin-

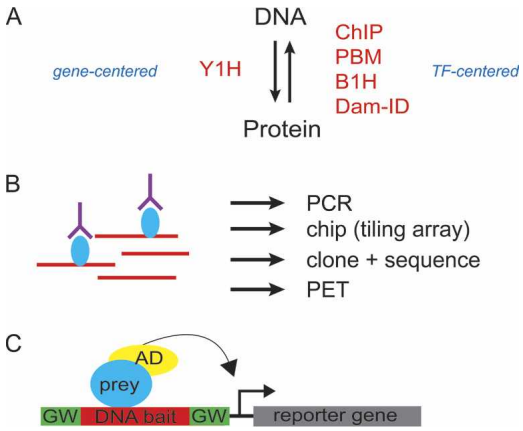

Figure 4. High-throughput methods for protein–DNA interaction mapping. (*A*) protein–DNA interactions can be mapped using either TF- or gene-centered methods, as indicated by the arrows. Y1H, yeast one-hybrid assays; ChIP, chromatin-immunoprecipitations; PBM, protein binding microarray; B1H, bacterial one-hybrid system; Dam-ID, DNA adenine methytransferase-ID. (*B*) ChIP is the most commonly used TF-centered method. It is based on the precipitation of a TF (blue) and its associated DNA fragments (red) using an anti-TF antibody (purple). Multiple readouts of the precipitated DNA can be used, including PCR with specific primers, tiling microarrays (chip), cloning and sequencing, and paired-end ditag sequencing. (*C*) Y1H assays are based on interactions of hybrid "prey" proteins with a DNA "bait" of interest. The hybrid protein consists of a protein that can bind DNA (blue) and a heterologous transcription activation domain (AD, yellow). The use of such a domain enables the identification of both activators and repressors of transcription. The readout for a protein–DNA interaction is the expression of one or more reporter genes. Prey identity is determined by PCR and sequencing. In high-throughput Y1H assays, vectors containing Gateway recombination sites (GW) are used to enable standardized cloning from promoterome resources.

immunoprecipitation (ChIP) (for review, see Elnitski et al. 2006). In ChIP assays an anti-TF antibody is used to precipitate DNA bound by the TF in vivo (Fig. 4B). These DNA fragments can subsequently be identified and quantified using a variety of readouts, including PCR, microarrays (referred to as ChIP-on-chip), and cloning/sequencing as in SAGE-like methods (serial analysis of gene expression, Fig. 4B; for review, see Blais and Dynlacht 2005; Elnitski et al. 2006). For yeast ChIP-on-chip assays, endogenous TFs were replaced by hybrid proteins in which the TFs were fused to a universal protein tag (Lee et al. 2002). Thus, almost 200 individual yeast strains were created, each carrying a different TF-TAG fusion protein. The advantage of this strategy is that the same antibody can be used for each TF. ChIP-on-chip has been used to identify target sequences for most yeast TFs under standard laboratory growth conditions (Lee et al. 2002). In addition, target binding has been examined under multiple experimental conditions for a subset of these TFs (Harbison et al. 2004; Workman et al. 2006). In addition to yeast, ChIP-on-chip has been used for a variety of mammalian TFs by using tissue culture cells (Cawley et al. 2004; Carroll et al. 2005; Bieda et al. 2006). So far, only a few studies focused on the DNA binding of metazoan TFs in their natural environment. For instance, in a pioneering study, endogenous promoters bound by HNF1a, HNF4a, and HNF6 within human liver and pancreas were identified (Odom et al. 2004). Similarly, by a combination of computational target prediction and ChIP-on-chip, many target promoters bound by CREB were identified in both tissue culture cells and primary hepatocytes (Zhang et al. 2005). ChIP-on-chip was also used to identify promoters bound by the TFs OCT4, SOX2, and NANOG in human embryonic stem cells (Boyer et al. 2005). Finally, ChIP-cloning (i.e., the cloning and sequencing of precipitated DNA) was used to identify in vivo target genes for the *C. elegans* FOXO TF, DAF-16 (Oh et al. 2005).

In DamID, a TF is fused to *Escherichia coli* DNA adenine methyltransferase (Dam) and expressed in tissue culture cells or intact model organisms (van Steensel and Henikoff 2000). Upon binding of the TF to DNA, the surrounding nucleotides are methylated. This methylation can be detected by PCR or microarrays after immunoprecipitation of methylated DNA. DamID has mainly been used to identify the DNA targets of general chromatin-binding proteins, but has also been used to dissect the *Drosophila* Myc TF network (Orian et al. 2003).

In protein-binding microarrays, a TF is fused to GST, expressed in bacteria or yeast, purified, and hybridized to a double-stranded DNA array that contains DNA sequences of interest (Mukherjee et al. 2004). To date, this method has been used to find targets for the yeast TFs Abf1, Rap1, and Mig1. The target sequences were then used to identify the consensus TF binding sites for each of these factors. DIP-ChIP can also be used to identify consensus TF binding sites. This method uses naked genomic DNA and a purified TF. Briefly, after incubation of the DNA with the factor, an immunoprecipitation is performed and TF-associated DNA is identified by microarray analysis (Liu et al. 2005). Although both DIP-chip and PBM are carried out in vitro, the TF binding sites obtained were in very good agreement with data obtained using in vivo methods, suggesting that they are effective and rapid methods to identify TF binding specificities, and, perhaps, affinities (Mukherjee et al. 2004; Liu et al. 2005).

In bacterial one-hybrid assays, a plasmid encoding a TF of interest is transformed into bacteria containing a library of random DNA elements (Meng et al. 2005). Binding of the TF to a specific element is selected on specific media and positive colonies are analyzed by sequencing. After aligning multiple sequences bound by an individual TF, its recognition sequence can be derived. This sequence can then be used to search the genome to identify putative TF target genes. Since TF binding sites are generally short, many of such sequences will occur in a genome, only some of which will likely be functional.

Gene–centered protein–DNA interaction mapping

Eukaryotic genomes encode hundreds of putative TFs, of which only a handful has been analyzed by TF-centered methods. The identification of protein–DNA interactions involving uncharacterized, predicted TFs has recently been facilitated by the development of high-throughput, gene-centered protein–DNA interaction mapping methods, such as yeast one-hybrid (Y1H) assays. The Y1H system was first developed to facilitate the identification of proteins that can bind to multiple copies of a short DNA sequence of interest (the "DNA bait") (Li and Herskowitz 1993; Wang and Reed 1993). This method is not suitable for the unbiased, comprehensive mapping of protein–DNA interactions with longer DNA fragments because the *cis*-regulatory elements that contribute to gene expression are only known for a few genes, and because the system was based on traditional, restriction enzyme-based cloning methods. To enable the unbiased, large-scale detection of protein–DNA interactions, we developed a high-throughput version of the Y1H system (Fig. 4C; Deplancke et al. 2004). This system is compatible with Gateway cloning, a recombinational cloning system by which many fragments (i.e., DNA baits) can be cloned simultaneously (Hartley et al. 2000; Walhout et al. 2000). This Y1H system can be used with single copy gene promoters as DNA baits and, therefore, allows the unbiased identification of TF-promoter interactions without prior knowledge about the *cis*-regulatory elements that reside within the promoter. The system is compatible with "promoterome" resources, collections of Gateway-cloned promoters, for the high-throughput cloning of DNA baits (Dupuy et al. 2004). The Gateway-compatible Y1H system also makes use of Gateway-compatible "protein prey" resources. For instance, mini-libraries consisting solely of predicted TFs can be created and screened successfully. This is important as TFs that are expressed at low levels or in only a few cells in an organism are difficult to retrieve from standard cDNA libraries (Deplancke et al. 2004). Recently, we used the Gateway-compatible Y1H system to map a first *C. elegans* gene-centered protein–DNA interaction network, containing 283 protein–DNA interactions between 72 promoters and 117 proteins, 107 of which encode predicted *C. elegans* TFs and 10 of which may be novel DNA binding proteins (Deplancke et al. 2006).

As with any large-scale, high-throughput method, protein–protein and protein–DNA interactions will be missed and wrongly identified by each of the methods discussed. Some of these methods identify interactions that do occur in vivo (e.g., ChIP with endogenous TFs) and others find interactions that can occur (e.g., in vitro methods, yeast two-hybrid and yeast one-hybrid assays). Protein–DNA interactions that occur infrequently, i.e., in a few cells or during a short time period in development or disease, will likely be missed by the first methods but may be found by the second. However, interactions found by the second do not necessarily occur in vivo. To assure the generation of high-quality data sets, it is desirable to filter protein–protein and protein–DNA interaction data, and to only include high-confidence interactions, i.e., interactions that are likely rel-

evant. Such criteria have previously been used for large-scale protein–protein interaction maps, generated by high-throughput yeast two-hybrid assays (Li et al. 2004; Rual et al. 2005; Stelzl et al. 2005), and we have recently devised stringent criteria to filter Y1H data (Deplancke et al. 2006). In summary, both sensitivity and specificity are important issues to consider when choosing a protein–protein or protein–DNA interaction identification method, and the choice depends on the question being addressed. As the various methods are highly complementary, it is desired, in the long term, to use a multitude of techniques for comprehensive, high-quality protein–protein and protein–DNA interaction mapping.

Emerging concepts and future challenges

Large sets of protein–protein and protein–DNA interactions can be visualized as network models using various freely available software packages, including Cytoscape (Shannon et al. 2003) and N-browse (Lall et al. 2006).

Network models serve multiple purposes. For instance, they provide a great tool for the visualization and navigation of large interaction data sets. In addition, networks can be analyzed at different levels, i.e., at the level of the network as a whole, the level of subgraphs and network motifs, and the level of individual nodes or edges. By doing so, they enable the derivation of hypotheses regarding different levels of gene expression.

Network analysis

Once visualized, networks can be analyzed topologically using different network parameters such as connectivity, path length, clustering coefficient, etc. (For review, see Barabasi and Oltvai 2004). As has been observed for other networks, transcription regulatory networks are highly connected and display a scale-free degree distribution (Albert et al. 2000), i.e., they contain a small number of disproportionately highly-connected nodes, or hubs, and many less-well connected nodes (Guelzim et al. 2002; Lee et al. 2002; Luscombe et al. 2004; Deplancke et al. 2006). Transcription regulatory networks potentially contain two types of hubs: TF hubs (TFs that bind many promoters) and promoter hubs (promoters that interact with many TFs). Interestingly, transcription regulatory networks predominantly contain TF hubs, rather than promoter hubs (Guelzim et al. 2002; Deplancke et al. 2006). As in other networks, such hubs provide integrity to the network: When nodes are randomly removed, the network stays connected. However, when hubs are sequentially removed the network disintegrates rapidly (for review, see Barabasi and Oltvai 2004). The biological implication of this became apparent when it was demonstrated that TF hubs have a higher tendency to be essential for the organism (Jeong et al. 2001; Yu et al. 2004; Deplancke et al. 2006).

We recently mapped a protein–DNA interaction network of genes expressed or involved in the *C. elegans* digestive tract (Deplancke et al. 2006). By visualizing and analyzing this network, we can derive hypotheses at different levels of gene regulation. For instance, we observed that the network is highly connected, contains several TF hubs, and is enriched for TFs expressed in the digestive tract (Fig. 5A; Deplancke et al. 2006). In addition, we found that most promoters are bound by a combination of TF hubs and less well-connected TFs, some of which may be master regulators. This led to a model in which we propose that *C. elegans* transcription is regulated by a layered organization of TF function (Fig. 5A; Deplancke et al. 2006). The digestive tract is predominantly composed of the pharynx and intestine, each of which is derived from distinct germlayers. We found that TF hubs interact with both pharyngeal and intestinal genes. This suggests that these TFs function as global regulators of gene expression and leads to the prediction that they interact with promoters of large numbers of genes that are expressed in other tissues as well.

In addition to hypotheses regarding gene regulation at the level of an entire network and system, one can also derive hypotheses by zooming into network subgraphs.

Network subgraphs

Network subgraphs can be network modules, motif clusters, or other network neighborhoods. A network module can be defined as a subgraph consisting of highly interconnected nodes that may fulfill a particular biological function. Network modularity has been observed in yeast regulatory networks (Ihmels et al. 2002; Bar-Joseph et al. 2003; Segal et al. 2003; Luscombe et al. 2004), although there are few modules that can be clearly separated from the main network component (Babu et al. 2004). This may be because individual yeast TFs may function in multiple, apparently unrelated pathways. These observations suggest that regulatory networks of higher eukaryotes such as *C. elegans* may be organized in modules as well but that these modules share multiple TFs. This hypothesis is in agreement with our observation that many *C. elegans* TFs are expressed in multiple tissues (Deplancke et al. 2006).

Figure 5B shows an example of a subgraph of the *C. elegans* digestive tract protein–DNA interaction network that can be used to derive specific biological hypotheses. This subgraph is composed of multiple bifan motifs (see below for network motifs) of the TF hubs DIE-1 and ZTF-1 and their target promoters. Interestingly, these TFs share 22 promoters, which is 73% of their combined targets (Fig. 5B). This leads to the prediction that DIE-1 and ZTF-1 have a similar biochemical function: For instance, they may have similar DNA binding motifs. The observation that they do not share all of their targets suggests that the motifs are not completely identical. Interestingly, these proteins share no homology in their primary amino acid sequence but both proteins do contain two pairs of C2H2 zinc fingers that are separated by a long amino acid sequence. The observation of shared targets also leads to the prediction that these two TFs may share biological functions. However, while knockdown of DIE-1 is lethal, ZTF-1 is dispensable for the function of the organism. In contrast, we could not create stable transgenic lines expressing ZTF-1, suggesting that overexpression of this protein may be lethal (Deplancke et al. 2006). Whereas DIE-1 can activate gene expression (Deplancke et al. 2006), the transcriptional function of ZTF-1 remains to be elucidated.

Network motifs

Network motifs are the building blocks of networks (Milo et al. 2002). Several motifs are overrepresented in experimentally derived transcription regulatory networks compared with random networks (Milo et al. 2002; Shen-Orr et al. 2002). Such motifs provide insights into the properties of networks and the propagation of regulatory signals. As such, the analysis of network motifs may help to uncover the biochemical functions of both TFs and their target genes. For instance, feed forward loops are overrepresented in transcription regulatory networks of various organisms (Milo et al. 2002; Shen-Orr et al. 2002). This may not

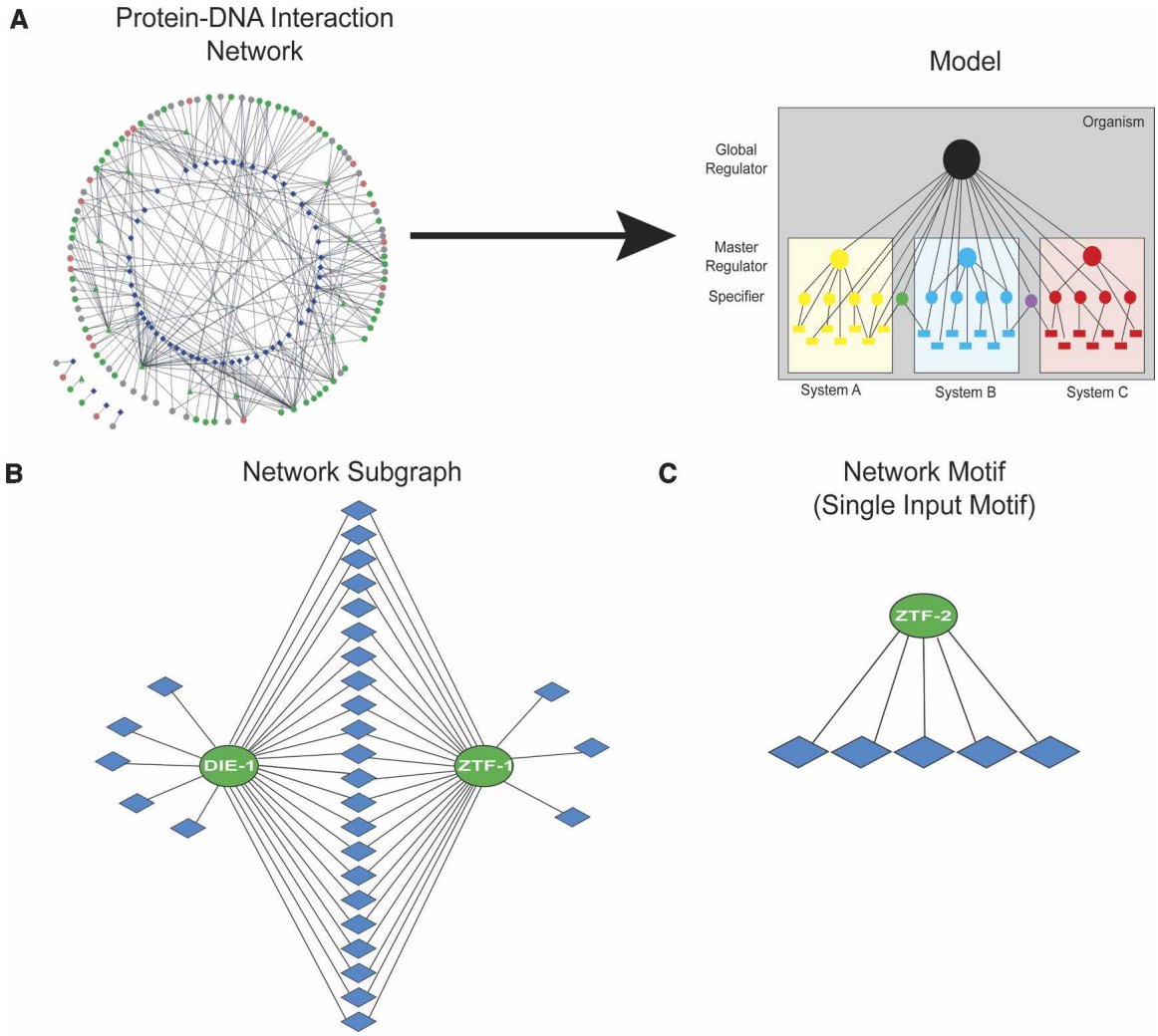

Figure 5. Deriving biological hypotheses from regulatory networks. (*A*) A protein–DNA interaction network of *C. elegans* digestive tract genes was used to derive a three-layered model of transcription regulation. Reprinted with permission from Elsevier © 2006, Deplancke et al. 2006. (*B*) Example of a protein–DNA interaction network subgraph. (*C*) Example of a protein–DNA interaction network motif. See main text for details.

be surprising as such loops offer a rapid gene expression output, for example in response to outside signals. In contrast, feed-back or autoregulatory loops can either reinforce or diminish a transcriptional output, whereas single input motifs can confer strong coexpression of downstream target genes (Shen-Orr et al. 2002). The analysis of each TF and the network context in which it functions will be important to unravel how each factor contributes to differential gene expression.

Network motifs can also be used to derive specific biological hypotheses, either for individual promoters, or TFs. For instance, we found a single input motif in which ZTF-2 interacts specifically with the promoters of five pharyngeal genes (Deplancke et al. 2006). This led to the prediction that ZTF-2 is a regulator of pharyngeal gene expression and that these promoters share a pharyngeal gene element to which ZTF-2 binds. We tested these hypotheses experimentally and found that ZTF-2 is itself expressed in the pharynx (and elsewhere), and that a knockdown of *ztf-2* results in a pharyngeal phenotype. Furthermore, we used the five promoter sequences to define a ZTF-2 binding motif and found that it is highly similar to a previously described pharyngeal gene element. Finally, we demonstrated that ZTF-2 represses expression of its pharyngeal targets and that it can bind the pharyngeal element in vivo. Taken together, the mapping, analysis and deconvolution of a protein–DNA interaction network into subgraphs and motifs can be used to derive biological hypotheses regarding differential gene expression at different levels.

Future challenges

The transcription regulatory networks that are currently available are likely to be a small representation of all the interactions that occur in vivo. Even in yeast, where binding of each TF has been examined under standard laboratory conditions and binding of a few under multiple conditions (Lee et al. 2002; Harbison et al. 2004; Workman et al. 2006), the regulatory information is likely far from complete. This is because many conditions remain to be tested and because TFs that bind DNA with low specificity or affinity may be difficult to analyze. The transcription regulatory networks that have been mapped in higher eukaryotes represent an even smaller sample of the entire network. This is because so

far (1) ChIP-on-chip assays mainly utilized arrays containing probes corresponding to promoter regions and, thus, TF binding to *cis*-regulatory modules located elsewhere in the genome will be missed; (2) only very few TFs have been examined by TF-centered methods; (3) <1% of all promoters in *C. elegans* have been examined by gene-centered methods (Deplancke et al. 2006), and, finally, not all *cis*-regulatory elements and TF binding sites have been identified either computationally (Xie et al. 2005) or experimentally (Mukherjee et al. 2004; Meng et al. 2005).

Cis-regulatory elements or TF binding sites are often found in intergenic regions. When intergenic regions are short (i.e., in yeast and *C. elegans*) and reside between genes that are transcribed from opposite strands, perhaps by bidirectional promoters, it is difficult to infer which of the two genes will be affected through such regulatory sequences. Similarly, in the genomes of higher eukaryotes, *cis*-regulatory modules can be located far from a transcribed unit and it may be difficult to infer the gene that is controlled by the *cis*-regulatory module. In the future, it will be important to scan larger genomic regions, both in *cis* and in *trans* for genes that can be affected by different individual *cis*-regulatory modules.

From protein–DNA interaction networks to transcription regulatory networks

Networks that are solely based on protein–protein and protein–DNA interactions do not contain regulatory information because the protein–protein and protein–DNA interaction detection methods discussed above do not provide insight into the consequences of physical interactions (e.g., activation or repression of transcription). The transcriptional consequences of protein–DNA interactions need to be superimposed onto protein–DNA interaction networks by integrating interaction data with other data types (Deplancke et al. 2006). This can be done either at the level of individual interactions using detailed and often labor-intensive methods such as quantitative RT-PCR and RNAi (Baugh et al. 2005; Oh et al. 2005; Deplancke et al. 2006) or at the network level by integration with other large data sets, such as expression profiles (Lee et al. 2002; Segal et al. 2003; Yu et al. 2003; Luscombe et al. 2004). In the future, it will be important to develop methods that can be used to map, at a large scale, the transcriptional activity of each TF.

Spatio–temporal network modeling

Protein–protein and protein–DNA interaction networks and transcription regulatory networks are static models of all the transcriptional events that can occur in a system of interest. To fully understand how such networks contribute to system development, function, and pathology, it is important to unravel where and when which parts of the network are active and what the biological consequences of this activity are (Davidson et al. 2002). Such analysis has again been pioneered in yeast. For instance, the binding of the cell cycle regulatory TF complexes SBF and MBF has been analyzed during different phases of the cell cycle (Horak et al. 2002). It was found that these TFs bind to and regulate many other TF-encoding genes that are involved in cell cycle progression and/or differentiation. In addition, more extensive networks that are active under different endogenous and exogenous experimental conditions were compiled (Luscombe et al. 2004). Surprisingly, it was found that these different subnetworks have different topological properties and motifs that may reflect their particular function. In the future, it will be important to extrapolate where and when which parts of transcription regulatory networks are active in higher eukaryotes as well.

Longer term, transcription regulatory networks need to be integrated to model more comprehensive regulatory networks in which transcription regulation of the expression of both protein-coding and microRNA-encoding genes is combined with gene regulation by both RNA binding proteins and microRNAs (Fig. 2E). Such networks need themselves to be integrated with spatio-temporal information about gene expression and TF/microRNA activity and with phenotypes conferred by TFs and microRNAs to obtain a comprehensive picture about regulatory networks and how they control the development, function, and pathology of complex metazoan systems.

Acknowledgments

I thank members of the Walhout laboratory for discussions, J. Lieb for information on intergenic region length in yeast, S. Ryder for advice about RNA binding proteins, I. Barrasa for help with yeast TF–TF dimer retrieval, and B. Deplancke, V. Vermeirssen, N. Martinez, and J. Dekker for critical reading of the manuscript. The Walhout laboratory is supported by NIH grants CA097516, DK068429, and DK071713.

References

Aerts, S., van Loo, P., Thijs, G., Moreau, Y., and de Moor, B. 2003. Computational detection of *cis*-regulatory modules. *Bioinformatics* **19:** ii5–ii14.

Albert, R., Jeong, H., and Barabasi, A.-L. 2000. Error and attack tolerance of complex networks. *Nature* **378:** 378–381.

Ambros, V. 2004. The functions of animal microRNAs. *Nature* **431:** 350–355.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. 2004. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14:** 283–291.

Barabasi, A.L. and Oltvai, Z.N. 2004. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5:** 101–113.

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.

Baugh, L.R., Hill, A.A., Claggett, J.M., Hill-Harfe, K., Wen, J.C., Slonim, D.K., Brown, E.L., and Hunter, C.P. 2005. The homeodomain protein PAL-1 specifies a lineage-specific regulatory network in the *C. elegans* embryo. *Development* **132:** 1843–1854.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. 2004. Ultraconserved elements in the human genome. *Science* **304:** 1321–1325.

Bieda, M., Xu, X., Singer, M.A., Green, R., and Farnham, P.J. 2006. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res.* **16:** 595–605.

Blais, A. and Dynlacht, B.D. 2005. Constructing transcriptional regulatory networks. *Genes & Dev.* **19:** 1499–1511.

Blanchette, M., Bataille, A.R., Chen, X., Poitras, C., Laganiere, J., Lefebvre, D., Deblois, G., Giuere, V., Ferretti, V., Bergeron, D., et al. 2006. Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Res.* **16:** 656–668.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309:** 1559–1563.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38:** 626–635.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W.L., Hestermann, E.V., Geistlinger, T.R., et al. 2005. Chomosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122:** 33–43.

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.

Cliften, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., Waterston, R.H., and Johnston, M. 2001. Surveying *Saccharyomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res.* **11:** 1175–1186.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. 2003. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301:** 71–76.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L., and Myers, R.M. 2006. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16:** 1–10.

Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G., and Collins, F.S. 2006a. DNase-chip: A high-resolution method to identify DNaseI hypersensitive sites using tiled microarrays. *Nat. Methods* **3:** 503–509.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., et al. 2006b. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* **16:** 123–131.

Davidson, E.H. 2001. *Genomic regulatory systems: Development and evolution.* Academic Press, San Diego.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, 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.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A.J.M. 2004. A Gateway-compatible yeast one-hybrid system. *Genome Res.* **14:** 2093–2101.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A.M., Grove, C.A., Martinez, N.J., Sequerra, R., Doucette-Stam, L., Reece-Hoyes, J.S., Hope, I.A., et al. 2006. A gene-centered *C. elegans* protein–DNA interaction network. *Cell* **125:** 1193–1205.

DeRisi, J.L., Iyer, V., and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278:** 680–686.

Dorschner, M.O., Hawrylycz, M., Humbert, R., Wallace, J.C., Shafer, A., Kawamoto, J., Mack, J., Hall, R., Goldy, J., Sabo, P.J., et al. 2004. High-throughput localization of functional elements by quantitative chromatin profiling. *Nat. Methods* **1:** 219–225.

Du, T. and Zamore, P.D. 2005. microPrimer: The biogenesis and function of microRNA. *Development* **132:** 4645–4652.

Dupuy, D., Li, Q., Deplancke, B., Boxem, M., Hao, T., Lamesch, P., Sequerra, R., Bosak, S., Doucette-Stam, L., Hope, I.A., et al. 2004. A first version of the *Caenorhabditis elegans* promoterome. *Genome Res.* **14:** 2169–2175.

Elnitski, L., Jin, V.X., Farnham, P.J., and Jones, S.J.M. 2006. Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques. *Genome Res.* (this issue).

ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia of DNA elements) project. *Science* **306:** 636–640.

Fields, S. and Song, O. 1989. A novel genetic system to detect protein–protein interactions. *Nature* **340:** 245–246.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., et al. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440:** 631–636.

Glazov, E.A., Pheasant, M., McGraw, E.A., Bejerano, G., and Mattick, J.S. 2005. Ultraconserved elements in insect genomes: A highly conserved intronic sequence implicated in the control of *homothorax* mRNA splicing. *Genome Res.* **15:** 800–808.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. Life with 6000 genes. *Science* **274:** 546, 563–567.

Guelzim, N., Bottani, S., Bourgine, P., and Kepes, F. 2002. Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* **31:** 60–63.

Gupta, M. and Liu, J.S. 2005. *De novo cis*-regulatory module elicitation for eukaryotic genomes. *Proc. Natl. Acad. Sci.* **102:** 7079–7084.

Hall, D.A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., and Snyder, M. 2004. Regulation of gene expression by a metabolic enzyme. *Science* **306:** 482–484.

Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. 2006. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124:** 47–59.

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.

Hartley, J.L., Temple, G.F., and Brasch, M.A. 2000. DNA cloning using in vitro site-specific recombination. *Genome Res.* **10:** 1788–1795.

Hieronymus, H. and Silver, P.A. 2004. A systems view of mRNP biology. *Genes & Dev.* **18:** 2845–2860.

Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., Piccirrillo, S., Gerstein, M., and Snyder, M. 2002. Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes & Dev.* **16:** 3017–3033.

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.

Imanishi, T., Itoh, T., Suzuki, Y., O'Donovan, C., Fukuchi, S., Koyanagi, K.O., Barrero, R.A., Tamura, T., Yamaguchi-Kabata, Y., Tanino, M., et al. 2004. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.* **2:** e162.

Jeong, H., Mason, S.P., Barabasi, A.-L., and Oltvai, Z.N. 2001. Lethality and centrality in protein networks. *Nature* **411:** 41–42.

Keene, J.D. and Lager, P.J. 2005. Post-transcriptional operons and regulons co-ordinating gene expression. *Chromosome Res.* **13:** 327–337.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423:** 241–254.

Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M., et al. 2005a. Direct isolation and identification of promoters in the human genome. *Genome Res.* **15:** 830–839.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. 2005b. A high-resolution map of active promoters in the human genome. *Nature* **436:** 876–880.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440:** 637–643.

Kummerfeld, S.K. and Teichmann, S.A. 2006. DBD: A transcription factor prediction database. *Nucleic Acids Res.* **34:** D74–D81.

Lall, S., Grun, D., Krek, A., Chen, K., Wang, Y.-L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., MacMenamin, P., et al. 2006. A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* **16:** 460–471.

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 Tjian, R. 2003. Transcription regulation and animal diversity. *Nature* **424:** 147–151.

Li, J.J. and Herskowitz, I. 1993. Isolation of the ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* **262:** 1870–1874.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.-O., Han, J.-D.J., Chesneau, A., Hao, T., et al. 2004. A map of the interactome network of the metazoan *C. elegans*. *Science* **303:** 540–543.

Liu, X., Noll, D.M., Lieb, J.D., and Clarke, N.D. 2005. DIP-chip: Rapid and accurate determination of DNA-binding specificity. *Genome Res.* **15:** 421–427.

Luscombe, N.M., Madan Babu, 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.

Maston, G.A., Evans, S.K., and Green, M.R. 2006. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7:** 29–59.

Meng, X., Brodsky, M.H., and Wolfe, S.A. 2005. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat. Biotechnol.* **23:** 988–994.

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

Mukherjee, S., Berger, M.F., Jona, G., Wang, X.S., Muzzey, D., Snyder, M., Young, R.A., and Bulyk, M.L. 2004. Rapid analysis of the DNA-binding specificities of transcription factors with DNA

microarrays. *Nat. Genet.* **36:** 1331–1339.
Newman, J.R.S. and Keating, A.E. 2003. Comprehensive identification of human bZip interactions with coiled-coil arrays. *Science* **300:** 2097–2101.
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.
Oh, S.W., Mukhopadhyay, A., Dixit, B.L., Raha, T., Green, M.R., and Tissenbaum, H.A. 2005. Identification of direct targets of DAF-16 controlling longevity, metabolism and diapause by chromatin immunoprecipitation. *Nat. Genet.* **38:** 251–257.
Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W.M., 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.
Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., and Walhout, A.J.M. 2005. A compendium of *C. elegans* regulatory transcription factors: A resource for mapping transcription regulatory networks. *Genome Biol.* **6:** R110.
Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. 2005. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437:** 1173–1178.
Ruvinsky, I. and Ruvkun, G. 2003. Functional tests of enhancer conservation between distantly related species. *Development* **130:** 5133–5142.
Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. 2006. Genome-scale mapping of DNaseI sensitivity in vivo using tiling DNA microarrays. *Nat. Methods* **3:** 511–518.
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.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13:** 2498–2504.
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:** i283–i291.
Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31:** 64–68.
Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm and yeast genomes. *Genome Res.* **15:** 1034–1050.
Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al. 2005. A human protein–protein interaction network: A resource for annotating the proteome. *Cell* **122:** 957–968.
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.
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.
Walhout, A.J.M., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. 2000. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287:** 116–122.
Wang, M.M. and Reed, R.R. 1993. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature* **364:** 121–126.
Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3:** e7.
Workman, C.T., Mak, H.C., McCuine, S., Tagne, J.B., Agarwal, M., Ozier, O., Begley, T.J., Samson, L.D., and Ideker, T. 2006. A systems approach to mapping DNA damage response pathways. *Science* **312:** 1054–1059.
Xie, X., Lu, J., Kulkobas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. 2005. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434:** 338–345.
Yu, H., Luscombe, N.M., Qian, J., and Gerstein, M. 2003. Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet.* **19:** 422–427.
Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., and Gerstein, M. 2004. Genomic analysis of essentiality within protein networks. *Trends Genet.* **20:** 227–231.
Zhang, X., Odom, D.T., Koo, S.-H., Conkright, M.D., Canettieri, G., Best, J., Chen, H., Jenner, R., Herbolsheimer, E., Jacobsen, E., et al. 2005. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc. Natl. Acad. Sci.* **102:** 4459–4464.