



Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation

Amy S. Weinmann and Peggy J. Farnham*

McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, 1400 University Avenue, Madison, WI 53706, USA

Accepted 15 January 2002

Abstract

The standard chromatin immunoprecipitation (ChIP) assay is used to examine the specific association of transcription factors with DNA in the context of living cells. Here we review two modifications to this protocol which are designed to identify novel target genes of transcription factors in mammalian cells. The main advantage to both of these approaches is that only DNA sequences directly bound by a factor within the context of a living cell will be identified. Therefore, artifacts associated with overexpression and/or alterations in signaling pathways are avoided. The first modification we describe, a ChIP cloning strategy, can be used to isolate any genomic fragment specifically associated with a particular factor. It requires no special equipment or reagents other than a high-affinity antibody to be used for immunoprecipitation of the factor of interest. However, it is most useful for the isolation of a small number of genomic targets. In contrast, the second modification, which combines ChIP with specialized CpG microarrays, is ideal for a more global analysis of target genes. Advantages, common problems, and detailed protocols for these two ChIP techniques are discussed. © 2002 Elsevier Science (USA). All rights reserved.

Keywords: Chromatin immunoprecipitation; Transcription factors; Target genes; CpG microarrays; Shotgun cloning; E2F

1. Introduction

As the sequencing of the human genome nears completion, the challenge that faces the scientific community is to decipher the underlying meaning behind these precisely ordered nucleotides. One portion of this sequence will provide the genetic information to create the large number of proteins required for maintaining the critical functions for diverse cell types. In addition, a significant fraction of the genome will provide the information required to direct the precise timing and pattern of expression for these proteins. The precise regulation of gene expression deserves great attention because the inappropriate expression of a single gene product can result in dramatic consequences which can include uncontrolled cellular proliferation leading to cancer.

New strategies focused on understanding gene regulation are being developed to exploit the vast amount of

information now available from the human genome sequencing efforts. Computer-assisted genome inspection strategies are being used to predict genes and regulatory regions. Although these approaches can provide valuable information, it is important to remember that primary DNA sequence provides only a small fraction of the information actually contained within the nuclear environment. Recent advances have highlighted the critical role that the chromatin environment plays in regulating gene expression. The modification of histone proteins can provide information to target transcription factors to specific regions of DNA [1]. For example, the specific acetylation or methylation of the core histone proteins can influence transcription factor binding. It is widely believed that hyperacetylated regions of the genome are more accessible to protein binding than hypoacetylated regions. Therefore, the same primary DNA sequence can be recognized and bound by proteins in one case (i.e., hyperacetylated nucleosomes) whereas in the opposing case (i.e., hypoacetylated nucleosomes) it is unavailable for protein recognition. This simple example makes it clear that to understand the regulatory regions contained within the human

* Corresponding author. Fax: (608) 262-2824.

E-mail address: farnham@oncology.wisc.edu (P.J. Farnham).

genome, strategies to examine transcription factor–DNA interactions within the context of living cells will provide the most accurate information possible.

It has been predicted that at least 2000 transcriptional activators are encoded by the human genome [2]. To make use of this information, it is now important to determine the sets of genes regulated by each of these factors. A common approach used to identify the target genes that are regulated by an individual factor is to couple the overexpression or underexpression of that factor to cDNA microarray analysis [3]. Although this approach allows for the isolation of a large set of potential target genes, the data need to be interpreted with caution for several reasons. First, the genes identified may not be direct target genes of the overexpressed factor, but instead may be isolated as the result of indirect regulation due to overall alterations of gene expression patterns. In addition, it is unclear that the genes regulated by a factor at levels vastly greater than normal, biologically relevant concentrations are in fact true target genes. Therefore, our studies have focused on the development of new approaches designed to examine the direct targets of a site-specific transcription factor in the context of physiologically relevant conditions.

In this article, we describe the development of two methods designed to identify direct target genes of mammalian transcription factors. Each method is based on the chromatin immunoprecipitation (ChIP) procedure, which allows for an examination of protein–DNA interactions in the context of living cells. Briefly, in the standard ChIP procedure, cells are treated with formaldehyde to crosslink proteins that are in close association with DNA, and as the procedure proceeds, specific protein–DNA complexes are isolated by immunoprecipitation. Following reversal of the crosslinks and purification of the DNA specifically associated with the protein of interest, specific DNA sequences can be examined by PCR with gene-specific primers. Therefore, when using the standard ChIP procedure, one must first suspect that a promoter might be bound by the transcription factor of interest to be able to design primers to a specific DNA sequence. This approach is of great use when confirming that a protein is bound to a gene previously characterized by other means, such as a mutational analysis of a promoter. However, this standard ChIP protocol cannot be used to identify unknown target promoters associated with a given factor.

To modify the ChIP procedure for the isolation of novel target sites, a method needed to be developed to examine DNA sequences specifically precipitated with an antibody to a desired protein with no prior knowledge of its target genes. For this means, we have developed two separate procedures: the first allows for the isolation of individual target genes [4] and the second provides a more global approach [5]. Both strategies have been designed to identify target genes that are

directly bound by the factor of interest in the context of the natural cellular environment.

2. Description of methods

2.1. ChIP cloning

Our first modification to the chromatin immunoprecipitation procedure was designed to clone individual promoter or enhancer fragments bound by a human transcription factor (Fig. 1). Although gene-specific primers are commonly used to analyze the precipitated chromatin, the precipitated samples contain a large subset of the genomic fragments bound by a given factor. Therefore, we reasoned that the preparation of a plasmid library containing the precipitated fragments would allow for the identification of novel binding sites. However, several changes to the standard chromatin immunoprecipitation protocol were required when moving from a primer-specific analysis to a shotgun cloning strategy.

The first problem that needed to be solved was eliminating as much of the nonspecific DNA in the immunoprecipitation reaction as possible. In the standard ChIP procedure highly abundant repeat regions of the genome are precipitated nonspecifically, as illustrated by their presence in immunoprecipitation reactions that do not contain an antibody. This is generally not a problem because gene-specific primers are used to amplify by polymerase chain reaction the desired target, and repetitive DNA elements will have no effect on this analysis. However, in the cloning procedure, nonspecific DNA will have an equal chance of being cloned. Therefore, we first performed two sequential immunoprecipitations with aliquots of the same transcription factor-specific antibody in an attempt to decrease the amount of nonspecific DNA. It is important to confirm that the second immunoprecipitation was successful before proceeding with the cloning procedure. We have found that some antibodies are unable to efficiently recognize protein complexes following elution and redilution. It is possible that some proteins do not renature appropriately for antibody recognition. Therefore, this step must be closely monitored prior to attempting the cloning portion of the procedure.

The amount of DNA following the second immunoprecipitation step is very small. Some chromatin immunoprecipitation cloning protocols modified for use in the yeast system have used PCR amplification steps to increase the amount of DNA available for cloning [6]. We have chosen not to perform a PCR amplification step to avoid cloning only easily amplified sequences. In general, it is very difficult to amplify sequences with a high GC content and a significant percentage of mammalian promoter regions are GC-rich [7]. Therefore, a

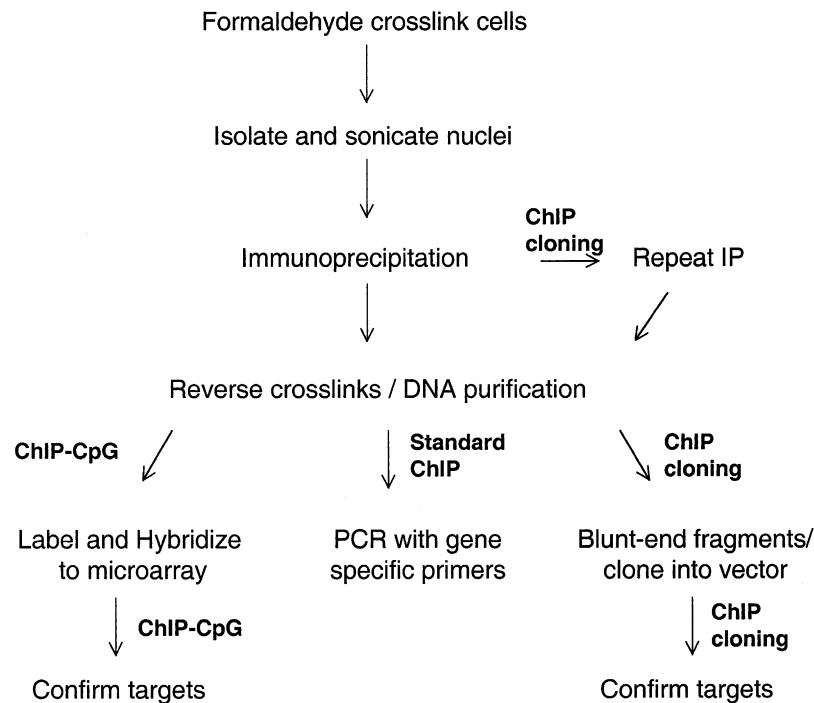


Fig. 1. Chromatin immunoprecipitation cloning schematic. Flowchart outlining the uses for the chromatin immunoprecipitation assay including the standard ChIP protocol and the two modifications designed to identify novel target genes.

PCR amplification step may preferentially amplify AT-rich, nonpromoter sequences, ultimately creating a false abundance of these sequences in the cloning pool. To avoid this potential bias, we feel it best not to include a PCR amplification step. Therefore, to compensate for the low yield after the second immunoprecipitation step, several identical, parallel immunoprecipitation reactions are performed and then pooled following the DNA purification steps.

Our next alteration to the ChIP procedure was to modify the immunoprecipitated DNA fragments to become competent for cloning. During the sonication step, the DNA is sheared which creates random overhangs at the 5' and 3' ends. These overhangs need to be modified for efficient cloning. One potential method is to digest the DNA with a restriction enzyme. However, restriction enzyme digestion will exclude from cloning any DNA fragment that does not contain two sites for that enzyme that flank the binding site of interest. To avoid this problem, we used T4 DNA polymerase to create blunt-ended DNA fragments that could then be cloned into a blunted vector for further characterization.

Another consideration in adapting the standard ChIP protocol for cloning target genes is the size of the chromatin. In a standard ChIP experiment, it is desirable to shear the chromatin to a relatively small size to ensure monitoring of only the binding sites located in close proximity to the gene-specific primers. In our initial ChIP cloning experiments, we found that the majority of DNA fragments cloned were very small in size (200–300 bp).

When these fragments were sequenced, the vast majority corresponded to AT-rich repetitive sequences. The validity of these sequences was tested in standard ChIP experiments and they were shown to be nonspecifically precipitated (as determined by having a high signal in the no antibody reaction (e.g., see Fig. 2A; nonspecific). Therefore, it appeared that the small fragments were most likely due to nonspecific precipitation of highly repetitive elements. Although a large fraction of the nonspecific DNA was removed in the second immunoprecipitation step, it is very difficult to completely eliminate it. In an attempt to more efficiently distinguish specific from nonspecific clones, we prepared larger chromatin (1–2 kb) and then analyzed only cloned DNA fragments of at least 500 bp. Using this criterion as an initial screen in an E2F ChIP cloning experiment, 11 of 14 genomic fragments were confirmed to be specifically bound by E2F family members in living cells [4]. Thus, size selection appears to aid in the screening procedure.

It is extremely important to validate each clone obtained in the ChIP cloning method. The first step to validate the clones is to perform independent, standard ChIP experiments using clone-specific primers. This analysis will eliminate false positives which can be isolated by random chance or due to nonspecific precipitation (Fig. 2A). As noted above, it is impossible to completely eliminate all repetitive elements and some false positives will inevitably be due to the isolation of these elements. Despite these shortcomings, the addition of the second immunoprecipitation step and the selec-

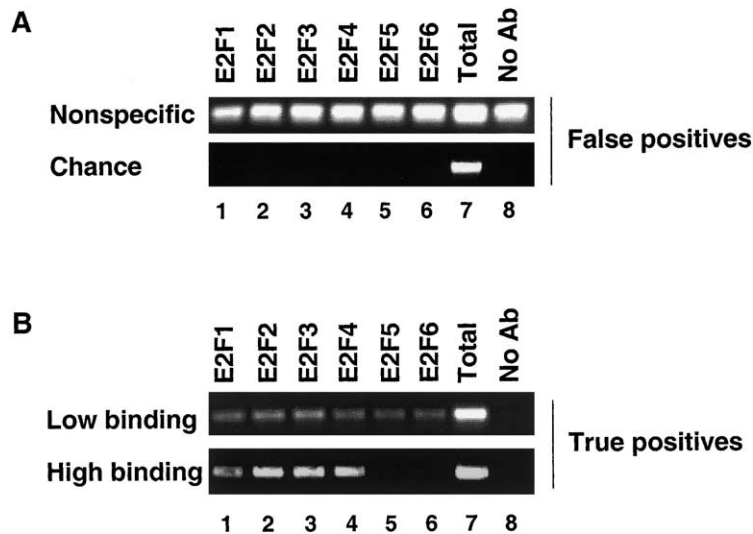


Fig. 2. Examples of false and true positives obtained using the ChIP cloning technique. The gel images are reproduced from [4] and represent the four categories of potential clones that can be encountered in the ChIP cloning procedure. Clone-specific primers were made to potential E2F clones isolated using the ChIP cloning procedure. All potential clones were size selected as described in the text. Shown in (A) and (B) are the standard ChIP assays used to test these clones. Clones that represent false positives are shown in (A) and clones that represent true positives are shown in (B). Consistent with previous findings, very little E2F family member preference is noted for E2F binding to targets in the context of living cells. (A) Nonspecific refers to DNA fragments that are highly abundant and found in the no antibody control. Chance refers to a single copy gene that is isolated due to random chance in the procedure but does not confirm in subsequent experiments. (B) Although binding affinity cannot be conclusively determined using this assay, for simplicity low binding refers to a PCR signal obtained in the specific antibody IP reaction which is higher than the no antibody control signal but less than a standardized aliquot (0.2%) of the total input chromatin. We define high binding as a PCR signal obtained in the specific antibody IP that is greater than or equal to the standardized aliquot of the total.

tion of large cloned inserts did allow for the successful isolation of genomic fragments specifically bound by members of the E2F transcription factor family in living cells [4]. The confirmed positives in these experiments were of two types: low- and high-level binding (see Fig. 2B). We define high-level binding as those clones that have a specific IP signal intensity greater than or equal to that of a standardized aliquot (0.2%) of the total. Low-level binding is defined as a clone that has a reproducible signal in the specific IP sample greater than that of the no antibody control but less than that of the standardized aliquot of the total. In addition, we found that 3 of the 11 confirmed genomic fragments mapped to promoter regions [4]. Thus, we were able to successfully isolate promoters specifically bound by E2F family members using the ChIP cloning method. Although the isolation of novel transcription factor targets was accomplished, it became clear that the strategy used was limited to the identification of only a small number of target genes because of the difficulty encountered in the screening procedure. If a more global analysis of transcription factor target genes is the goal of an analysis, a new method for screening the isolated targets needed to be developed.

2.2. ChIP-CpG microarray

The most promising high-throughput screening method is a microarray-based approach. Recently, the

coupling of chromatin immunoprecipitation with genomic microarray analysis has been successfully performed in the yeast system to uncover a significant number of genomic sites directly bound by specific transcription factors in the context of living cells [6,8,9]. However, a similar analysis in mammalian cells is more difficult due to the lack of a comparable genomic microarray because of the vastly greater size associated with mammalian genomes. Therefore, to perform a similar analysis in mammalian cells, a suitable genomic microarray first needed to be determined.

Most commercially available microarrays contain cDNA sequences. Therefore, if the goal is to isolate regulatory regions (i.e., promoters or enhancers) bound by a specific factor, a cDNA microarray will not be useful. To circumvent this problem, we have used a CpG island microarray as the source for a promoter-enriched microarray. The rationale in using this approach is that CpG islands are often associated with the promoter and first exon of a gene [7,10,11]. Therefore, it is likely that a significant portion of clones contained on a CpG island microarray will be in close proximity to gene regulatory regions. Thus, by probing a CpG island microarray with chromatin immunoprecipitated with an antibody to a human transcription factor, it should be possible to isolate a large, although not completely encompassing, subset of promoters bound by that factor *in vivo* [5].

In addition to the CpG microarray, it will also be possible to use other genomic microarrays for these

types of studies. For instance, the Dynlacht group has created a microarray by spotting it with PCR fragments derived from a percentage of known promoter regions [12]. This type of microarray approach will also allow for the examination of a large set of mammalian promoters. In addition, it is also theoretically possible to create an array of PCR fragments representing 1 kb of DNA upstream of all known and predicted genes. If the goal is to examine regulatory sites within this immediate promoter region, such an array would provide an ideal setting. However, if the factor of interest does not bind within the proximal promoter region, another strategy would need to be employed. One potential strategy to examine distant regulatory regions is the use of an array containing bacterial artificial chromosome (BAC) sequences. The main disadvantage associated with this approach at present is that because of the large size of BAC clones, one will not be able to localize binding resolution to finer detail than the size of the spotted BAC clone. Instead, intensive follow-up studies would be required to further localize binding within that large genomic region.

As with the ChIP cloning strategy, it is important to confirm factor binding with independent experiments. It is worth noting that repetitive elements are not a significant problem in this system because the no antibody control will identify these sequences (i.e., red and green hybridization signals will be equal). Also, by performing two hybridizations and analyzing only the spots that are positive in both experiments, many false positives will be eliminated. However, it is still imperative to confirm the positives. First, an independent ChIP analysis should be used to demonstrate that indeed the targets are reproducibly bound by the factor of interest in living cells. When designing gene-specific primers

based on the positive CpG clones, it is important to realize that the binding site identified may not reside directly in the CpG clone spotted on the microarray (Fig. 3). This possibility is due to the size of the chromatin generated in the ChIP procedure (which will be used to hybridize to the microarray). For example, if the average chromatin size is 1 kb, the factor binding site responsible for recruiting the transcription factor to that DNA fragment could be located anywhere within that 1 kb. If the CpG island represents one end of the fragment and the factor binding site is found at the opposite end, it is possible that the factor binding site could be located up to 1 kb in either direction of the CpG island, leading to a 2-kb area that must be analyzed. It may at first seem advantageous to decrease the size of the chromatin used in the immunoprecipitations to limit the size of the region that must be analyzed. However, it must be remembered that the entire human genome is not spotted on the microarray. Rather, only 200 bp to 1 kb CpG-rich fragments are spotted. The advantage of hybridizing the array with larger chromatin is that a factor binding site does not necessarily need to be found within the CpG island itself. Rather, a factor binding site within relatively close proximity to the CpG island can be identified using this approach. It is fairly simple to monitor binding to a 2-kb fragment in the initial confirmation experiments. Essentially, four sets of primer pairs will allow for the rough localization of the site if the chromatin is approximately 500 bp in the confirmation experiments (Fig. 3).

It is also important to remember for both ChIP cloning techniques that when true positives are identified and confirmed in independent ChIP assays, further studies are required to determine the functional consequence of this interaction. For instance, promoter–

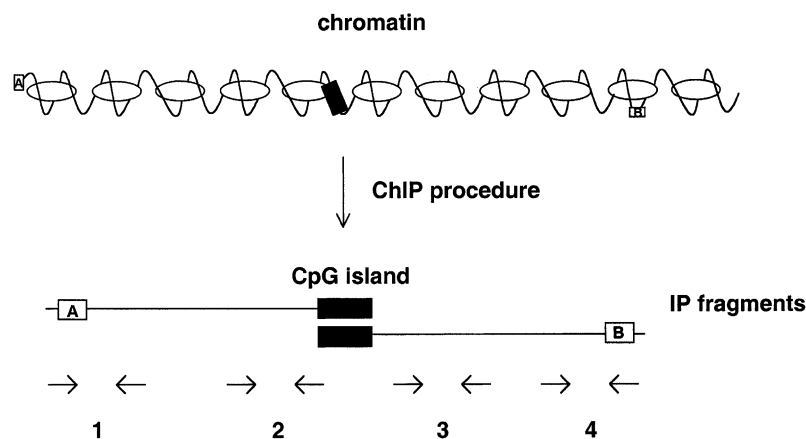


Fig. 3. Schematic representation of the monitoring of factor binding sites in proximity to a CpG island. A view of the chromatin in the nucleus is shown at the top of the schematic with nucleosomes represented by ovals. A CpG island is represented in the figure by a black box and the farthest possible upstream (A) and downstream (B) locations of the transcription factor binding site relative to the CpG island are represented by white boxes. The chromatin obtained in the ChIP assay that will be hybridized to the CpG island microarray is shown as a thin line. As noted in the text, the use of chromatin sheared to an average length of 1 kb can result in the need to monitor 2 kb of the genome for binding. However, a set of four primer pairs can cover the region if chromatin is sheared to an average of 500 bp in the confirmation ChIP experiments.

reporter constructs can be generated to examine the importance of various binding elements in an individual promoter. Deletion and substitution mutations can be used to localize the DNA elements that are important for promoter activity. This type of analysis will help uncover a subset of factor binding sites responsible for promoter regulation. However, some important regulatory elements may be missed in standard promoter analyses due to the lack of typical chromatin structure on transiently transfected constructs. Therefore, more sophisticated approaches for examining the influence an individual factor has on promoter activity may be needed. Detailed protocols and examples of appropriate studies for determining the role an individual factor plays in the regulation of transcription can be found in a comprehensive book written by Carey and Smale [13]. Other articles in this issue [15,16], also provide appropriate strategies and protocols.

2.3. Methodology: ChIP cloning protocols

2.3.1. Day 1

1. Add formaldehyde directly to tissue culture medium (in fume hood) to a final concentration of 1%. We generally use 1×10^7 cells per antibody per time point. For cloning, multiple, identical IPs are performed and pooled at later steps. We generally start with 1×10^8 cells (to be used for 10 IP reactions). Incubate adherent cells on a shaking platform and suspension cells on a stir plate for 10 min at room temperature. (*Note:* Formaldehyde is very toxic and steps using formaldehyde should be performed in a fume hood according to safety guidelines. Dispose of formaldehyde waste according to safety guidelines.)

2. Stop the crosslinking reaction by adding glycine to a final concentration of 0.125 M. Continue to rock or spin at room temperature for 5 min.

3. For adherent cells, pour off medium and rinse plates twice with cold $1 \times$ phosphate-buffered saline (PBS). For suspension cells, centrifuge at 1500 rpm and wash cell pellet twice with cold $1 \times$ PBS (for suspension cells then proceed to step 6).

4. For adherent cells, add an appropriate volume (try 5 ml per 500-cm² dish) of $1 \times$ PBS or a 0.05% trypsin solution (a 1:5 dilution of a $1 \times$ trypsin–EDTA tissue culture grade solution diluted in $1 \times$ PBS). Incubate at 37 °C for 10 min if using trypsin. The trypsin step is useful for cells that are difficult to swell. Thus, for cell types that are easily swelled, this step may not be necessary.

5. Following addition of trypsin or $1 \times$ PBS, scrape adherent cells from dishes. If using trypsin, inactivate the trypsin by adding a small amount of serum. Centrifuge scraped adherent cells at 1500 rpm and wash pellet once with $1 \times$ PBS plus phenylmethylsulfonyl fluoride (PMSF, 10 µl/ml).

6. Resuspend cell pellet in cell lysis buffer plus the protease inhibitors PMSF (10 µl/ml), aprotinin (1 µl/ml), and leupeptin (1 µl/ml). The final volume of cell lysis buffer should be sufficient so that there are no clumps of cells. Incubate on ice for 10 min. Cells can also be homogenized on ice with a Dounce homogenizer (B pestle) several times to aid in release of nuclei. The cell lysis conditions, buffers, and homogenizing may need to be optimized for isolation of nuclei from individual cell types. Cell lysis can be monitored by examining the cells under a microscope.

7. Centrifuge at 5000 rpm for 5 min at 4 °C to pellet the nuclei.

8. Discard supernatant and resuspend nuclear pellet in 1 ml nuclei lysis buffer plus protease inhibitors. Incubate on ice for 10 min. If the nuclei are too dense, resuspend nuclei in a larger volume of nuclei lysis buffer.

9. Sonicate chromatin to an average length of approximately 1–2 kb. Keep samples cold during the sonication procedure. The sonication time and number of pulses will vary depending on sonicator, cell type, and extent of crosslinking. As a starting point, we have performed the sonication step using four pulses of 15 s each at setting 7 on a Fisher Model 60 sonic dismembrator. Between pulses, samples should be allowed to cool on ice. To aid in sonication, it may also be helpful to add 0.1 g of glass beads (212–300 micron; Sigma, Catalog No. G-1277) prior to sonication. After sonication, microcentrifuge samples at 14,000 rpm for 10 min at 4 °C. At this point, chromatin can be snap-frozen in liquid nitrogen and stored at –70 °C.

10. Carefully remove the supernatant and transfer to a new tube. Preclear chromatin by adding 50 µl of pre-blocked Staph A cells (for 10^8 crosslinked cells). The preparation and blocking of Staph A cells are described in the appendix.

11. Incubate on a rotating platform at 4 °C for 15 min. Microcentrifuge at 14,000 rpm for 4 min.

12. Transfer supernatant to a new tube and divide equally among IP samples. If starting with 1×10^8 cells, then divide into 10 samples: 9 to be precipitated with the antibody used for cloning and 1 for a no antibody or preimmune serum control. However, if performing a CpG microarray analysis (see below), divide the samples equally between antibody and no antibody samples. Also include a “mock” sample which contains IP dilution buffer without chromatin (no antibody and mock samples are critical to monitor for nonspecific interactions and DNA contamination of IP and wash solutions). Adjust the final volume of each sample with two times the chromatin volume of IP dilution buffer (plus protease inhibitors). The volume of each sample should be between 200 and 600 µl. Add 1 µg of the specific antibody to the appropriate samples. However, the optimal antibody concentration may vary for individual

antibodies; initial tests can be performed with 0.5, 1, 2, and 5 μg of antibody per sample.

13. Incubate on a rotating platform at 4 °C overnight (or for at least 3 h).

2.3.2. Day 2

14. If you are not using rabbit polyclonal antibodies (i.e., you are using a monoclonal antibody, or a goat polyclonal antibody), add 1 μg of an appropriate secondary antibody and incubate for an additional 1 h. We have found that rabbit polyclonal antibodies appear to associate well with Staph A cells; therefore secondary antibodies should be from a rabbit.

15. Add 10 μl of blocked Staph A cells to each sample. Incubate on a rotating platform at 4 °C for 15 min.

16. Microcentrifuge samples at 14,000 rpm for 4 min. Save the supernatant from the “no antibody” sample as “total input chromatin.” For other samples, discard supernatant.

17. Wash pellets twice with 1.2 ml of 1 \times dialysis buffer (*if you are using a monoclonal antibody, omit the Sarkosyl) and four times with 1.2 ml of IP wash buffer (*pH 8.0 for monoclonal antibodies). For each wash, dissolve the pellet in 600 μl of buffer. Then, add an additional 600 μl of buffer and incubate samples on a rotating platform for 3 min. Next, microcentrifuge samples at 14,000 rpm for 4 min. Try to remove as much buffer as possible after each wash without aspirating the Staph A cells.

If performing a standard single IP, omit steps 18–24, proceed to step 25, and elute as described. If proceeding to a CpG microarray analysis, the samples are treated as a standard, single IP at this stage. Therefore, omit steps 18–24, proceed to step 25, and elute as described. Also, for the no antibody and mock samples used in the ChIP cloning experiment, omit steps 18–24 and elute as described in step 25 as well. However, for the experimental samples used for ChIP cloning, proceed to step 18.

18. After the last wash, microcentrifuge samples at 14,000 rpm for 4 min. Remove supernatant and repeat spin for 1–2 min. Remove the last traces of buffer. Elute antibody/protein/DNA complexes by adding 30 μl of IP elution buffer. Shake on vortexer for at least 30 min at setting “Vortex 3.” Microcentrifuge samples at 14,000 rpm for 4 min. Transfer supernatants to new tubes.

19. Microcentrifuge collected samples at 14,000 rpm for 4 min to remove any traces of Staph A cells. Transfer supernatants to new tubes.

20. Add 270 μl of dilution buffer (plus protease inhibitors) for each 30 μl of eluted samples. Combine samples into approximately 600 μl per tube. Add 1 μg of the same antibody used in the first precipitation and incubate at 4 °C overnight (or for at least 3 h).

2.3.3. Day 3

21. If you are not using rabbit polyclonal antibodies (e.g., you are using a mouse monoclonal antibody or a goat polyclonal antibody), add 1 μg of an appropriate secondary antibody and incubate for an additional 1 h.

22. Add 10 μl of blocked Staph A cells to each sample. Incubate on a rotating platform at 4 °C for 15 min.

23. Microcentrifuge samples at 14,000 rpm for 4 min.

24. Wash pellets twice with 1.2 ml of 1 \times dialysis buffer (*if you are using a monoclonal antibody, omit the Sarkosyl) and four times with 1.2 ml of IP wash buffer (*pH 8.0 for monoclonal antibodies). For each wash, dissolve the pellet in 600 μl of buffer. Add an additional 600 μl of buffer and then incubate samples on a rotating platform for 3 min. Next, microcentrifuge samples at 14,000 rpm for 4 min. Try to remove as much buffer as possible after each wash without aspirating the Staph A cells.

25. After the last wash, microcentrifuge and remove the last traces of buffer. Elute antibody/protein/DNA complexes by adding 150 μl of elution buffer. Shake on vortexer for at least 15 min at setting “Vortex 3.” Microcentrifuge at 14,000 rpm for 4 min. Transfer supernatants to new tubes. Repeat and combine both elutions in the same tube.

26. After the second elution, microcentrifuge samples at 14,000 rpm for 4 min to remove any traces of Staph A cells. Transfer supernatants to new tubes. Add 1 μl of high-concentration RNase A (Roche, Catalog No. 1579681, 10 mg/ml) and 12 μl of 5 M NaCl to a final concentration of 0.3 M. Remember to include the “total” sample at this point to reverse crosslinks. For the total sample, use only 20% of starting volume and increase the volume of the sample to 600 μl with elution buffer. Add 24 μl of 5 M NaCl and 2 μl RNase A to the total sample. Incubate samples at 67 °C for 4–5 h to reverse formaldehyde crosslinks. After the 4- to 5-h incubation, add 2.5 vol of ethanol and precipitate at –20 °C overnight.

2.3.4. Day 4

27. Microcentrifuge samples at 14,000 rpm for 15–20 min at 4 °C. Discard supernatant, and respin to remove residual ethanol. Allow pellets to air-dry completely.

28. Dissolve each pellet in 100 μl of TE. Add 25 μl of 5 \times PK buffer and 1.5 μl of proteinase K (25 mg/ml) to each sample. The “total” sample will be viscous and may have to be dissolved in a larger volume. Incubate at 45 °C for 1–2 h.

29. Add 175 μl of TE to each sample. Extract once with 300 μl of phenol/chloroform/isoamyl alcohol and once with 300 μl chloroform/isoamyl alcohol. Total input samples may need to be extracted twice.

30. Add 30 μl of 5 M NaCl, 5 μg of glycogen (Roche, Catalog No. 901393), and 750 μl of ethanol to each

sample. Precipitate at -20°C overnight. At this point, the CpG array method proceeds to step 31B.

2.3.5. Day 5

31A. Microcentrifuge samples at 14,000 rpm for 20 min at 4°C . Allow pellets to air-dry. Resuspend the experimental samples in 30 μl water and combine the multiple, identical experimental IP samples into one tube. Also resuspend the total sample in 30 μl water. Save an aliquot (10–15 μl) of the experimental sample to analyze known targets by PCR to confirm the efficiency of the IP. For the confirmation, use 2 μl of the IP sample, 2 μl of the no antibody sample, and 2 μl of the mock sample for the PCRs. Prior to PCR, dilute the total sample an additional 100-fold (therefore total represents 0.2% of starting input chromatin) and use 2 μl for the PCR. Do not proceed with the cloning procedure if samples do not have efficient immunoprecipitation (as determined by the specific enrichment of a known target gene in the antibody samples relative to the no antibody control).

32A. To create blunt-ended DNA fragments for cloning purposes, a DNA polymerase reaction is necessary. Add 60 μl of the experimental sample, 20 μl of 1 mM dNTPs, 20 μl of $10\times$ T4 DNA polymerase buffer, 2 μl bovine serum albumin (BSA, 10 mg/ml), 92 μl of water, and 6 μl of T4 DNA polymerase. Incubate at 37°C for 30 min. The amount of sample added to the polymerase reaction may need to be varied depending on the concentration of the DNA in the precipitated sample.

33A. Perform a phenol:chloroform extraction on the sample. Add 200 μl of phenol:chloroform:isoamyl alcohol (25:24:1) and vortex. Spin samples at 14,000 rpm for 5 min. Save the aqueous phase in a new tube.

34A. Extract the samples with chloroform. Add 200 μl of chloroform:isoamyl alcohol (24:1) and vortex. Spin samples at 14,000 rpm for 5 min. Save the aqueous phase in a new tube.

35A. Precipitate the DNA with 0.1 vol 3 M NaOAc and 2.5 vol ethanol. Incubate samples at -20°C for at least 2 h or on dry ice for approximately 10 min.

36A. Spin samples at 14,000 rpm for 10–20 min. Discard supernatant and dry DNA pellet. Resuspend pellet in 10 μl of water.

37A. Ligate the sample into a vector previously digested with an enzyme which creates blunt ends. For this step, we have used the pUC19 vector digested with *HincII*. Following digestion, the vector was treated with alkaline phosphatase to help prevent the religation of the vector alone. Standard ligation reactions are performed in a 10- μl volume and a range of chromatin concentrations should be used (2, 1, 0.5, and 0.1 μl). We have found that using increasing concentrations of ChIP DNA in the ligation reaction increases the probability of cloning larger fragments. However, the increasing DNA

concentrations also can decrease overall ligation efficiency. Therefore, a balance needs to be found to obtain optimal ligation efficiency while retaining the greater insert size. Ligations are performed at 16°C overnight.

2.3.6. Day 6

38A. Perform standard transformation reactions using entire ligation mixture and competent bacterial cells. We have used TOP10 competent cells (Invitrogen) and a heat shock method for transformation. Plate cells under appropriate selection and incubate overnight at 37°C .

2.3.7. Day 7

39A. Pick colonies and grow for minipreps. Cultures can be grown for a minimum of 6–8 h or overnight. Isolate vector using a miniprep plasmid purification protocol (such as those provided by Qiagen). Digest the vector with appropriate restriction enzymes to release the cloned DNA insert (remember that the restriction enzyme site that was cloned into will be destroyed). Run the digest on a 1% agarose gel to determine the insert size. We have found that cloned fragments of 500 bp or greater have been more likely to contain real clones instead of nonspecific DNA such as repeat regions. Therefore, we only proceed to analyze cloned inserts greater than 500 bp.

2.3.8. Days 8–10

40A. Sequence the cloned inserts using vector-specific primers (it is best to sequence the inserts from both directions).

41A. Design clone-specific primers to analyze in independent chromatin immunoprecipitation experiments to confirm that the cloned fragment is specific to the antibody treatment and not due to the nonspecific precipitation of the fragment.

2.4. ChIP-CpG microarray method

To obtain sufficient DNA for two microarray hybridizations, it is best to perform at least 28 single IPs using the specific antibody. Quantitation will be required to determine the exact amount of DNA obtained in the pooled ChIP samples. As a control, at least 28 single IPs with either no antibody or preimmune serum should also be performed. Rather than performing 56 reactions at once, we have found that it is more manageable to perform smaller ChIP experiments on separate days to accumulate sufficient IP numbers. However, it is best to perform the same number of antibody and no antibody IPs in each individual experiment until sufficient DNA is obtained to probe the microarray twice with 2 μg of DNA each time.

Perform steps 1–17 as described for the ChIP cloning protocol, but do not perform the second immunopre-

precipitation reaction (skip steps 18–24). Next, perform steps 25–30. However, stop after step 30 and proceed as described below.

31B. Microcentrifuge samples at 14,000 rpm for 10–20 min. Discard supernatant and allow pellets to air-dry completely. Resuspend DNA in 10 μ l of water per IP sample.

32B. Before performing the CpG island microarray analysis, separately combine the experimental and the no antibody samples from individual experiments and check an aliquot by PCR for the efficiency of the ChIP experiment. Proceed only if the positive control demonstrates the ChIP experiment was efficient. For the confirmation PCR, use 2 μ l of a 1:3 dilution of the pooled IP or the no antibody sample. This can be compared with a PCR containing 2 μ l of the 0.2% total sample.

33B. Quantitate DNA and label at least 2 μ g of the specific antibody ChIP sample as well as the no antibody control with amino-allyl dUTP (aa-dUTP, Sigma). The Bioprime DNA labeling system (Life Technologies) can be used for this step. Cy5 or Cy3 dyes are next coupled to the aa-dUTP labeled DNA (one dye for each type of sample). The labeled samples are then cohybridized to an appropriate microarray. Following hybridization, the arrays are washed and then scanned to determine the positive clones. It is best to perform this type of analysis in cooperation with a laboratory or facility that specializes in microarray analysis. Alternatively, it is best to work out the labeling and hybridization conditions prior to attempting the ChIP-microarray analysis. The protocols given here for the microarray portion of the procedure are brief and not intended to be followed independently. Detailed microarray protocols can be found on a public website maintained by the DeRisi lab at <http://www.microarrays.org>.

3. Potential problems and solutions

It is critical to ensure the efficiency of the immunoprecipitation reaction prior to beginning the modified ChIP procedures described in this article. Several controls can be performed that will provide confidence that the standard ChIP assay is working. For example, not all antibodies are efficient in immunoprecipitation reactions. Therefore, numerous antibodies may need to be tested to find a specific antibody with the ability to immunoprecipitate the desired protein complexes. It is best to first examine factor binding to a known target gene before proceeding with the ChIP cloning protocol. In the absence of testing a known target gene (e.g., if the factor was recently identified and no targets are known), the first steps of the standard ChIP assay can be performed followed by Western blot analysis. This

will allow for the confirmation that the antibody can work in the IP portion of the procedure. In addition, to confirm that the technical aspects of the ChIP assay are working efficiently, the association of RNA polymerase II or histones with promoter regions can be monitored. Signals equal to or greater than that of 0.2% of the starting chromatin (i.e., a PCR signal intensity obtained with 2 μ l of the IP sample which is equal to or greater in intensity than the signal obtained using 2 μ l of the diluted input sample) should be seen with either the RNA polymerase II or histone antibodies. If the ChIP assay is not working efficiently using these antibodies, do not proceed with the procedure because this indicates a technical problem with the standard ChIP assay.

After the standard ChIP procedure is working efficiently, the cloning procedures can be attempted. As noted above, one common problem associated with the ChIP cloning protocol is the isolation of a large number of nonspecific clones. This suggests that the nonspecific DNA has not been sufficiently cleared from the specific IP samples. Increasing the stringency of the washing conditions may help decrease the nonspecific association with the Staph A cells. Alternatively, it may be worth attempting the IP portion of the procedure with protein A/G agarose or Sepharose beads. In addition, it may be worthwhile to perform the ChIP-CpG microarray procedure instead to increase the chances of isolating CpG-rich regulatory regions.

The high-throughput ChIP-CpG microarray strategy may increase the efficiency of isolating target promoters because the promoter and first exon of many genes are CpG islands [7]. However, it is possible that a factor will selectively regulate promoters that are not associated with CpG islands. In this case, an alternative array may need to be used or, instead, the ChIP cloning strategy may need to be employed. It is also worth noting that neither of the modified ChIP strategies is currently designed to be used as an exhaustive search for target genes. Rather, they are intended to be used to find a set of genes directly regulated by the factor of interest. In addition, as mentioned above, further analysis is necessary to demonstrate the functional consequence of factor binding. Therefore, it is possible, and perhaps likely, that a subset of targets found using these methods will be bound by, but not regulated by, the transcription factor of interest.

4. Concluding remarks

We have described two modifications of the ChIP protocol that allow for the identification of target genes of mammalian transcription factors. These methods both have two major advantages over the commonly used approach which couples the overexpression or

underexpression of a factor with cDNA microarray analysis. First, the ChIP protocols identify DNA sequences directly bound by the factor of interest. Therefore, the ChIP protocols eliminate the identification of indirect targets that are only isolated due to the deregulation of a cascade of signaling events and are not directly regulated by the factor of interest. Recent work in the yeast system has emphasized the need to distinguish direct versus indirect targets. For example, a common mode of coordinating a series of precisely ordered events in the cell is to have key transcription factors regulate the gene expression of other transcription factors [14]. Although biologically interesting, the indirect regulation of a series of genes by a transcription factor could be misinterpreted in a cDNA microarray experiment as direct regulation by the overexpressed transcription factor.

A second advantage in using the ChIP procedures to identify novel target genes is that the assay is performed in the context of the natural cellular environment. When a study involves overexpression, a significant fraction of the genes isolated may not be targets when the protein is present at physiological concentrations. Artifacts due to overexpression could be the result of a factor interacting with sites not normally regulated by a protein because of limiting amounts of the protein in the normal cell. Alternatively, forced protein–protein interactions that do not normally occur could lead to the recruitment of the overexpressed protein to a DNA sequence not normally bound by the protein. All these potential problems due to overexpression are eliminated in the ChIP cloning protocols because the experiment is performed using the endogenous protein binding to the endogenous gene in the context of the nucleus.

In summary, we suggest that the described modifications to the ChIP protocol can be used to identify promoters that are directly bound by a specific transcription factor under normal physiological conditions. The cloning protocols described above have been successfully used to isolate novel E2F target genes [4,5] and more recently other transcription factor targets (Weinmann, Kanjo, and Farnham, unpublished data). Therefore, we suggest that one or both of these techniques should be useful in the study of many mammalian transcription factors. It is likely that the use of these unbiased approaches will provide new insight into gene regulation. For instance, it is possible that different consensus sequences will emerge when comparing the *in vivo* versus *in vitro* binding results. In support of this hypothesis, we note that many E2F target promoters identified using the ChIP-CpG approach do not contain a site closely resembling the known E2F consensus binding site [5]. Finally, we suggest that the ChIP-CpG profiling of a factor using several cell types may provide novel insight into biological function and tissue-specific gene regulation.

Acknowledgments

Public Health Service Grants CA45250 and CA07175 (a NCI Cancer Center Core grant) support P.J.F. and Training Grant CA09681 from the National Institutes of Health supports A.S.W. We thank members of the Farnham laboratory for critical reading of this article.

Appendix A. Reagents

Note. Review safety precautions associated with all chemicals and reagents to be used in the procedure prior to attempting this protocol. Some reagents are hazardous.

Cell Lysis buffer

5 mM PIPES, pH 8.0
85 mM KCl
0.5% Nonidet P-40
Protease inhibitors (add fresh)

Nuclei Lysis buffer

50 mM Tris–Cl, pH 8.1
10 mM EDTA
1% sodium dodecyl sulfate (SDS)
Protease inhibitors (add fresh)

IP Dilution buffer

0.01% SDS
1.1% Triton X-100
1.2 mM EDTA
16.7 mM Tris–Cl, pH 8.1
167 mM NaCl
Protease inhibitors (add fresh)

1× Dialysis buffer

2 mM EDTA
50 mM Tris–Cl, pH 8.0
0.2% Sarkosyl (omit for monoclonal antibodies)
PMSF (add fresh)

IP Wash buffer

100 mM Tris–Cl, pH 9.0 (8.0 for monoclonal antibodies)
500 mM LiCl
1% NP-40
1% Deoxycholic acid
PMSF (add fresh)

Elution buffer

50 mM NaHCO₃
1% SDS

5× PK buffer

50 mM Tris–Cl, pH 7.5
25 mM EDTA
1.25% SDS

Protease Inhibitors

100 mM PMSF (Sigma, Catalog No. P-7626) in 2-propanol.

10 mg/ml aprotinin (Sigma, Catalog No. A-1153) in 0.01 M Hepes, pH 8.0.

10 mg/ml leupeptin (Sigma, Catalog No. L-2884) in water.

Staph A cells

Resuspend 1 g of lyophilized protein A-positive *Staphylococcus aureus* cells, whole cells (Cowan 1 strain), heat-killed, fixed in formalin (Staph A cells, Roche, Catalog No. 100061) in 10 ml of 1× dialysis buffer. Centrifuge at 10,000 rpm for 5 min at 4 °C. Discard supernatant and repeat wash. Resuspend in 3 ml of 1× PBS plus 3% SDS and 10% BME. Boil for 30 min. Centrifuge at 10,000 rpm for 5 min. Wash in 1× dialysis buffer and centrifuge at 10,000 rpm for 5 min. Repeat wash. Resuspend in 4 ml of 1× dialysis buffer. Divide into 100-µl aliquots, snap-freeze, and store in liquid nitrogen. Before use in each ChIP experiment, thaw one tube (100 µl) of cells for approximately every 5×10^7 starting cells. Add 10 µl of herring sperm DNA (10 mg/ml) and 10 µl of BSA (10 mg/ml) to each tube of Staph A cells. Incubate on a rotating platform at 4 °C for at least 3 h or overnight. Before using, microcentrifuge for 3 min. Remove supernatant and wash pellet twice in 1× dialysis buffer. Resuspend cells in 1× dialysis buffer (equal to starting volume of thawed Staph A cells).

References

- [1] T. Jenuwein, C.D. Allis, *Science* 293 (2001) 1074–1080.
- [2] R. Tupler, G. Perini, M.R. Green, *Nature* 409 (2001) 832–833.
- [3] H. Muller, A.P. Bracken, R. Vernell, M.C. Moroni, F. Christians, E. Grassilli, E. Prosperini, E. Vigo, J.D. Oliner, K. Helin, *Genes Dev.* 15 (2001) 267–285.
- [4] A.S. Weinmann, S.M. Bartley, M.Q. Zhang, T. Zhang, P.J. Farnham, *Mol. Cell. Biol.* 21 (2001) 6820–6832.
- [5] A.S. Weinmann, P.S. Yan, M.J. Oberley, T.H.-M. Huang, P.J. Farnham, *Genes Dev.* 16 (2002) 235–244.
- [6] B. Ren, F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T.L. Volkert, C.J. Wilson, S.P. Bell, R.A. Young, *Science* 290 (2000) 2306–2309.
- [7] Ioshikes, Zhang, *Nat. Genet.* 26 (2000) 61–63.
- [8] J.D. Lieb, X. Liu, B. Botstein, P.O. Brown, *Nat. Genet.* 28 (2001) 327–334.
- [9] V.R. Iyer, C.E. Horak, C.S. Scafe, D. Botstein, M. Snyder, P.O. Brown, *Nature* 409 (2001) 533–538.
- [10] F. Antequera, A. Bird, *Proc. Natl. Acad. Sci. USA* 90 (1993) 11995–11999.
- [11] S. Hannenhalli, S. Levy, *Bioinformatics* 17 (2001) S90–S96.
- [12] B. Ren, H. Cam, Y. Takahashi, T. Volkert, J. Terragni, R.A. Young, B.D. Dynlacht, *Genes Dev.* 16 (2002) 245–256.
- [13] M. Carey, S.T. Smale, in: J. Cuddihy, P. Barker, D. deBruin, S. Schaefer (Eds.), *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000.
- [14] I. Simon, J. Barnett, N. Hannett, C.T. Harbison, N.J. Rinaldi, T.L. Volkert, J.J. Wyrick, J. Zeitlinger, D.K. Gifford, T.S. Jaakkola, R.A. Young, *Cell* 106 (2001) 697–708.
- [15] D. DeRyckere, J. DeGregori, *Methods* 26 (2002) 57–75.
- [16] D.J. Segal, *Methods* 26 (2002) 76–83.