(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 May 2020 (14.05.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/097294 A1

(51) International Patent Classification:
C12N 15/10 (2006.01) *G01N 33/536* (2006.01)
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/US2019/060212

(22) International Filing Date:
07 November 2019 (07.11.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/757,584 08 November 2018 (08.11.2018) US

(71) Applicant: **UNIVERSITY OF MIAMI** [US/US]; 1951 NW 7th Avenue, Miami, FL 33136 (US).

(72) Inventors: **SUN, Jun**; 1951 NW 7th Avenue, Miami, FL 33136 (US). **LIPPMAN, Marc, E.**; 1951 NW 7th Avenue, Miami, FL 33136 (US).

(74) Agent: **BRASHEAR, Jeanne, M.**; Marshall, Gerstein & Borun LLP, 233 S. Wacker Drive, 6300 Willis Tower, Chicago, IL 60606-6357 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

(54) Title: MATERIALS AND METHODS FOR CHROMATIN IMMUNOPRECIPITATION

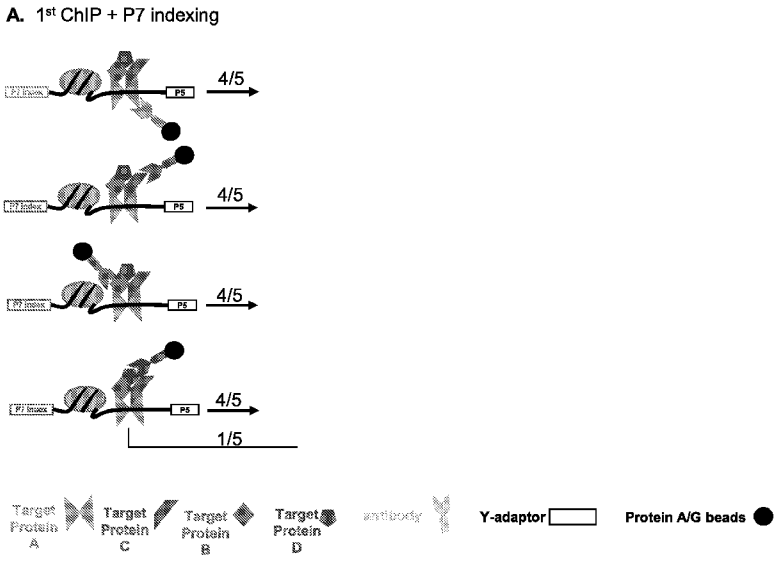


Figure 1

(57) Abstract: The present disclosure provides materials and methods for analyzing a chromatin sample. In one aspect, described herein is a method of analyzing a chromatin sample, the method comprising adding a biotin-labeled non-mammalian DNA and an anti-biotin antibody to a sample comprising chromatin fragments, wherein the biotin-labeled DNA and the antibiotin antibody form a complex in the sample; immunoprecipitating a target binding protein in the sample with an antibody that binds the target binding protein; reverse crosslinking of DNA bound to the target binding protein and disassociating the non-mammalian DNA bound to biotin; and purifying the DNA of step (c).

WO 2020/097294 A1

[Continued on next page]

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

MATERIALS AND METHODS FOR CHROMATIN IMMUNOPRECIPITATION

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/757,854, filed on November 8, 2018, the entire contents of which are fully incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The present disclosure is directed to materials and methods for analyzing chromatin samples.

BACKGROUND

[0003] Chromatin immunoprecipitation (ChIP) is a powerful tool to study protein-DNA interactions (Dedon et al., 1991). Coupled with next generation sequencing, ChIP-seq has been widely used to identify genome-wide binding of transcription factors and co-factors as well as histone modifications at chromatin (Mardis 2007). Normalization in ChIP-seq based on the total number of mapped reads sometimes does not reflect the signal strength accurately since only a small portion of mapped reads is derived from specific binding of an antibody and the IP efficiency can vary significantly. Analogous to RNA-seq analysis for global gene expression, normalization without a proper spike-in strategy could result in erroneous conclusions (Loven et al., 2012). A proper normalization strategy for ChIP-seq is required to make results comparable from different antibodies or different treatment conditions. The importance for properly normalizing ChIP-seq data was highlighted by the ChIP-seq analyses of histone marks after drug treatment that caused the global reduction of the modified histones (Orlando et al., 2014). Some strategies for normalization of the ChIP-seq data have been developed (Bonhoure et al., 2014; Orlando et al., 2014; Grzybowski et al., 2015; Egan et al., 2016). The commonly used commercially available strategy from Active Motif is to add drosophila chromatin and an antibody against drosophila specific H2A.V histone variant into the ChIP process and to normalize ChIP-Seq data based on the drosophila DNA reads from each reaction. Recently, a normalization strategy based on a parallel factor ChIP serving as an internal control is proposed (Guertin et al., 2018).

[0004] Sequential ChIP-seq (re-ChIP-seq) has been employed to investigate genome-wide co-localization of histone marks or transcription factors (Kinkley et al., 2016; Luna-Zurita et al., 2016). To overcome some technical challenges, combinatorial indexed ChIP-seq assay has been developed to investigate combinations of histone marks and co-localization of

transcription factor and histone marks (Lara-Astiaso et al., 2014; Weiner et al.,2016). But, applying this method to investigate multiple components in one transcriptional complex at chromatin remains challenging due to the low yield in library construction and the lack of proper normalization strategy.

SUMMARY

[0005] In one aspect, described herein is a method of analyzing a chromatin sample, the method comprising adding a biotin-labeled non-mammalian DNA and an anti-biotin antibody to a sample comprising chromatin fragments, wherein the biotin-labeled DNA and the anti-biotin antibody form a complex in the sample; immunoprecipitating a target binding protein in the sample with an antibody that binds the target binding protein; reverse crosslinking of DNA bound to the target binding protein and disassociating the non-mammalian DNA bound to biotin; and purifying the DNA of step (c). In some embodiments, the method further comprises the step of (e) sequencing the purified DNA in the sample. In some embodiments, the method further comprises the step of quantifying the amount of the non-mammalian DNA in the sample. The quantifying step is optionally performed by qPCR.

[0006] In another aspect, described herein is a method of analyzing a chromatin sample, the method comprising adding a biotin-labeled non-mammalian DNA and an anti-biotin antibody to a sample comprising chromatin fragments, wherein the biotin-labeled non-mammalian DNA and the anti-biotin antibody form a complex in the sample; a first step of immunoprecipitating a target binding protein in the sample with an antibody that binds the target binding protein; performing chromatin indexing using a substrate; eluting the immunoprecipitated proteins from the substrate to produce an eluted sample; reverse crosslinking of DNA bound to the target binding protein and disassociating the non-mammalian DNA bound to biotin in the eluted sample; a second step of immunoprecipitating a target binding protein in the eluted sample with an antibody that binds the target binding protein; reverse crosslinking of DNA bound to the target binding protein; purifying the DNA; and sequencing the purified DNA of step (h).

[0007] In some embodiments, the non-mammalian DNA is a synthetic nucleic acid sequence. In other embodiments, the non-mammalian DNA is a fragment of a nucleic acid sequence derived from firefly luciferase, GFP or CAS9.

[0008] In some embodiments, the chromatin fragments are generated by sonication.

[0009] In some embodiments, the chromatin indexing comprises incubating beads with the immunoprecipitated target proteins at 55°C for at least 10 minutes.

[0010] In some embodiments, the method further comprises reserving a sample of the chromatin indexing substrate as an input sample.

[0011] In some embodiments, the method comprises the step of quantifying the amount of the non-mammalian DNA in the input sample. In some embodiments, the method comprises the step of quantifying the amount of the non-mammalian DNA after the second immunoprecipitating step. In some embodiments, the quantifying step is performed by qPCR.

BREIF DESCRIPTION OF THE FIGURES

[0012] Figure 1. The schematic diagram for combinatory ChIP-seq assay. A. The first ChIP is performed with each target antibody in a separate reaction with biotin-labeled spike-in DNA molecules and an antibody against biotin included (not shown for simplicity). A Y-adaptor containing an index at P7 side is ligated to ChIP DNA on beads. Each antibody target is represented by a unique P7 index. 1/5 of each ChIP product is taken out as input control. The rest is subjected to antibody dissociation and pooling all the samples together (B). C. The pooled ChIP sample is aliquoted into 4 equal portions with each portion subjected to a second ChIP with one target antibody. After purification of the ChIP DNA, an index-containing P5 primer is used with P7 primer to amplify ChIP DNA for DNA sequencing using the templates from the second IP or ChIP input control DNA. The P5 indices represent ChIP input samples or the antibodies used in the second IP.

DETAILED DESCRIPTION

[0013] Chromatin immunoprecipitation (ChIP) is an analytical method used to investigate the interactions between proteins and DNA in vivo. Variations in the efficiency of the immunoprecipitation and losses of material during the purification of the DNA are sources of variability that reduce the accuracy of the results and impair the use of ChIP as a quantitative tool. We have developed a simple method to improve the quantification of ChIP data based on the use of an external reference. Fixed amounts of biotin-labeled DNA and an anti-biotin antibody are spiked into the chromatin extract at the beginning of the ChIP assay. The amounts of biotin-labeled DNA recovered in each tube are measured at the end of the assay and used to normalize the results obtained with the antibodies of interest. Synthetic DNA or DNA of bacterial origin (or other origin without homology with eukaryotic sequences) were

used. Using this biotin-labeled DNA as an external reference, the variability between individual ChIP samples was strongly reduced, which increased the accuracy and the statistical resolution of the data.

[0014] The present disclosure provides a spike-in strategy for ChIP-seq that is also applicable for ChIP-qPCR. The method described herein is simple, cost effective, up scalable (e.g., multiple spike-ins can be added to avoid any significant deviation caused by one control), and suitable for complexed assays such as the sequential indexed ChIP analysis. Unlike the drosophila genomic DNA, the spike-in non-mammalian DNA molecules used herein have no alignment against the mammalian genomes such as human and mouse. ChIP-seq is routinely performed by single-end DNA sequencing.

[0015] With biotin-labeled non-mammalian DNAs spiked into the ChIP and re-ChIP process, each step can be monitored by PCR for these spike-in non-mammalian DNA. This enables quantitative comparisons among ChIP-signals derived from different target antibodies to normalize different sequencing depth from each sample and different signal/noise ratios from each antibody used.

[0016] Efficiency to ligate the indexed sequencing adaptor to ChIP DNA on beads is very low with the protocol provided in the art (Lara-Astiaso et al., 2014; Weiner et al.,2016), which can be challenging for analyzing transcriptional cofactors due to their low yield in ChIP. As described in the Examples, a single step with 55°C incubation with the DNA-bound beads for 10 minutes before adaptor ligation increased the yield for ChIP-seq DNA library over 25 fold, which made it possible to analyze transcription co-factors. The re-ChIP process greatly increases the signal/noise ratio compared to the first ChIP, resulting much larger number of called peak. The spike-in non-mammalian DNA also normalizes the signals for the ChIP-seq and the re-ChIP-seq which have completely different signal versus noise ratios. Because the ChIP and re-ChIP have different signal/noise ratios, the spike in counts are not used to compare the ChIP and re-ChIP directly. Also, re-ChIP-seq could be an improved method for ChIP-seq assay for low affinity antibodies and/or high background in ChIP-seq.

EXAMPLES

[0017] *Materials and Methods*

[0018] *Generating spike-in DNA fragments*: DNA fragments derived from GFP, Cas9 or fire fly luciferase gene were amplified by PCR using a PCR Master Mix (Promega). The

amplified DNAs were purified with a PCR Purification Kit (Qiagen) and were verified in an agarose gel. DNA quantification was performed with Qubit dsDNA HS Assay Kit (Invitrogen). For generating biotin-labeled DNAs, 1:1 ratio of biotin-16-AA-2'-dCTP (TriLink) to dCTP was used in PCR reaction. The primers used for PCR are listed in the followings. For GFP DNA fragment (163bp), forward: TATATCATGGCCGACAAGCAG; reverse: ACTGGGTGCTCAGGTAGTGGT. For luciferase DNA fragment (132 bp), forward: ATTCTTCGCCAAAAGCACTCT; reverse: CGTATCCCTGGAAGATGGAAG. For CAS9 DNA fragment (144 bp), forward: GAACCGCCAGAAGAAGATACAC; reverse: CGTGCTTCTTATCCTCTTCCAC.

[0019] *Cell fixation*: MCF-7 cells were grown in a 150 mm culture dish to 80-90% confluency in IMEM (Invitrogen) with 5% FBS (Hyclone). Cells were fixed with 1% formaldehyde (Sigma) at room temperature for 15 min on a shaker. Glycine (final concentration of 125mM) was added to quench the reaction for 5 min with rotation. Cells were washed with cold PBS and collected into a 15 ml conical tube on ice. Collect the cell pellet after centrifugation the tube at 1000 g for 5 min at 4 ºC. The cell pellet was resuspended in 10 ml of PBS-0.5% NP40 and incubated on ice for 10 min. Collect the cell pellet as the previous step and resuspend it in 10 ml of PBS-0.5% NP40 with 1mM PMSF and incubated it on ice for 10 min. Centrifuge again and carefully remove all supernatant. The pellet was kept at -80 ºC for later process.

[0020] *ChIP-qPCR and ChIP-Seq*: The ChIP was performed using Magna ChIP™ HiSens Chromatin Immunoprecipitation Kit (Millipore, 17-10460) with some modifications. The cell pellet was resuspended in 0.5 ml of SCW buffer containing 2.5 µl of 200x Protease Inhibitor Cocktail III. The cell lysate was sonicated using Bioruptor (Diogenode) for 12 cylces (30 sec on/off). The sheared chromatin DNA was verified in the range of ~150 to 500 bp. 30 µg of sheared chromatin DNA was used in each immunoprecipitation and incubated for 4 hr at 4 ºC in 200 µl volume with 2 µg of an indicated antibody. Following the washing steps, beads were resuspended in 50 µl ChIP Elution Buffer with Proteinase K followed by incubation at

65 °C for 6 hr or overnight. The supernatant ChIP DNA sample was diluted 10x with H_2O before qPCR analysis using Perfecta SYBR® Green FastMix (Quantabio) in a real-time PCR machine (Roche 480). The PCR primers using for one target protein A binding site within target protein C locus are forward: GAGAGGGTGGTGACACTTGG; reverse: AGCTGACAGAGGAGACAAAACG. For the ChIP-seq analysis, ChIP DNA was purified with Ampure XP beads (Beckman Coulter) and quantified with Qubit dsDNA HS Assay Kit before the DNA sequencing library construction using NEBNext UltraII DNA Library Prep Kit for Illumina (NEB). The ChIP-seq library DNA was analyzed with DNA Screen Tape (Agilent) on a Tapestation 4200 (Agilent) and was sequenced using Illumina NextSeq 500 with 76 bp cycle. The antibodies to target protein A and target protein B are from Santa Cruz; to target protein C are from Bethyl , Abcam, and our self-made antibody stocks; to target protein D is from Millipore (05-389).

[0021] *Combinatory indexed ChIP-seq*: The protocol of Weiner (Weiner et al., 2016) was adapted as follows (Fig.1).

[0022] *The first immunoprecipitation*: 50 µg of sheared chromatin DNA was used in each immunoprecipitation as in the ChIP-seq protocol with addition of 5 pg biotin-labeled DNA and 0.5 µg antibody to biotin (Bethyl, A150-109A-11). For the maximum ChIP DNA recovery, two or three rounds of IP were performed with 2 µg antibodies against target protein A, target protein B, target protein C or target protein D with at least 4 hr incubation for each round. The recovered magnetic beads after each immunoprecipitation were pooled together and washed 3 times with 200 µl of 10 mM Tris (pH 8.0).

[0023] *Chromatin indexing*: NEBNext® ChIP-Seq Library Prep Master Mix Set for Illumina® (NEB, E6240) components were used to add sequencing index to ChIP DNA on beads. To end repair of ChIP DNA, the washed beads were incubated with constant shaking in 4 µl of End Repair Buffer, 1 µl of End Repair Enzyme Mix and 35 µl of H_2O at room temperature for 30 min. The supernatant was removed and the beads were washed 3 times with 200 µl of 10 mM Tris (pH 8.0). 4 µl of dA-Tailing Reaction Buffer and 35 µl of H_2O were added to the beads followed by 55 °C incubation for 10 min. After cooling down, 1 µl of

Klenow Fragement (3'->5' exo$^-$) was added to the mixture and incubated at 37 ºC for 30 min. After removal of the liquid, the beads were washed 3 times with 200 µl of 10 mM Tris (pH 8.0). 6 µl of Quick Ligation Reaction Buffer, 19 µl of H2O, 4 µl of Quick T4 DNA Ligase and 1 µl of Y-adaptor index (1.5 µM) were added to the beads and incubated at room temperature for 30 min with constant shaking. After removal of the supernatant, the indexed chromatin on beads was washed with 500 µl of 10 mM Tris (pH 8.0). One fifth of the beads with indexed chromatin were kept as ChIP input sample and the rest beads were subjected to chromatin release.

[0024] *Input ChIP-seq*: The beads kept as input were subjected to 2 times wash each with high salt buffer (0.1% SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris pH 8, 500 mM NaCl) and Li buffer (0.25 M LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris pH 8.0) followed by 2 times wash with TE (pH 8.0). The beads were resuspended in 50 µl ChIP Elution Buffer with 1 µl of Proteinase K included. The mixture was incubated at 65 ºC for 6 hr. The ChIP DNA in the supernatant was purified with Ampure XP beads (Beckman Coulter). After quantification with Qubit dsDNA HS Assay Kit, the ChIP DNA was amplified with primers P7 and an indexed P5 primer in PCR using NEBNext® Q5 Hot Start HiFi PCR Master Mix (NEB). The PCR product was purified with Ampure XP beads and analyzed with DNA Analysis Screen Tape before pooling samples for DNA sequencing using Illumina NextSeq 500.

[0025] *Chromatin release and the second immunoprecipitation*: After taking 1/5 of the beads as input, the rest beads were subjected to antibody dissociation. 15 µl of 100 mM DTT (Invitrogen) was added to the beads and incubated at 37 ºC for 15 min with shaking. 15 µl of 2x chromatin release buffer (1 M NaCl, 4% SDS, 2% sodium deoxycholate) was added to the mixture. After thorough mixing, the sample was incubated at 42 ºC for 15 min with shaking. Then, the supernatant was collected and the beads were washed twice with 300 µl of TE (pH 8) which was later combined with the supernatant followed by addition of 200 µl of PBS-0.5% BSA-0.5% Tween 20. The diluted supernatant was subjected to concentration using Amicon Ultra Centrigugal Filter with 10-Kda cutoff (Millipore). After the volume of concentrated chromatin was reduced to about 100 µl, the sample was collected into a new tube. The filter was washed twice with 50 µl of TE each and combine them with the chromatin sample (~ 200 µl). Each chromatin sample from 4 different antibodies against target proteins (i.e., target protein A, target protein B, target protein C or target protein D) was pooled together and then aliquoted into four equal portions. Each portion was subjected to a second round of immunoprecipitation

incubated at 4ºC overnight with 20 µl of Protein A/G beads pre-coated with 2 µg antibodies against the target proteins (i.e., target protein A, target protein B, target protein C or target protein D) and 0.5 µg of anti-biotin antibody. The beads were processed the same way as the input ChIP-seq procedure. The recovered DNA was usually too low for Qubit assay. A small portion of sample was used in qPCR to estimate the cycle number needed for DNA sequencing library prep using PCR.

[0026] *Primers for DNA sequencing.*

Y primer: ACACTCTTTCCCTACACGACGCTCTTCCGATC*T

*indicates phosphorothioate

Y-index primer:

pGATCGGAAGAGCACACGTCTGAACTCCAGTCAC**NNNNNN**ATCTCGTATGCCGTCTT CTGCTTG

p: 5' phosphorylation

NNNNNN: i7 primer index; represents the antibody used in the first IP.

Y primer and Y-index primer are annealed to form Y-adaptor to ligate ChIP DNA.

P7: CAAGCAGAAGACGGCATACGAGAT

P5-index primer:

AATGATACGGCGACCACCGAGATCTACAC**XXXXXXXX**ACACTCTTTCCCTACACGACGCTCT TCCGATCT

XXXXXXXX: i5 primer index; represents the antibody used in re-IP or the first IP input control.

[0027] *ChIP Sequencing data analysis*: All ChIP sequencing reads were aligned to the human reference genome hg19 using Bowtie2 (Langmead et al., 2012) version 2.2.6 with default parameters. The unmapped sequencing reads were used to count the spike-in DNA molecules using Bowtie2. The enriched regions for each factor were identified using MACS2 (Zhang et al., 2008) version 2.1.1 with the shifting model and q-value 0.01 followed by the subtraction of known dark regions. The enriched sequencing motifs were searched in a 100

bp region centered around the binding peaks using HOMER (Heinz et al., 2010). The common or differential binding peaks among the four factors were derived through bedtools (Quinlan et al., 2010) version 2.17.0. The comparisons for genome-wide binding patterns for these factors were performed using DeepTools (Ramirez et al., 2016) version 3.0.2. For ChIP-seq signal normalization, the number of mapped reads for each factor was adjusted according to the spike-in DNA reads. The ChIP-seq signal from target protein A served as a reference, which was set as 1 billion wigsum in BigWig file transformation.

[0028] For the re-IP signal normalizations, due to the much increased signal to noise ratio after the second IP compared to the first IP, a two steps procedure was utilized. First, each factor after its re-IP was adjusted to the same level as its input control. The top 10,000 binding peaks for ER, 1,000 peaks for target protein B, target protein C or target protein D was used for these adjustments. Based on these, an adjustment factor for spike-in DNA counts linked to each factor with its own antibody in re-IP was achieved. Second, for each re-IP with one target antibody, the signals for three other targets were normalized to the adjusted target signals using their respective adjusted spike-in DNA counts.

Example 1 – Non-mammalian DNAs as spike-ins in ChIP-seq library prep

[0029] The number of DNA sequencing reads generated in ChIP-seq is arbitrary. ChIP-seq results are usually analyzed after normalizing the number of the mapped sequencing reads. Different antibodies have varying affinities towards their epitopes and immunoprecipitate non-identical amounts of both specific and non-specific cross-linked DNA molecules in ChIP. In ChIP-qPCR, only the specific target DNA fragments are amplified for quantification and the efficiencies of IP can be compared using the input materials. In ChIP-seq, all the immunoprecipated DNA molecules are sequenced including the background noises. Thus, it is not valid to quantitatively compare the ChIP signal strength from different antibodies through the regular analysis after ChIP-seq.

[0030] A normalization procedure was tested in ChIP-seq to make the quantitative comparisons valid. A series of ChIP was performed with a variety of antibodies against

different target proteins using a sonicated chromatin sample prepared from MCF-7 cells. 1 pg of each PCR-generated DNA fragment derived from firefly luciferase, GFP or CAS9, respectively, was added to the purified ChIP DNAs before generating libraries for DNA sequencing. Since these DNA sequences cannot be aligned to the mammalian genome, the unmapped sequencing reads against human genome were used to count these spiked-in non-mammalian DNA molecules. In a smaller scale of sequencing run, the total number of mapped reads against human genome in the range of 5.4 to 7.3 million reads was achieved. The spike-in non-mammalian DNA reads were counted from 1.9 to 49 k (~25-fold range after the adjustment according to the total number of mapped reads against the human genome). Among the three different spike-in non-mammalian DNAs tested, their counts are highly correlated to each other ($R2 > 0.99$), starting from 1,851 within 6.1 million mapped sequencing reads. The sequencing run was repeated to generate about 20 million reads from these samples and recounted the spike-in non-mammalian DNA molecules (in the range of 6 to 173 k). The spike-in non-mammalian DNA molecules remained highly correlated to each other ($R2 > 0.99$). The counts of these spike-in non-mammalian DNA reads were negatively correlated to the input ChIP DNA amounts used for the sequencing library construction, which were small amounts (in ng range) and difficult to quantify with high accuracy.

Example 2 - Evaluating antibodies against target protein C and target protein D for ChIP-seq

[0031] Several antibodies against protein C were generated and tested with some commercial antibodies against the same target in ChIP-seq using sonicated chromatin samples prepared from MCF-7 cells. Luciferase DNA fragment was spiked into the ChIP DNA samples before the sequencing library construction. The counts of luciferase DNA molecule were used to normalize the ChIP-seq signals derived from target protein C antibodies. All these antibodies generated highly similar target protein C binding patterns, consistent with the previous finding that target protein C binding sites are largely co-localized to the target protein A binding sites. The signal intensities from the target protein C antibodies vary in ~20x range after normalization and correlate to their signal intensities in Western blot analysis (not shown). It was observed that target protein C binding signals were not as strong as the ones generated from the target protein A antibody, which is likely due to its indirect binding to DNAs.

[0032] A small portion of target protein D is present in the nucleus in MCF-7 cells. But, its function at chromatin have not been characterized. To investigate any potential role of target

protein D in target protein A action at chromatin, we tested in ChIP-seq assay using an antibody against target protein D, with the luciferase DNA added to the ChIP DNA as a spike-in control. Surprisingly, chromatin associated target protein D signals in ChIP-seq was detected. A large amount of target protein D binding sites co-localize to the target protein A binding sites. The signal intensity for target protein D is not as strong as the one for target protein A or target protein C, reflecting non-direct target protein D-DNA interaction and/or lower affinity of the antibody. Target protein D binding in an MCF-7 derivative cell line that lost target protein A expression was tested. The target protein D binding sites that co-localized to the target protein A binding sites in MCF-7 cells disappear in target protein A-negative MCF-7 cells. Thus, data provided herein demonstrates that target protein D is associated with chromatin in a cell context dependent manner.

Example 3 - Biotin-labeled non-mammalian DNA as spike-in controls for ChIP-seq

[0033] In order to control the whole immunoprecipitation process for ChIP-seq, biotin-labeled spike-in non-mammalian DNA molecules were generated using PCR. 5 pg of biotin-labeled DNA was added with an anti-biotin antibody and the target antibody (i.e., antibody against target protein A, target protein B, target protein C or target protein D) to each ChIP using 30 ug sonicated chromatin from MCF-7 cells. After ChIP-seq, the number of biotin-labeled non-mammalian DNA molecules were found highly correlated to each other (the correlation coefficient > 0.99). Thus, the biotin-labeled non-mammalian DNA molecules serve as good controls for ChIP-seq assay and is suitable to track targets in a complexed assay such as sequential ChIP.

Example 4 - Quantitative combinatory indexed ChIP-seq assay

[0034] Both target protein D and target protein C were detected at the target protein A binding sites through ChIP-seq. Neither target protein C nor target protein D possessed a DNA binding domain to bind DNA directly. Therefore, target protein C potentially associated with chromatin through a protein complex containing target protein A that binds DNA through its DNA binding domain.

[0035] To address whether both target protein D and target protein C are in the target A complex at chromatin, the combinatory indexed sequential ChIP assay first developed for histone marks (Fig. 1) (Lara-Astiaso et al., 2014; Weiner et al., 2016) was modified. In the first round of ChIP, equal amounts of sonicated chromatin prepared from formaldehyde fixed MCF-7 cells were immunoprecipitated with antibodies against target protein A, target protein

B, target protein C or target protein D. The bona fide target protein A coactivator target protein B, amplified in MCF-7 cell and directly interacts with target protein A, was included as a positive control. Biotin-labeled non-mammalian DNA molecules and an anti-biotin antibody were also added to each reaction to track the whole process and serve for normalizing ChIP-seq signals. When chromatin/antibody complex and biotin-labeled DNA/biotin antibody complex were immobilized to Protein A/G magnetic beads, DNA adaptor with a specific index for sequencing was ligated to the chromatin DNA fragments as well as the biotin-labeled DNA molecules. Thus, the target chromatin DNAs for each antibody as well as the associated spike-in non-mammalian DNA molecules were given a unique molecular identifier. Efforts have been taken to achieve the maximum yields from each antibody in the first round of ChIP. One fifth of the each ChIP products were kept aside as input controls for each antibody reaction. The rest ChIP products were subjected to the inactivation of antibody and to the release of the barcoded chromatin and spike-in non-mammalian DNAs from the beads. All the released products were pooled together and then split into four equal portions, followed by a second round of ChIP for each portion using one target antibody and the anti-biotin antibody. After DNA purification, the second round ChIP DNA products and the first round ChIP control DNA samples were subjected to DNA sequencing library construction using PCR with one primer containing a second sequencing index which represents the antibody used in the second ChIP or the input control samples. After DNA sequencing and mapping analysis, the spike-in non-mammalian DNA counts were used to normalize the ChIP signals. The signals from the first ChIP were quantitatively compared to the signals from the re-ChIP assays (Fig.1).

Example 5 - Input ChIP-seq analysis

[0036] ChIP-seq for the input samples was analyzed using the antibodies against target protein A, target protein B, target protein C, and target protein D. Peak calling program MACS2 was used to find the binding peaks (with $q<0.01$) for each factor in each reaction. The common peaks for each factor derived from two independent ChIP-seq assays were kept for further analyses. Binding sites of 26,776, 11,506, 7,191, and 4,455 for target protein A, target protein C, target protein B and target protein D, respectively, was achieved. The binding signals for each factor in two replicates were highly consistent; with Pearson coefficients as 0.99 for target protein A, target protein C or target protein B; 0.97 for target protein D.. 7,175 target protein B binding sites (99.8%) are bound by target protein A, indicating target protein A and target protein B were co-localized at chromatin. The majority

of target protein C binding sites (10,608, 92.2%) were shared with the target protein A binding sites. Most target protein D binding sites were bound by target protein A (3,882, 87.1%); 573 target protein D sites (12.9%) are not associated with target protein A binding. Motif analysis for the DNA sequences at the target protein D binding sites indicates the same motifs that are found among target protein A binding sites except the REST binding motif. When the motif analysis was performed on 573 target protein D binding sites with no target protein A binding, the major enriched motif identified was the REST motif (data not shown). Interestingly, only the REST motif appeared enriched when motif analysis was performed on the target protein D binding sites derived in a target protein A-negative MCF-7 derivative cell line. Target protein A has the strongest signal resulting the highest number of binding sites identified; target protein C and target protein B showed similar binding intensities overall from their respective antibodies; target protein D signal was the weakest among these four factors, resulting the smallest number of binding sites identified. Since a portion of target protein D binding sites are not shared with target protein A, these sites are not bound with target protein C or target protein B.

[0037] The ChIP signal intensities between each two-factor pair, respectively, were compared. Among 7,175 common binding sites for target protein A and target protein B, the correlation coefficient is 0.98. It is not surprising that target protein A and target protein B binding are highly correlated since target protein B is a bona fide coactivator of target protein A and the high correlation reflects direct protein interaction between target protein A and target protein B. The correlation coefficient between target protein A and target protein D binding among 3,098 common binding sites for target protein A and target protein D is 0.83, which was likely contributed at least in part by the overall weaker target protein D ChIP signals and indirect interaction between target protein A and target protein D since the correlation between target protein B and target protein D is a little higher (0.87) for the same binding sites. Nevertheless, target protein D is likely a component within the target protein A transcriptional complex at chromatin in addition to target protein B. But, the correlation between target protein A and target protein C binding signals at their common binding sites are weak (correlation coefficient 0.73); it is even weaker between target protein C and target protein B (0.65) or between target protein C and target protein D (0.56), despite the fact that target protein C binding intensities are similar to the ones from target protein B and stronger than the ones from target protein D.

[0038] Furthermore, the common binding sites between target protein A and each other factor were separated into 20% segments ranked by the target protein A binding strength and compared their correlations within each segment. Among the top 20% target protein A binding sites shared with target protein B, the correlation between target protein A and target protein B is close to perfect within one set of samples or across the two different sets of samples (0.97 to 0.99). With the target protein A binding signals falling, their correlation coefficients slowly drop to the lowest level at 0.63. The correlations between target protein A and target protein D are similar except that they start with a little lower values (0.82 to 0.87 for the top 20% binding sites) and their correlation drop a little faster with the decline of target protein A binding signal, which are consistent to the fact that target protein D has lower binding signals than target protein B.

[0039] Interestingly, among the top 20% target protein A binding sites shared with target protein C, their correlation is the lowest (0.67 to 0.70) comparing to the ones between target protein A and target protein B or between target protein A and target protein D, despite the fact that target protein C binding intensity is much stronger than target protein D. In addition, among the remaining 80% target protein A binding sites, their correlations remain stably low (0.33-0.48). Therefore, there are significant differences between the target protein A/target protein B/target protein D complexes and the target protein C complexes occupying the same genomic loci. The target protein A complex might not be the only complex at the target protein A binding sites.

Example 6 - Increased sensitivity for genomic binding sites detection after re-ChIP

[0040] The consistency between two independent re-ChIP assays was found to be very high, with correlation coefficients for the sequential target protein A, target protein C, target protein B or target protein D ChIPs as 0.99, 0.97, 0.98 or 0.94, respectively (Fig. 3A). After the re-ChIP, increased numbers of peaks for each antibody were detected, largely due to the increased signal/noise ratio after the re-ChIP (data not shown).

[0041] Furthermore, most of the peaks identified in the first ChIP-seq (>98.5%) were recovered. For the binding sites identified only in the re-ChIP-seq assays, the binding sites share the similar DNA binding motifs as the ones identified in the first ChIP (not shown), indicating the validity of the additional peaks detected through re-ChIP but were too weak to be identified through the first ChIP-seq. From the first round of ChIP-seq analyses, 898 (7.80%) target protein C sites, 16 (0.22%) target protein B sites, 573 (12.86%) target protein

D sites do not overlap with the target protein A binding sites. When the target protein A binding sites derived from the target protein A re-ChIP-seq were used for comparison, 92 (0.80%) target protein C sites, 15 (0.21%) target protein B sites, 490 (11.00%) target protein D sites do not overlap the target protein A binding sites. In terms of binding sites per se, target protein C and target protein D are closer to target protein A than target protein D.

Example 7 - Distinct transcriptional complexes at the ER binding sites at chromatin

[0042] The ChIP-seq signals for the first and re-ChIP-seq samples were normalized and their quantitative relationships for all four factors were compared. The target protein C enhancer region has multiple binding sites for target protein A, target protein C, target protein B or target protein D. For the re-ChIP assay using the anti-target protein A antibody, all the input amounts for target protein A, target protein B or target protein D could be recovered; however, less than half of the target protein C inputs could be recaptured. It was similar when anti-target protein B antibody or anti-target protein D ntibody was used in the re-ChIP. On the contrary, anti-target protein C antibody in the re-ChIP could recover all the target protein C inputs, as well as those inputs for target protein A, target protein B or target protein D. These data suggest that target protein A, target protein C, target protein B and target protein D are all present in one complex at these sites and a large portion of target protein C is not associated with the target protein A complex at these loci. In other words, there are distinct complexes containing target protein C present at these sites.. In general, we could fully recover the same factor as in the input after the re-ChIP process, as indicated by their signals accumulated around the equal line (the red line indicates the equal signals between the first ChIP and the re-ChIP) with high correlations (0.98 for target protein A and target protein A-target protein A; 0.96 for target protein C and target protein C-target protein C; 0.94 for target protein B and target protein B-target protein B; 0.89 for target protein D and target protein D-target protein D; targeting antibodies used are indicated sequentially). The relatively lower correlation for target protein D ChIP and its re-ChIP is likely due to the weaker affinity for this antibody. Any target protein A, target protein B or target protein D antibody used in re-ChIP could recapture all the input signals of target protein A, target protein B and target protein D, but only partial signals from the target protein C inputs. But correlations between target protein C input and any re-ChIP using target protein A, target protein B or target protein D antibody remain very high (0.95 to 0.96) compared to the correlation 0.96 between target protein C input and target protein C re-ChIP. Therefore, genome-wide distributions between the different target protein C-containing complexes at the

most sites remain relatively constant. At some specific loci, the target protein A/target protein C/target protein B/target protein D complex could be much higher or lower than the average level compared to the other target protein C-containing complex, indicating the local chromatin environment could dictate the equilibrium between different target protein C-containing complexes. To test the noise level in the re-ChIP process, the sites bound only by target protein D, but not by target protein A, target protein B, or target protein C was checked. In general, sites bound by target protein D only yield no detective peaks or very low signals from the re- ChIP using the antibodies against target protein A, target protein C or target protein B. These validate the sequential ChIP assay. The antibody carryover from the first ChIP into the re-ChIP is therefore not an issue of concern.

References cited:

[0043] 1. Dedon PC, Soults JA, Allis CD, Gorovsky MA. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. Analytical biochemistry. 1991;197(1):83-90. Epub 1991/08/15.

[0044] 2. Mardis ER. ChIP-seq: welcome to the new frontier. Nat Methods. 2007;4(8):613-4. Epub 2007/08/01.

[0045] 3. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320-34. Epub 2013/04/16.

[0046] 4. Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, Fritz CC, Bradner JE, Guenther MG. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep. 2014;9(3):1163-70. Epub 2014/12/02.

[0047] 5. Bonhoure N, Bounova G, Bernasconi D, Praz V, Lammers F, Canella D, Willis IM, Herr W, Hernandez N, Delorenzi M. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. Genome Res. 2014;24(7):1157-68. Epub 2014/04/09.

[0048] 6. Grzybowski AT, Chen Z, Ruthenburg AJ. Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. Mol Cell. 2015;58(5):886-99. Epub 2015/05/26.

[0049] 7. Egan B, Yuan CC, Craske ML, Labhart P, Guler GD, Arnott D, Maile TM, Busby J, Henry C, Kelly TK, Tindell CA, Jhunjhunwala S, Zhao F, Hatton C, Bryant BM,

Classon M, Trojer P. An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition. PLoS One. 2016;11(11):e0166438. Epub 2016/11/23.

[0050] 8. Guertin MJ, Cullen AE, Markowetz F, Holding AN. Parallel factor ChIP provides essential internal control for quantitative differential ChIP-seq. Nucleic Acids Res. 2018;46(12):e75. Epub 2018/04/20.

[0051] 9. Daly B, Olopade OI, Hou N, Yao K, Winchester DJ, Huo D. Evaluation of the Quality of Adjuvant Endocrine Therapy Delivery for Breast Cancer Care in the United States. JAMA Oncol. 2017;3(7):928-35. Epub 2017/02/06.

[0052] 10. Lippman ME, Bolan G. Oestrogen-responsive human breast cancer in long term tissue culture. Nature. 1975;256(5518):592-3. Epub 1975/08/14.

[0053] 11. Frasor J, Danes JM, Komm B, Chang KC, Lyttle CR, Katzenellenbogen BS. Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology. 2003;144(10):4562-74. Epub 2003/09/10.

[0054] 12. Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011;145(4):622-34. Epub 2011/05/10.

[0055] 13. Onate SA, Tsai SY, Tsai MJ, O'Malley BW. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. Science. 1995;270(5240):1354-7. Epub 1995/11/24.

[0056] 14. Voegel JJ, Heine MJ, Zechel C, Chambon P, Gronemeyer H. TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. Embo j. 1996;15(14):3667-75. Epub 1996/07/15.

[0057] 15. Chen H, Lin RJ, Schiltz RL, Chakravarti D, Nash A, Nagy L, Privalsky ML, Nakatani Y, Evans RM. Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. Cell. 1997;90(3):569-80. Epub 1997/08/08.

[0058] 16. Anzick SL, Kononen J, Walker RL, Azorsa DO, Tanner MM, Guan XY, Sauter G, Kallioniemi OP, Trent JM, Meltzer PS. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science. 1997;277(5328):965-8. Epub 1997/08/15.

[0059] 17. Thenot S, Charpin M, Bonnet S, Cavailles V. Estrogen receptor cofactors expression in breast and endometrial human cancer cells. Mol Cell Endocrinol. 1999;156(1-2):85-93. Epub 1999/12/28.

[0060] 18. Carroll JS, Brown M. Estrogen receptor target gene: an evolving concept. Molecular endocrinology (Baltimore, Md). 2006;20(8):1707-14. Epub 2006/01/07.

[0061] 19. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan X, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan Y. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009;462(7269):58-64. Epub 2009/11/06.

[0062] 20. Welboren WJ, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, Stunnenberg HG. ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. Embo j. 2009;28(10):1418-28. Epub 2009/04/03.

[0063] 21. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, Ali S, Chin SF, Palmieri C, Caldas C, Carroll JS. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012;481(7381):389-93. Epub 2012/01/06.

[0064] 22. Zwart W, Theodorou V, Kok M, Canisius S, Linn S, Carroll JS. Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer. Embo j. 2011;30(23):4764-76. Epub 2011/10/18.

[0065] 23. Mohammed H, D'Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, Rueda OM, Holmes KA, Theodorou V, Robinson JL, Zwart W, Saadi A, Ross-Innes CS, Chin SF, Menon S, Stingl J, Palmieri C, Caldas C, Carroll JS. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep. 2013;3(2):342-9. Epub 2013/02/14.

[0066] 24. Ghosh MG, Thompson DA, Weigel RJ. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. Cancer Res. 2000;60(22):6367-75. Epub 2000/12/05.

[0067] 25. Sun J, Nawaz Z, Slingerland JM. Long-range activation of GREB1 by estrogen receptor via three distal consensus estrogen-responsive elements in breast cancer cells. Molecular endocrinology (Baltimore, Md). 2007;21(11):2651-62. Epub 2007/08/02.

[0068] 26. Deschenes J, Bourdeau V, White JH, Mader S. Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. J Biol Chem. 2007;282(24):17335-9. Epub 2007/04/28.

[0069] 27. Hnatyszyn HJ, Liu M, Hilger A, Herbert L, Gomez-Fernandez CR, Jorda M, Thomas D, Rae JM, El-Ashry D, Lippman ME. Correlation of GREB1 mRNA with protein expression in breast cancer: validation of a novel GREB1 monoclonal antibody. Breast cancer research and treatment. 2010;122(2):371-80. Epub 2009/10/21.

[0070] 28. Rae JM, Johnson MD, Scheys JO, Cordero KE, Larios JM, Lippman ME. GREB 1 is a critical regulator of hormone dependent breast cancer growth. Breast cancer research and treatment. 2005;92(2):141-9. Epub 2005/06/30.

[0071] 29. Payapilly A, Malliri A. Compartmentalisation of RAC1 signalling. Curr Opin Cell Biol. 2018;54:50-6. Epub 2018/05/04.

[0072] 30. Kazanietz MG, Caloca MJ. The Rac GTPase in Cancer: From Old Concepts to New Paradigms. Cancer Res. 2017;77(20):5445-51. Epub 2017/08/16.

[0073] 31. Goka ET, Lippman ME. Loss of the E3 ubiquitin ligase HACE1 results in enhanced Rac1 signaling contributing to breast cancer progression. Oncogene. 2015;34.

[0074] 32. Rosenblatt AE, Garcia MI, Lyons L, Xie Y, Maiorino C, Desire L, Slingerland J, Burnstein KL. Inhibition of the Rho GTPase, Rac1, decreases estrogen receptor levels and is a novel therapeutic strategy in breast cancer. Endocr Relat Cancer. 2011;18.

[0075] 33. Kinkley S, Helmuth J, Polansky JK, Dunkel I, Gasparoni G, Frohler S, Chen W, Walter J, Hamann A, Chung HR. reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells. Nature communications. 2016;7:12514. Epub 2016/08/18.

[0076] 34. Luna-Zurita L, Stirnimann Christian U, Glatt S, Kaynak Bogac L, Thomas S, Baudin F, Samee Md Abul H, He D, Small Eric M, Mileikovsky M, Nagy A, Holloway Alisha K, Pollard Katherine S, Müller Christoph W, Bruneau Benoit G. Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. Cell. 2016;164(5):999-1014.

[0077] 35. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, Keren-Shaul H, Mildner A, Winter D, Jung S, Friedman N, Amit I. Chromatin state dynamics during blood formation. Science. 2014;345(6199):943-9.

[0078] 36. Weiner A, Lara-Astiaso D, Krupalnik V, Gafni O, David E, Winter DR, Hanna JH, Amit I. Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. Nat Biotechnol. 2016;34(9):953-61. Epub 2016/07/28.

[0079] 37. Oesterreich S, Zhang P, Guler RL, Sun X, Curran EM, Welshons WV, Osborne CK, Lee AV. Re-expression of estrogen receptor alpha in estrogen receptor alpha-negative MCF-7 cells restores both estrogen and insulin-like growth factor-mediated signaling and growth. Cancer Res. 2001;61(15):5771-7.

[0080] 38. Greene GL, Gilna P, Waterfield M, Baker A, Hort Y, Shine J. Sequence and expression of human estrogen receptor complementary DNA. Science. 1986;231(4742):1150-4. Epub 1986/03/07.

[0081] 39. Kumar V, Green S, Stack G, Berry M, Jin JR, Chambon P. Functional domains of the human estrogen receptor. Cell. 1987;51(6):941-51. Epub 1987/12/24.

[0082] 40. Tikkanen MK, Carter DJ, Harris AM, Le HM, Azorsa DO, Meltzer PS, Murdoch FE. Endogenously expressed estrogen receptor and coactivator AIB1 interact in MCF-7 human breast cancer cells. Proc Natl Acad Sci U S A. 2000;97(23):12536-40. Epub 2000/10/26.

[0083] 41. McCloskey ML, Stoger R, Hansen RS, Laird CD. Encoding PCR products with batch-stamps and barcodes. Biochemical genetics. 2007;45(11-12):761-7. Epub 2007/10/24.

[0084] 42. Barrera-Chimal J, Andre-Gregoire G, Nguyen Dinh Cat A, Lechner SM, Cau J, Prince S, Kolkhof P, Loirand G, Sauzeau V, Hauet T, Jaisser F. Benefit of Mineralocorticoid Receptor Antagonism in AKI: Role of Vascular Smooth Muscle Rac1. J Am Soc Nephrol. 2017;28(4):1216-26. Epub 2017/01/15.

[0085] 43. 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(16):5221-31. Epub 2008/08/08.

[0086] 44. Wu Y, Zhang Z, Cenciarini ME, Proietti CJ, Amasino M, Hong T, Yang M, Liao Y, Chiang HC, Kaklamani VG, Jeselsohn R, Vadlamudi RK, Huang TH, Li R, De

Angelis C, Fu X, Elizalde PV, Schiff R, Brown M, Xu K. Tamoxifen Resistance in Breast Cancer Is Regulated by the EZH2-ERalpha-GREB1 Transcriptional Axis. Cancer Res. 2018;78(3):671-84. Epub 2017/12/08.

[0087] 45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. Epub 2012/03/06.

[0088] 46. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. Epub 2008/09/19.

[0089] 47. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89. Epub 2010/06/02.

[0090] 48. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

[0091] 49. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research. 2016;44(W1):W160-W5.

What is claimed is:

1. A method of analyzing a chromatin sample, the method comprising

(a) adding a biotin-labeled non-mammalian DNA and an anti-biotin antibody to a sample comprising chromatin fragments, wherein the biotin-labeled DNA and the anti-biotin antibody form a complex in the sample;

(b) immunoprecipitating a target binding protein in the sample with an antibody that binds the target binding protein;

(c) reverse crosslinking of DNA bound to the target binding protein and disassociating the non-mammalian DNA bound to biotin antibody; and

(d) purifying the DNA of step (c).

2. The method of claim 1, further comprising the step of (e) sequencing the purified DNA in the sample.

3. The method of claim 1, further comprising the step of quantifying the amount of the non-mammalian DNA in the sample.

4. The method of claim 3, wherein the quantifying step is performed by qPCR.

5. The method of claim 1, wherein the non-mammalian DNA is a synthetic nucleic acid sequence.

6. The method of claim 1, wherein the non-mammalian DNA is a fragment of a nucleic acid sequence derived from firefly luciferase, GFP or CAS9.

7. The method of claim 1, wherein the chromatin fragments are generated by sonication or enzyme digestion.

8. A method of analyzing a chromatin sample, the method comprising

(a) adding a biotin-labeled non-mammalian DNA and an anti-biotin antibody to a sample comprising chromatin fragments, wherein the biotin-labeled non-mammalian DNA and the anti-biotin antibody form a complex in the sample;

(b) a first step of immunoprecipitating a target binding protein in the sample with an antibody that binds the target binding protein;

(c) performing chromatin indexing using a substrate;

(d) eluting the immunoprecipitated proteins from the substrate to produce an eluted sample;

(e) reverse crosslinking of DNA bound to the target binding protein and disassociating the non-mammalian DNA bound to biotin antibody in the eluted sample;

(f) a second step of immunoprecipitating a target binding protein in the eluted sample of step (e) with an antibody that binds the target binding protein;

(g) disassociating the DNA bound to the target binding protein of step (f);

(h) purifying the DNA of step (g);

(i) sequencing the purified DNA of step (h).

9. The method of claim 8, wherein the non-mammalian DNA is a synthetic nucleic acid sequence.

10 The method of claim 8, wherein the non-mammalian DNA is a fragment of a nucleic acid sequence derived from firefly luciferase, GFP or CAS9.

11. The method of claim 8, wherein the chromatin fragments are generated by sonication or enzyme digestion.

12. The method of claim 8, wherein the chromatin indexing comprises incubating beads with the immunoprecipitated target proteins at 55°C for at least 10 minutes.

13. The method of claim 8, further comprising reserving a sample of the chromatin indexing substrate as an input sample.

14. The method of claim 13, further comprising the step of quantifying the amount of the non-mammalian DNA in the input sample.

15. The method of claim 8, further comprising the step of quantifying the amount of the non-mammalian DNA after the second immunoprecipitating step.

16. The method of claim 14 or claim 15, wherein the quantifying step is performed by qPCR.

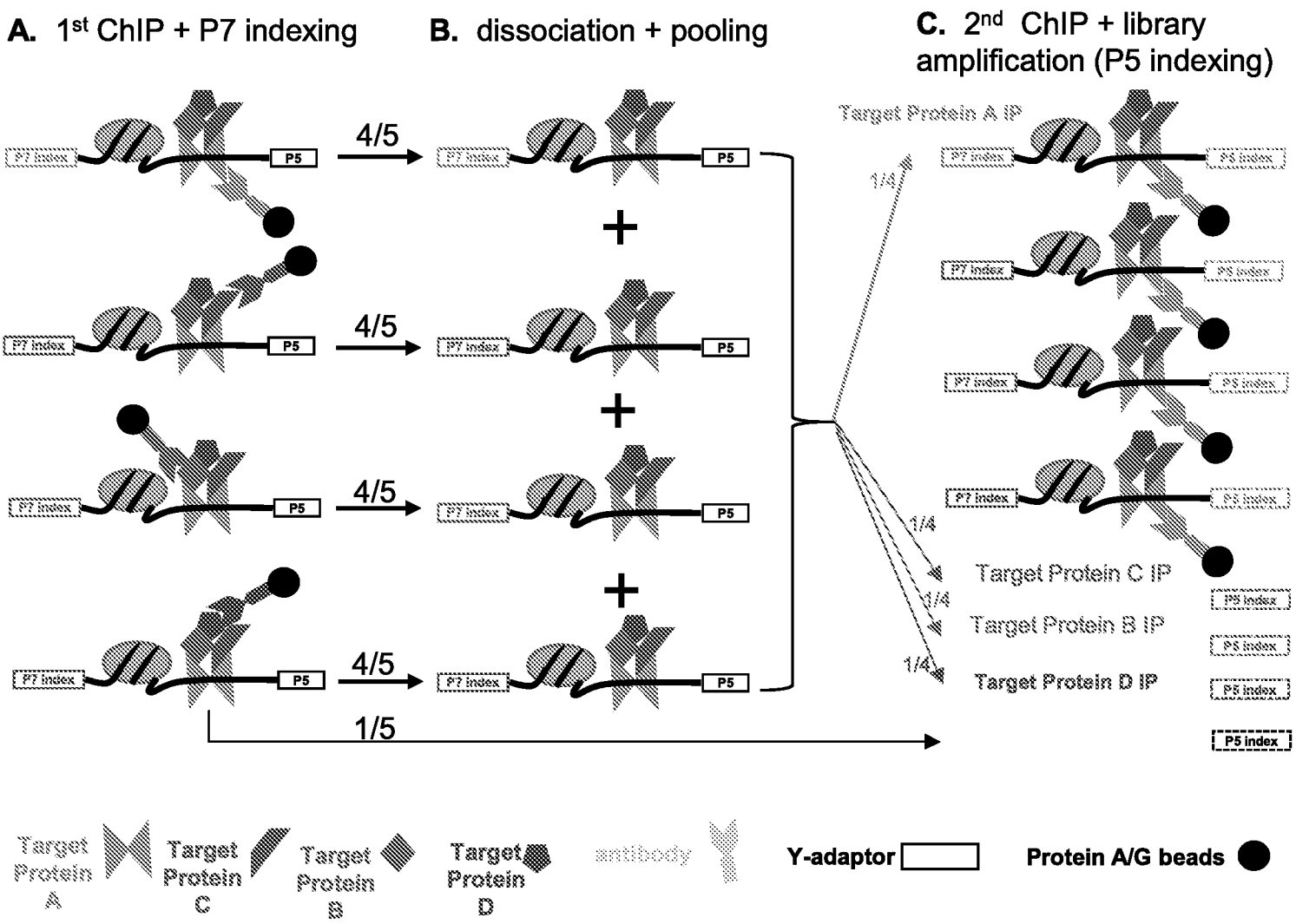

A. 1st ChIP + P7 indexing
B. dissociation + pooling
C. 2nd ChIP + library amplification (P5 indexing)
P7 index
P5
4/5
1/5
1/4
Target Protein A IP
Target Protein C IP
Target Protein B IP
Target Protein D IP
P5 index
Target Protein A
Target Protein C
Target Protein B
Target Protein D
antibody
Y-adaptor
Protein A/G beads

Figure 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/060212

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

ISA/225 was mailed on 21 November 2019. No approved electronic sequence listing was submitted in response to the ISA/225.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/060212

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 15/10; C12Q 1/68; G01N 33/536 (2020.01)
CPC - C12N 15/10; C12Q 1/6804; C12Q 1/6876; G01N 33/6875 (2020.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/6.11; 530/391.3 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2013/070167 A1 (VISA et al) 16 May 2013 (16.05.2013) entire document	1, 3-5, 7 --- 2, 6, 8-16
Y	US 2016/0341743 A1 (THE UNIVERSITY OF CHICAGO) 24 November 2016 (24.11.2016) entire document	2, 8-16
Y	US 2014/0141442 A1 (VERDELHO TRINDADE VAN GERVEN et al) 22 May 2014 (22.05.2014) entire document	6, 10
P, X	US 2018/0335424 A1 (THE TRUSTEES OF PRINCETON UNIVERSITY) 22 November 2018 (22.11.2018) entire document	1-16
A	WO 2015/159295 A1 (YEDA RESEARCH AND DEVELOPMENT CO. LTD.) 22 October 2015 (22.10.2015) entire document	1-16

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 January 2020	20 FEB 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774