A global transcriptional regulatory role for c-Myc in Burkitt’s lymphoma cells

Zirong Li*, Sara Van Calcar*, Chunxu Qu*, Webster K. Cavenee†, Michael Q. Zhang‡, and Bing Ren*§

*Ludwig Institute for Cancer Research and Departments of †Cellular and Molecular Medicine and ‡Medicine, University of California at San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0653; and §Cold Spring Harbor Laboratory, 1 Bungtown Road, P.O. Box 100, Cold Spring Harbor, NY 11724

Communicated by Thomas Maniatis, Harvard University, Cambridge, MA, May 8, 2003 (received for review March 24, 2003)

Overexpression of c-Myc is one of the most common alterations in human cancers, yet it is not clear how this transcription factor acts to promote malignant transformation. To understand the molecular targets of c-Myc function, we have used an unbiased genome-wide location-analysis approach to examine the genomic binding sites of c-Myc in Burkitt’s lymphoma cells. We find that c-Myc together with its heterodimeric partner, Max, occupy >15% of gene promoters tested in these cancer cells. The DNA binding of c-Myc and Max correlates extensively with gene expression throughout the genome, a hallmark attribute of general transcription factors. The c-Myc/Max heterodimer complexes also colocalize with transcription factor IID in these cells, further supporting a general role for overexpressed c-Myc in global gene regulation. In addition, transcription of a majority of c-Myc target genes exhibits a similar mechanism of c-Myc-mediated repression in rodent model systems has shown that overexpression of c-Myc can cause malignant transformation, and that sustained high levels of c-Myc due to a chromosomal translocation (34). We show that the overexpressed c-Myc binds to a large number of gene promoters in these cancer cells, amounting to nearly 15% of the loci tested. Furthermore, the DNA binding by c-Myc possesses an N-terminal transactivation domain. Deletions or mutations in this domain result in loss of c-Myc transformation (15). This model implies that c-Myc may cause transformation by activating a select set of genes that in turn play key roles in malignant transformation (3). Throughout the years, a large number of genes targeted by c-Myc regulation either directly or indirectly have been found. However, such “transformation” genes remain elusive (16).

Some evidence suggests that c-Myc may promote transformation through different mechanisms. First, the transcriptional activation potential of c-Myc does not always correlate with its ability to transform rodent fibroblast cells (4). For example, several studies showed that mutations in the Myc box II domain within c-Myc can abrogate its transformation capacity without affecting c-Myc activation of reporter gene constructs (17, 18). Second, c-Myc also acts as a transcriptional repressor. The mechanism of c-Myc-mediated repression is not entirely clear but in some cases may involve the association of c-Myc/Max with transcriptional activators (19–21). Because some of the genes normally repressed by c-Myc are key cell-cycle regulators, it is conceivable that c-Myc-mediated repression of these genes may also contribute to tumorigenesis (22).

More recently, it was shown that c-Myc could directly activate RNA polymerase (pol) III promoters (23). Because most pol III promoters lack the canonical c-Myc binding sites, it was demonstrated that c-Myc-regulated transcription from pol III promoters occurs through its association with the transcription factor IIIB (TFIIIB) complex, which is a pol III-specific general transcription factor. This suggests a potentially broad role for c-Myc in regulating gene expression and raises the possibility that c-Myc may use a similar mechanism in regulating pol II promoters.

Identifying the genomic binding sites of c-Myc in cancer cells should help resolve the long-standing questions regarding mechanisms of oncogenic transformation by c-Myc. As an initial effort to characterize c-Myc DNA binding in vivo, we used a genome-wide location-analysis approach that allows the determination of transcription factor-binding sites throughout the genome (24, 25). This method has allowed identification of the DNA-binding sites in the yeast genome for proteins involved in transcription, chromatin modification, and DNA replication (26–31). More recently, the same approach has been used to reveal the promoters directly bound and regulated by the E2F transcription factors in human cells (32, 33).

Here we report experiments designed to identify genomic binding sites of c-Myc in Burkitt’s lymphoma cells, which express high levels of c-Myc due to a chromosomal translocation (34). We show that the overexpressed c-Myc binds to a large number of gene promoters in these cancer cells, amounting to nearly 15% of the loci tested. Furthermore, the DNA binding by c-Myc generally correlates with global gene transcription activities and largely coincides with binding of the general transcription factor TFIIIB throughout the genome. A majority of the c-Myc target genes we identified are expressed in other cell types in a manner correlated with c-Myc mRNA levels. Taken together, these results suggest a rather general role for c-Myc in the regulation of genome expression in many cancer cells.

Methods

Design and Manufacturing of Human Promoter Microarrays. To conduct a comprehensive analysis of the genomic sites of c-Myc in human cells, we developed a DNA microarray (henceforth referred to as the hu6K array) that contained PCR products spanning the proximal promoters of 4,839 human genes chosen from the NCBI Refseq database. These genes were selected because their promoters were best-annotated, and there was a clear function associated with each gene. The DNA fragments have an average size of 900 bp and typically cover the sequence from 650 bp upstream to 250 bp downstream of the transcription start site in a gene. The choice of these regions is based on previous observations that human transcription factors frequently bind to proximal promoter sequences (32). The hu6K

Abbreviations: pol, RNA polymerase; TF, transcription factor; hu6K, DNA microarray of ~6,000 human genomic DNA fragments; ChIP, chromatin immunoprecipitation.

*To whom correspondence should be addressed. E-mail: biren@ucsd.edu.
array also includes 729 coding sequence and 221 genomic regions 
>1 kb upstream of the transcription start site of a gene. The 
latter two categories of sequences serve as internal controls for 
the location analysis. The description of all the sequences on the 
hu6K array can be found in Array Description, which is published 

We designed oligonucleotide primers to amplify the genomic 
regions discussed above using the PCR. After PCR amplification, 
we purified each DNA fragment, verified the product by 
agarose gel electrophoresis, and then spotted the purified 
DNA to GAPSII glass slides (Corning) using a contact printer 
(Cartesian Technologies, Irvine, CA). After UV crosslinking, 
the glass slides were stored under vacuum until use.

**Genome-Wide Location Analysis.** A detailed protocol for genome-
wide location analysis with mammalian factors can be found in 
Supporting Materials and Methods. Briefly, Daudi cells (a gift 
from William Sugden, University of Wisconsin, Madison) were 
grown in flasks with RPMI medium 1640 supplemented with 
10% FBS/100 units/ml penicillin/100 μg/ml streptomycin to a 
density of 10⁶ per liter at 37°C in 5% CO₂. A total of 10⁸ 
proliferating Daudi cells were fixed with formaldehyde, har-
vested, and disrupted by sonication. To enrich for target genes 
bound to a transcription factor, we immunoprecipitated the 
resulting chromatin fragments with polyclonal antibodies that 
specifically recognize c-Myc (sc-764), Max (sc-197), E2F1 (sc-
193), TAFⅱ250 (sc-735), or pol II (sc-9001), obtained from Santa 
Cruz Biotechnology. After reversal of crosslinks, the purified 
and enriched DNA was amplified by ligation-mediated PCR and 
subsequently labeled with the Cy5 fluorophore by random 
priming. For purposes of normalization, we also performed 
ligation-mediated PCR on DNA that was not enriched by 
immunoprecipitation and labeled the amplified sample with a 
second fluorescent dye, Cy3. Chromatin immunoprecipitation 
(ChIP)-enriched and nonenriched (total input) pools of DNA 
were mixed with Cot-1 DNA to suppress annealing of repetitive 
sequences and hybridized under stringent conditions to a hu6K 
array. DNA microarray hybridization was carried out as de-
scribed in ref. 32. The microarray then was analyzed by using a 
GenenPix 4000B scanner (Axon Instruments, Foster City, CA).

**Data Analysis.** Analysis of microarray scanning images was per-
formed according to published protocols (24) with modifications 
(see Supporting Materials and Methods, which is published as 
supporting information on the PNAS web site). Data from 
independent replicate experiments were combined (24). An 
average enrichment ratio was calculated for each DNA species 
on the array from at least three replicate data sets. The binding 
of a factor to DNA was deemed significant if the average P 
value was <0.001. By using these criteria, no DNA achieved signifi-

Fig. 1. Genome-wide location analysis of c-Myc- and Max-bound promoters in Burkitt’s lymphoma cells. Five independent experiments were performed to 
identify c-Myc or Max bound DNA in formaldehyde-crosslinked Daudi cells. (A) A scatter plot showing the fluorescent intensities for each spot on the hu6K 
promoter array in one of the c-Myc location-analysis experiments. A few previously known c-Myc target genes are highlighted. (B and C) Scatter plots 
corresponding to one of the Max location-analysis experiments, or the control experiment, in which ChIP was performed in the absence of primary antibodies. 
(D) Venn diagram comparing the DNA binding of c-Myc and Max. The data from five experiments were averaged (see Methods), and the genes with a P value 
<0.001 were counted as targets. Overall, 876 and 931 promoters are bound by c-Myc and Max, respectively. Marked with red in A–C are 776 promoters that are 
bound by both. (E) A comparison of the fraction of promoter or coding regions represented on the hu6K array that are bound by c-Myc, Max, or both in Daudi 
cells. (F) Conventional ChIP confirms the enrichment of c-Myc/Max target genes identified in the genome-wide location-analysis experiments. ChIP 
was performed with chromatin from Daudi cells by using the indicated primary antibodies, and enriched DNA was amplified with primers corresponding to a random 
select number of genes. As a negative control, magnetic beads lacking primary antibodies were used.
cancer in control experiments where input DNA was compared with the same DNA. The Supporting Raw Data Set containing processed microarray data is published as supporting information on the PNAS web site.

Conventional ChIP. Confirmation of a select number of c-Myc/Max target sites by the conventional ChIP method was performed according to ref. 32. The primers used for amplification of genomic sequences corresponding to each binding site were MARS (5'-aagtggacttgcctaaaa-3' and 5'-ccatcgactgggactaca-3'), ATF4 (5'-ctccctttgagccctaa-3' and 5'-ccgacagaaacactca-3'), TFF2 (5'-aattgaagactgtgctca-3' and 5'-ccagctgacacca-3'), GKN1 (5'-gctcttaactctctctt-3' and 5'-ccatgtttgctccagtt-3'), and control (5'-tgggctagagaccaata-3' and 5'-cttgctgtctgctgta-3'). The control PCR product lies 9,595 bp upstream of the transcription start site in the CAD gene.

Results

Widespread DNA Binding by c-Myc/Max in Burkitt's Lymphoma Cells. We first examined the promoter occupancy by c-Myc in Daudi cells, an established cell line originally derived from a Burkitt's lymphoma patient (34, 35). These cells express a high level of c-Myc and have been used extensively as an experimental system to study the role of c-Myc in tumorigenesis. We performed five independent c-Myc location-analysis experiments and averaged the results using an error model described in Supporting Materials and Methods (Fig. 1A). The analysis indicates that c-Myc is associated with the promoters of 876 genes based on a P value ≤0.001. Because c-Myc binds to DNA as a heterodimer with Max, we also performed five independent location-analysis experiments using antibodies that specifically recognize Max. A total of 931 gene promoters was found to be associated with Max in vivo (Fig. 1B). As a control, we carried out immunoprecipitation in the absence of specific antibodies. Only 32 gene promoters were enriched under such conditions, most likely because of their nonspecific interactions with the magnetic beads used in these experiments (Fig. 1C).

Several measures suggest that our location analysis is very robust and highly reliable. First, there is a significant overlap between the set of gene promoters bound by c-Myc and those bound by Max. A total of 776 gene promoters bound both proteins, accounting for 88% of the observed c-Myc binding sites and 83% of the observed Max binding sites (Fig. 1D). Second, consistent with previous knowledge, c-Myc and Max rarely bind to the coding sequences represented on the hu6K array. There is only 1.1% occupancy by c-Myc and Max for these regions compared with the 15% occupancy on gene promoters by both proteins (Fig. 1E). Assuming that observed binding of c-Myc and Max to coding sequences is false positives, we then estimate that the false-positive rates for the c-Myc/Max complex binding results is 7%, which is in line with previous estimates of false-positive rates of genome-wide location-analysis experiments (26). Third, we used the conventional ChIP method with a select set of c-Myc/Max target genes and confirmed all their in vivo protein–DNA interactions (Fig. 1F).

Based on this analysis, we estimate that at least 93% of the gene promoters identified as bound by c-Myc and Max in this study are true positives. Therefore, the c-Myc/Max complexes occupy at least 721 of 776 (93%) gene promoters, or nearly 15% of the 4,839 gene promoters examined, in Daudi cells. Because these genes are an unbiased selection of all human genes, c-Myc/Max complexes probably bind close to 15% or 4,500 of the human genes in these Burkitt's lymphoma cells.

The DNA Binding of c-Myc and Max Corresponds to Genome Transcription Levels. In most cell types, only a small fraction of the genome is expressed. Because c-Myc binds to almost 15% of the human genes in Daudi cells and has a known function as a transcription activator, one prediction is that c-Myc DNA binding would correlate with global gene expression in these cells. To test this hypothesis, we analyzed the correlation of c-Myc or Max DNA binding and gene expression in Daudi cells. As a measure of the gene expression, we used previously reported transcription data obtained from the Daudi cells (36). We divided the total gene promoters represented by the hu6K array into 17 groups based on their transcription activity and plotted the

Fig. 2. DNA binding of c-Myc and Max correlates to genome transcription levels and resembles the behavior of a general transcription factor. (A) Line chart showing the correlation between gene-expression levels and promoter binding by c-Myc, Max, pol II, and E2F1 in Daudi cells. The 4,839 gene promoters in the hu6K arrays were divided into 17 groups based on their transcription activities (see Supporting Data Set for Fig. 2, which is published as supporting information on the PNAS web site). The lower and upper boundaries of log_{10}-transformed gene-expression values for these groups are distributed evenly between –1 and 5. The fraction of promoters bound by each factor in every group is plotted with regard to the log_{10}-transformed transcription activity for the group. (B) Table showing the pairwise comparison of genomic binding profiles of c-Myc, Max, pol II, and E2F1 with respect to genome expression. P values were calculated based on two-factor ANOVA analysis without replicate.

Fig. 3. c-Myc and Max colocalize with TFIIID in Burkitt's lymphoma cells. A Venn diagram compares the gene promoters bound by the c-Myc/Max complexes and that by TFIIID in Daudi cells.
fraction of c-Myc-bound promoters in each group with regard to transcription activities for genes in the group (Fig. 2A). The analysis indicates a strong correlation between c-Myc DNA binding and global gene transcription ($\rho = 0.96$). Nearly 50% of the highly active promoters bound c-Myc in Daudi cells, and the promoter occupancy gradually decreases in gene groups with lower promoter activities. Similarly, there is also a strong correlation ($\rho = 0.95$) between the DNA binding by Max and genome transcription activities (Fig. 2A).

The extensive correlation between c-Myc’s DNA binding and global gene transcription not only confirms our prediction of c-Myc function in Daudi cells but also reveals an extensive role for this oncogenic transcription factor in global gene regulation, a role that usually is associated with general transcription factors. To assess the similarity between the role of c-Myc in genome transcription and that of a general transcription factor, we performed genome-wide location-analysis experiments to identify the genomic binding sites of pol II in Daudi cells. As a control, we also examined the DNA binding of a typical transcriptional activator, E2F1, in these cells (32). In Fig. 2A, the percentage of promoter occupancy by pol II and E2F1 is plotted with regard to levels of promoter activities in the 17 gene groups discussed above. Although the DNA binding for both proteins seems to correlate with genome transcription activities, the DNA-binding profile of E2F1 with respect to global gene expression is markedly different from that of pol II ($P = 0.0047$), c-Myc ($P = 0.0003$), and Max ($P = 0.0002$) when analyzed by using a two-factor ANOVA method (Fig. 2B). By contrast, the DNA-binding profile of pol II is indistinguishable from that of c-Myc ($P = 0.50$) and Max ($P = 0.44$) by using the same analysis. This result strongly suggests that c-Myc and Max act more like a general transcription factor than a sequence-specific transcription factor in Daudi cells.

c-Myc and Max Colocalize with TFIID on Gene Promoters. The similar DNA-binding profiles with respect to genome expression between c-Myc and pol II supports the model that c-Myc may play a general role in gene transcription in Daudi cells. To further confirm this and explore the mechanisms of its action in Daudi cells, we tested whether the DNA binding of c-Myc coincides with other general transcription factors. In particular, because c-Myc was shown to interact directly with TATA-box-binding protein (TBP) (37, 38), the core component of the general transcription factor TFIID, we tested whether c-Myc colocalizes with TFIID on gene promoters.

We performed three independent genome-wide location-analysis experiments to identify the gene promoters associated with TAF$_{II}$250, the largest subunit of TFIID, in Daudi cells (39, 40). We found that c-Myc bound to the promoters of genes that were also bound by TAF$_{II}$250, confirming that c-Myc and Max act more like a general transcription factor than a sequence-specific transcription factor in Daudi cells.

Fig. 4. Expression of c-Myc/Max target genes correlates with levels of c-myc transcript in human tissues and cell lines. (Left) Two-way clustering was performed on the expression data corresponding to 453 c-Myc/Max target genes in 46 human tissues and cell lines. The expression data for each gene were log-transformed and normalized such that the median log expression value is 0. The gene-expression values were represented by using a red-green color scheme, with red corresponding to higher-than-median expression values and green corresponding to lower-than-median expression values. The expression data corresponding to different genes are arranged from left to right, and different tissues are arranged from top to bottom. (Right) The c-myc mRNA levels in the corresponding samples, as measured by using Affymetrix GeneChip, are shown on a bar graph.
Expression of c-Myc/Max Target Genes Correlates with c-myc mRNA Levels in Diverse Tissues and Cell Lines. We next asked whether the promoters bound by c-Myc/Max are expressed in a c-Myc-dependent fashion. For this purpose, we compared the expression of these genes in a variety of human tissues and cell lines in which c-Myc transcript levels range from undetectable to very high (36). Among these c-Myc/Max targets, some genes were not included in the previous study, and many could not be reliably detected in the various samples. Therefore, these genes were excluded from this analysis. The remaining 453 c-Myc target genes were clustered based on their expression across these tissues and cell lines. As shown in Fig. 4, gene expression from a majority of the c-Myc/Max target promoters corresponds with c-myc mRNA levels (Class A). Genes in this category include known c-Myc target genes such as CAD, LDHA, p53, and JTV1. Interestingly, there is a small group of c-Myc/Max target genes with expression that inversely correlates with c-myc mRNA levels (Class B). Among these genes is the gene encoding the cdk inhibitor p27, a known target of c-Myc repression. The identities of these genes can be found in Supporting Data Set 4 for Fig. 4, which is published as supporting information on the PNAS website.

### Program of Tumorigenesis in Cancer Cells.

This extensive role of c-Myc in gene regulation might underlie tumorigenesis in cells that overexpress the oncogenic protein. Perhaps in primary cells or normal tissues, where c-Myc levels are low, the transcription factor only acts in a sequence-specific manner on a small number of genes and influences a limited spectrum of cellular behaviors. In Burkitt’s lymphoma cells or other cancer cells expressing high levels of c-Myc, on the other hand, c-Myc may boost the global

### Discussion

A General Role for c-Myc in Global Transcriptional Regulation in Cancer Cells. Overexpression of the c-Myc gene has been associated with a large number of human malignancies (5). Although extensive research has been focused on the molecular function of this protein, it remains unknown how c-Myc promotes malignant transformation. In this study we attempted to identify the genomic binding sites for c-Myc and Max complexes in Burkitt’s lymphoma cells. The results strongly suggest that c-Myc plays a general role in global gene regulation in these cells. First, we showed that the c-Myc/Max complexes bind to a large number of gene promoters in Burkitt’s lymphoma cells, accounting for nearly 15% of genes tested. Second, DNA binding of c-Myc strongly correlates with transcription activities throughout the genome and resembles the genomic binding profile of pol II but not that of the sequence-specific transcription factor in the same cells. Third, c-Myc and Max extensively colocalize with the general transcription factor TFIIID. Finally, expression from most c-Myc/Max target promoters corresponds to c-myc transcript levels in a diverse panel of human tissues and cell lines. Based on these observations, we propose a model whereby c-Myc plays a general role in the regulation of global gene expression in Burkitt’s lymphoma cells. This model is consistent with the recent findings that the Drosophila dMyc/Max/Mnt network is involved in regulating ~15% of the genome (41).
gene expression and influence a wide spectrum of cellular pathways by acting on a large number of genes. Many of the genes we identified as c-Myc/Max targets play key roles in signaling, cell-cycle regulation, DNA replication, protein biosynthesis, and energy metabolism (Table 1), and their increased expression as a result from c-Myc overexpression would be expected to lead to dramatic changes in multiple processes such as an increase in the overall rate of cell-mass accumulation and shortened cell cycle. Interestingly, some c-Myc/Max target genes we identified have been implicated previously in apoptosis. Activation of these genes, including p53, Daxx, and Dap3, most likely will sensitize cells to apoptotic stimuli and necessitate the accumulation of mutations in these pathways during expansion of the lymphoma cell. It is possible also that an increased rate of metabolism in Daudi cells may result in higher levels of reactive oxygen species and DNA damage, which may further sensitize cells for apoptotic signals (42). Subsequent loss of function of p53 or BLC2 then could lead to unchecked cell growth and malignant transformation (1).

**Models for c-Myc DNA-Binding Specificity in Cancer Cells.** Our analysis of the sequences of the c-Myc-bound promoters in Burkitt’s lymphoma cells indicates that the c-Myc recognition motif is not present in many of these target promoters. For example, only ~26% of the promoters contain the CACGTG motif between 1,000 bp upstream of the transcription start site and the transcription start site itself, where it would be expected. Therefore, CACGTG motif-independent mechanisms may be used by c-Myc in its binding to gene promoters in Daudi cells. Two models could explain the widespread DNA binding of c-Myc in Burkitt’s lymphoma cells.

In the first model, a high level of c-Myc protein may allow it to bind to sequences that are different from the CACGTG motif and are not occupied by c-Myc in normal cells. These weak c-Myc binding sites could be variants of the CACGTG motif or completely different sequences and occur more frequently in the genome than the canonical c-Myc-binding sites. Although this model is simple and can explain the increased DNA binding of c-Myc in Daudi cells, it does not explain the mechanisms of specificity of c-Myc DNA binding or the strong correlation between c-Myc DNA binding and global transcription levels.

In the second model, c-Myc may bind to DNA through its association with other sequence-specific transcription factors or general transcription factors. Indeed, c-Myc has been shown previously to directly interact with TATA-box-binding protein (TBP) (37, 38), and we also demonstrate here that c-Myc DNA binding coincides with TFIIID occupancy on gene promoters. Taken together, these results suggest that c-Myc may be recruited to gene promoters as part of general transcription complexes, especially in the absence of the canonical c-Myc-binding sites. A similar mechanism has been implicated recently in c-Myc’s activation of pol III promoters (23). In this case, c-Myc binds to TFIIIB, a pol III-specific general transcription factor, and directly activates pol III transcription.

In conclusion, by analyzing the in vivo binding of c-Myc and Max to ~5,000 gene promoters, we observed that c-Myc is involved in the regulation of a surprisingly large number of gene promoters in Burkitt’s lymphoma cells. Such global transcriptional regulatory function by c-Myc may be a key mechanism that this oncogenic protein uses to promote tumorigenesis.

We thank Dr. Richard Kolodner for the generous contribution toward the making of hu6K arrays; Dr. William Sugden for the Daudi cell line; Drs. Tae Hoon Kim and Frank Furnari for helpful discussions; and Ms. Leah Ortiz-Luis Barrera and Mr. Joseph Hegler for the initial bioinformatics analysis. We also thank Drs. Robert N. Eisenman and Herman J. Bussemaker for making their manuscript available to us before publication. This research was supported by the Ludwig Institute for Cancer Research, a Sidney Kimmel Foundation for Cancer Research Scholar Award (to B.R.), and National Institutes of Health Grant R01-GM05151 (to M.Q.Z.).
Supporting Materials and Methods

Crosslinking of Cells, Fragmentation, and Immunoprecipitation of Chromatin

Formaldehyde Crosslinking. Cells (10^9) suspended in growth medium are transferred as 40-ml aliquots into 50-ml tubes and placed on ice for 10 min. Then 1/10 volume of crosslinking solution (11% formaldehyde/0.1M NaCl/1mM Na-EDTA/0.5 mM Na-EGTA/50mM Heps, pH 8.0) is added directly to each tube. After 10-min incubation on ice, 1/20 vol of a 2.5-M glycine solution is added to each tube to stop the crosslinking reaction. The cells are harvested by centrifugation at 2,000 \times g for 10 min at 4°C. The cell pellets are resuspended in cold PBS (137 mM NaCl/2.7 mM KCl/10 mM Na2HPO4/2 mM KH2PO4) and washed twice. The final cell pellet is snap frozen in liquid nitrogen.

Extraction of Chromatin. The frozen cell pellet is resuspended in 30 ml of lysis buffer 1 [0.05 M Hepes-KOH, pH 7.5/0.14 M NaCl/1 µM EDTA/10% glycerol/0.5% NP-40/0.25%, Triton X-100, with protease inhibitor cocktail from Roche Applied Science (catalogue no. 1836170)] with a pipette and mixed for 10 min at 4°C on a rocking platform. After centrifugation at 2,000 \times g for 10 min at 4°C, the cell pellet is resuspended in 24 ml of lysis buffer 2 (0.2 M NaCl/1 µM EDTA/0.5µM EGTA/10 µM Tris, pH 8/protease inhibitor cocktail) with a pipette and mixed gently at room temperature for 10 min on a rocking platform. After centrifugation again at 2,000 \times g for 10 min at 4°C, the pellet is resuspended in 10 ml of lysis buffer 3 (1 µM EDTA/0.5 µM EGTA/10 µM Tris-HCl, pH 8/protease inhibitor cocktail).

Fragmentation of Chromatin. The mixture is divided into 5-ml aliquots and placed into 15-ml tubes. These tubes are then placed into 50-ml tubes containing ice. An ultrasonic sonicator (Branson Sonifier 450, with power setting at 5) is used to break down the cell and nucleus membranes and fragment the chromatin. The sonicator probe tip is first immersed in the mixture followed by 25 s of continuous sonication. Subsequently, the tube is placed on ice for at least 1 min to avoid accumulation of heat. The sonication is repeated until the chromatin fragments are of desired length. The chromatin fragment size can be examined by agarose gel electrophoresis analysis of 10 µl of cell extracts, digested with protease K for 1 h. The number of sonication cycles varies with different cell types and crosslinking conditions, and pilot tests are recommended. Usually, 10 cycles of sonication are necessary to achieve the desired fragmentation size. Finally, the chromatin solution is adjusted to 0.5% Sarkosyl (sodium lauryl sarcosine) and gently mixed for 10 min at room temperature on a rocking platform. The chromatin solution is then transferred to a centrifuge tube and spun for 10 min at 10,000 \times g to remove cell debris. The supernatant is collected for chromatin immunoprecipitation. At this step, the DNA concentration of the solution should be around 1-2 mg/ml. The solution can be stored at –80°C as 1-ml aliquots.

Immunoprecipitation of Chromatin. Magnetic beads (Dynal, Oslo) precoupled to the primary antibodies are used to immunoprecipitate the DNA associated with the protein of interest. To prepare the magnetic beads, 100 µl of sheep anti-rabbit IgG-conjugated
dynabeads (Dynal; catalogue no. 112.04) or goat anti-mouse IgG-conjugated beads (Dynal; catalogue no. 110.05) are first washed three times with cold PBS containing 5 mg/ml BSA (Sigma; catalogue no. A-7906) and then resuspended in 5 ml of cold PBS. Rabbit polyclonal (or mouse monoclonal) antibody (10 µg) is added to the mixture and incubated overnight on a rotating platform at 4°C. After collecting the magnetic beads by centrifugation and washing three times with cold PBS containing 5 mg/ml BSA, the beads are resuspended in 100 µl of cold PBS with 5 mg/ml BSA and are ready for immunoprecipitation.

In an Eppendorf tube, 2 mg of soluble chromatin solution is first adjusted to 0.1% Triton X-100, 0.1% sodium deoxycholate, and 1 mM PMSF, then mixed with 100 µl of magnetic beads precoupled with the antibody. The mixture is incubated at 4°C overnight in a rotating platform. The magnetic beads are then collected using a magnet MPC-E (Dynal), and the supernatant is removed by aspiration. To remove materials nonspecifically bound to the beads, 1 ml of RIPA buffer (50 mM Hepes, pH 7.6/1 mM EDTA/0.7% DOC/1% NP-40/0.5 M LiCl/protease inhibitor cocktail) is added to the tube, and the beads are gently resuspended on a rotating platform in a cold room. The magnetic beads are again collected with MPC-E. After washing with RIPA buffer for a total of five times, and once with 1 ml of TE buffer, the beads are collected by centrifugation at 2,000 \( \times g \) for 3 min and resuspended in 50 µl of elution buffer (50 mM Tris, pH 8/10 mM EDTA/1% SDS). To elute precipitated chromatin from the beads, the tubes are incubated at 65°C for 10 min with constant agitation, then centrifuged for 30 s at 2000 \( \times g \). Supernatant (40 µl) is taken out and mixed with 120 µl of TE buffer with 1% SDS. This solution is incubated at 65°C overnight to reverse the crosslinks. As a control, 100 µg of chromatin is mixed with 120 µl of TE with 1% SDS in a separate tube and is incubated at 65°C overnight.

**Purification of DNA.** After reversal of crosslinks, proteins in the DNA sample are removed by incubation with 120 µl of proteinase K solution (2% glycogen/5% 20 mg/ml proteinase K stock solution/TE buffer) for 2 h at 37°C. The sample is then extracted twice with phenol (Sigma; catalogue no. P-4557), once with 24:1 chloroform/isoamyl alcohol (Sigma; catalogue no. C-0549). The sample is adjusted to 200 mM NaCl. After ethanol precipitation, the DNA is dissolved in 30 µl of TE buffer containing 10 µg of DNase-free RNase A (Sigma; catalogue no. 6513) and incubated for 2 h at 37°C. The DNA at this step can be further purified with Qiagen PCR kit (Qiagen, Valencia, CA; catalogue no. 28106).

**Blunting, Ligation of Linker to DNA, and Amplification by PCR**

The immunoprecipitated DNA is usually in so small a quantity (1-10 ng of total) that it is necessary to amplify them significantly for the subsequent labeling step and DNA microarray analysis. To achieve this, a ligation-mediated PCR (LM-PCR) procedure is used. Because the DNA at this step contains uneven ends as a result of the physical shearing process, it is first treated with T4 DNA polymerase to form blunt ends. Then a linker is ligated to the DNA fragments. The added linker allows the DNA to be amplified by PCR using a universal oligonucleotide primer.
**Blunting.** In an Eppendorf tube, the 40 µl of immunoprecipitated DNA (or 20 ng of control input DNA) is mixed with 11 µl of (10×) T4 DNA polymerase buffer (New England Biolabs; catalogue no. 007-203), 0.5 µl of BSA (10 mg/ml) (New England Biolabs; catalogue no. 007-BSA), 0.5 µl of dNTP mix (20 mM each), 0.2 µl of T4 DNA pol (3 units/µl) (New England Biolabs; catalogue no. 203L), and dH2O to a total volume of 112 µl. After 20 min incubation at 12°C, 1/10 vol of 3 M sodium acetate (pH 5.2) and 1 µg of glycogen (Roche Applied Sciences; catalogue no. 0901393) is added to the tube, and the DNA sample is extracted with phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma; catalogue no. P-3803) once. After ethanol precipitation, the final DNA is dissolved in 25 µl of dH2O.

**Ligation.** The blunt-ended DNA is mixed with 8 µl of dH2O, 10 µl of 5× ligase buffer (Invitrogen; catalogue no. 46300-018), 6.7 µl of annealed linkers (oligo-1: GCGGTGACCCGGGAGATCTGAATTC, oligo-2: GAATTCAGATC, annealed to make a 15-µM solution), 0.5 µl of T4 DNA ligase (New England Biolabs; catalogue no. 202L), and distilled water in a 50.2-µl total volume. The ligation reaction is allowed to continue overnight at 16°C. After the reaction, the DNA is purified by ethanol precipitation and dissolved in 25 µl of dH2O.

**PCR.** The DNA that has been ligated to the linker oligo is mixed with 4 µl of 10× ThermoPol reaction buffer (New England Biolabs; catalogue no. B9004S), 4.75 µl of ddH2O, 5 µl of 10× dNTP mix (2.5mM each dATP, dTTP, dGTP, and dCTP), 1.25 µl of oligo-1 (40 µM stock) in a final volume of 40 µl in a 500-µl-thin-wall Eppendorf PCR tube. The tube is first incubated at 55°C for 2 min on a thermal cycler; then 10 µl of an enzyme mix [8 µl of dH2O, 1 µl of Taq DNA polymerase (5 units/µl), 1 µl of ThermalPol reaction buffer, and 0.025 unit of Pfu polymerase (Stratagene; catalogue no. 600250-51)] is added. Subsequently, the following PCR cycle is performed:

- step 1: 72°C for 5 min;
- step 2: 95°C for 2 min;
- step 3: 95°C for 1 min;
- step 4: 60°C for 1 min;
- step 5: 72°C for 1 min;
- step 6: go to step 3 for 22 times;
- step 7: 72°C for 5 min;
- step 8: 4°C indefinitely.

After the PCR, the DNA is purified using the Qiaquick PCR purification kit (Qiagen; catalogue no. 28106) and eluted in 60 µl of elution buffer provided with the kit.

**Labeling of DNA and Microarray Hybridization**

**Labeling of Amplified DNA.** In a colored Eppendorf tube, 200 ng of DNA from the previous step is mixed with 20 µl of 2.5× random primer solution (BioPrime kit,
Invitrogen; catalogue no. 18094-011) and dH2O in a final volume of 42.5 µl. The mixture is boiled for 5 min and then cooled on ice for 5 min. Subsequently, 5 µl of 10× low dCTP mixture (2.5 mM each for dATP, dTTP, and dGTP, and 0.6 mM for dCTP), 1.5 µl of Cy5-dCTP (Amersham; catalogue no. PA55021) or Cy3-dCTP (Amersham; catalogue no. PA53021), and 40 units of Klenow DNA polymerase are added to tube. The tube is incubated at 37°C for 2 h. After the reaction, the labeled DNA is purified using the Qiagen PCR kit (Qiagen; catalogue no. 28106).

**DNA Microarray Hybridization.** In a new Eppendorf tube, 2.5 µg of Cy5-labeled chromatin immunoprecipitation (ChIP) DNA is mixed with 2.5 µg of Cy3-labeled genomic DNA and 36 µg of human Cot-1 DNA (Invitrogen; catalogue no. 15279-011). Sodium acetate (1/10 vol of 3 M) is added to the tube along with 2 vol of ethanol. After ethanol precipitation, the DNA is dissolved in 22.4 µl of hybridization buffer 1 (2.2× SSC/0.22% SDS). The mixture can be heated for 10 min at 37°C to facilitate the DNA to dissolve. Then 20 µl of hybridization buffer 2 (70% formamide/3× SSC/14.3% dextran sulfate) is added to the mixture, and the tube is first heated at 95°C for 5 min to denature DNA, then incubated at 42°C for 2 min. At this step, 4 µl of yeast tRNA (Sigma; catalogue no. R9001 at 10 µg/µl) and 3 µl of 2% BSA are added to the mixture, which is then spotted to a DNA microarray slide that has been incubated with the prehybridization solution for 40 min at 42°C. A 25-mm × 60-mm cover slip is then gently placed on top of the sample, and the hybridization is carried on in a hybridization chamber (Corning; catalogue no. 07-200-271) at 60°C overnight in a water bath.

**Washing Microarrays.** After the hybridization, the microarray slide is washed once with washing buffer 1 (2× SSC/0.1% SDS) at 60°C for 5 min in a glass slide staining dish. This is followed by a wash with buffer 2 (0.2× SSC/0.1% SDS) for 10 min at room temperature and three times with buffer 3 (0.2× SSC) at room temperature. The slide is then dried by a brief spinning at 1,000 × g in a table-top centrifuge.

**Microarray Analysis and Identification of In Vivo DNA-Binding Sites**

**Microarray Scanning and Initial Analysis.** The microarray slides are scanned using GenePix 4000B scanner from Axon Instruments (Foster City, CA), and each microarray image is first analyzed with the image analysis software GENEPIX PRO 3.0 to derive the Cy3 and Cy5 fluorescent intensity and background noise for all the spots on the array. The intensities for both Cy3 and Cy5 channels are first adjusted by subtracting the background intensity of each spot by using the formula \( I_{\text{channel}} - B_{\text{channel}} \), where channel represents either Cy5 or Cy3. The intensities are further adjusted by subtracting the median intensity of all the blank spots on the array. If any of these values was lower than 10, it was then raised to 10.

**Normalization.** The normalization factor is calculated based on the intensities of spots that are considered good (more than 65% of pixels have intensities higher than the background intensity plus 1 SD). The median of the intensity ratios, \( I_{635}/I_{532} \) is then used to adjust the Cy3 channel intensity to the same level as that of the Cy5 channel.
**Error Model.** The quantitative amplification of small amounts of DNA generates some uncertainty in values for the low intensity spots. In order to track that uncertainty and average repeated experiments with appropriate related weights, a single array error model (1) is used. The significance of a measured ratio at a spot is defined by a statistic \( X \), which is formulated as:

\[
X = \frac{a_2-a_1}{\sqrt{a_1^2 + a_2^2 + f^2(a_1^2 + a_2^2)}}
\]  

where \( a_{1,2} \) are the intensities measured in the two channels for each spot, \( \sigma_{1,2} \) are the uncertainties due to background subtraction, and \( f \) is a fractional multiplicative error such as would come from hybridization nonuniformities, fluctuations in the dye incorporation efficiency, scanner gain fluctuations, etc. The distribution of \( X \) can be found to be close to Gaussian distribution in experiments where Cy3 and Cy5 samples are identical. The significance of a change of magnitude \( X \) is then calculated using a one-sided probability model as follows:

\[
P = 1 - \text{ErfP}(\frac{X - \mu}{\sigma})
\]  

where \( \mu \) is the average of \( X \), and \( \sigma \) is the standard deviation of \( X \). Because the intensities are normalized, \( \mu \) should be near 0. The \( \text{Erf}(x) \) function is the standard normal accumulative distribution function corresponding to standard normal curve areas.

If the Cy3 and Cy5 samples are not identical, the Gaussian distribution can be skewed, because the ChIP can result in many DNA spots with significantly higher intensities in one channel than in the other. Since the input DNA is always present, the intensity distribution of those nonenriched spots can be used to obtain the parameters of \( X \) distribution. First, the spots whose \( X \) is a value less than 0 are identified. These spots should be on the left half of the Gaussian distribution. Their mirror spots are then generated by using \( X^+ = -X^- \). The mean and SD of this new \( X \) can be calculated.

**Combining Replicate Datasets.** When the genomic binding sites of a protein are investigated using the above method, it is routine to perform several independent experiments so that the results are more reliable. To combine replicate data sets, each sample is first analyzed individually using the above single array error model. The average binding ratio and associated \( P \) value from these multiple experiments are then calculated using a weighted averaging analysis method (1). For each spot, the uncertainty in the \( \log(\text{Ratio}) \) is defined as

\[
\sigma = \log(a_2 / a_1) / X_{\text{norm}}
\]  

Where \( a_{1,2} \) are the intensities measured in the two channels for each spot, and \( X_{\text{norm}} \) is the normalized \( X \) for each spot. The weights of each spot are then defined as
\[ w_i = 1/\sigma_i^2 \]  

[4]

The averaged log(Ratio) is then calculated using the following formula

\[ x = \frac{\sum_{i=1,n} w_i x_i}{\sum_{i=1,n} w_i} \]  

[5]

Where \( n \) is the total number of experiments, \( x_i \) and \( w_i \) are the log(Ratio) and weight for each experiment of a particular spot. \( x \) is normally distributed, and the averaged \( P \) value is then calculated using Eq. 2, where the variable for Erf(x) function is \( \bar{x} \) instead. Target genes are selected based on a significant \( P \) value (e.g., <0.001).