International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2017; 5(5): 2039-2044

Received: 27-07-2017
Accepted: 28-08-2017

Sidharth Prasad Mishra
Ph.D. Scholar, Department of Animal Genetics & Breeding, WBUAFS, Kolkata, West Bengal, India

Chinmoy Mishra
Assistant Professor, Department of Animal Breeding & Genetics, OUAT, Bhubaneswar, Odisha, India

Siddhant Sekhar Sahoo
Ph.D. Scholar, Department of Animal Genetics & Breeding, NDRI, Karnal, India

Kamdev Sethy
Assistant Professor, Department of Animal Nutrition, OUAT, Bhubaneswar, Odisha, India

Kumaresh Behera
Assistant Professor, Department of Livestock Management and Production, OUAT, Bhubaneswar, Odisha, India

Correspondence
Sidharth Prasad Mishra
Ph.D. Scholar, Department of Animal Genetics & Breeding, WBUAFS, Kolkata, West Bengal, India

Chromatin Immunoprecipitation (ChIP) Assay - A Tool for Functional Analysis of Gene Expression

Sidharth Prasad Mishra, Chinmoy Mishra, Siddhant Sekhar Sahoo, Kamdev Sethy and Kumaresh Behera

Abstract
The Whole Genome Sequencing Project developed enormous amount of data about the genetic architecture of organisms. But mere structural characterization of genome will ultimately not be the concluding stage; a functional characterization is essentially required for the complete understanding of genome biology. The temporal and spatial regulation of gene expression though occurs at several levels of central dogma process but regulation at transcriptional level is most important in vertebrates. Identification of transcription factors controlling gene expressions is becoming utmost important for understanding the regulation of gene expression. The Chromatin Immunoprecipiation (ChIP) assay provides a powerful technique for identifying the *in vivo* association of transcription factors with regulatory elements of a gene. ChIP allows enrichment of genomic regions which are associated with specific transcription factors, histone modifications, and indeed any other epitopes which are present on chromatin. The original ChIP methods used site-specific PCR and Southern blotting to confirm which regions of the genome were enriched, on a candidate basis. The combination of ChIP with genomic tiling arrays (ChIP-chip) and next generation sequencing now allowed a more unbiased approach to map ChIP-enriched sites.

Keywords: ChIP, ChIP chip, ChIP seq, ChIP exo, Gene expression, Immunoprecipitation.

Introduction
Gene expression is a very complex process, regulated at different stages (i.e. transcription, translation and post-translational stage). Transcriptional stage of regulation of gene expression is principally important in vertebrates which is controlled by different proteins (transcription factors) binding to specific regions of genes (Rhee and Pugh, 2011) [39]. The incessant approach to spot the specific binding sites of different transcription factors in a genome is always going on. Some recent techniques, such as SAGE and DNA microarrays, allow large numbers of possible target genes to be screened at once (Duggan *et al*., 1999) [10]. However, these techniques identify both direct and indirect targets genes. Because these techniques usually require over expression or deletion of the transcription factor of interest, they do not necessarily identify genes that are targets under unperturbed, physiological conditions. *In silico* motif searches are insufficient to identify all *in vivo* binding locations for a protein because putatively bound locations either lack an obvious motif or contain multiple degenerate versions of the motif (Cawley *et al*., 2004; Walter and Biggin, 1994) [7, 48]. Protein-binding microarrays have been a proven method to identify intrinsic specificity of DNA-binding domains of proteins *in vitro*. However, *in vivo*, such specificity may be altered, vetoed, or guarded due to presence of thousands of other proteins (Li *et al*., 2008) [25]. Other methods, such as electrophoretic gel mobility shift assay (EMSA) have been developed to study different aspects of protein-DNA interactions. But such methods do not take into account of the *in vivo* DNA-protein interactions and requires a priori knowledge of identified sequence to which protein binds (Jothi *et al*., 2008) [21]. Additional problem may arise when different proteins bind to the same sequence and some low-occupancy binding sites become available in genomes with significant physiological importance (Li *et al*., 2008) [25].
Chromatin immunoprecipitation (ChIP) is a method designed to analyze the protein–DNA interactions *in vivo* (Solomon and Varshavsky, 1985) [45]. In the ChIP assay, proteins are crosslinked to their DNA-binding sites *in vivo* and then immunopurified from fragmented chromatin. Subsequently, the bound DNA is identified genome-wide by microarray hybridization (ChIP-chip) was explained by Ji *et al*., 2008 [18] or deep sequencing (ChIP seq)

by Albert *et al*., 2008; Johnson *et al*., 2008; Ren *et al*., 2000 [2, 20, 36]. This method has its own limitations. In particular, it requires antibodies of a quality that is sufficient for a specific interaction and at the same time can tolerate stringent binding and washing conditions. Because unbound DNA contaminates the immunoprecipitate, ChIP only provides a set of statistically enriched high-occupancy binding regions, rather than a complete and precise set of bound locations (Zang *et al*., 2009; Rozowsky *et al*., 2009; Tuteja *et al*., 2009) [51, 42, 47]. A sizeable fraction of this DNA may represent false positives (erroneous calls), and many other lower-affinity sites may be missed (false negatives). Moreover, size heterogeneity of randomly sheared ChIP DNA technically limits mapping resolution, and thus cannot distinguish binding among clusters of neighboring sites.

ChIP Assay Preparation

ChIP assay is a popular technique for fine-mapping the location of modified histones, transcription factors, and nonhistone chromosomal proteins (Robyr *et al*., 2002) [40] In step wise manner the assay involves DNA-protein crosslinking, sonication, immunoprecipitation (IP) of the crosslinked DNA-protein complexes, capture of these complexes, DNA recovery from the precipitated product and DNA analysis by polymerase chain reaction (PCR), quantitative real-time PCR (Q-PCR), hybridization on microarrays (ChIP-ChIP), or direct sequencing (ChIP-seq).

Crosslinking of DNA-Protein

DNA-bound proteins can be crosslinked to DNA by treatments with UV light, dimethyl sulfate or a variety of other agents (Gilmour and Lis, 1984) [14]. But the use of these treatments damage DNA extensively upon prolonged exposure. On the other hand, formaldehyde produces DNA-protein (Dedon *et al*., 1991) [9] as well as RNA-protein (Niranjanakumari *et al*., 2002) [30], protein-protein (Kurdistani *et al*., 2003) [24] crosslink both *in vitro* and *in vivo* (Yamanaka *et al*., 1993) [49] preventing any large-scale redistribution of cellular components upon prolonged fixation and at the same time displays virtually no reactivity toward free double-stranded DNA (McGhee and Hippel, 1975; Trifonov *et al*., 1967) [27]. Formaldehyde reacts simultaneously with arginine and lysine side chains of proteins and purine and pyrimidine moieties of DNA. Formaldehyde-induced crosslinks, in particular DNA-protein crosslinks, can be reversed under relatively mild conditions (Jackson and Chalkley, 1981) [17]. DNA is typically digested with micrococcal nuclease before isolating protein bound DNA fragments that are then sequenced. Sonication is an alternative method to it (Orlando, 2000) [32].

Sonication

Sonication is the act of applying sound energy (usually ultrasound energy) to stir up particles in a sample through a sonicator. During sonication, DNA samples are subjected to shearing of size less than 300-500 bp (Reneker *et al*., 1991) [37]. The sonication condition is dependent upon the types of sonicators, degree sonication (length of desired DNA fragments) and location of the target DNA sequence which can be determined by conducting sonication experiments for assorted time lengths and the size of the DNA fragments is determined by agarose gel electrophoresis (Hendrickson, 1985) [15]. The efficiency of sonication is reliant on the extent of crosslinking, the volume of a sample and the depth to which the sonication probe is placed in the sample affect. Cooling the sample between sonication pulses and placing the sample in an ice water bath during sonication decrease the incidence of foaming.

Immunoprecipitation of the crosslinked DNA-protein complexes

Krig *et al*., 2007 [23] described that immunoprecipitation (IP) is the technique of precipitating an antigen using an antibody that specifically binds to it. Several types of immunoprecipitation techniques have been developed which includes Individual protein Immunoprecipitation (IP), Protein complex immunoprecipitation (Co-IP), Chromatin immunoprecipitation (ChIP) and RNA immunoprecipitation (RIP). Out of them Chromatin immunoprecipitation is the desired method of choice for analysis of DNA sequence of regulatory region (Spencer *et al*., 2003) [46]. The protein of our interest (i.e. transcription factors) along with the DNA associated with it can be isolated by using an antibody specific to that protein. The cleared lysate obtained after sonication can be incubated with a primary antibody and the antibody-antigen complexes can be captured with protein A- or protein G-Sepharose depending on the nature of the antibody (Egelhofer *et al*., 2011) [11]. The Sepharose is washed several times with buffers containing different salt and detergent concentrations and the antibody-antigen complexes can be eluted from the protein A/G with a high detergent elution buffer (Egelhofer *et al*., 2011) [11]. The protein complement of the immune complexes can be digested away and the DNA can be isolated by phenol-chloroform-isoamyl alcohol followed by ethanol precipitation.

Controls for ChIP

To ensure the reliability of ChIP data, two control samples specific for the ChIP experiment should be included: the input sample and the 'no-antibody' (NoAb) control sample. The input sample, an aliquot taken from the chromatin before preclearing will be indicative for the presence and amount of chromatin used in the ChIP reaction. The chromatin aliquot is decrosslinked and DNA is isolated. This DNA serves as a positive control (Nowak *et al*., 2005) [31].

Optimizing the signal to noise ratio

Several options are available to optimize the signal-to-noise ratio. To prevent non-specific binding of chromatin to the protein A/G agarose beads, the beads are blocked with BSA and non-specific blocking DNA before they are used in ChIP assays. To further reduce background, BSA, non-specific serum and unrelated DNA can be added to the pre-incubation step of chromatin with beads (Bernstein *et al.,* 2005) [5]. Options for preventing chromatin from binding to plastic tubes are pre-incubation of the tubes with BSA and unrelated DNA (e.g. salmon sperm DNA), or the use of siliconized tubes. The background can also be reduced by lowering the amount of input chromatin, the amount of protein A/G agarose beads or the amount of antibody, but this may lower signal levels as well.

Analysis of precipitated material

The detection of specific DNA fragments can be done by conventional PCR, quantitative PCR, microarray analysis and slot blotting. Microarray analysis is useful when studying the genome wide distribution of histone modifications (Rozowsky *et al*., 2009) [42]. Analysis by slot blotting is feasible when dealing with highly repetitive sequences (Day *et al*., 2010) [8], but appears not sensitive enough to detect single copy

sequences in ChIP samples. Earlier ChIP precipitates were analyzed by conventional PCR which requires lots of testing per individual primer set to ensure that measurements are taken in the linear range of amplification (Rozowsky *et al*., 2009) [42]. If this condition is not fulfilled, the resulting data cannot be considered quantitative, impeding data interpretation. Analyzing ChIP precipitates by quantitative real-time PCR (QPCR) has several advantages over conventional PCR. The QPCR technique quantifies the initial amount of template DNA is calculated from the kinetics of the PCR reaction (Day *et al*., 2010) [8].

Types of Chip Assay

ChIP-chip

The cells are fixed with formaldehyde, harvested, and disrupted by sonication. The DNA fragments cross-linked to a protein of interest are enriched by immunoprecipitation with a specific antibody (Egelhofer *et al*., 2011) [11]. After reversal of the cross-links, the enriched DNA is amplified and labeled with a fluorescent dye (Cy5) with the use of ligation-mediated-polymerase chain reaction (LM-PCR). A sample of DNA that is not enriched by immunoprecipitation is subjected to LM-PCR in the presence of a different fluorophore (Cy3), and both immunoprecipitation (IP)-enriched and -unenriched pools of labeled DNA are hybridized to a single DNA microarray containing all intergenic sequences (Acevedo *et al*. 2007; Sikes *et al*., 2009) [1, 44]. Hybridization of the immunoprecipitated DNA to tiling or promoter microarrays (ChIP-chip) allowed extension of ChIP from single-gene studies to the whole genome (Ji *et al*., 2008) [18].

Due to their size and more repetitive nature, higher eukaryotic genomes are a challenge for tiling microarray design. Most of the repetitive sequence cannot be interrogated with high confidence, whereas direct sequencing can reveal binding events located in repetitive regions in the mammalian genome was demonstrated by Kaufmann *et al*., 2010 [22].

ChIP Seq

The combination of ChIP with high-throughput sequencing technology has allowed more comprehensive surveys of genome occupancy, greater resolution, and lower cost for whole genome coverage (Jothi *et al*., 2008) [21]. This approach was termed ChIP-seq and offers tremendous advantages over ChIP-chip, such as single-base pair resolution, much lower starting material requirements and the absence of DNA-hybridization-related sensitivity issues (Park, 2009) [33]. A number of high-throughput sequencing technology platforms have been developed that are suitable for ChIP-seq, including the Genome Analyzer (Illumina, formerly Solexa), SOLiD (Applied Biosystems), 454-FLX (Roche) and HeliScope (Helicos) (Park, 2009) [33]. The Illumina Genome Analyzer and the ABI SOLiD sequencers produce shorter reads but give a higher number of sequencing reads per run, whereas the 454-FLX sequencer gives longer yet fewer sequencing reads per run (Jothi *et al*., 2008) [21]. Sequencing depth is a critical factor in identifying weaker binding positions and it has been shown that millions of mapped sequencing tags are needed to detect enrichments significantly higher than twofold (Bao *et al*., 2011) [4]. The conventional way of analyzing this data is to identify sequencing peaks along the chromosomes that are significantly higher than the read background. Mapping of the sequencing reads to the genome reveals positions where high numbers of reads pile up to create peaks, indicating protein binding sites (Bao *et al*., 2011) [4]. For histone modifications and other epigenetic marks, it is often preferable to find a characteristic region of enrichment in sequencing reads relative to gene annotations. Peaks are often located far from loci or do not contain binding motifs and yet are clearly not artifactual. Because of this, accurate target gene assignment is currently one of the major problems with TF ChIP-seq.

ChIP-Exo

Chromatin immunoprecipitation (ChIP-chip and ChIP-seq) assays identify where proteins bind throughout a genome (Hong *et al*., 2005) [16]. However, DNA contamination and DNA fragmentation heterogeneity produce false positives (erroneous calls) and imprecision in mapping. Consequently, stringent data filtering produces false negatives (missed calls). ChIP-exo, involves an exonuclease trims ChIP DNA to a precise distance from the crosslinking site (Rhee and Pugh, 2012) [38]. Bound locations are detectable as peak pairs by deep sequencing. Contaminating DNA is degraded or fails to form complementary peak pairs with the single bp accuracy provided by ChIP-exo. The binding sites become unambiguous and reveal diverse tendencies governing *in vivo* DNA-binding specificity that include sequence variants, functionally distinct motifs, motif clustering, secondary interactions, and combinatorial modules within a compound motif (Rozowsky *et al*., 2009) [42].

This involves chromatin immunoprecipitation (ChIP) combined with lambda exonuclease digestion followed by high-throughput sequencing. ChIP-exo allows us to identify a nearly complete set of binding locations of DNA-bound proteins at near single nucleotide resolution with almost no background (Rhee and Pugh, 2011) [39]. The process is initiated by cross-linking DNA and associated proteins. Chromatin is then isolated from nuclei and subjected to sonication. Subsequently, an antibody against a specific protein is used to immunoprecipitate specific DNA-protein complexes. ChIP DNA is purified, sequencing adaptors are ligated, and digested by lambda exonuclease (Tuteja *et al*., 2009) [47]. High-throughput sequencing generates 25 to 50 nucleotides sequences. The sequence of the DNA fragments is mapped back to the reference genome for determination of the binding locations. 5′ ends of DNA fragments on the forward strand indicate the left border of DNA-protein boundaries and 5′ ends of DNA fragments on the reverse strand indicate the right border of DNA-protein boundaries (Rhee and Pugh, 2012) [38].

Double Chromatin Immunoprecipitation

The conventional ChIP analysis for individual factors (single ChIP) does not provide information on co-occupancy of two interacting TFs on target genes, even if both bind to the same chromatin regions (Geisberg and Struhl, 2004) [12]. Double ChIP analysis involves sequential (double) immunoprecipitation of two chromatin-binding proteins and can be used to study co-occupancy of two or more factors on specific regions of the same DNA allele. Furthermore, by including a cell type-specific protein in double-ChIP, target co-occupancy in a specific cell type can be studied even if the other partner is more widely expressed (Peng and Chen, 2013) [34].

Limitations

ChIP assays suffer from inherent problems that can often lead to misleading, or even erroneous, interpretation of the ChIP data (Machanick and Bailey, 2011) [26]. These problems arise during IP, the most crucial part in the ChIP assay. Two main

components of IP determine the quality and quantity of the ChIP DNA: the antibody and the solid support used for binding the antigen/antibody complex. The ChIP antibody may contribute to nonspecific background signal because of their possible cross-reactivity with other antigens. The ChIP antibody may also be responsible for a low yield of the recovered DNA due to a low affinity for the protein bound to the DNA (Pillai *et al*., 2009) [35]. However, the bulk of nonspecific binding to the solid support, in particular to agarose-based matrices such as Protein A Sepharose beads commonly used in IP and ChIP assays, is thought to be due to the DNA/protein complexes reacting with diverse chemical groups on the surface of the Sepharose. Some antibodies are sensitive to inhibitory factors present in the input chromatin sample, resulting in a decrease in binding efficiency of the antibody when increasing the amount of input (Li *et al*., 2008) [25].

As, it is difficult to shear chromatin into fragments of less than several hundred base pairs, standard ChIP is not a suitable assay for high-resolution characterization of binding sites (Yan *et al*., 2004) [50]. At present, *in vitro* gel mobility shift analysis using cellular extract is the standard assay for determining which specific nucleotides within a region of DNA constitute a binding site for a particular factor. Although gel mobility shift analysis is a fast and easy method for identifying which nucleotides are required for binding under *in vitro* conditions, *in vivo* binding conditions are often difficult to recreate *in vitro* (Pillai *et al*., 2009) [35]. For example, binding of proteins that require DNA secondary structure, such as looping, to bring two distal binding sites into close proximity often cannot be studied using this method. Additionally, sites that require multiprotein complex formation to stabilize protein-DNA interactions are difficult to study *in vitro* due to dissolution of the complex during gel electrophoresis. Finally, if a transcription factor is recruited to DNA via interaction with another, perhaps unknown, DNA-binding protein *in vitro* assay conditions may not be optimized for binding of the factor that actually contacts DNA.

Softwares for ChIP Data analysis

A massive and heterogeneous datasets were obtained from these studies and analyzing these data poses several challenges, including effective data visualization, seamless connection of low-level close to raw data and high-level close to answering biological questions analysis, integration of data from multiple technological platforms, and flexibility of the data to customize the analysis method so that specific biological concept can be solved. Although there are several recently developed programs that target some of the individual steps, an integrated tool that can satisfy all basic needs in ChIP data analyses (Nicol *et al*., 2009) [29].

CisGenome

Genome-wide chromatin immunoprecipitation (ChIP) data can be analyzed by CisGenome software. CisGenome is designed to analyse all types of ChIP data including visualization, data normalization, peak detection, false discovery rate computation, gene-peak association, sequence and motif analysis (Giardine *et al*., 2005) [13]. Along with the ChIP-chip analysis methods, the software implemented with advanced statistical methods that especially designed to analyse ChIP-seq data (Bailey *et al*., 2011) [3]. The modular design of CisGenome enables it to support interactive analyses through a graphic user interface as well as customized batch-mode computation for advanced data mining. A built-in browser allows visualization of array images, signals, gene structure, conservation, and DNA sequence and motif information (Shim *et al*., 2008) [43].

Starr

Starr is an extension of the Ringo package for the analysis of ChIP-chip projects. Whereas the latter is specialized to the processing of Nimblegen and Agilent arrays, the former provides all corresponding features for Affymetrix arrays (Buck and Lieb, 2004) [6]. Data can be read in from Affymetrix CEL-fles, or from text fles in gff format. Standard quality assessment and data normalization tools are available. Starr uses the Bioconductor Expression Set class for data storage. The probeAnno class from Ringo serves as a mapping of the ChIP signals onto the genomic position. Consequently, all functions from Ringo that operate on either ExpressinoSet or probeAnno can be used without modifcation. These include smoothing operations, peak-fnding, and quality control plots. Starr adds new options for high-level analysis of ChIP-chip data (Nakato *et al*., 2013) [28]. Starr's facilitate an an experiment that compares DNA binding under two different conditions. Three chips have been produced, two contain the actual immunoprecipitated DNA, and the other one is a control experiment.

ChIPOTle

Analysing chromatin immunoprecipitation (ChIP) data with ChIPOTle creates several output sheets naming as summary sheet, significant regions, significant negative regions, chromosomes aveP, peaks, and description. The summary sheet entered with all the input data that used to run ChIPOTle, thereafter get sorted by chromosome and start coordinate. For each window specified by the user, the significant regions sheet contains the following function i.e., center coordinate, chromosome assignment, number of independent arrayed elements within each window, and names of the arrayed elements that comprise the window (Buck and Lieb, 2004) [6]. Similarly, a Significant Negative Regions is present which contains all of the windows that meet the significance criteria but are sign-flipped. The number of windows reported in this sheet can be used as an estimate of the number of false-positive findings expected for the selected or estimated cutoff. Moreover, Chromosome aveP carries the names of all the arrayed elements that encompass each window (with the chromosome, center coordinate and value of all windows) regardless of whether they satisfy the significance criterion or not. In Peaks sheet the data were reported similar to those in Significant Regions, except that all neighboring windows meeting the significance criteria are collapsed into a single peak (Bailey *et al*., 2011) [3]. Therefore, a peak is defined as any window with a P value that meets the significance criterion defined by the user and all neighboring windows that also meet the significance criteria (Buck and Lieb, 2004) [6]. So, in peak sheet, each peak is determined in order of its occurrence along the chromosome, the highest window for each peak, the highest raw log 2 ratio for any element within the peak, start coordinate of the peak, the width of portion of the peak above the significance cutoff, 'array density' of the peak, and the P value for that peak. The phenomenon array density value explains that the average number of arrayed elements used to calculate the window values for all windows that comprise the peak. Therefore, the array density value provides an estimate of the number of actual raw data measurements that underlie

in each peak. Lastly, the description sheet contains a summary of the ChIPOTle execution parameters, which includes date and time, the selected window size, the step size, the significance method chosen and corresponding parameters, the number of significantly enriched peaks, and the total number of windows.

Database of ChIP

hmChIP is a database of genome-wide chromatin immunoprecipitation (ChIP) data in human and mouse. Currently, the database contains more than 2000 samples from more than 500 ChIP-seq and ChIP-chip experiments, representing a total of more than 170 proteins and more than 10,000,000 protein-DNA interactions (Ji, 2010) [19]. A web server provides interface for database query. Protein-DNA binding intensities can be retrieved from individual samples for user-provided genomic regions (Rosenbloom *et al*., 2015) [41]. The retrieved intensities can be used to cluster samples and genomic regions to facilitate exploration of combinatorial patterns, cell type dependencies, and cross-sample variability of protein-DNA interactions (Rosenbloom *et al*., 2015) [41].

Conclusion

Chromatin immunoprecipitation (ChIP) is a well-established procedure to investigate interactions between proteins and DNA. Long-range chromatin interacts between specific DNA regulatory elements play an important role in gene expression regulation. These interactions are characterized by the formation of 3-dimensional chromatin structure which is pivotal in understanding signaling networks and cell states. Chromatin immunoprecipitation and digestion with deoxyribonuclease and micrococcal nuclease together combined with high-throughput sequencing to provide detailed nucleosome occupancy maps and chromosome conformation will get capture. Its variants have illustrated that higher-order chromatin structure involves with long-range loop formation between distant genomic elements. Coupled with whole-genome DNA microarrays, ChIP assay allow one to determine the entire spectrum of in-vivo DNA binding sites for any given protein. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a high-throughput antibody-based method to study genome-wide protein–DNA binding interactions (Park, 2009) [33]. ChIP-seq technology allows obtaining more accurate data providing genome-wide coverage with less starting material and in shorter time compared to older ChIP-chip experiments.

References

1. Acevedo LG, Iniguez AL, Holster HL, Zhang X, Green R, Farnham PJ. Genome-scale ChIP-chip analysis using 10,000 human cells. Biotechniques. 2007; 43:791-797.
2. Albert I, Wachi S, Jiang C, Pugh BF. GeneTrack- a genomic data processing and visualization framework. Bioinformatics. 2008; 24:1305-1306.
3. Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics. 2011; 27:1653-1659.
4. Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ. Evaluation of next-generation sequencing software in mapping and assembly. J Hum Genet. 2011; 56:406-414.
5. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, *et al*. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell. 2005; 120:169-181.
6. Buck MJ, Lieb JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics. 2004; 83:349-360.
7. Cawley S. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell. 2004; 116:499-509.
8. Day DS, Luquette LJ, Park PJ. Estimating enrichment of repetitive elements from high-throughput sequence data. Genome Biol. 2010; 11:69.
9. Dedon PC, Soults JA, Allis CD, Gorovsky A. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein–DNA interactions. Anal Biochem. 1991; 197:83-90.
10. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. Nat Genet. 1999; 21:10-14.
11. Egelhofer TA, Minoda A, Klugman S, Lee K, Kolasinska-Zwierz P, Alekseyenko AA, *et al*. An assessment of histone-modification antibody quality. Nat Struct Mol Biol. 2011; 18:91-93.
12. Geisberg JV, Struhl K. Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions *in vivo*. Nucleic Acid Res. 2004; 32:151
13. Giardine B. Galaxy: A platform for interactive large-scale genome analysis. Genome Res. 2005; 15:1451-1455.
14. Gilmour DS, Lis JT. Detecting protein-DNA interactions *in vivo*: Distribution of RNA polymerase on specific bacterial genes. Proc Natl Acad Sci. 1984; 81:4275-4279.
15. Hendrickson W. Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay, Bio Techniques. 1985; 3(3):198-207.
16. Hong P. A boosting approach for motif modeling using ChIP-chip data. Bioinformatics. 2005; 21:2636-2643.
17. Jackson V, Chalkley R. A new method for the isolation of replicative chromatin: selective deposition of histone on both new and old DNA. Cell. 1981; 23:121-134.
18. Ji H, Jiang H, Ma W. An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol. 2008; 26:1293-1300.
19. Ji H. Computational analysis of ChIP-seq data. Methods Mol Biol. 2010; 674:143-159.
20. Johnson DS. Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. Genome Res. 2008; 18:393-403.
21. Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res. 2008; 36:5221-5231.
22. Kaufmann K, Muino JM, Østeras M, Farinelli L, Krajewski P, Angenent GC. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc. 2010; 5:457-472.
23. Krig SR, Jin VX, Bieda MC, O'Geen H, Yaswen P, Green R, *et al*. Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays. J Biol Chem. 2007; 282: 9703-9712.

24. Kurdistani SK, Grunstein M. *In vivo* protein-protein and protein-DNA cross-linking for genomewide binding microarray. Methods. 2003; 31:90-95.
25. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, *et al*. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. PLoS Biol. 2008; 6: 27.
26. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics. 2011; 27:1696-1697.
27. McGhee JD, von Hippel PH. Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. Biochemistry. 1975; 14:1281-1296.
28. Nakato R, Itoh T, Shirahige K. DROMPA: easy-to-handle peak calling and visualization software for the computational analysis and validation of ChIP-seq data. Genes Cells. 2013; 18:589-601.
29. Nicol JW, Helt GA, Blanchard SGJ, Raja A, Loraine AE. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics. 2009; 25:2730-2731.
30. Niranjanakumari S, Lasda E, Brazas R, Garcia-Blanco MA. Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions *in vivo*. Methods. 2002; 26:182-190.
31. Nowak DE, Tian B, Brasier AR. Two-step cross-linking method for identification of NF kappaB gene network by chromatin immunoprecipitation. Biotechniques. 2005; 39:715-725.
32. Orlando V. Mapping chromosomal proteins *in vivo* by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem Sci. 2000; 25:99-104.
33. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009; 10:669-680.
34. Peng GH, Chen S. Double Chromatin Immunoprecipitation: Analysis of Target Co-occupancy of Retinal Transcription Factors. Retinal Degeneration Methods in Molecular Biology. 2013; 935:311-328.
35. Pillai S, Dasgupta P, Chellappan SP. Chromatin immunoprecipitation assays: analyzing transcription factor binding and histone modifications *in vivo*. Methods Mol Biol. 2009; 523:323-339.
36. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, *et al*. Genome-wide location and function of DNA binding proteins. Science. 2000; 290:2306-2309.
37. Reneker JS, Brotherton TW. Discrete regions of the avian betaglobin gene cluster have tissue-specific hypersensitivity to cleavage by sonication in nuclei. Nucleic Acids Res. 1991; 19(17):4739-4745.
38. Rhee HS, Pugh BF. ChIP-exo: A Method to Identify Genomic Location of DNA-binding proteins at Near Single Nucleotide Accuracy, Curr Protoc Mol Biol. 2012 21:10.
39. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell. 2011; 147:1408-1419.
40. Robyr D, Suka Y, Xenarios I, Kurdistani SK, Wang A, Suka N, *et al*. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. Cell. 2002; 109:437-446.
41. Rosenbloom KR, Armstrong J, Barber GP. The UCSC Genome Browser database: update. Nucleic Acids Res. 2015; 43:670-681.
42. Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, *et al*. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol. 2009; 27:66-75.
43. Shim H, Keles S. Integrating quantitative information from ChIP-chip experiments into motif finding. Biostatistics. 2008; 9:51-65.
44. Sikes ML, Bradshaw JM, Ivory WT, Lunsford JL, McMillan RE, Morrison CR. A streamlined method for rapid and sensitive chromatin immunoprecipitation. J Immunol Methods. 2009; 344:58-63.
45. Solomon MJ, Varshavsky A. Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. Proceedings of the National Academy of Sciences. 1985; 82:6470-6474.
46. Spencer VA, Sun JM, Li L, Davie JR. Chromatin immunoprecipitation: a tool for studying histone acetylation and transcription factor binding, Methods. 2003; 31:67-75.
47. Tuteja G, White P, Schug J, Kaestner KH. Extracting transcription factor targets from ChIP-Seq data. Nucleic Acids Res. 2009; 37:113.
48. Walter J, Dever CA, Biggin MD. Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in *Drosophila* embryos. Genes & Dev. 1994; 8:1678-1692.
49. Yamanaka K, Tezuka M, Kato K, Hasegawa A, Okada S. Crosslink formation between DNA and nuclear proteins by *in vivo and in vitro* exposure of cells to di-methylarsinic acid. Biochem Biophys Res Commun. 1993; 191:1184-1191.
50. Yan Y, Chen H, Costa M. Chromatin immunoprecipitation assays. Methods Mol Biol. 2004; 287:9-19.
51. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics. 2009; 25:1952-1958.