Published in final edited form as:

Methods Mol Biol. 2009 ; 590: 223–233. doi:10.1007/978-1-60327-378-7_14.

Chromosome-wide analysis of protein binding and modifications

Kevin D. Sarge^{1,*}, Hongyan Xing, and Ok-Kyong Park-Sarge²

¹ Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, KY 40536

² Department of Physiology, University of Kentucky, Lexington, KY 40536

Abstract

In order to fully understand the functions of a DNA-binding protein it is necessary to identify all of its binding sites in chromosomes, so that the role of each site in the overall biological function of the factor can be assessed. An approach called ChIP-on-Chip, which combines the chromatin immunoprecipitation technique with chromosomal DNA microarray analysis, has proven to be a powerful means for the chromosome-wide identification of protein binding sites. This approach can also be used to characterize chromosome-wide variations in patterns of post-translational protein modifications, for example histone modifications. This chapter presents methodologies for the ChIP-on-Chip analysis, using as an example the identification of chromosome-wide binding sites for the TATA-binding protein in mitotic cells.

Keywords

Chromatin immunoprecipitation; ChIP-on-Chip; chromosome-wide; TBP; epigenetics; method

1. Introduction

One method for identifying new binding sites of factors for which a consensus recognition sequence has been determined is to perform database searches to identify matches to this sequence in the genome. The chromatin immunoprecipitation technique could then be used to test whether these identified sites are actually bound by the factor. However, this approach is not only inefficient, in that it would likely involve analyzing many sites that turn out not to be actual binding sites for the protein, but would also likely miss many actual binding sites whose sequences are similar to but don't exactly match the test sequence(s) used to perform the database searching. In the ChIP-on-Chip approach, DNA fragments isolated by chromatin immunoprecipitation using an antibody against the protein of interest are used to probe a DNA Chip containing an array of genomic DNA sequences. Thus, this experimental methodology has the advantages of being able a) to identify binding sites of a protein in a way that is not biased with respect to the DNA sequence bound, and b) to simultaneously identify binding sites over very large regions of genomic DNA. This technique can not only provide important new insights into the genomic DNA binding sites of DNA-binding proteins, but also into the variations in patterns of post-translational modifications of proteins such as histones throughout chromosomes (1–8). In this chapter we describe the key experimental details involved in performing ChIP-on-Chip experiments. To illustrate the types of data that can be obtained using these methodologies we present figures showing the results of chromatin immunoprecipitation and ChIP-on-Chip identification of chromosome-wide binding sites of the general transcription factor TATA-binding protein (TBP) in mitotic cells (9).

*To whom correspondence should be addressed: kdsarge@uky.edu, Phone: (859) 323–5777.

2. Materials

2.1. Chromatin Immunoprecipitation Analysis

1. Cells expressing the protein of interest.
2. Tissue culture media and associated reagents appropriate for the cells being used.
3. Formaldehyde (37% stock solution).
4. 1 M Glycine.
5. PMSF (phenylmethylsulfonyl fluoride).
6. NP40 lysis buffer (50 mM Tris (pH 8.0), 85 mM KCl, 0.5% NP40, 1 mM PMSF (added fresh)).
7. Phosphate-Buffered Saline (PBS): 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, and 1.5 mM KH₂PO₄.
8. SDS lysis buffer (50 mM Tris (pH 8.0), 1% SDS, 10 mM EDTA, 1 mM PMSF).
9. ChIP assay kit (Upstate Biotechnology, Lake Placid, NY).
10. Sonicator.
11. ChIP Dilution buffer (20 mM Tris (pH 8.0), 0.01% SDS, 1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl).
12. Protein G-sepharose or protein A-sepharose.
13. Primary antibody capable of immunoprecipitating protein of interest and species-matched non-specific IgG.
14. Low salt wash buffer (20 mM Tris (pH 8.0), 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl).
15. High salt wash buffer (20 mM Tris (pH 8.0), 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl).
16. LiCl wash buffer (10 mM Tris (pH 8.0), 250 mM LiCl, 1% NP40, 1% deoxycholic acid, 1 mM EDTA).
17. ChIP Wash buffer (100 mM Tris (pH 8.0), 500 mM LiCl, 1% NP40, 1% deoxycholic acid).
18. 1 × TE (10 mM Tris-HCl (pH 8.0), 1 mM EDTA).
19. ChIP elution buffer (1% SDS, 100 mM NaHCO₃).
20. DNA purification column designed for DNA fragments (e.g. Qiagen PCR kit).
21. Quantitative PCR machine, or regular PCR machine and native gel electrophoresis equipment and solutions.

2.2. Probing Genomic DNA Chips with fluorescent-labeled DNAs from ChIP assay

1. Random primers for linear PCR amplification.
2. DNA biotin labeling kit.
3. Human cot1 DNA
4. Access to Microarray hybridization incubators and detection instruments (e.g. Affymetrix) or a Microarray Facility.

5. Software to analyze the resulting data (e.g. GeneChip Operating Software (GCOS) and TAS software (Affymetrix))

3. Methods

3.1. Chromatin Immunoprecipitation Analysis

In this first part of the ChIP-on-Chip experiment, cells are subjected to the chromatin immunoprecipitation assay using an antibody against the protein of interest in order to obtain protein-bound DNA fragments that will later be labeled and used to probe the Chip arrays containing genomic DNA sequences in order to identify new binding sites for the protein. In addition, by performing PCR analysis of the resulting DNA fragments obtained from the samples isolated by the antibodies against the protein vs. non-specific control antibodies using primers that amplify a known binding site for the protein as well as a site which is known not to bind the factor, one can also verify the specificity of the ChIP part of the procedure. This is important to prevent wasting substantial time and effort that would result from using sub-optimal DNA fragment sets to probe the DNA Chip arrays in the second part of the ChIP-on-Chip experiment. Figure 1 shows an example of such PCR analysis, in this case quantitative PCR analysis, to show the specific binding of TBP to the histone H4 promoter, a known binding site for this protein, in mitotic cells.

1. Grow tissue culture cells are grown in media appropriate for that cell line. Cells capable of growing in suspension, such as human Jurkat cells, are often chosen for the reasons described in Note 1, but adherent cells have also been used successfully.
2. Cross-link exponentially growing cells (at least 10×10^6 cells per sample) by adding formaldehyde directly to the tissue culture medium to a final concentration of 1%.
3. After incubation for 10 minutes at 25°C with gentle rotation, quench (stop) crosslinking by adding glycine to a final concentration of 125 mM.
4. Pellet cross-linked cells by centrifuging at 3,000g for 5 minutes.
5. Wash cross-linked cells 3 times with cold PBS, each time collecting cells by centrifuging at 3,000g for 5 minutes.
6. Incubate washed cross-linked cells in NP40 lysis buffer (50 mM Tris (pH 8.0), 85 mM KCl, 0.5% NP40, 1 mM PMSF; 0.2 ml/ 1×10^6 cells) for 10 minutes on ice.
7. Pellet insoluble material containing cross-linked protein-DNA by centrifugation at 6,000 rpm for 10 minutes at 4°C.
8. Incubate pellets in SDS lysis buffer (50mM Tris (pH 8.0), 1% SDS, 10 mM EDTA, 1 mM PMSF; 0.2 ml/ 1×10^6 cells) for 10 minutes on ice and solubilize by sonication to an average DNA length of 500–1000bp (see Note 2).
9. Dilute sonicated chromatin using ChIP Dilution buffer (16.7 mM Tris (pH 8.0), 0.01% SDS, 1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl).
10. Pre-clear the soluble chromatin by adding 80 µl of a 50% slurry of salmon sperm DNA saturated-protein A agarose beads and 10 µl non-specific (control) rabbit IgG and incubating for 1 h at 4°C with rotation.

¹Non-adherent cells are often chosen because a) it is easier to grow the large numbers of cells required for these types of experiments, for example using spinner flasks, compared to adherent cells, and b) it is likely that cross-linking is more efficient in such cells because more of their surface is exposed to the medium.

²The sonication conditions used, such as the energy and duration, are dependent on the sonicator employed and must be empirically tested to find the conditions that yield fragments in this size range as observed on polyacrylamide gels stained with ethidium bromide.

11. Centrifuge at 3,000g for 1 minute.
12. Take approximately 30µl of the supernatant (pre-cleared chromatin) to be used later as the “Input” DNA sample.
13. To half of the rest of the supernatant (precleared chromatin) add the primary antibody against the protein of interest, and to the other half add species-matched non-specific control IgG, and incubate with rotation at 4°C for 16 hours.
14. To these tubes add 60 µl salmon sperm DNA saturated-protein A agarose beads and incubate for 1 hour at 4°C.
15. Centrifuge at 3,000g for 1 minute to collect beads.
16. Wash beads once for 5 minutes at RT with low salt wash buffer (20 mM Tris pH 8.0, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl).
17. Wash beads once for 5 minutes at RT with high salt wash buffer (20 mM Tris pH 8.0, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl).
18. Wash beads once for 5 minutes at RT with LiCl wash buffer (10 mM Tris (pH 8.0), 250 mM LiCl, 1% NP40, 1% deoxycholic acid, 1 mM EDTA).
19. Wash beads three times for 5 minutes at RT with ChIP Wash buffer (100 mM Tris (pH 8.0), 500 mM LiCl, 1% NP40, 1% deoxycholic acid).
20. Wash beads twice for 5 minutes at RT with TE.
21. Elute immune complexes by doing two successive incubations of the beads with 75µl ChIP elution buffer (1% SDS, 100 mM NaHCO₃), and then pool elutions. In addition, the Input DNA sample collected in step 12 above should be made 1% in SDS and 100 mM in NaHCO₃ using stock solutions of each.
22. Reverse cross-links by adding NaCl to a concentration of 200 mM, and then heating at 67°C for at least 5 hours.
23. Concentrate and purify immunopurified and Input DNAs using a column designed for DNA fragments (such as Qiagen PCR purification kit). DNA is eluted in 30 µl 10 mM Tris-HCl (pH 8.5).
24. To test the specificity of the DNA fragments obtained from the chromatin IP, subject aliquots of 1 µl of the immunoprecipitated DNAs as well as input samples obtained prior to immunoprecipitation by PCR, using primers that amplify a known binding region and primers that amplify a region where the factor is known not to bind. Quantitative PCR analysis is preferred, but if not available then PCR products can be separated by agarose gel electrophoresis and visualized by ethidium bromide staining.

3.2 Probing Genomic DNA Chips with fluorescent-labeled DNAs from ChIP assay

If the results of the analysis described at the end of part 3.1 above shows the specificity of the ChIP DNAs, these DNA fragments can then be used to probe the DNA Chip arrays containing the genomic sequences. In the first part of this second half of the ChIP-on-Chip, the DNA fragments isolated by immunoprecipitation using antibodies against the protein of interest, and the control Input DNA sample, is amplified by PCR and then labeled with biotin. These labeled DNAs are then incubated with the Chips containing the genomic DNA arrays and the DNA spots that hybridize with the probes identified using specialized instruments designed for these purpose (e.g. Affymetrix equipment). The binding sites for the factor of interest are then determined by analyzing the data using computer software. Figures 2, 3, and 4 illustrate the types of data output that can result from ChIP-on-Chip studies, in this case from our studies

that characterized the binding sites of the TBP protein within chromosomal DNA of mitotic cells(9).

1. Subject 15–30 µl of the immunoprecipitated DNA and Input DNA samples to the first round of random primer PCR to amplify the DNA fragments (see Note 3).
2. Subject the DNA from the first round of amplification to a second round of amplification using a different primer (see Note 4).
3. Label the amplified DNAs with biotin following the manufacturer's protocol (e.g. Affymetrix).
4. Hybridize 10 µg of the labeled DNAs to the genomic DNA Chip arrays probes chosen for analysis (See Note 5) in the presence 15 µg human cot1 DNA, using a specialized incubator and following the manufacturer's instructions (Note 6). This step is typically done by a Microarray Facility, as is the following step.
5. Chips are washed, stained with fluorescent streptavidin-phycoerythrin, and scanned using specialized instruments (e.g. Affymetrix) are typically done in a dedicated Microarray Facility that has the required specialized machines.
6. Identify binding sites for the factor of interest by analyzing the resulting data using software from the manufacturer (see Note 7).

Acknowledgments

We are very grateful to David Rodgers for allowing us to use computers in his lab to analyze the data from our ChIP-on-Chip experiments.

References

1. Dietz SC, Carroll JS. Interrogating the genome to understand oestrogen-receptor-mediated transcription. *Expert Rev Mol Med* 2008;10:e10. [PubMed: 18377699]

³In the Affymetrix protocol we followed, the initial round of linear amplification was carried out using random primer A (5'-GTTTCCCAGTCACGGTC(N)₉-3'), and utilized 4 cycles of: 95°C, 4 minutes, 10°C, 5 minutes, ramp from 10°C to 37°C over 9 minutes, followed by 8 minutes at 37°C, using sequenase.

⁴In the second round of amplification, in the Affymetrix protocol the DNA from the first round of amplification is then amplified using primer B (5'-GTTTCCCAGTCACGGTC-3') by Taq polymerase PCR using 32 cycles of: 95°C, 30s, 45°C, 30s, 55°C, 30s, 72°C, 1 minute.

⁵For our experiments we used the Affymetrix Genechip Human tiling array 2.0 c, which contains genomic DNA spanning human chromosomes 3, 21, 22, X, and Y, including intergenic regions, promoters, untranslated regions, exons, introns etc. However, DNA Chips are also available in which only known promoter regions are included, which may be desirable if you are studying a transcription factor and your goal is just to determine which promoters are bound by the factor of interest. The benefit of such a selective genomic DNA Chip is that one Chip can contain all the sequences one is interested in examining, whereas you would have to buy and hybridize many different Chips if you want to examine all of the genomic sequence. Of course, the disadvantage of using a selective genomic DNA sequence Chip is that it would miss any binding sites for your factor that are not in the promoter sequences selected for inclusion on the Chip (e.g. not-yet-identified promoters, control DNA sequences that are within introns, or far upstream of proximal promoter regions, etc.). To have confidence in the results of the probing, it is suggested that 3 of each Chip be probed in each experiment (i.e. probing done in triplicate).

⁶The steps of the procedure involving Chip hybridization, staining with fluorescent streptavidin-phycoerythrin, and scanning are typically done in a dedicated Microarray Facility that has the required specialized machines. For example, in our study the University of Kentucky Microarray Facility performed these steps using equipment including an Affymetrix GeneChip hybridization incubator 640, Fluidics Station 450, and GeneChip Scanner 3000 7G.

⁷In our study data were collected and analyzed using GeneChip Operating Software (GCOS) and TAS software (Affymetrix). The results (from three TBP IP DNA-probed chips and three genomic input DNA-probed chips) were quantile-normalized within treatment/control replicate groups. A Wilcoxon Rank Sum test was applied to the transformation log₂ for data, testing the null hypothesis of the equality of the two population distribution functions against the alternative of a positive difference on location between the probability distribution of the treatment and that of the control. The Wilcoxon test was applied in a sliding window across the chromosomes. The chromosomal positions bound by TBP were identified based on a *P* value cut off of 0.005. Results were visualized using IGB software (Affymetrix) and the UCSC GenomeViewer, and compared to the RefSeq database (NCBI).

2. Trelle MB, Jensen ON. Functional proteomics in histone research and epigenetics. *Expert Rev Proteomics* 2007;4:491–503. [PubMed: 17705707]
3. Wade JT, Struhl K, Busby SJ, Grainger DC. Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol Microbiol* 2007;65:21–26. [PubMed: 17581117]
4. Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* 2007;8:424–436. [PubMed: 17486122]
5. Elnitski L, Jin VX, Farnham PJ, Jones SJ. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. *Genome Res* 2006;16:1455–1464. [PubMed: 17053094]
6. Wu J, Smith LT, Plass C, Huang TH. ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res* 2006;66:6899–6902. [PubMed: 16849531]
7. DeAngelis JT, Farrington WJ, Tollefsbol TO. An overview of epigenetic assays. *Mol Biotechnol* 2008;38:179–183. [PubMed: 17943463]
8. Callinan PA, Feinberg AP. The emerging science of epigenomics. *Hum Mol Genet* 2006;15(Spec No 1):R95–101. [PubMed: 16651376]
9. Xing H, Vanderford NL, Sarge KD. The TBP-PP2A mitotic complex bookmarks genes by preventing condensin action. *Nat Cell Biol* 2008;10:1318–1323. [PubMed: 18931662]
10. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007;316:1484–1488. [PubMed: 17510325]

Fig. 1.

Analysis of specificity of ChIP TBP promoter binding in mitotic cells. Chromatin immunoprecipitation (ChIP) assay was performed on mitotic (nocodazole-blocked) Jurkat cells using TBP antibodies or control IgG antibodies, followed by quantitative PCR assay using primer pairs specific to the H4 and β -globin gene promoters. Data are means $\pm$ s.d. (n=3, * P <0.0002). This figure is from Xing et al., 2008 (9).

Fig. 2.

TBP is bound to many chromosomal loci during mitosis. DNA fragments immunoprecipitated by TBP antibodies by ChIP assay using mitotic cells, with input genomic DNA as control, were hybridized to the Affymetrix Genechip human tiling array 2.0c, which includes genomic DNA arrays spanning human chromosomes 3, 21, 22, and X. Shown at the top of each panel are the locations of TBP binding in regions representing 10 million base-pair segments of the indicated chromosome. In the lower part of each panel are close-up views of TBP binding to a gene within that region, indicating the association of TBP with the promoter regions of each gene. The genes are CYRR1 (panel a), DRG1 (panel b), TOMM70A (panel c), KLHL6 (panel d), FMR1NB (panel e), and OTUD5 (panel f). This figure is from Xing et al., 2008 (9).

Fig. 3. Correlation between TBP binding and Transcriptome. The top parts of panels a and b show the data for TBP binding to a 56 million bp region of chromosome 3 (vertical lines represent locations of TBP binding), along with a depiction of the RefSeq genes and Transcriptome data (10) for this region. In the bottom parts of panels a and b are shown the same information for two different sub-portions of this 56 million bp region, one that is approximately 9 million bp long and the other 6 million bp in length. This figure is from the Supplementary Data of Xing et al., 2008 (9).

Fig. 4.

Lack of strong mitotic TBP binding in a chromosomal region containing pseudogenes, genes expressed in a tissue-restricted manner, and large regions lacking annotated genes. Panel a shows the data for TBP binding to the KLHL6 gene from Figure 5d, but compressed so that its size is displayed relative to the region shown in panel b. The asterisk indicates the TBP peak associated with this gene's promoter region, and the dotted line relates the height of this TBP signal to the TBP signals in panel b. In panel b is shown a region of chromosome 21 that contains a number of pseudogenes, genes whose expression is tissue-restricted (C21orf99: testis and muscle; ANKRD21: testis; LOC441956: mammary gland), as well as regions that are devoid of any annotated genes. The position of each pseudogene is indicated by the vertical bar next to its identifier. The sizes of the two regions in panels a and b are displayed on the same relative scale, as are the magnitudes of the TBP binding signals. This figure is from the Supplementary Data of Xing et al., 2008 (9).