Transcriptional regulation of protein complexes within and across species

Kai Tan, Tomer Shlomi, Hoda Feizi, Trey Ideker, and Roded Sharan

PNAS published online Jan 16, 2007;
doi:10.1073/pnas.0606914104

Transcriptional regulation of protein complexes within and across species

Kai Tan*, Tomer Shlomi†, Hoda Feizi*, Trey Ideker*, and Roded Sharan†‡

*Department of Bioengineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093; and †School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel

Edited by Philip P. Green, University of Washington School of Medicine, Seattle, WA, and approved November 27, 2006 (received for review August 9, 2006)

Yeast two-hybrid and coimmunoprecipitation experiments have defined large-scale protein–protein interaction networks for many model species. Separately, systematic chromatin immunoprecipitation experiments have enabled the assembly of large networks of transcriptional regulatory interactions. To investigate the functional interplay between these two interaction types, we combined both within a probabilistic framework that models the cell as a network of transcription factors regulating protein complexes. This framework identified 72 putative coregulated complexes in yeast and allowed the prediction of 120 previously uncharacterized transcriptional interactions. Several predictions were tested by new microarray profiles, yielding a confirmation rate (58%) comparable with that of direct immunoprecipitation experiments. Furthermore, we extended our framework to a cross-species setting, identifying 24 coregulated complexes that were conserved between yeast and fly. Analyses of these conserved complexes revealed different conservation levels of their regulators and provided suggestive evidence that protein–protein interaction networks may evolve more slowly than transcriptional interaction networks. Our results demonstrate how multiple molecular interaction types can be integrated toward a global wiring diagram of the cell, and they provide insights into the evolutionary dynamics of protein complex regulation.

data integration | network alignment | network evolution

This decade has seen an enormous amount of data on molecular interactions released into the public domain. Although many types of molecules comprise the cell and can interact with one another, the two types that have been measured at largest scale are protein–protein interactions (PPIs) and transcriptional interactions (TIs). The two-hybrid system (1) and coimmunoprecipitation (co-IP) followed by mass spectrometry (2) have been the two most popular technologies to obtain large-scale PPI data. For transcriptional interactions, ChIP coupled with whole-genome DNA microarray (ChIP-chip) allow one to determine the entire spectrum of *in vivo* DNA-binding sites for any given protein (3, 4).

The availability of large-scale PPI and TI data from multiple species has made it possible to study how these two interaction types are combined toward a coordinated cellular response. Previously, PPI and TI data have been integrated to infer hybrid network motifs (5), sets of interacting genes that are differentially expressed (6) and causal pathways that explain differential gene expression (7). In other recent studies (8, 9), yeast TIs were mapped onto known protein complexes as recorded in the Munich Information Center for Protein Sequences (MIPS) database (10). Although these studies did not consider PPI data directly, they established that proteins within the same complex are often encoded by genes that are regulated by the same transcription factors (TFs). Protein complexes were further shown to exhibit expression coherency (11) and to include synergistic TF pairs (12).

A major problem confounding these analyses is that, depending on the underlying technology, interaction data can be noisy. Errors may arise in the two-hybrid system from self-activators, in the co-IP system from abundant protein contaminants, and in both systems from weak or nonspecific interactions (13, 14). Aside from issues of noise, it has become clear that molecules in the cell are very highly connected, such that it is possible to traverse from one molecule to any other by stepping through only a small number of interaction partners (15). These two fundamental problems, noise and high connectivity, make it challenging to organize interaction networks into discrete models of functional pathways and complexes.

Integrative approaches have met with some success in addressing both of these problems (16). Efforts to integrate molecular interaction data for inferring protein machinery can be divided roughly into two types of analyses. The first type infers a network of functional linkages between proteins, based on a weighted combination of independent sources of evidence for pairwise association (e.g., physical interaction, synthetic lethality, coexpression, cocitation) (8, 9, 11, 17–20). Advantages of this type of analysis are its sensitivity (a large number of predictions can be made by drawing from different sources) and versatility (new functional linkages can be established without in-depth requirements on, or understanding of, the individual data sources). In contrast, in the second type of analysis, the inferred functional linkages are guided by specific knowledge of the structure of the different data sources and their inter-relationships (21–24). For example, protein–protein and synthetic-lethal networks each embed dense interaction subnetworks which tend to be orthogonal rather than overlapping (21, 25). Where coverage of both interaction types is high, the predicted functional linkages achieve high specificity.

Here, we integrate PPI and TI data in a single model which simultaneously detects protein complexes and their transcriptional regulators. We apply this model to map coregulated protein complexes in yeast, as well as coregulated protein complexes that are conserved across the networks of both yeast and fly. Numerous instances of yeast/fly conservation are identified, demonstrating that the specific mechanisms regulating protein complexes can be conserved over vast evolutionary distances. For complexes regulated by Rpn4, the implied transcriptional interactions are validated by microarray profiling of an *rpn4*Δ knockout strain.

Results

Identifying Coregulated Protein Clusters. Previous studies have demonstrated a significant correlation between yeast protein complexes and the transcriptional network (8, 9). However, when testing specific complexes, only few display significant associations to TFs [9 of 78 in our analysis; see supporting information (SI) Table 3].

Author contributions: K.T., T.I., and R.S. designed research; K.T., H.F., and R.S. performed research; K.T., T.S., T.I., and R.S. contributed new reagents/analytic tools; K.T., T.S., and R.S. analyzed data; and K.T., T.I., and R.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: co-IP, coimmunoprecipitation; FDR, false discovery rate; GO, gene ontology; MIPS, Munich Information Center for Protein Sequences; PPI, protein–protein interaction; TF, transcription factor; TI, transcriptional interaction; DBD, DNA-binding domain.

[‡] To whom correspondence should be addressed. E-mail: roded@tau.ac.il.

This article contains supporting information online at www.pnas.org/cgi/content/full/0606914104/DC1.

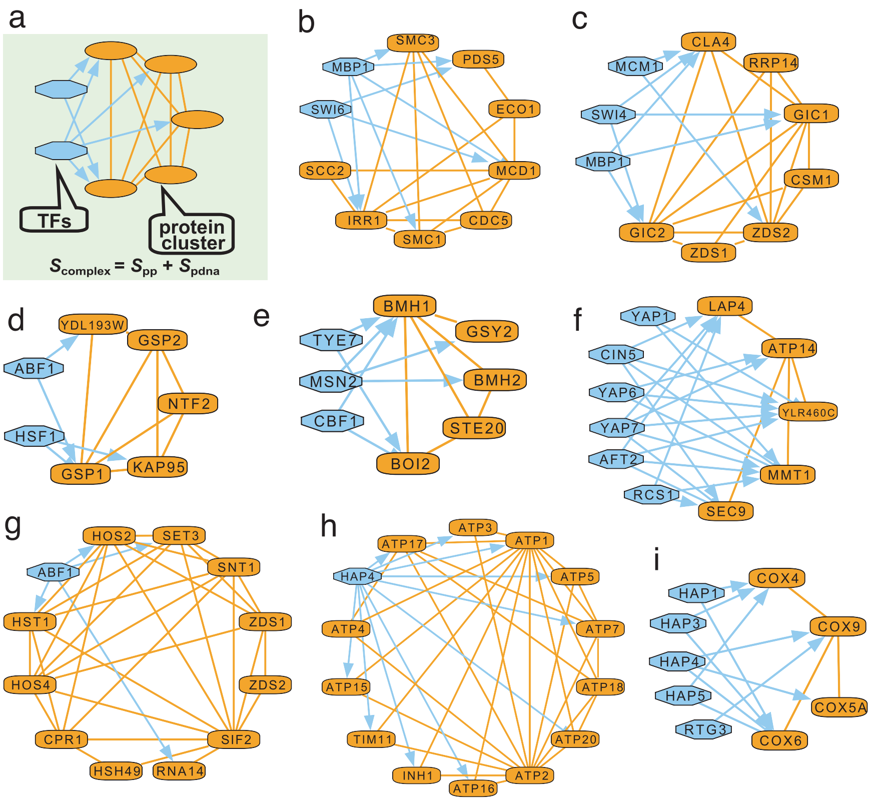


Fig. 1. Coregulated protein clusters in yeast. (*a*) A typical coregulated cluster and its scoring scheme. Orange ovals, protein cluster members; blue octagons, TFs; orange lines, PPI; blue arrows, TI. (*b–i*) Representative examples of coregulated protein clusters. Shown are enriched Gene Ontology (GO) biological processes ($P < 0.05$) of clusters: cell cycle (*b*); budding (*c*); cytoplasmic transport (*d*); cell shape and size regulation (*e*); mitochondrial membrane transport (*f*); histone deacetylation (*g*); hydrogen transport (*h*); and energy pathways (*i*).

To map this relatively unexplored world of coregulated protein complexes, we developed an algorithm for detecting dense protein clusters in the PPI network whose members are coregulated by one or more TFs. As described further in *Materials and Methods*, this approach is based on integrating the protein–protein and transcriptional interaction networks of a species, and searching for sets of proteins that densely interact in the PPI network and whose gene promoters are targeted by the same TFs in the TI network. Such protein sets are termed coregulated protein clusters, and their identification relies on a statistical model for coregulated protein complexes and on efficient search techniques.

We first applied this model to integrate and analyze the PPI and TI networks of yeast, identifying 72 significant coregulated protein clusters. Representative examples of these clusters are shown in Fig. 1, whereas a summary of all clusters is shown in SI Fig. 5. Detailed statistics on the size and composition of these clusters are given in SI Table 4.

To validate our predictions, we tested the coregulated clusters for functional enrichment, expression coherency and conservation coherency of their members (Table 1 and *SI Materials and Methods*). The last two measures quantify the extent to which the genes in a cluster are correlated in their expression levels under multiple conditions, or in their phylogenetic profiles. As a baseline, we compared the performance of the coregulated cluster collection under these measures to that of a collection of 452 protein clusters inferred by using the PPI data only (ignoring the TI data, see *SI Materials and Methods*). We also included in the comparison two collections of complexes derived by co-IP experiments (26, 27). We found that the coregulated clusters exhibited substantially higher expression coherency and conservation coherency levels than the experimentally derived complexes and the baseline clusters (Table 1). Furthermore, 100% of the clusters were functionally enriched, slightly higher than the baseline clusters (99%) and markedly higher than the experimentally derived complexes (61% in ref. 26 and 77% in ref. 27).

Comparing our results with previous work, it is clear that an

Table 1. Validation of yeast clusters by functional enrichment, expression coherency, and conservation coherency of their members*

	Complex source	GO enrichment, %	Expression coherency, %	Conservation coherency, %
MS-derived complexes	Ho *et al.* (26)	61	8	24
	Gavin *et al.* (27)	77	9	36
Protein clusters	Current study	99	26	22
Coregulated clusters	Current study	100	45	59

*All analyses were restricted to clusters of size at least 7, although the same trends were observed over a wide range of cluster size cutoffs.

integrated search of the PPI and TI networks finds significant signal in the data, whereas a fixed search of TI interactions versus MIPS complexes yields a much weaker association. As mentioned above, only 9 of the 78 manually curated MIPS complexes were found to have a significant association to a TF. Of the 72 coregulated clusters we identified, 50 (69%) had no overlap with any of these 9 MIPS complexes, demonstrating the usefulness of our approach in identifying previously uncharacterized regulated complexes.

Prediction of Novel Transcriptional Interactions in Yeast. A coregulated protein cluster, which involves direct transcriptional regulation of some cluster members by a specific TF (or more than one), supports the prediction that the same TF directly regulates other members of the cluster. To prioritize these predictions, we assessed the extent to which the predicted TF targets had correlated expression and phylogenetic conservation with the respective TF, as well as the presence of known TF-binding sites in their promoters. All three measures were combined within a logistic regression classifier to assign a quantitative confidence score to each potential transcriptional interaction (*Materials and Methods*). This classifier attained high sensitivity (82%) and specificity (91%) levels in 10-fold cross validation (Fig. 2*a*). Overall, combining the classifier scores with the coregulated cluster information, we predicted 120 previously uncharacterized transcriptional interactions involving 23 TFs and 99 protein cluster members (SI Table 5).

To evaluate the accuracy of these predictions, we experimentally tested 12 predicted transcriptional interactions for the TF Rpn4. Although these interactions were assigned high confidence scores by our approach, none had been detected in the large-scale chIP-chip study of Harbison *et al*. (28), in which Rpn4 interactions were measured in yeast growing under uninduced condition (30°C in YPD media) and under oxidative stress. However, previous studies have established that Rpn4 could be activated by multiple types of cellular stresses, including heat shock (29). We compared the expression profiles of wild type and *rpn4* gene deletion strains under heat-shock-induced stress (*Materials and Methods*). Seven (*RPN5*, *RPN12*, *CCT8*, *PRE5*, *RPT5*, *PRE10*, *PRE2*) of the 12 newly predicted Rpn4 targets exhibited differential expression ($P < 0.05$ by VERA, see *Materials and Methods*). This fraction (58%) was significant when compared with the fraction of differentially expressed genes overall (11%; $P = 1.2 \times 10^{-4}$); it was also much higher than that attained when predictions were made by the classifier alone (19%). The newly confirmed interactions may have been missed in the Harbison ChIP-chip study because they are specific to heat shock. Notably, the fractions of differentially expressed genes in Harbison Rpn4 targets (53%) and our predicted Rpn4 targets (58%) were similar (Fig. 2*c*), suggesting that our prediction method attains an accuracy level that is comparable with the high-throughput experimental approach.

Many of our other predictions seem to be bona-fide targets of the respective TFs given that the predicted targets are enriched for functional categories consistent with the function(s) of the regulating TF. For instance, the predicted targets of the cell cycle regulator Fkh2 are enriched for the cell cycle and DNA processing functions ($P = 7.93 \times 10^{-6}$). Another example is Dig1 whose set of predicted targets is enriched for genes involved in cell differentiation ($P = 6.47 \times 10^{-5}$), in agreement with its functional role (30, 31). Overall, for all 10 TFs having more than five predicted targets, their target sets were significantly enriched for functional categories consistent with the function(s) of the regulating TF (SI Table 6).

Conserved Protein Complexes. Next, we questioned whether the coregulated clusters we had identified might be conserved through evolution. To tackle this question, we extended our approach to identify coregulated clusters that are conserved across two species (*Material and Methods*). Although PPI networks are now available for a variety of model species such as the yeast and fly, at the present time large number of TIs have been mapped for yeast only.

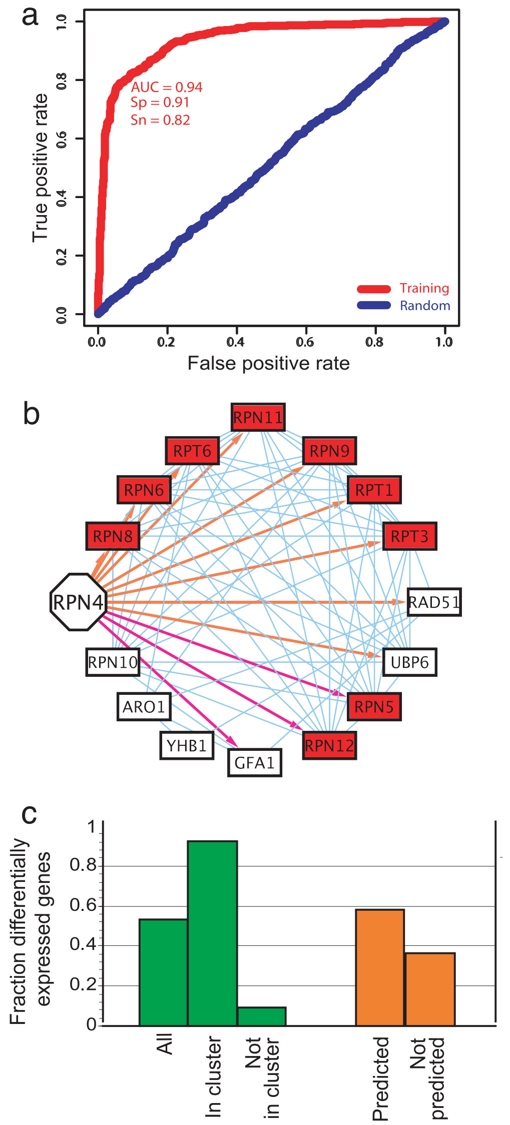


Fig. 2. Transcriptional interaction prediction in yeast. (*a*) Receiver operating characteristics curve of the logistic regression classifier. AUC, area under the curve; Sn, sensitivity; Sp, specificity. (*b*) An example of a predicted cluster regulated by Rpn4. Orange arrows, known Rpn4 TIs from Harbison *et al*. (28); purple, newly predicted Rpn4 TIs. Shades of red represent *P* values (≤0.05) for differential gene expression. (*c*) Fraction of differentially expressed genes in various gene sets. Green, genes bound by Rpn4 from the Harbison data; orange, genes in cluster models but not bound by Rpn4 based on the Harbison data.

Accordingly, we applied our algorithm to identify coregulated protein clusters that were conserved across the PPI networks of both yeast and fly but were supported by TIs in yeast only. We hypothesized that such an approach would reveal protein clusters that maintain coregulation also in fly, even though no fly TI data were used in their construction.

Overall, we identified 24 significant coregulated conserved clusters. These protein clusters were highly functionally enriched, both in yeast (92%; SI Table 7) and in fly (88%, SI Table 7), supporting their biological significance. Examples of coregulated conserved clusters are shown in detail in Fig. 3; an overview of these clusters is given in SI Fig. 5, and their general characteristics are described in SI Table 4.

To examine the contribution of transcriptional interactions in the cross-species setting, we compared the coregulated conserved

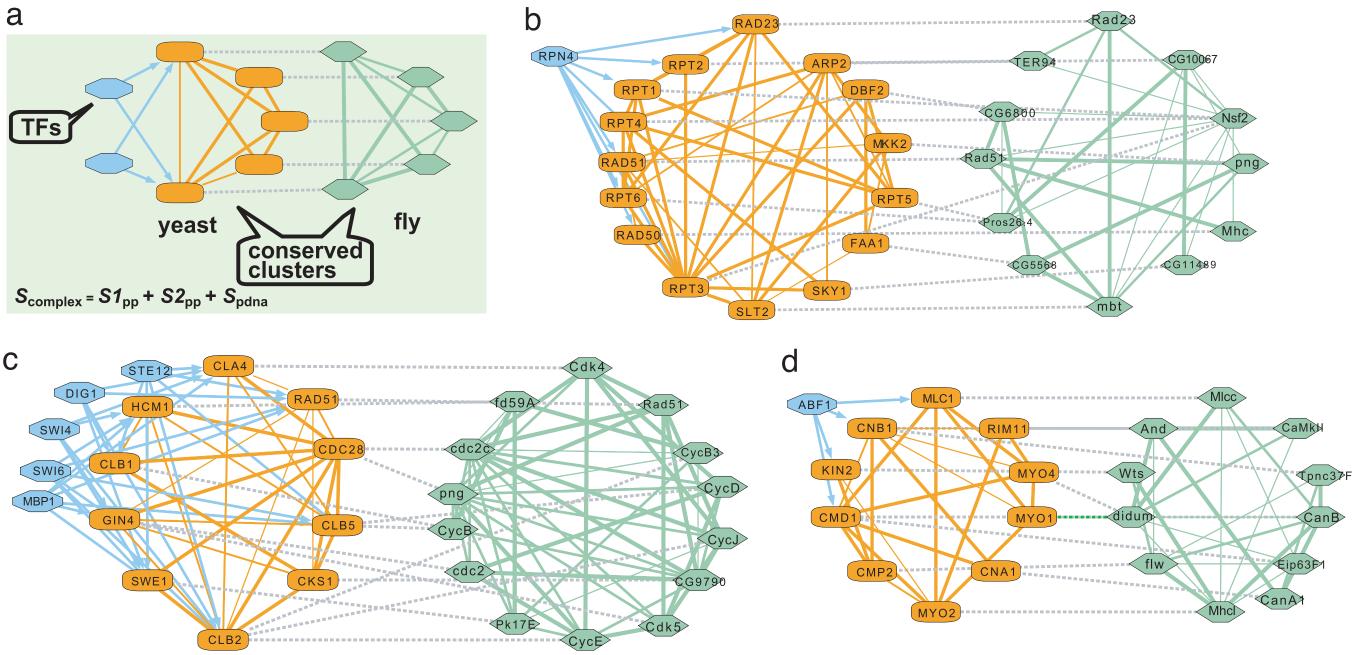
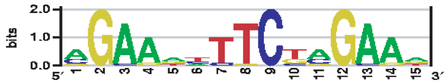
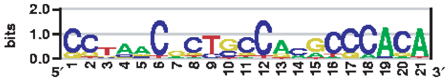
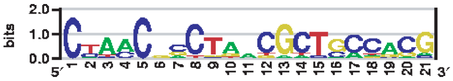
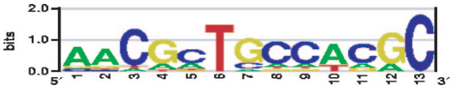
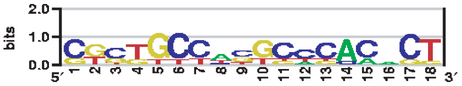


Fig. 3. Conserved coregulated protein clusters between yeast and fly. (*a*) A typical conserved coregulated cluster and its scoring scheme. Protein cluster members: orange ovals, yeast; green hexagons, fly. Blue octagons, transcription factors; orange/green lines, PPI; blue arrows, TI. Horizontal dotted links indicate cross-species sequence similarity between proteins (BLAST *E* value $< 10^{-7}$). (*b*–*d*) Representative examples of conserved protein clusters across yeast and fly. Conserved GO biological processes of clusters ($P < 0.05$): protein catabolism (*b*); cell cycle (*c*); motor (*d*). Proteins are connected by direct (thick line) or indirect (connection via a common network neighbor; thin line) protein interactions.

clusters to a collection of 181 conserved protein clusters that were predicted based on PPI data only (see *SI Materials and Methods*). We found that the coregulated clusters displayed markedly higher levels of expression coherency and conservation coherency, and similar levels of functional enrichment (SI Table 8).

Enrichment of DNA-Binding Motifs in Conserved Fly Clusters. To further study the transcriptional regulation of the fly clusters, we searched for known and previously uncharacterized DNA sequence motifs upstream of their member genes. In total, we identified five enriched DNA motifs spanning 12 of the 24 conserved clusters (Table 2). One of the motifs (Motif 1, enriched in clusters 1 and 12) was a known motif of the fly TF Hsf1. This motif is almost identical to the yeast Hsf1 motif, which was also the yeast TF associated with these clusters. In addition to this known motif, we identified four previously uncharacterized motifs. Motifs 2 (enriched in cluster 4) and 3 (enriched in cluster 5) were identified by searching sets of fly promoters

Table 2. Enriched DNA motifs in conserved fly clusters

Motif ID	Promoter source	Fraction with motif*	Motif logo	*P* value
1	Clusters 1 and 12	34/34		4.02×10^{-7}
2	Cluster 4	16/20		0.009
3	Cluster 5	12/12		0.01
4	d.Fkh2†	18/18		0.004
5	d.Mbp1†	22/26		0.01

*The number of sequences predicted to have the motif out of the total number of input sequences.

†d.Fkh2 and d.Mbp1: Fly regulons, i.e., sets of fly cluster members whose yeast counterparts are regulated by yeast Fkh2 and Mbp1, respectively.

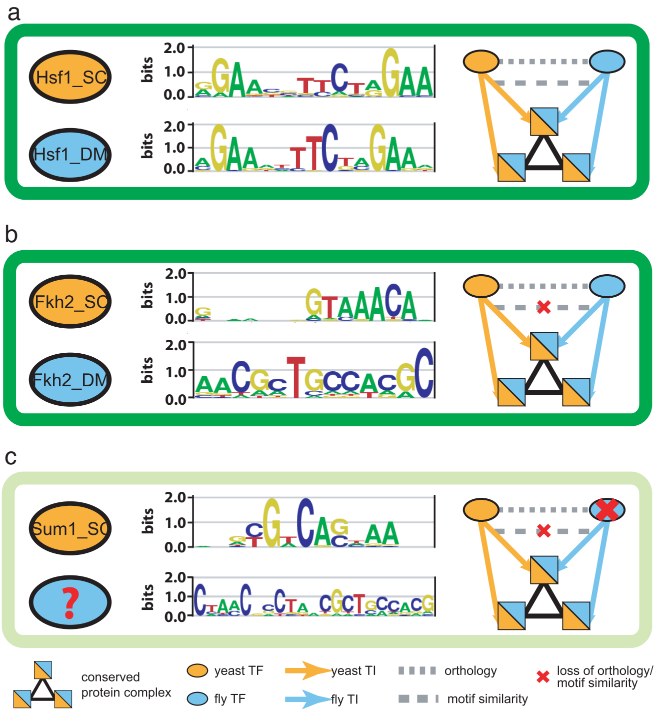


Fig. 4. Conservation levels of control mechanisms regulating conserved clusters. (*a*) Both the TF and its DNA-binding motif are conserved. (*b*) The TF is conserved as an ortholog but not its DNA-binding motif. (*c*) The TF regulating the conserved complex has likely changed.

corresponding to conserved coregulated clusters. Motifs 4 (enriched in putative fly regulon d.Fkh2) and 5 (enriched in putative fly regulon d.Mbp1) were identified by searching sets of promoters of putative fly regulons (*Materials and Methods*). Notably, all four previously uncharacterized motifs bear no similarity to the known motifs of the corresponding yeast TFs (see below).

Evolution of the Regulatory Mechanisms of Protein Complexes. We found that the 24 conserved coregulated clusters could be categorized into three levels of conservation depending on their regulatory mechanisms. At the highest level of conservation (Fig. 4*a*), there were two clusters that showed conservation in both the regulating TF and its DNA-binding motif. Both of these clusters (1 and 12) involved the heat shock factor Hsf1 as the regulating yeast TF, which has a clear ortholog in fly. Moreover, a previous study has demonstrated that heat shock factors and their DNA motifs are highly conserved from yeast to human (32). Indeed, for Hsf1 the DNA-binding domain (DBD) of the orthologous TFs are 99% alignable and the amino acid residues involved in contacting DNA sites are highly conserved (33). In addition, their DNA motifs were also alignable over their entire length.

At a middle level of conservation (Fig. 4*b*), there were eleven clusters in which the TF was conserved as an ortholog but not the DNA-binding motif. Yeast TFs regulating these clusters included Abf1, Cbf1, Fkh2, Mcm1, and Yhp1. In these cases, the DBDs of the orthologous fly TFs may have diverged to such an extent that their DNA-binding specificities have also changed (on average, the fly DBD is alignable to the corresponding yeast DBD at only 69% of the positions). For instance, the fly motif associated with clusters regulated by Fkh2 is very different from that of its yeast homolog (motifs alignable at 3 of 12 positions, Fig. 4*b*). Although it is possible that a nonorthologous fly TF might be responsible for the coregulation (nonorthologous displacement), to the best of our knowledge, this scenario is very rare and mostly observed in bacterial systems because of horizontal gene transfer (34).

Finally, at the lowest level of conservation (Fig. 4*c*), the TF regulating the conserved cluster has likely changed, as there is no detectable fly ortholog of the yeast TF. Eighteen clusters were in this category, involving the yeast TFs Dig1, Mbp1, Ndd1, Rcs1, Reb1, Rpn4, Ste12, Sum1, and Swi4. In these cases, the conserved clusters are probably regulated by nonhomologous TFs. For instance, in clusters 4 and 5 the yeast TFs Ndd1 and Sum1 are not conserved in fly, and the fly DNA motifs (Table 2 and Fig. 4*c*) do not match any known yeast motif.

Discussion

A strength of integrative approaches, such as the one proposed here, is their ability to cope with high levels of noise in any single data set. To cope with false positives, we have estimated the reliabilities of the protein–protein interactions we considered, and incorporated these estimates into our probabilistic model for protein complexes (see *SI Materials and Methods*). In addition, network comparison itself serves to reduce false positives because spurious protein interactions are generally not reproducible across species (23). False negatives were handled in two ways: first, by integrating data from two independent sources (protein–protein and transcriptional interactions); and second, by employing a "soft" definition of a complex in our probabilistic model, i.e., not forcing a certain interaction density for a complex, but rather measuring the likelihood that it fits our model of a protein complex against the likelihood that it arose at random.

The coregulated protein clusters reinforce the idea of combinatorial regulation as a primary mechanism for achieving fine-tuned transcriptional control (35). Many of the clusters were regulated by more than one TF (13/24 clusters were associated with two to six TFs). Included in this set are many well known examples of coregulation, such as the cohesin complex involved in chromosome segregation (cluster 6, regulated by Swi6 and Mbp1, Fig. 1*b*) (36) and the actin cap complex involved in budding (cluster 62, regulated by Swi4, Mbp1 and Mcm1, Fig. 1*c*) (4). We also uncovered many putative complexes with combinatorial regulation, such as complexes involved in mRNA transport (cluster 53 regulated by Hsf1 and Abf1, Fig. 1*d*) and regulation of cell size and shape (cluster 25 regulated by Cbf1, Msn2, Tye7, Fig. 1*e*).

Our cross-species analysis revealed that most conserved clusters (22/24) had regulatory mechanisms that were divergent between yeast and fly, either at the motif level or at the TF level. Hence, it was tempting to conclude that transcriptional interactions are, overall, less conserved than protein–protein interactions across the same evolutionary distance. To investigate this hypothesis, we used two different measures to compare the conservation levels of the PPI and TI networks. As a first measure, we computed the fraction of all large-scale yeast interaction data (protein–protein or transcriptional) in which both participating genes/proteins had orthologs in fly. For each protein we identified its best BLAST match and considered it a putative ortholog if their BLAST E value was $<1 \times 10^{-7}$. Overall, we obtained a 46% conservation rate for PPI data and a 14% conservation rate for TI data. The same trend was observed by Yu *et al.* (37) by using a similar approach.

As a second measure, we evaluated the level of conservation of the protein clusters we identified versus that of the coregulated clusters. Specifically, we estimated the conservation level of the PPI network as the fraction of protein clusters that overlapped conserved protein clusters. For the TI network, our measure was based on the fraction of coregulated protein clusters that overlapped conserved ones (see *SI Materials and Methods*). For an overlap threshold of 3 proteins, the PPI and TI network conservation rates were 90% and 51%, respectively (the same trend was observed for other thresholds). Thus, by using either measure, it seems that protein–protein interactions are in fact more conserved than their transcriptional counterparts. This finding is consistent with previous observation that transcriptional regulatory network is evolutionarily flexible and is a major driving force for phenotypic variations (38, 39).

In summary, our analysis reveals many coregulated protein complexes that are conserved over large evolutionary distances. It further suggests that transcriptional regulatory mechanisms diverge faster than protein–protein interactions. Further studies are needed to combine additional large-scale networks, such as genetic, metabolic, and coexpression, into more complete models of cellular machinery. Such models could shed light on the interplay and coevolution of molecular networks within the cell.

Materials and Methods

Protein Cluster Scoring and Discovery. We developed a statistical model for a cluster of proteins that are regulated by a set of transcription factors. Under this model, members of a complex are assumed to interact with high probability and each TF in the regulating set is assumed to bind each of the genes in the complex with high probability. We contrast this model against a background model, which assumes that both the transcriptional network and the physical network were chosen uniformly at random from the corresponding collection of all networks with the same degree sequence. For each suggested cluster with a set of regulators, we thus compute a log likelihood ratio score, based on its fit to each of the models. We search for protein clusters using a greedy approach that starts from high scoring seeds and refines them by using local search. The significance of the suggested cluster is computed by comparing its score to that of random clusters that were obtained by applying our methodology to randomly shuffled networks. Full details on the data and the algorithm are available in *SI Materials and Methods*.

For discovering conserved coregulated clusters, we let the protein nodes of the integrated network denote pairs of yeast and fly proteins with significant sequence match (BLAST E value $< 10^{-7}$). The log likelihood ratio score of a conserved cluster is computed as the sum of the log likelihood ratio scores within each species network (using only the PPI term for fly).

For validation purposes, we used a simplified method to identify significant protein clusters or conserved clusters, regardless of their regulation. Specifically, the likelihood ratio score was modified to account only for the PPI data, and the search was restricted to the PPI networks.

Prediction of Novel Transcriptional Interactions in Yeast. We combined three types of measures: expression coherency (EC), conservation coherency (CC), and presence of TF-binding sites (BS), to predict transcriptional interactions. A logistic regression classifier was trained on these three types of measures and used to predict new TIs. See *SI Materials and Methods* for details.

Gene Expression Microarray. Expression profilings were performed with the wild type haploid BY4741 (ATCC, Manassas, VA) and *rpn4* deletion (Research Genetics, Huntsville, AL) strains. For each strain, log-phase haploid yeast cultures growing in synthetic complete media were heat shocked by shifting growth temperature from 30°C to 37°C for 30 min. Duplicated microarray experiments were processed to obtain normalized log expression ratios for each of the ≈6,200 genes represented on the microarray along with P values of differential expression generated by using the VERA package (40). See *SI Materials and Methods* for details.

Enriched DNA Motifs in Conserved Fly Clusters. We searched for both known and previously uncharacterized fly motifs in the promoter regions of fly genes participating in the conserved coregulated clusters. Known fly motifs (46 in total) were taken from the TRANSFAC (41) and JASPAR (42) databases. An enrichment P value for each motif in gene promoters of the conserved clusters was calculated by using the hypergeometric distribution. To discover previously uncharacterized motifs, we generated two data sets: (*i*) sets of promoters for fly genes in each conserved cluster; (*ii*) putative fly regulons, created by pooling all fly genes in the conserved clusters whose yeast counterparts were regulated by the same TF (putative fly regulons). For each set of promoter sequences, we searched for significantly enriched DNA motifs by using the program PhyloCon (43). See *SI Materials and Methods* for details.

We thank Mark Miller and members of the T.I. and R.S. laboratories for helpful discussion. T.I. was supported by National Institute on Environmental Health Sciences (NIEHS) Grant ES014811 and the Packard Foundation. R.S. was supported by an Alon Fellowship and by a research grant from the Ministry of Science and Technology of Israel. K.T. was funded by a postdoctoral fellowship from IBM Corporation, through its Institutes of Innovation program. T.S. was supported by the Tauber Fund.

1. Fields S (2005) *FEBS J* 272:5391–5399.
2. Aebersold R, Mann M (2003) *Nature* 422:198–207.
3. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Hannett N, Kanin, E, *et al.* (2000) *Science* 290:2306–2309.
4. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) *Nature* 409:533–538.
5. Yeger-Lotem E, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter RY, Alon U, Margalit H (2004) *Proc Natl Acad Sci USA* 101:5934–5939.
6. Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) *Bioinformatics* 18(Suppl 1):S233–S240.
7. Yeang CH, Mak HC, McCuine S, Workman C, Jaakkola T, Ideker T (2005) *Genome Biol* 6:R62.
8. Zhang LV, King OD, Wong SL, Goldberg DS, Tong AH, Lesage G, Andrews B, Bussey H, Boone C, Roth FP (2005) *J Biol* 4:6.
9. Simonis N, van Helden J, Cohen GN, Wodak SJ (2004) *Genome Biol* 5:R33.
10. Mewes HW, Frishman D, Mayer KF, Munsterkotter M, Noubibou O, Pagel P, Rattei T, Oesterheld M, Ruepp A, Stumpflen V (2006) *Nucleic Acids Res* 34:D169–D172.
11. Simonis N, Gonze D, Orsi C, van Helden J, Wodak SJ (2006) *J Mol Biol* 363:589–610.
12. Manke T, Bringas R, Vingron M (2003) *J Mol Biol* 333:75–85.
13. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) *Nature* 417:399–403.
14. Bader JS, Chaudhuri A, Rothberg JM, Chant J (2004) *Nat Biotechnol* 22:78–85.
15. Barabasi AL, Oltvai ZN (2004) *Nat Rev Genet* 5:101–113.
16. Sharan R, Ideker T (2006) *Nat Biotechnol* 24:427–433.
17. Lee I, Date SV, Adai AT, Marcotte EM (2004) *Science* 306:1555–1558.
18. de Lichtenberg U, Jensen LJ, Brunak S, Bork P (2005) *Science* 307:724–727.
19. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M (2003) *Science* 302:449–453.
20. Beyer A, Workman C, Hollunder J, Radke D, Moller U, Wilhelm T, Ideker T (2006) *PLoS Comput Biol* 2:e70.
21. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang, M, *et al.* (2004) *Science* 303:808–813.
22. Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, Stockwell BR, Ideker T (2003) *Proc Natl Acad Sci USA* 100:11394–11399.
23. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T (2005) *Proc Natl Acad Sci USA* 102:1974–1979.
24. Herrgard MJ, Lee BS, Portnoy V, Palsson BO (2006) *Genome Res* 16:627–635.
25. Kelley R, Ideker T (2005) *Nat Biotechnol* 23:561–566.
26. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier, K, *et al.* (2002) *Nature* 415:180–183.
27. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat, CM, *et al.* (2002) *Nature* 415:141–147.
28. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo, J, *et al.* (2004) *Nature* 431:99–104.
29. Owsianik G, Balzi, L, Ghislain M (2002) *Mol Microbiol* 43:1295–1308.
30. Olson KA, Nelson C, Tai G, Hung W, Yong C, Astell C, Sadowski I (2000) *Mol Cell Biol* 20:4199–4209.
31. Cook JG, Bardwell L, Kron SJ, Thorner J (1996) *Genes Dev* 10:2831–2848.
32. Liu XD, Liu PC, Santoro N, Thiele DJ (1997) *EMBO J* 16:6466–6477.
33. Vuister GW, Kim SJ, Orosz A, Marquardt J, Wu C, Bax A (1994) *Nat Struct Biol* 1:605–614.
34. Koonin EV, Mushegian AR, Bork P (1996) *Trends Genet* 12:334–336.
35. Beer MA, Tavazoie S (2004) *Cell* 117:185–198.
36. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1998) *Science* 282:699–705.
37. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, Vidal M, Gerstein M (2004) *Genome Res* 14:1107–1118.
38. Lozada-Chavez I, Janga SC, Collado-Vides J (2006) *Nucleic Acids Res* 34:3434–3445.
39. Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB (2004) *PLoS Biol* 2:e398.
40. Ideker T, Thorsson V, Siegel AF, Hood LE (2000) *J Comput Biol* 7:805–817.
41. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer, K, *et al.* (2006) *Nucleic Acids Res* 34:D108–10.
42. Vlieghe D, Sandelin A, De Bleser PJ, Vleminckx K, Wasserman WW, van Roy F, Lenhard B (2006) *Nucleic Acids Res* 34:D95–D97.
43. Wang T, Stormo GD (2003) *Bioinformatics* 19:2369–2380.