Exploring Transcriptional Regulatory Networks in the Worm

Illés J. Farkas,[1] Qasim K. Beg,[2] and Zoltán N. Oltvai[2,*]
[1]Department of Biological Physics and HAS Group, Eötvös University, Budapest 1117, Hungary
[2]Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15213, USA
*Contact: oltvai@pitt.edu
DOI 10.1016/j.cell.2006.06.002

Using a yeast one-hybrid assay, Deplancke et al. (2006) report in this issue a protein-DNA interaction network for 72 gene promoters and 117 regulatory proteins expressed in cells of the nematode digestive tract. This study is a first step toward mapping transcriptional regulatory networks in distinct metazoan cell lineages and organs using a "gene-centered" approach.

Living cells continuously monitor their environment for external cues and based on this information and their own internal state mount appropriate responses that often include transcriptional regulation. During transcriptional regulatory events, transcription factors, RNA binding proteins, and regulatory RNAs alter gene expression to achieve new metabolic states or to initiate cellular programs such as the cell cycle or differentiation. In the yeast *Saccharomyces cerevisiae* and in a few mammalian cell types, there has emerged a large amount of data related to transcriptional regulation from experiments using chromatin immunoprecipitation (ChIP) assays, a method that provides information on the in vivo localization of chromatin-bound proteins (such as transcription factors and chromatin modifiers) under selected environmental conditions (Harbison et al., 2004; Pokholok et al., 2005; Wei et al., 2006; Workman et al., 2006). However, the disadvantages of this powerful "protein-centered" technique are that it relies on known DNA binding proteins, requires highly specific antibodies against them, and can be limited by the scarcity of cells and the low concentrations of many regulatory proteins that are typically found in complex metazoan cell lineages.

In this issue of *Cell*, Walhout and colleagues (Deplancke et al., 2006) demonstrate the benefits of an alternative "gene-centered" approach for the large-scale identification of proteins that bind to gene promoters. For their model system, they chose cells of the digestive tract of the nematode *Caenorhabditis elegans*. Their screen uses the yeast one-hybrid method—a variant of the well-known yeast two-hybrid system—developed to identify potential DNA-protein interactions. In this assay, a promoter (obtained in part from the *C. elegans* promoterome resource [http://vidal.dfci.harvard.edu/promoteromedb]) is first cloned upstream of selection and reporter genes (in this case *His3*, *Ura3*, and *LacZ*, respectively) to create "bait-reporter" constructs that are subsequently integrated into the chromosome of yeast cells. Following selection and the elimination of self-activating bait-reporter strains, the remaining clones are screened against a cDNA or transcription factor mini-library (obtained in part from the *C. elegans* ORFeome resource [http://worfdb.dfci.harvard.edu/]). β-galactosidase positive strains are subsequently sequenced to identify all promoter bound gene products that initiated *lacZ* expression (Deplancke et al., 2004).

Starting with 167 gene promoters, this protocol and a subsequent filtering scheme identified 209 protein-DNA interactions in 110 distinct yeast strains bearing the integrated reporter constructs with *C. elegans* promoters. An additional directed yeast one-hybrid screen that used all transcription factor gene products available in the *C. elegans* ORFeome to test binding to each of the 110 promoters resulted in 74 more protein-DNA interactions. Thus, together the two protocols yielded a total of 283 protein-DNA interactions among 72 promoters and 117 gene products. Of these, ten were proteins that had not previously been recognized as transcription factors. Follow-up ChIP experiments confirmed DNA binding for eight out of the ten putative transcription factors, demonstrating that the gene-centered approach is indeed useful for discovering new regulatory proteins. In addition, the ability to screen up to one million colonies simultaneously allows the one-hybrid system to identify interactions with proteins of very low abundance. A bacterial version of the system is already used for the identification of DNA binding motifs (Meng et al., 2005) and, with some modifications, the system may also be used to identify RNA binding proteins or for analyses of structure-function for either the DNA binding motif or the DNA binding protein.

But there are disadvantages, too. In its current format, the yeast one-hybrid method cannot detect the potential cooperative and synergistic activity of transcription factors. Moreover, some transcription factors alone can have nonspecific DNA binding activity but bind specifically to target sequences in vivo by interacting with other transcription fac-

tors. Thus, some of the interactions discovered by the yeast one-hybrid method may represent binding to nonauthentic target promoters. Another limiting factor is that if the promoter of an intestinal gene is used in the yeast one-hybrid system and an interaction with a transcription factor that is expressed in a handful of pharyngeal neurons is obtained, this will be provisionally recorded as a physiologically relevant interaction. However, this interaction would be relevant only if the target gene shows a change in its expression upon altered expression of the transcription factor, and if both are found in the same cell.

In a series of experiments, Deplancke et al. (2006) obtained worm deletion strains for some of the transcription factors that were identified as binding to intestinal gene promoters and compared expression of their putative gene targets with that in wild-type animals. For example, using quantitative RT-PCR they observed expression changes among some of the target genes of DAF-3. To demonstrate coexpression of a transcription factor and its target gene, they then created transgenic worms expressing GFP under the control of the *mdl-1* gene promoter, one of the targets identified for DAF-3. RNAi-induced knockdown of DAF-3 expression resulted in an increased *mdl-1* promoter activity. Moreover, a ChIP assay using an anti-DAF-3 antibody, showed that DAF-3 binds directly to the endogenous *mdl-1* promoter, clearly indicating that it is a direct transcriptional repressor of *mdl-1* expression in vivo.

The ChIP and yeast one-hybrid methods are thus highly complementary to each other and their combined use will be of great value in mapping transcriptional regulatory networks in metazoan cell lineages. However, neither method provides temporal resolution nor allows one to determine the activity of a bound transcription factor a priori. Moreover, the binding and activity of transcription factors are often contextual and combinatorial (Ao et al., 2004; Harbison et al., 2004) requiring detailed systematic follow-up studies (Deplancke et al., 2006; Workman et al., 2006). Yet it is clear that experimental biology is undergoing a major transition: the small-scale hypothesis-driven experiments of individual research groups are increasingly complemented by focused large-scale systematic data collection projects requiring extended collaborations. This paradigm shift—initiated by the various genome projects—is transforming both structural and functional genomics (e.g., protein-protein interactions, protein-DNA interactions, transcriptome profiling, RNAi). The advantages of high-throughput methods are difficult to deny: they are fast, relatively low cost, and can be organized along modern industrial principles. Moreover, the data they provide has an internal consistency that is usually missing from a collection of individual experiments. Their disadvantage, however, is that they allow one to investigate a system only under relatively few conditions or, as with the yeast one-hybrid system, without any functional context.

Compared to the large-scale data collection projects of for example, astronomy or particle physics, the data sets produced by molecular biology may not be the largest, but they are certainly among the most complex. Each entry whether it is a molecule, an interaction, or a molecular complex can have a surprisingly large number of properties (such as phosphorylation or methylation states, and binding affinities), which is further complicated by the large variety of experimental techniques and mathematical interpretation tools (cut-offs, probabilities, vectors, and matrices) used to capture these properties. Thus, consulting all primary data sources can be prohibitively difficult for a single research lab, and indeed, a significant portion of published data on individual molecular interactions becomes rapidly invisible to the individual researcher (Cokol et al., 2005). To retain this large amount of highly structured information as active knowledge, expert-based curation "pipelines" (such as the BIND, BioGRID, and RegulonDB databases, and several industrial databases) are currently the best choice as they compile data sets for both the academic and industrial research communities. However, given that PubMed alone indexes more than 16 million research articles, the number of professional curators is relatively small leading to a bottleneck in the process of mediating experimental data to the 'users', the biomedical community, which may leave large amounts of context-specific information behind. The approach of Deplancke et al. is among several new techniques contributing to the rapid expansion of repositories of molecular biological data on transcriptional regulatory interactions. It is evident that databases will play an increasingly critical role in the coherent organization of collective biomedical knowledge. Yet, as long as data integration and curation pipelines remain strictly expert-based, segments of existing context-specific information about each individual molecule and its interactions may remain uncurated. This difficulty may be alleviated in part through community-level open access annotation projects such as ORegAnno established for transcriptional interactions (Montgomery et al., 2006). Such efforts require a well-designed entry format, online and university-based tutorials for data entry, assessment of confidence levels, and computer-readable content. Yet, once up and running, such a comprehensive integration of biomedical knowledge would be of benefit to all.

REFERENCES

Ao, W., Gaudet, J., Kent, W.J., Muttumu, S., and Mango, S.E. (2004). Science *305*, 1743–1746.

Cokol, M., Iossifov, I., Weinreb, C., and Rzhetsky, A. (2005). Nat. Biotechnol. *23*, 1243–1247.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A.J.M. (2004). Genome Res. *14*, 2169–2175.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A.M., Grove, C.A., Martinez, N.J., Sequerra, R., Doucette-Stamm, L., Reece-Hoyes, J.S., Hope, I.A., et al. (2006). Cell, this issue.

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). Nature *431*, 99–104.

Meng, X., Brodsky, M.H., and Wolfe, S.A. (2005). Nat. Biotechnol. *23*, 988–994.

Montgomery, S.B., Griffith, O.L., Sleumer, M.C., Bergman, C.M., Bilenky, M., Pleasance, E.D., Prychyna, Y., Zhang, X., and Jones, S.J. (2006). Bioinformatics *22*, 637–640.

Pokholok, D.K., Harbison, C.T., Levine, S.S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Cell *122*, 517–527.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., et al. (2006). Cell *124*, 207–219.

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). Science *312*, 1054–1059.

HIV Pathogenesis: Nef Loses Control

John L. Foster[1,*] and J. Victor Garcia[1,*]
[1]Division of Infectious Diseases, Department of Internal Medicine, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390, USA
*Contact: john.foster@utsouthwestern.edu, victor.garcia@utsouthwestern.edu
DOI 10.1016/j.cell.2006.06.005

In this issue of *Cell*, Schindler et al. (2006) show that the Nef protein from nonpathogenic strains of simian immunodeficiency virus (SIV) induces the downregulation of host T cell receptor/CD3, whereas Nef from human immunodeficiency virus (HIV-1) does not. This loss of function in the Nef of HIV-1 may partly explain the pathogenicity of this virus.

Despite their relatively small size, the *nef* genes of HIV-1, HIV-2, and SIV pack a big punch. Nef is a highly complex and multifunctional protein that intersects multiple host cellular pathways essential for cell survival and function. Despite extensive in vitro characterization, the role of Nef during in vivo infection remains unclear. In this issue of *Cell*, Schindler et al. (2006) present intriguing evidence suggesting that, during evolution, the Nef protein of HIV-1 lost its ability to downregulate host T cell receptor/CD3 (TCR-CD3). This new characteristic of Nef may have provided HIV-1 with a dramatic pathogenic potential in its human host.

The Nef protein of HIV was originally viewed as a negative regulator of the long terminal repeat transcriptional promoter of HIV. As *nef* is one of the HIV-1 genes not required for viral production, it was named for its presumed negative role (*ne*gative *f*actor). This notion was dismissed with the finding that Nef from the SIVmac strain was essential for the maintenance of high virus load and disease progression in rhesus macaques (Kestler et al., 1991). Subsequently, similar observations were made with respect to HIV-1 Nef in a cohort of Australian patients infected after transfusions of blood containing a Nef-defective virus (Deacon et al., 1995). These and other studies emphasize the importance of Nef in both HIV and SIV pathogenesis, and since then, multiple possible explanations for this dramatic in vivo phenotype have been proposed (Simmons et al., 2001). Thus, Nef has been labeled a pathogenic factor.

In general, Nefs from all primate lentiviruses (HIV-1, HIV-2, and SIV) share multiple activities in vitro. These include downregulation of CD4, MHC class I proteins, and CD28 in host cells; enhanced infectivity of the viral particle; and activation of cellular kinases (Arora et al., 2002). In their new work, Schindler et al. (2006) focus their attention on the one in vitro activity that is absent from all HIV-1 Nefs tested, namely the ability to induce host-cell TCR-CD3 downregulation. It is well established that HIV-1 Nef does not induce TCR-CD3 downregulation, whereas HIV-2 and most SIV Nefs do. The Schindler et al. (2006) study now makes an important distinction regarding Nefs from different species. Specifically, the Nef protein of SIVcpz, the primate lentivirus that is most closely related to HIV-1, also does not induce host cells to downregulate TCR-CD3, whereas Nefs from SIVsmm, the closest primate virus to HIV-2 retain this ability.

Although sooty mangabeys are the natural host of SIVsmm, this strain of virus does not cause disease in a majority of these animals (Rey-Cuille et al., 1998), but it does in rhesus macaques (Schindler et al., 2006). However, a small percentage of infected mangabeys do experience a decline in the number of their $CD4^+$ T cells that correlates with increased activation of the immune system. Nefs from the viruses present in mangabeys that experienced $CD4^+$ T cell loss were slightly defective in their ability to downregulate TCR-CD3. This result suggested that Nef may play an important protective role by maintaining low levels of immune activation in naturally infected primates. Therefore, the results from Schindler et al. (2006) support the notion of a crucial role for