Genome-Wide Location and Function of DNA Binding Proteins

Bing Ren, François Robert, John J. Wyrick, Oscar Aparicio, Ezra C. Jennings, Itamar Simon, Julia Zeitlinger, Jörg Schreiber, Nancy Hannett, Elenita Kanin, Thomas L. Volkert, Christopher J. Wilson, Stephen P. Bell, Richard A. Young

Understanding how DNA binding proteins control global gene expression and chromosomal maintenance requires knowledge of the chromosomal locations at which these proteins function in vivo. We developed a microarray method that reveals the genome-wide location of DNA-bound proteins and used this method to monitor binding of gene-specific transcription activators in yeast. A combination of location and expression profiles was used to identify genes whose expression is directly controlled by Gal4 and Ste12 as cells respond to changes in carbon source and mating pheromone, respectively. The results identify pathways that are coordinately regulated by each of the two activators and reveal previously unknown functions for Gal4 and Ste12. Genome-wide location analysis will facilitate investigation of gene regulatory networks, gene function, and genome maintenance.

Many proteins bind to specific sites in the genome to regulate genome expression and maintenance. Transcriptional activators, for example, bind to specific promoter sequences and recruit chromatin modifying complexes and the transcription apparatus to initiate RNA synthesis (1–3). The reprogramming of gene expression that occurs as cells move through the cell cycle, or when cells sense changes in their environment, is effected in part by changes in the DNA binding status of transcriptional activators. Distinct DNA binding proteins are also associated with origins of DNA replication, centromeres, telomeres, and other sites, where they regulate chromosome replication, condensation, cohesion, and other aspects of genome maintenance (4, 5). Our understanding of these proteins and their functions is limited by our knowledge of their binding sites in the genome.

The genome-wide location analysis method we have developed allows protein-DNA interactions to be monitored across the entire yeast genome (6). The method combines a modified chromatin immunoprecipitation (ChIP) procedure, which has been previously used to study protein-DNA interactions at a small number of specific DNA sites (7), with DNA microarray analysis. Briefly, cells were fixed with formaldehyde, harvested, and disrupted by sonication. The DNA fragments cross-linked to a protein of interest were enriched by immunoprecipitation with a specific antibody. After reversal of the cross-links, the enriched DNA was amplified and labeled with a fluorescent dye (Cy5) with the use of ligation-mediated–polymerase chain reaction (LM-PCR). A sample of DNA that was not enriched by immunoprecipitation was subjected to LM-PCR in the presence of a different fluorophore (Cy3), and both immunoprecipitation (IP)-enriched and -unenriched pools of labeled DNA were hybridized to a single DNA microarray containing all yeast intergenic sequences (Fig. 1). A single-array error model (8) was adopted to handle noise associated with low-intensity spots and to permit a confidence estimate for binding (P value). When independent samples of 1 ng of genomic DNA were amplified with the LM-PCR method, signals for greater than 99.8% of genes were essentially identical within the error range (P value ≤10⁻⁴). The IP-enriched/unenriched ratio of fluorescence intensity obtained from three independent experiments was used with a weighted average analysis method to calculate the relative binding of the protein of interest to each sequence represented on the array.

To investigate the accuracy of the genome-wide location analysis method, we used it to identify sites bound by the transcriptional activator Gal4 in the yeast genome. Gal4 activates genes necessary for galactose metabolism and is among the best characterized transcriptional activators (1, 9). We found 10 genes to be bound by Gal4 (P value ≤0.001) and induced in galactose using our analysis criteria (Fig. 2A). These included seven genes previously reported to be regulated by Gal4 (GAL1, GAL2, GAL3, GAL7, GAL10, GAL80, and GCY1). The MTH1, PCL10, and FUR4 genes were also bound by Gal4 and activated in galactose. Each of these results was confirmed by conventional ChIP analysis (Fig. 2B) (6), and MTH1, PCL10, and FUR4 activation in galactose was found to be dependent on Gal4 (Fig. 2C). Both microarray and conventional ChIP showed that Gal4 binds to GAL1, GAL2, GAL3, and GAL10 promoters under glucose and galactose conditions, but the binding was generally weaker in

---

1Whitehead Institute for Biomedical Research, Cambridge Center, Cambridge, MA 02142, USA. 2Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 3Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 4Program in Molecular Biology University of Southern California, Los Angeles, CA 90089—1340, USA. 5Corning, Inc., Corning, NY 14834, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: young@wi.mit.edu

---

1 Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, USA. 2 Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 3 Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 4 Program in Molecular Biology University of Southern California, Los Angeles, CA 90089—1340, USA. 5 Corning, Inc., Corning, NY 14834, USA.

* These authors contributed equally to this work.
† To whom correspondence should be addressed. E-mail: young@wi.mit.edu

---

1 Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, USA. 2 Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 3 Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 4 Program in Molecular Biology University of Southern California, Los Angeles, CA 90089—1340, USA. 5 Corning, Inc., Corning, NY 14834, USA.

* These authors contributed equally to this work.
† To whom correspondence should be addressed. E-mail: young@wi.mit.edu
glucose (6). The consensus Gal4 binding sequence that occurs in the promoters of these genes (CGGNN_{1–3}CCG) can also be found at many sites through the yeast genome where Gal4 binding is not detected; therefore, sequence alone is not sufficient to account for the specificity of Gal4 binding in vivo. Previous studies of Gal4-DNA binding have suggested that additional factors such as chromatin structure contribute to specificity in vivo (10, 11).

The identification of MTH1, PCL10, and FUR4 as Gal4-regulated genes reveals previously unknown functions for Gal4 and explains how regulation of several different metabolic pathways can be coordinated (Fig. 2D). MTH1 encodes a transcriptional repressor of certain HXT genes involved in hexose transport (12). Our results suggest that the cell responds to galactose by increasing the concentration of its galactose transporter at the expense of other transporters. In other words, while Gal4 activates expression of the galactose transporter gene GAL2, Gal4 induction of the MTH1 repressor gene leads to reduced levels of glucose transporter expression. The Pcl10 cyclin associates with Pho85p and appears to repress the formation of glycogen (13). Thus, the observation that PCL10 is Gal4-activated suggests that reduced glycogenesis occurs to maximize the energy obtained from galactose metabolism. FUR4 encodes a uracil permease (14), and its induction by Gal4 may reflect a need to increase intracellular pools of pyrimidines to permit efficient uridine 5'-diphosphate (UDP) addition to galactose catalyzed by Gal7.

We next investigated the genome-wide binding profile of the transcription activator Ste12, which functions in the response of haploid yeast to mating pheromones (15). Activation of the pheromone-response pathway by mating pheromones causes cell cycle arrest and transcriptional activation of more than 200 genes in a Ste12-dependent fashion (8, 15). However, it is not clear which of these genes is directly regulated by Ste12 and which are regulated by other ancillary factors. The genome-wide binding profile of epitope-tagged Ste12, determined before and after pheromone treatment in three independent experiments, indicates that 29 pheromone-induced genes are regulated directly by Ste12. Figure 3A lists the yeast genes whose promoter regions are bound by Ste12 at the 99.5% confidence level (i.e., P value <0.005) and whose expression is induced by α factor. These 29 genes are likely to be directly regulated by Ste12 because (i) all have promoter regions bound by Ste12, (ii) exposure to pheromone causes an increase in their transcription, and (iii) pheromone induction of transcription is dependent on Ste12.

Of the genes that are directly regulated by Ste12, 11 are already known to participate in various steps of the mating process (Fig. 3B). Fus3 and STE12 encode components of the signal transduction pathway involved in the response to pheromone (16); AFRI and GIC2 are required for the formation of mating projections (17–19); FIG2, AGA1, FIG1, and FUS1 are involved in cell fusion (20–23); and CIK1 and was quantified by phosphorimaging. (D) Model summarizing the role of Gal4 in galactose-dependent cellular regulation. The products of genes newly identified as directly regulated by Gal4 are shown as green circles; those previously identified are shown in blue.

![Fig. 2](https://www.sciencemag.org/science/vol290/issue2423/)

**Fig. 2.** Genome-wide location of Gal4 protein. (A) Genes whose promoter regions were bound by myc-tagged Gal4 (P value <0.001) and whose expression levels were induced at least twofold by galactose are listed. The weight-averaged ratios and P values are shown for Gal4 binding in galactose and glucose. Binding ratios are also displayed using a blue and white color scheme and expression ratios of galactose/glucose are displayed using a red and green color scheme. (B) Confirmation of microarray data for each gene in panel A using conventional chromatin IP procedure. Strains with (+) or without (−) a myc-tagged Gal4 protein were grown in galactose. Amplification of the unenriched DNA (I) and IP-enriched DNA (P) is shown. ARN1 (control) was used as a negative control. (C) Galactose-induced expression of FUR4, MTH1, and PCL10 is Gal4-dependent. Samples from wild-type and gal4− strains were taken before and after addition of galactose. The expression of FUR4, MTH1, and PCL10 was monitored by quantitative reverse transcriptase–PCR (RT-PCR)
and KAR5 are required for nuclear fusion (24). Furthermore, Fus3 and Far1 are required for pheromone-induced cell cycle arrest (25, 26) (Fig. 3B). Among the Ste12 target genes identified in this study that were not previously reported to be involved in mating, many are involved in processes likely to be relevant to mating: CHS1, PCL2, ERG24, SPC25, HYM1, and PGM1 encode proteins involved in cell wall biosynthesis, cell morphology, membrane biosynthesis, nuclear congression and regulation of gene expression (Fig. 3B). Furthermore, YER019W, YOR129C, and SCH9 are among the genes that are cell cycle regulated (27).

The genes that are regulated by Ste12 can be divided into two classes: those bound by Ste12 both before and after pheromone exposure (e.g., STE12, PCL2, Fig2, and Fus1) and those bound by Ste12 only after exposure to pheromone (e.g., Cki1 and CHS1) (Fig. 3A). The first class of genes is induced immediately after pheromone exposure, most likely by a mechanism that converts an inactive DNA-bound Ste12 protein to an active transcriptional activator. This could take place by removal of repressors of Ste12 such as Dig1/Rst1 and Dig2/Rst2 (28). In the second class of genes, induction of transcription is relatively slow. In this case, the binding of Ste12 appears to be limited before pheromone exposure. It is also possible that the epitope tag on Ste12 is masked at these promoters before pheromone treatment, perhaps due to the presence of additional regulatory proteins.

We have shown that a combination of genome-wide location and expression analysis can identify the global set of genes whose expression is controlled directly by transcriptional activators in vivo. The application of location analysis to two yeast transcriptional activators revealed how multiple functional pathways are coordinately controlled in vivo.
Inhibition of Eukaryotic DNA Replication by Geminin Binding to Cdt1

James A. Wohlschlegel, Brian T. Dwyer, Suman K. Dhar, Christin Cvetic, Johannes C. Walter, Anindy V Dutta

In all eukaryotic organisms, inappropriate firing of replication origins during the G2 phase of the cell cycle is suppressed by cyclin-dependent kinases. Multicellular eukaryotes contain a second putative inhibitor of re-replication called geminin. Geminin is believed to block binding of the mini-chromosome maintenance (MCM) complex to origins of replication, but the mechanism of this inhibition is unclear. Here we show that geminin interacts tightly with Cdt1, a recently identified replication initiation factor necessary for MCM loading. The inhibition of DNA replication by geminin that is observed in cell-free DNA replication extracts is reversed by the addition of excess Cdt1. In the normal cell cycle, Cdt1 is present only in G1 and S, whereas geminin is present in S and G2 phases of the cell cycle. Together, these results suggest that geminin inhibits inappropriate origin firing by targeting Cdt1.

Eukaryotic cells maintain a tight control on the initiation of DNA replication to ensure that chromosome duplication in S phase and chromsome segregation in M phase strictly alternate with each other so that daughter cells do not receive too many or too few chromosomes. Fusion of cells in the G2 phase of the cell cycle to those in S phase revealed that the G2 nuclei were unable to replicate their DNA, although this situation was also observed in Xenopus eggs. Although geminin was shown to block DNA replication by preventing the loading of MCM complex on chromatin (7), the target of geminin in this inhibitory reaction was unclear, but its expression in S and G2 led to the suggestion that it provides an additional mechanism to prevent re-replication in the normal cell cycle. To identify the target of geminin, we raised antibodies to bacterially expressed His6-geminin (Fig. 1A) (8). Immunoblotting of extracts from 293T cells with this antibody recognized the 33-kD geminin protein in cell extracts (Fig. 1A). Immunoprecipitation of 35S-methionine labeled cell extracts indicated that the 33-kD geminin protein co-immunoprecipitated with cellular proteins of 65 and 130 kD (Fig. 1B) (9). Re-immunoprecipitation with antibody to geminin under denaturing conditions precipitated geminin but not the 65- and 130-kD polypeptides (Fig. 1B). These results suggest that the 65- and 130-kD polypeptides were present in the native anti-geminin immunoprecipitates and that geminin was released from Cdt1 in the presence of denaturant.

References and Notes
29. We thank P. Cosma, K. Nasmyth, V. Myer, F. Holstege, and F. Lewitter for reagents, protocols, and helpful discussions. Supported by funds from Corn ing, Inc., NIH (R.A.Y. and S.P.B.), Helen Hay Whitney Foundation (B.R.), National Cancer Institute of Canada (F.R.), NSF (J.J.W.), Burroughs-Wellcome Fund (O.A.), Howard Hughes Medical Institute (E.G.) and S.P.B., European Molecular Biology Organization (I.S.), and the Human Frontier Science Program (J.Z.).

18 September 2000; accepted 16 November 2000