ChIP-chip Comes of Age for Genome-wide Functional Analysis

Jiejun Wu,[1,2] Laura T. Smith,[2] Christoph Plass,[1,2] and Tim H-M. Huang[2]

Departments of [1]Molecular Genetics and [2]Molecular Virology, Immunology, and Medical Genetics, Division of Human Cancer Genetics, Comprehensive Cancer Center, Ohio State University, Columbus, Ohio

Abstract

In the post-genome era, attention has focused on the functions of genome sequences and how they are regulated. The emerging epigenomic changes and the interactions between *cis*-acting elements and protein factors may play a central role in gene regulation. To understand the crosstalk between DNA and protein on a genome-wide scale, one emerging technique, called ChIP-chip, takes the strategy of combining chromatin immunoprecipitation with microarray. This new high-throughput strategy helps screen the targets of critical transcription factors and profile the genome-wide distribution of histone modifications, which will enable the feasibility of conducting a large-scale study, such as the Human Epigenome Project. (Cancer Res 2006; 66(14): 6899-902)

Introduction

The completion of the Human Genome Project provides a road map for thorough interrogation of gene functions. This new endeavor has also made possible further technological advances. In addition to identifying novel transcription factor targets, current studies may shift our attention to genome-wide characterization of histone modifications and DNA methylation. The importance of this type of study is further echoed by a recent proposal for the Human Epigenome Project (1). This requires genome-wide technologies with high-throughput capability. One of these techniques, combining chromatin immunoprecipitation (ChIP) with microarray (ChIP-chip), has been used extensively as a genome-wide tool to screen the binding position of protein factors (2). Because the use of ChIP-chip is beginning to gain popularity in the research community, the present review provides a timely overview of recent developments and applications.

The Strategy and Development of ChIP-chip

ChIP-chip combines ChIP and microarray techniques (Fig. 1*A*; refs. 3–5). Briefly, DNA and protein are cross-linked *in vivo* with formaldehyde and immunoprecipitated with specific antibodies against the protein of interest. DNA bound by protein is sheared by sonication to fragments of ∼0.2 to 2 kb. It is necessary to apply consistent variables to DNA sonication so that the sizes of sheared DNA are of similar range. The pulled-down DNA and appropriate controls are fluorescently labeled and applied to microscopic slides for microarray analysis. With input DNA as background, the comparison of immunoprecipitated DNA with the background control helps to profile the binding position of specific proteins in the genome. Ren et al. (3) identified novel targets of the yeast transcription factors Gal4 and Ste12 first with ChIP-chip on a yeast intergenic DNA array. A subsequent study profiled the binding sites of the cell-cycle transcription factor SBF and MBF in yeast (4). Compared with the PCR amplification of specific target sequences from immunoprecipitated material, ChIP-chip is a genome-wide "reverse-genetic" approach. Another advantage of ChIP-chip is that it targets those genes directly bound by the protein factor. The classic expression arrays cannot distinguish between directly regulated genes and those changed secondarily.

The extent of ChIP-chip application depends, in part, on the development of DNA microarray technology, especially the availability of arrayed slides for these organisms. In humans, one of the first ChIP-chip experiments adopted was the use of a CpG island array for screening novel E2F4 targets (6). Other arrays were then developed for ChIP-chip including selected promoter array and DNA tiling arrays (2). The advantages and limitations of these array have been discussed elsewhere (2). Recently, human genome-wide array with PCR amplicons printed has further been improved to cover >90% of human nonrepetitive DNA sequences (7). Unlike a PCR amplicon–based strategy, a new oligonucleotide array technique helped build another human array representing all human nonrepetitive regions with resolution at 100 bp (8). A microarray, currently not available, containing the entire human genome sequence would be ideal for ChIP-chip.

Critical for ChIP-chip experiments is the amount of starting material required for successful microarray hybridization. The number is highly variable depending on the quality of the antibody, binding frequency of protein to DNA, and other possible unknown factors. In one previous report, 1 to 10 ng of ChIP DNA were pulled down from a total of 30 to 60 μg of *Drosophila* DNA (9). Up to 50 individual ChIP DNA samples have been pooled to acquire enough DNA for array hybridization (5). An alternative approach to enrich the starting material is through PCR-based amplification (2–4). A RNA polymerase–based method is known to linearly amplify ChIP DNA (10). However, a PCR bias may occur, especially with mammalian systems, where large amounts of repeat sequences may skew data.

Another concern is background DNA that is pulled down by nonspecific interactions of protein and DNA. In one typical ChIP-chip experiment based on pull-down by Suz12, >50% of the targets, with 3-fold enrichment compared with control group, were false-positives (11). Recently, optimization of ChIP-chip decreased the false positive rate to <1% and false negative rate to 20% to 25% (12). Attention should be paid to several key basics for ChIP-chip, such as antibody quality, immunoprecipitation handling, optimization of array hybridization conditions, and data normalization and analysis. It is also necessary to establish appropriate controls. Generally, genomic DNA is used as an input control and samples from no antibody or immunoglobulin G groups are used as negative controls. Other control designs, such as transformed cell lines versus empty vector cell lines, wild-type target versus
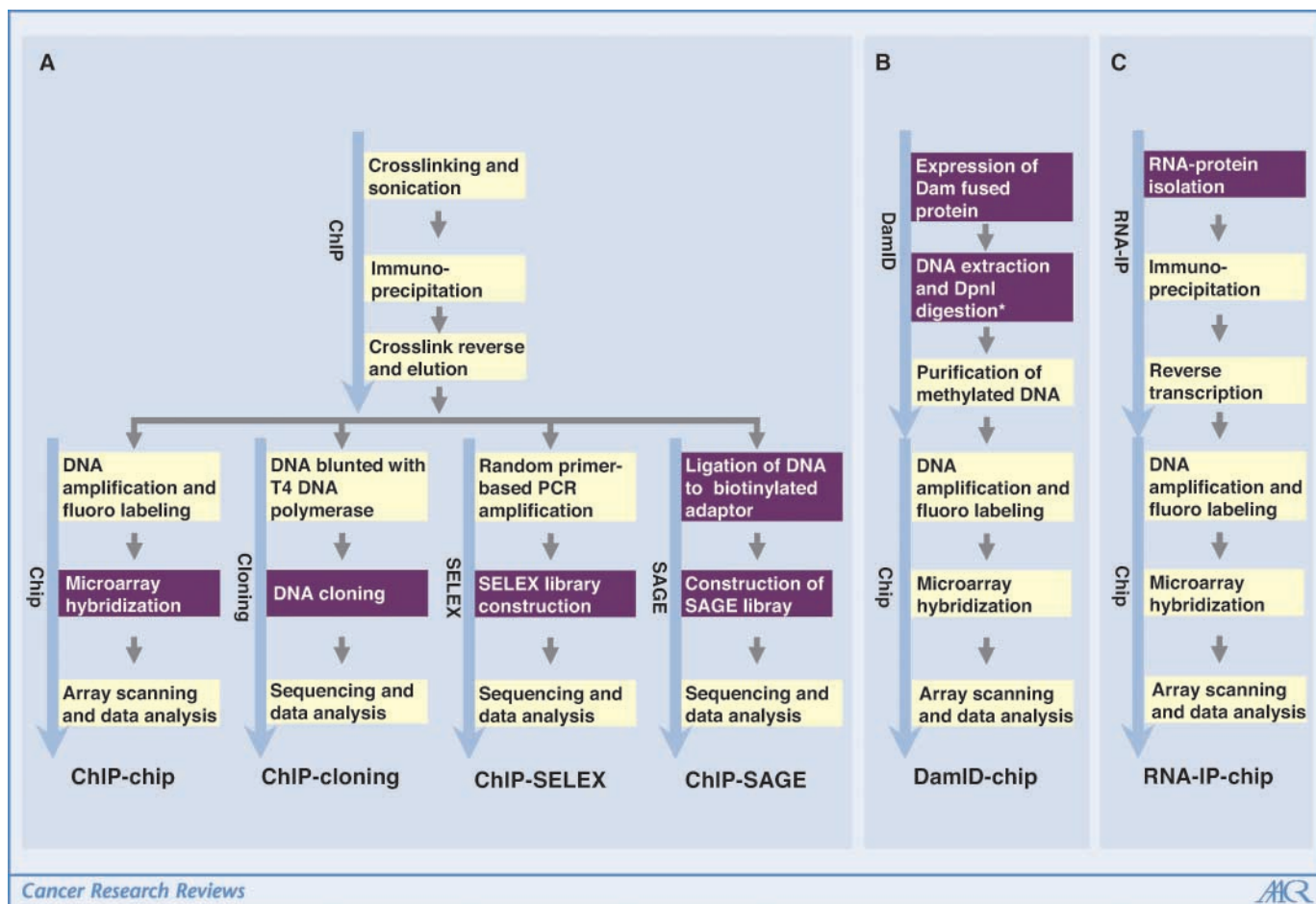
Figure 1. Brief comparison of the procedures of ChIP-chip with other high-throughput techniques studying DNA-protein or RNA-protein interactions. The procedures of (*A-C*) are general principles of ChIP-chip and others. Key steps are highlighted with red text box. *Dpn*I (*star*) is a methylation-sensitive restriction enzyme.

mutation target, and with versus without drug treatment, can all be considered as control options.

ChIP-chip Applications in Genome-wide Functional Analysis

The ChIP-chip studies discussed above indicate two strategies using ChIP-chip (3, 4). One is to screen and identify binding targets of protein factor without prior knowledge. Specific targets are then confirmed by ChIP-PCR. The other is to map protein binding location to provide a genome-wide binding profile, which may require more accurate raw data for statistical analysis because the confirmation of the top list of enriched targets by ChIP-chip cannot guarantee that the final map reflects the real spectrum of target distributions. Two major areas of research can be studied using ChIP-chip: first, the identification of sequences with specific histone modifications and, second, the identification of binding targets for nuclear protein factors.

Two strategies with ChIP-chip are used in histone modification studies. One is to detect the distribution of histone modifications using antibodies specially targeting these modifications (13). The other is using ChIP-chip to locate not the modified histones but the enzymes that catalyze the histone modification reactions (14). In yeast, when mutations are introduced to a specific histone modifying enzyme, the changes

of histone modifications can identify targets (13, 14). This kind of study can also take advantage of combining ChIP-chip with other high-throughput strategies, especially expression profiling, to establish the correlation of histone markers with transcription activity (13, 15). In humans, three different arrays have been applied in studying histone modifications with ChIP-chip. One is CpG island array that showed strong correlation between CpG methylation and histone modifications (16). The second is the cDNA array, which provided new information on the distribution of histone methylation patterns in the coding regions of human genes (17). Finally, the tiling array mapped H3 markers (dimethyl-K4 and trimethyl-K4, acetyl-H3K9, and acetyl-H3K14) to nonrepetitive regions of human chromosomes 21 and 22 (18). In these studies, ChIP-chip provided a wider view than gene-by-gene studies.

Conventional methods to identify DNA binding elements, such as footprinting and gel shift assays, can be laborious. Various algorithms, based on comparative "phylogenetic footprinting" or coregulation clustering, identify regulatory elements *in silico* (19). It is significant that directly and *in vivo*, the genome-wide binding data from ChIP-chip can be used for bioinformatic analysis to identify binding elements (20). One can further combine the data from ChIP-chip, expression profiling, comparative genomic, and published literature to reduce false findings and to identify true interactions (21). The information on binding

elements identified by ChIP-chip leads to the discovery of modules that can be used to build a global view of the regulation network (12). In ChIP-chip, formaldehyde can cross link not only interacting protein-DNA but also protein-protein. To avoid the interference from indirect DNA protein binding caused by formaldehyde, two alternative methods, DNA immunoprecipitation with microarray detection and protein binding microarrays, may be considered (22, 23).

ChIP-chip has also been extended or modified for other purposes. Some custom arrays can be applied to check the binding position of regulator or histone modifications pertaining to a single gene of interest (11). To study gene expression, DNA methylation, and histone acetylation in parallel, a progeny array panel, called expressed CpG island sequence tags, was designed from CpG island arrays (24). Another technique used antibodies directed against 5-methylcytidine for immunoprecipitation and subsequent hybridization onto arrays for DNA methylation profiling (25). ChIP-chip with antibodies against methyl-CpG binding proteins can also be applied to identify the methylated targets (26).

Other High-Throughput Methods Beyond ChIP-chip

Because the ideal array covering all human chromosomes is not available, alternative options have been used to address this limitation (Fig. 1*A*). One way is to digest genomic DNA with restriction enzymes creating blunt ends and to clone the DNA fragments precipitated by ChIP into a plasmid vector (27). One drawback of this cloning strategy is a high ratio of false-positive targets and high labor and expenditure cost. Another possible way is to establish a SELEX (systematic evolution of ligands by exponential enrichment) genomic library for sequencing, although similar questions of nonspecificity remain (Fig. 1*A*; ref. 28). A combination of ChIP and arbitrarily-primed PCR, called ChAP, is similar to ChIP-SELEX and uses DNA sequencing to identify protein-binding sites (29). Additionally, combining ChIP to modified serial analysis of gene expression, which can be termed ChIP-SAGE, can be used, although the requirement for extensive DNA sequencing makes it less convenient than ChIP-chip (Fig. 1*A*; ref. 30).

Other similar high-throughput techniques include DNA adenine methyltransferase identification (DamID) coupled with microarray and RNA-IP-chip, combining RNA immunoprecipitation with microarray (Fig. 1*B* and *C*). DamID uses the DNA methyltrasferase to "mark" the positions of DNA-protein interactions, which does not require antibody as ChIP-chip (31). However, DamID-chip has limitations when compared to ChIP-chip. First, it requires more time to express the fusion protein than formaldehyde cross linking of endogenous protein. Second, DamID cannot be applied for detecting posttranslational changes, such as histone modifications. Third, the resolution of DamID is less precise than that of ChIP-chip although both methods can produce similar binding maps (32). RNA-IP-Chip is applied to understand RNA-protein interactions (Fig. 1*C*; ref. 33).

Perspective on Future Development of ChIP-chip

Generally, the ChIP-chip protocol can be completed in three steps: ChIP, post-ChIP DNA handling, and microarray analysis. Some modifications can be made. For example, UV and other covalent trapping of protein-DNA complexes, such as the covalent binding of DNA methyltransferase and 5-aza-2′-deoxycytidine, may

be modified to provide another marker for *in vivo* mapping (34, 35). Unbiased amplification of ChIP DNA would be beneficial to data quality from ChIP-chip. Efforts to amplify genomic DNA homogeneously include carefully designed amplification processes, a chemical additive, and application of more efficient enzymes, such as $\Phi29$ DNA polymerase (36).

Currently, data analysis used for ChIP-chip analysis is largely adopted from expression microarray studies and, as such, similar assumptions are made, which may not always be reliable. For instance, the DNA pulled down, such as those against universal histone modifications, may be significantly changed so the assumption that the total signals from the test samples and the control should be equal is not valid, as is normal for data normalization of expression microarray. Currently, attention is focused on creating algorithms to use the ChIP-chip data; to combine ChIP-chip with other platforms; and to finally establish some predictable models of transcription regulation. In addition, the basic work of design, application, and analysis of ChIP-chip itself may be equally important to improve the reliability of these models.

The combination of ChIP-chip with other high-throughput methods may be beneficial. This entails both combining multiple ChIP-chips and incorporating ChIP-chip with other high-throughput methods, such as expression array. Recently, the yeast two-hybrid has been applied to create the network of protein-protein interactions, and this proteome map can be connected to the transcriptome (37). Theoretically, the information of protein-DNA crosstalk from ChIP-chip, combined with that from protein-protein crosstalk, can provide input signals to reconstruct the regulation network. The expression array may indicate the output signals from this network. It is promising that, along with the appropriate computational strategy, emerging high-throughput methods from different platforms can work cooperatively to produce a clearer picture of the regulation network and disease-related changes.

Conclusion on ChIP-chip

New biological questions continue to drive the development of new techniques. DNA sequencing, expression microarray, proteomic two-dimensional electrophoresis gel, and other system tools have helped to understand the structure and amount of cell components. New genome-wide, high-throughput tools, such as ChIP-chip, are necessary to study the activities of key components, like epigenetic modifications and DNA-protein binding in cells. ChIP-chip has been frequently used in basic biological studies and may be modified further and expanded to other aspects, such as human diseases. Lastly, the large amount of discoveries by ChIP-chip, and other high-throughput techniques, may be connected with emerging bioinformatics to add to our knowledge of life and diseases.

References

1. Jones PA, Martienssen R. A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop. Cancer Res 2005;65:11241–6.

2. Buck MJ, Lieb JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics 2004;83:349–60.

3. Ren B, Robert F, Wyrick JJ, et al. Genome-wide location and function of DNA binding proteins. Science 2000;290: 2306–9.

4. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 2001;409: 533–8.

5. Oberley MJ, Tsao J, Yau P, Farnham PJ. High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays. Methods Enzymol 2004; 376:315–34.

6. Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev 2002;16: 235–44.

7. Kim TH, Barrera LO, Qu C, et al. Direct isolation and identification of promoters in the human genome. Genome Res 2005;15:830–9.

8. Kim TH, Barrera LO, Zheng M, et al. A high-resolution map of active promoters in the human genome. Nature 2005;436:876–80.

9. Orlando V, Strutt H, Paro R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. Methods 1997;11:205–14.

10. Liu CL, Schreiber SL, Bernstein BE. Development and validation of a T7 based linear amplification for genomic DNA. BMC Genomics 2003;4:19.

11. Kirmizis A, Bartley SM, Kuzmichev A, et al. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev 2004;18: 1592–605.

12. Boyer LA, Lee TI, Cole MF, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 2005;122:947–56.

13. Bernstein BE, Humphrey EL, Erlich RL, et al. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A 2002;99:8695–700.

14. Reid JL, Iyer VR, Brown PO, Struhl K. Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Mol Cell 2000;6:1297–307.

15. Kurdistani SK, Tavazoie S, Grunstein M. Mapping global histone acetylation patterns to gene expression. Cell 2004;117:721–33.

16. Kondo Y, Shen L, Yan PS, Huang TH, Issa JP. Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. Proc Natl Acad Sci U S A 2004;101: 7398–403.

17. Miao F, Natarajan R. Mapping global histone methylation patterns in the coding regions of human genes. Mol Cell Biol 2005;25:4650–61.

18. Bernstein BE, Kamal M, Lindblad-Toh K, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 2005;120: 169–81.

19. Pennacchio LA, Rubin EM. Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet 2001;2:100–9.

20. Liu XS, Brutlag DL, Liu JS. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol 2002;20:835–9.

21. Harbison CT, Gordon DB, Lee TI, et al. Transcriptional regulatory code of a eukaryotic genome. Nature 2004;431:99–104.

22. Liu X, Noll DM, Lieb JD, Clarke ND. DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome Res 2005;15:421–7.

23. Mukherjee S, Berger MF, Jona G, et al. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nat Genet 2004;36:1331–9.

24. Shi H, Wei SH, Leu YW, et al. Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. Cancer Res 2003;63:2164–71.

25. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 2005;37:853–62.

26. Ballestar E, Paz MF, Valle L, et al. Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. EMBO J 2003;22:6335–45.

27. Weinmann AS, Bartley SM, Zhang T, Zhang MQ, Farnham PJ. Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol 2001;21: 6820–32.

28. Labhart P, Karmakar S, Salicru EM, et al. Identification of target genes in breast cancer cells directly regulated by the SRC-3/AIB1 coactivator. Proc Natl Acad Sci U S A 2005;102:1339–44.

29. Liang G, Lin JC, Wei V, et al. Distinct localization of histone H3 acetylation and H3-4 methylation to the transcription start sites in the human genome. Proc Natl Acad Sci U S A 2004;101:7357–62.

30. Colvis CM, Pollock JD, Goodman RH, et al. Epigenetic mechanisms and gene networks in the nervous system. J Neurosci 2005;25:10379–89.

31. van Steensel B, Henikoff S. Epigenomic profiling using microarrays. Biotechniques 2003;35:346–50, 52–4, 56–7.

32. Holland S, Ioannou D, Haines S, Brown WR. Comparison of Dam tagging and chromatin immuno-precipitation as tools for the identification of the binding sites for *S. pombe* CENP-C. Chromosome Res 2005;13:73–83.

33. Iyer VR. Exploring the post-transcriptional RNA world with DNA microarrays. Trends Biotechnol 2004; 22:498–500.

34. Zhang L, Zhang K, Prandl R, Schoffl F. Detecting DNA-binding of proteins *in vivo* by UV-crosslinking and immunoprecipitation. Biochem Biophys Res Commun 2004;322:705–11.

35. Liu K, Wang YF, Cantemir C, Muller MT. Endogenous assays of DNA methyltransferases: evidence for differential activities of DNMT1, DNMT2, and DNMT3 in mammalian cells *in vivo*. Mol Cell Biol 2003;23: 2709–19.

36. Dean FB, Hosono S, Fang L, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 2002;99: 5261–6.

37. Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature 2005;437:1173–8.

ChIP-chip Comes of Age for Genome-wide Functional Analysis

Jiejun Wu, Laura T. Smith, Christoph Plass, et al.

Updated version	Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/66/14/6899

Cited articles	This article cites 37 articles, 17 of which you can access for free at: http://cancerres.aacrjournals.org/content/66/14/6899.full#ref-list-1
Citing articles	This article has been cited by 13 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/66/14/6899.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/66/14/6899. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.