Chapter 18

Chromatin Immunoprecipitation-Based Analysis of Gene Regulatory Networks Operative in Human Embryonic Stem Cells

Marc Jung and James Adjaye

Abstract

Chromatin immunoprecipitation (ChIP) followed by microarray-based (ChIP-Chip) or next-generation sequencing-based (ChIP-Seq) analysis has been established as a powerful and widely used method to investigate DNA–protein interactions relative to a genomic location in vivo. Here, we present a ChIP-Chip protocol, which utilizes an alternative, easier amplification protocol and when using high-quality ChIP-grade antibodies, will generate enough material for hybridization or sequencing with negligible enrichment bias due to amplification.

Key words: Embryonic stem cells, Chromatin, Protein–DNA interaction, Immunoprecipitation, Transcription factor, OCT4, NANOG, Next-generation sequencing

1. Introduction

1.1. Chromatin Immunoprecipitation

The basic principle underlying chromatin immunoprecipitations (ChIPs) is that proteins (transcription factors) are cross-linked to the DNA double helix by using cross-linking agents like formaldehyde. Formaldehyde is a tight (2Å) cross-linking agent that efficiently produces both protein–nucleic acid and protein–protein cross-links in vivo. Formaldehyde is a very reactive dipolar compound in which the carbon atom acts as a nucleophilic center. Amino and imino groups of amino acids (lysines, arginines, and histidines) and of DNA (primarily adenines and cytosines) readily react with formaldehyde leading to the formation of a Schiffbase. This intermediate can further react with a second amino group and condense to give the final cross-link. These reactions take place in vivo within minutes after the addition of formaldehyde to living

Kursad Turksen (ed.), *Human Embryonic Stem Cells Handbook*, Methods in Molecular Biology, vol. 873,
DOI 10.1007/978-1-61779-794-1_18,

cells or embryos (1). Although other cross-linking reagents have been employed (2), formaldehyde remains the most widely used as the reaction can be reversed by heat. This is achieved primarily by protonation of imino groups at low pH in aqueous solution. After cross-linking the chromatin, the cells are either directly lysed or the nuclei are extracted. The chromatin is sheared into fragments of the desired size by sonication or through micrococcal nuclease digest to a size of usually 0.2–1.0 kb. For ChIP-Chip or ChIP-Seq applications, a smaller size is essential if a higher resolution of the subsequent analysis is desired. The fragments bound to the protein of interest are usually enriched by immunoprecipitation (IP) with an antibody against the respective protein under investigation. Protein-specific antibodies require optimizing for their application in IP conditions. Additionally, using different polyclonal antibodies for the same protein may show a different preference for epitopes, resulting in a possible distinct selection of cross-linked loci. Monoclonal antibodies would be preferable due to their specific epitope selection, but it is more difficult to obtain functional ChIP-grade monoclonal antibodies by this approach. The control cell is processed with the pre-immune serum from the host organism of the specific antibody used for the IP. This control identifies unspecific fragments enriched, e.g., by adhesion to the samples tubes. The formaldehyde cross-links are then reversed and the precipitated DNA fragments are purified. Yields from ChIP are usually low but sufficient for subsequent PCR or qPCR analysis. In this chapter, we describe a protocol which we adapted and expanded with a different amplification method from an earlier publication, focusing on the key pluripotency-associated transcription factors OCT4, NANOG, and SOX2 (3) and which we used for the analysis of OCT4-mediated gene regulatory networks necessary for maintaining self-renewal in embryonic stem cells and embryonal carcinoma cells (4).

1.2. Chromatin Immunoprecipitation Followed by Microarray Hybridization (ChIP-Chip) or Deep Sequencing (ChIP-Seq)

As the traditional methods had failed to create high-resolution, genome-wide maps of the interaction between a DNA-binding protein and DNA, the combination of ChIP and whole-genome promoter microarrays (ChIP-Chip) and next-generation sequencing-based (ChIP-Seq) circumvented these limitations by creating high-resolution genome-wide maps of the in vivo interactions between DNA-associated proteins and DNA.

The ChIP-Chip technique was first used to identify binding sites for individual transcription factors in *Saccharomyces cerevisiae* (1, 5, 6). More recently, a c-Myc epitope protein tagging system was used to map the genome-wide positions of 106 transcription factors in yeast (7).

For microarray- and next-generation sequencing-based detection of immunoprecipitated DNA, amplification of the DNA is generally necessary, as the DNA yield, obtained after the pulldown

is not sufficient for hybridization or sequencing. Ideally, the ChIP reactions are scaled up and amplifications are avoided. Three amplification methods have so far been widely used: randomly primed (8), ligation-mediated PCR (9) as well as amplification on the basis of T7 DNA polymerase (10). Before adding the antibodies for the pulldown reaction, a part of the fragmented chromatin will be retained as total genomic reference DNA. Although these samples usually give enough material for microarray hybridization, they should also be amplified to avoid any amplification bias. The enriched and the reference DNA are then fluorescently labeled. Although one color platforms, where both samples have the same label, e.g., Cy3 are hybridized on separate arrays the use of two color platforms is often preferred, as this minimizes the influence of microarray batch effects on the experimental results. In this case, the ChIP DNA is labeled with different fluorescent dyes and the samples are combined and hybridized to a single DNA microarray. The relative intensities of the two dyes allow the detection of the fragments that are enriched in the IP, thereby enabling the identification of protein–DNA interaction sites (see Fig. 1). For a comprehensive analysis, microarrays used in ChIP-Chip applications represent ideally the entire genome of the organism in the form of overlapping fragments. In this case, the limitation will be the obligatory selection of preferred probe sequences for optimal hybridization, which in turn defines the maximal resolution of the tiling array. Furthermore, for larger genomes such as for higher eukaryotes these are not available or only at very high monetary cost. Therefore, arrays are often custom designed for specific applications. The resolution of the identified binding sites depends on the size of the sheared DNA and the size and spacing of the probes on the arrays. For example, typical yeast experiments achieve a resolution of about 1 kb, which is sufficient to assign binding to the regulation of a single gene. Once the bound regulatory region is identified, the exact binding site can often be inferred by computational methods. In comparison, the ChIP-Seq approach offers an unbiased analysis regarding genomic loci, which have not been predefined by tilling arrays. This advantage will still come with a higher cost for deep sequencing projects.

2. Materials

All solutions should be prepared using ultrapure deionized water and analytical grade reagents. Prepare and store all solutions at room temperature unless indicated otherwise. For ChIP reaction:

1. NaCl, 137 mM; KCl, 2.7 mM; Na_2 HPO_4 , 10 mM; and $KHPO_4$, 2 mM of pH 7.2.
2. PBS/2% FBS/PMSF, 1 mM.

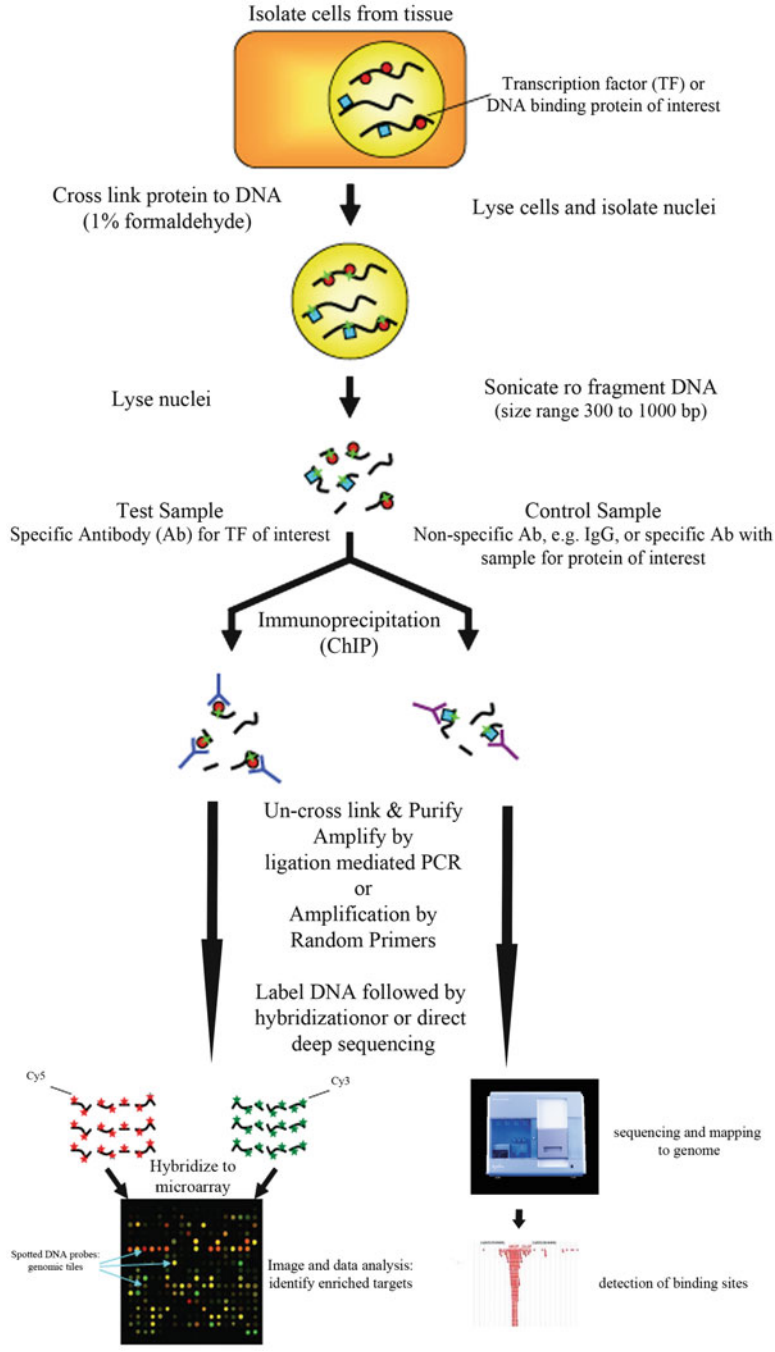

Fig. 1. An illustration of the principle of a ChIP-Chip or ChIP-Seq experiment (Adapted from Peter White, Ph.D.).

3. Formaldehyde 11% (25 ml): 7.45 ml 37% formaldehyde, 0.1 M NaCl, 1 mM EDTA, 0.5 mM EGTA, and 50 mM HEPES.
4. Glycine 2.5 M: 18.767 g in 100 ml sterile water.

5. Lysis buffer 1 (LB1): 50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton-X-100.
6. Lysis buffer 2 (LB2): 10 mM Tris–HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA.
7. Lysis buffer 3 (LB3): 10 mM Tris–HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% *N*-lauroylsarcosine.
8. Wash buffer (RIPA): 50 mM Hepes-KOH, pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate.
9. Elution buffer: 50 mM Tris–HCl, pH 8.0, 10 mM EDTA, 1% SDS.
10. TE: 10 mM Tris–HCl of pH 8.0, 2 mM EDTA. Supplemented with protease inhibitors prior to using.
11. Bovine serum albumin-Fraction V (BSA).
12. MiniElute purification Kit (Qiagen).
13. Wizard SV Gel and PCR clean-up System (Promega, USA).
14. Tris–hydrochloride (Merck, Germany).
15. KCl (Merck).
16. Tween20, nuclease free (Sigma-Aldrich, USA).
17. $MgCl_2$ (Merck).
18. Antibodies used here: Anti-OCT3/4 (H134), sc-9081-x (Santa Cruz, USA).
19. Dynabeads Protein G (Invitrogen, USA).
20. Dynabeads Protein A (Invitrogen).

For random PCR-based amplification:

21. For random_primer_mix:
    Make stocks of 100 pmol/ml of each Primer.
    random_primer_a: GTT TCC CAG TCA CGA TCN NNN NNN NA
    random_primer_t: GTT TCC CAG TCA CGA TCN NNN NNNNT
    random_primer_g: GTT TCC CAG TCA CGA TCN NNNNNN NG
    random_primer_c: GTT TCC CAG TCA CGA TCNNNN NNN NC
    where N can be any base. Dilute each of these to 40 pmol/ml working stock and store at 20°C. Pool in equal amounts (each primer 10 pmol/ml, total 40 pmol/ml). 10× PCR buffer (500 mM KCL, 150 mM Tris pH 8.3, 1% Tween, 15 mM MgCl).
22. Sequenase T7 DNA Polymerase Version 2.0, 13 U/μl (Amersham, UK).
23. Sequenase buffer (Amersham).

24. BSA (molecular biology grade) 500 μg/ml (Biolabs).
25. DTT 0.1 M, RNase free (Promega).
26. dNTPs (Promega).
27. Taq polymerase (Promega).
28. Pfu DNA Polymerase (Fermentas, USA).

3. Methods

3.1. Chromatin Immunoprecipitation

We recommend using approximately 5×10^7–1×10^8 cells (70–90% confluency for adhesion cells from two 15 cm^2 plates or 175 cm^2 flasks) for each IP reaction (see Note 1).

3.1.1. Formaldehyde Cross-linking of Cells

1. Add 1/10th volume of freshly prepared 11% formaldehyde solution to the plates (see Note 2).
2. Swirl plates briefly and allow to stand at room temperature for 10 min.
3. Add 1/20th volume of 2.5 M glycine to the plates to quench the formaldehyde.
4. Rinse cells twice with 5 ml 1× PBS. Harvest cells using a silicon scraper.
5. Pool cells in 50 or 15 ml conical tubes and spin at 1,350 × *g* for 5 min at 4°C (using a swinging bucket rotor). Discard supernatant and resuspend pellet in 10 ml 1× PBS per 10^8 cells.
6. Transfer 5×10^7–1×10^8 cells to a 15 ml conical tube and spin at 1,350 × *g* for 5 min at 4°C (using a swinging bucket rotor). Discard the supernatant.

3.1.2. Preblocking and Binding of Antibodies to Magnetic Beads

1. Add 100 μl Dynal magnetic beads to a microfuge tube. Add 1 ml of blocking solution. Set up one tube per IP reaction.
2. Collect the beads using a magnetic stand. Remove supernatant.
3. Wash the beads in 1.5 ml blocking solution two more times.
4. Resuspend the beads in 250 μl blocking solution and add 10 μg of antibody.
5. Incubate overnight at 4°C on a rotating platform.
6. On the next day, wash the beads (three times in 1 ml block solution).
7. As described above in step 3.
8. Resuspend the beads in 100 μl blocking solution.

3.1.3. Sonication of Cells

1. Resuspend each pellet of 5×10^7–10^8 cells in 5 ml of lysis buffer LB1. Place the tube on ice, turning the tubes every 2 min for a period of 10 min. Spin at 1,350 × *g* for 5 min at 4°C using a tabletop centrifuge.

2. Resuspend each pellet in 5 ml of lysis buffer LB2. Rock gently at room temperature for 10 min. Pellet nuclei using tabletop centrifuge by spinning at 1,350 × *g* for 5 min at 4°C.
3. Resuspend each pellet in each tube in 3 ml of lysis buffer LB3.
4. Transfer cells to a, 15 ml polypropylene conical tube, cut into two pieces at the 7 ml mark.
5. Sonicate suspension with a tapered microtip with a 6.5 mm diameter, attached to a BRANSON 250 and sonicate at power 3 for 11 min with 30% duty cycle at 4°C while samples are immersed in an ice bath. Sonication is a critical step of the whole experiment and should be planned with care (see Note 3).
6. Add 300 µl of 10% Triton X-100 to sonicated lysate. Split into two 1.5 ml centrifuge tubes. Spin at 20,000 × *g* for 10 min at 4°C to pellet debris.
7. Combine supernatants from the two 1.5 ml centrifuge tubes in a new 15 ml conical tube for IP.
8. Save 50 µl of the cell lysate from each sample as whole cell extract (WCE) DNA. Store at –20°C.

3.1.4. Preblocking and Binding of Antibody to Magnetic Beads

1. Add 100 µl Dynal magnetic beads to a microfuge tube. Add 1 ml blocking solution.
   Set up one tube per IP reaction (see Note 5).
2. Collect the beads using the magnetic stand. Remove the supernatant.
3. Wash beads in 1.5 ml blocking solution twice.
4. Resuspend beads in 250 µl blocking solution and add 1–10 µg of antibody (see Note 4).
   The exact amount of antibody needs to be tested first and can vary extensively between different antibodies, but also between different batches of the same antibody (see Note 7).
5. Incubate overnight at 4°C on a rotating platform.
6. Next day, wash beads as described above (three times in 1 ml blocking solution).
7. Resuspend in 100 µl blocking solution.

3.1.5. Reversal of Cross-links After Chromatin Immunoprecipitation

1. Add 100 µl of antibody/magnetic bead mix to cell lysates.
   With the OCT4 antibody used here, we could scale down the volume to a fifth of the total lysate and could still obtain sufficient enrichment signals.
2. Gently mix on rotator or rocker at 4°C overnight. Make sure that there is enough liquid in the tube to enable efficient rotation.

The following steps should be carried out in a 4°C cold room. Prechilled tubes should be used.

1. Transfer half the volume of an IP to a 1.5 ml microfuge tube (see Note 6).
2. Place tubes in the magnetic stand to concentrate and localize the beads. Remove supernatant and add the remaining IP. Repeat this step once.
3. Add 1 ml of wash buffer (RIPA) to each tube. Remove tubes from the magnetic stand and shake or agitate tube gently to resuspend the beads. Place the tubes again in the magnetic stand to concentrate the beads. Remove supernatant. Repeat this washing step 3–7 times.
4. Wash once with 1 ml TE containing 50 mM NaCl.
5. Spin at 1,000 × *g* for 3 min at 4°C and remove any residual TE buffer.
6. Add 200 µl of elution buffer.
7. Elute at 65°C for 20 min. Resuspend beads every 2 min with brief vortexing.
8. Spin down beads at 16,000 × *g* for 1 min at room temperature.
9. Remove 190 µl of supernatant and transfer to a new tube. Reverse cross-link the immunoprecipitated DNA by incubating at 65°C overnight.
10. Thaw 50 µl of WCE reserved after sonication, add 150 µl of elution buffer, and mix. Reverse cross-link of this WCE DNA by incubating at 65°C overnight.

3.1.6. Purification of DNA

Isolate IP and input DNA using the Qiagen PCR purification kit and incorporating these modifications:

1. Add 500 µl of PB buffer and allow to stand for 10 min.
2. After column purification, elute with 50 µl of elution buffer prewarmed to 60°C (see Note 8).

3.2. Amplification of ChIP and Input DNA (PCR-Based Amplification)

1. Linear amplification of ChIPed DNA and input control is based on a random primer amplification described by Bohlander et al. (8) and subsequently modified for ChIP applications (7) (see Note 9 and 10). The protocol is laid out in Table 1.
2. Amplified samples can be purified using the Wizard SV PCR purification kit following the manufacturer's instructions.
3. For assessing the quality of the DNA and possible amplification bias which might have occurred after the randomized amplification, the range of the distribution after the amplification has to be tested, as illustrated in Fig. 2. Furthermore, nonamplified DNA samples should be compared with amplified DNA samples employing qPCR. Figure 3 illustrates a typical comparison between the distribution of nonamplified

Table 1
Reagents and cycling parameters for the PCR-based amplification of ChIP-derived genomic DNA

Round A reaction	
1	Per reaction (µl)
ChIP DNA	7
5× Sequenase buffer	2
Random_primer_mix (40pmol/µl)	1
Total volume	*10 µl*
Program for thermocycler	
2 min at 94°C	
5 min at 10°C	
Then, add	
5× Sequenase buffer	1
dNTP (3 mM)	1.5
DTT (0.1 M)	0.75
BSA (500 µg/µl)	1.5
Sequenase (13 U/µl)	0.5
Total volume	*15.25 µl*
Program for thermocycler	
Ramp from 10 to 37°C over 8 min	
Hold at 37°C for 8 min	
Rapid ramp to 94°C	
2 min 94°C	
Rapid ramp to 10°C	
Hold 5 min at 10°C	
Add	
1.2 µl sequenase diluted 1:4 with buffer	
Ramp from 10 to 37°C over 8 min	
Hold at 37°C for 8 min	
Add H_2O to a total volume of 60 µl	
Round B template	(µl)
Round A template	15
10× PCR buffer	10
dNTP (25 mM each)	1
random_pimer_2 (100 pmol/µl)	1
Taq polymerase/Pfu (10:1)	1
H_2O	72
Total volume	*100 µl*
Add 85 µl of a premixed Mastermix, containing the PCR buffer, the dNTPs, random primers and the Taq polymerase/Pfu mixture to each 15 µl aliquoted reaction from Round A	
Program for thermocycler	
20 Cycles	94°C 30 s
	40°C 30 s
	50°C 30 s
	72°C 2 min

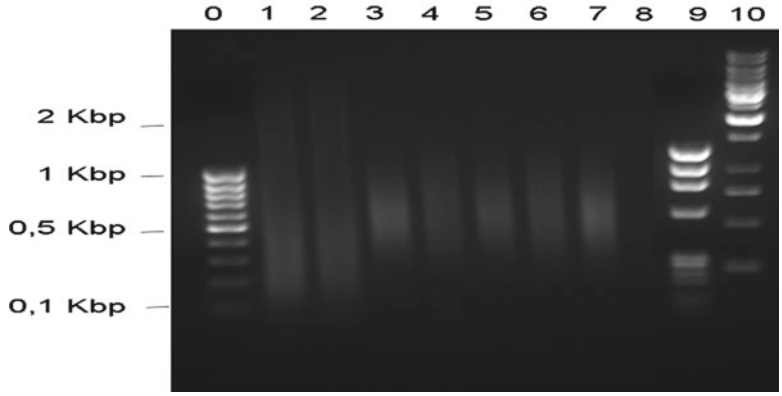

Fig. 2. An illustration of the distribution of DNA fragments before and after a random PCR-based amplification. The lanes are loaded with 300 ng of DNA. *Lanes 1* and *2*: DNA from whole cell extracts (WCE). *Lane 3*: input-antibody control, *Lane 4*: enrichment-antibody control, *Lane 5*: input OCT4-N19, *Lane 6*: enrichment OCT4-N19, *Lane 7*: input OCT4-H134, *Lane 8*: enrichment OCT4-H134, *Lane 9*: 1 kbp ladder (Fermentas), *Lane 0*: 100 bp ladder (Fermentas).

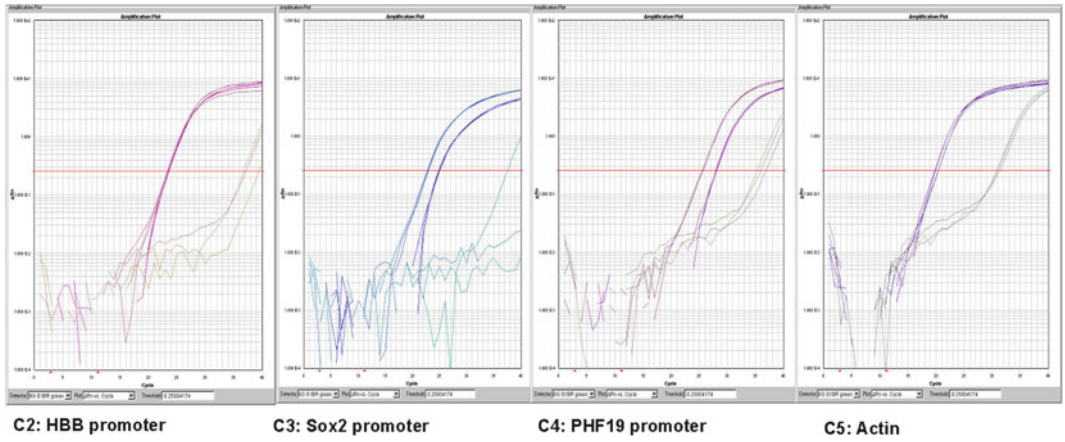

Fig. 3. Quantitative real-time PCR after random PCR-based amplification. *SOX2* and *PHF19* promoters harbor a binding motif for OCT4, whereas control promoter regions from hemoglobin beta (HBB) and *ACTIN* do not show enrichment after amplification.

and amplified DNA. This figure is based on the analysis carried out on our previously reported OCT4 binding sites and negative control loci (4).

4. Based on a comparative result as shown in Figs. 2 and 3, the DNA samples should now be ready for downstream processing for array-based hybridization (ChIP-Chip) or next-generation sequencing (ChIP-Seq).

4. Notes

1. Keep cell numbers constant in each experiment as well as incubation and sonication conditions. Chromatin can be stored at 4°C for up to 4 months.
2. The time needed for cross-linking proteins to chromatin as well as to other proteins should be constant for several samples. For optimal IP, each protein of interest might require different formaldehyde concentrations or reaction times.
3. Sonication conditions need to be optimized. Shearing varies greatly depending on cell type, growth conditions, quantity, volume, cross-linking, and equipment. In general, it is recommended to look for the lowest settings that result in sheared DNA that ranges from 100 to 600 bp in size, in order to prevent over shearing and still guarantee an optimal resolution for the tiling array or sequencing. We recommend a BRANSON 250 model but of course newer models, such as the DIGITAL Sonifier® UNITS, which can be programmed can be used. We adapted our method using a 1.7 microcentrifuge tube and 500 µl total volume with a program, using 20% sonification time for a total of 11 min. Important here is to prevent the sample from foaming and to make sure that the tube is sufficiently cooled during the whole duration of the sonication.
4. Antibodies need to be carefully selected and tested by Western blotting, immunofluorescence, or similar assays to guarantee specificity.
5. The exact type of Dynal beads depends on the antibody being used. Each type has preferences for different species and must be taken into account.
6. The exact number of washes depends on the quality of the antibody and may need to be optimized for each antibody.
7. For the immuno-enrichment and the washing steps use always siliconized tubes to be used for IPs. This is necessary if one aims to reduce nonspecific adherence of the antibodies to the reaction tubes.
8. The prediction of an optimal amount of antibody used is not possible. However, different ranges of concentrations can be tested and the enrichment can be subsequently analyzed for known binding sites by quantitative real-time PCR.
9. After elution, purified DNA can be stored in TE buffer and stored –20°C for long periods.
10. An alternative to random PCR amplification can be ligated mediated PCR. The advantage of the random PCR approach is that the ligation step can be skipped, without increasing bias.

To assess the reproducible amplification of enriched sequences, quantitative real-time PCR analysis should be employed (4). Here, negative genomic controls for PCR amplification should be used, for example a set of primers should be designed within a genomic sequence, adjacent to the binding site at which you would not expect an interaction with the protein under investigation. This reaction should not produce an amplicon. Re-check the specificity of the signal, when using SyberGreen-based real-time PCR, by analyzing the reaction on an agarose gel for the presence of a single amplicon.

Acknowledgments

We would like to thank Dr. Jenny Fischer for advice on random PCR amplifications. This work was supported by ENFIN, a Network of Excellence funded by the European Commission within its FP6 Programme, under the thematic area. "Life sciences, genomics and biotechnology for health," contract number LSHG-CT-2005-518254, and also the Max Planck Society.

References

1. Orlando V (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem Sci 25(3):99–104
2. Kurdistani SK, Grunstein M (2003) In vivo protein-protein and protein-DNA crosslinking for genomewide binding microarray. Methods 31(1):90–95
3. Boyer LA et al (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122(6):947–956
4. Jung M et al (2010) A data integration approach to mapping OCT4 gene regulatory networks operative in embryonic stem cells and embryonal carcinoma cells. PloS One 5(5):e10709
5. Lee TI et al (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298(5594):799–804
6. Lieb JD et al (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet 28(4): 327–334
7. Iyer VR et al (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409(6819):533–538
8. Bohlander SK et al (1992) A method for the rapid sequence-independent amplification of microdissected chromosomal material. Genomics 13(4):1322–1324
9. Mueller PR, Wold B (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246(4931):780–786
10. Liu CL, Schreiber SL, Bernstein BE (2003) Development and validation of a T7 based linear amplification for genomic DNA. BMC Genomics 4(1):19