Published in final edited form as:
Curr Protoc Mol Biol. 2010 October ; CHAPTER: Unit–21.20. doi:10.1002/0471142727.mb2120s92.

Genome-wide Location Analysis by Pull Down of in vivo Biotinylated Transcription Factors

Aibin He[1] and **William T. Pu**[1]

[1]Department of Cardiology, Children's Hospital Boston, Harvard Medical School, Boston, MA and Harvard Stem Cell Institute, Harvard University, Boston, MA

Abstract

Recent development of methods for genome-wide identification of transcription factor binding sites by chromatin immunoprecipitation (ChIP) has led to novel insights into transcriptional regulation and greater understanding of the function of individual transcription factors (Farnham, 2009). Chromatin immunoprecipitation requires highly specific antibody against the transcriptional regulator of interest, and availability of suitable antibodies is a significant impediment to broader application of this approach. This limitation can be circumvented by tagging the transcriptional regulator of interest with a short bio epitope, which is specifically biotinylated by the E. coli enzyme BirA. The biotinylated transcription factor can then be selectively pulled down on streptavidin beads under stringent conditions. This unit provides a detailed protocol for genome-wide location analysis of in vivo biotinylated transcription factors by streptavidin pull-down followed by high throughput sequencing (bioChIP-seq).

Introduction

Genome-wide identification of transcription factor binding sites provides insights into mechanisms of transcriptional regulation and an unbiased approach to identify genes and enhancers directly regulated by an individual transcription factor (Farnham, 2009). The standard approach involves immunoprecipitation of crosslinked chromatin (ChIP). The critical reagent for chromatin immunoprecipitation is the immunoprecipitating antibody. However, antibodies suitable for chromatin immunoprecipitation are not available for many transcription factors. In addition, an antibody may have non-specific binding that is idiosyncratic to the particular antibody preparation, and the need for a different antibody for each factor complicates comparison between factors. An alternative approach is to express the transcription factor fused to a short *bio* peptide tag, which is specifically biotinylated by the E. coli enzyme BirA (Beckett et al., 1999; de Boer et al., 2003). The transcription factor can then be pulled down with high affinity under stringent conditions using the biotin binding protein streptavidin. DNA bound to the tagged transcription factor can subsequently be identified by microarray hybridization ("bioChIP-chip") or high throughput sequencing ("bioChIP-seq"), thereby providing the transcription factor binding locations genome-wide (Kim et al., 2008).

Correspondence to: William T. Pu.

Internet Resources with Annotations

Sole-Search
Web-based tool for read alignment, peak calling, and peak annotation.
http://chipseq.genomecenter.ucdavis.edu/cgi-bin/chipseq.cgi

Primer3
Web-based tool for designing PCR primers.
http://frodo.wi.mit.edu/primer3/

In the basic protocol, we describe how to perform bioChIP-seq starting from cells expressing the biotinylated transcription factor. This protocol is similar in principle to ChIP protocols presented in Units 21.3, 21.9, and 12.13, but differs in the conditions used for pull-down on streptavidin beads versus immunoprecipitation. Additionally, the protocol is optimized for identification of pulled down DNA by high throughput sequencing rather than by PCR or microarray hybridization. The output from the basic protocol is purified DNA enriched for fragments bound by the transcription factor. In a support protocol, we describe how to convert this DNA into a library suitable for sequencing on an Illumina Genome Analyzer 2. We also provide support protocols for expression of in vivo biotinylated transcription factors, and validation of bioChIP-seq results by bioChIP followed by quantitative PCR (bioChIP-qPCR).

Basic Protocol: bioChIP-seq

This unit starts with adherent cultured cells that express the transcription factor of interest. Protein-DNA complexes are stabilized by crosslinking with formaldehyde. Complexes containing the protein of interest are then isolated by pull down on streptavidin beads followed by extensive washing. Purified DNA is recovered after crosslink reversal.

Materials List

- Cell type of interest, grown under appropriate cell culture conditions in appropriate medium, and expressing BirA and the *bio* tagged transcription factor.
- 37% formaldehyde (Fisher Scientific, Cat# F79-500)
- 2.5 M glycine (American Bioanalytic, Cat# AB00730)
- PBS
- PBS/1% BSA
- Hypotonic Buffer (see recipe)
- ChIP Dilution Buffer (see recipe)
- SDS Wash Buffer (see recipe)
- High Salt Buffer (see recipe)
- LiCl Buffer (see recipe)
- Low Salt Buffer (see recipe)
- TE Buffer (see recipe)
- SDS ChIP elution buffer (see recipe)
- Proteinase K, DNase-free (20 mg/ml)
- RNase A, DNase-free (10 mg/ml)
- Protein A magnetic beads (Invitrogen)
- M-280 Streptavidin magnetic beads (Invitrogen)
- Roche Protease Inhibitor Cocktail (Cat No. 11697498001)
- Quant-It PicoGreen dsDNA DNA reagent (Invitrogen)
- Cell Lifter
- 15 and 50 ml conical centrifuge tubes

- 1.7 ml microfuge tubes
- Siliconized non-stick microfuge tubes
- 15 cm tissue culture dishes
- Tissue culture centrifuge
- Refrigerated microfuge
- Cell culture incubator
- Culture incubator (37°C, 5% CO2)
- Nanodrop spectrophotometer
- Water baths: 37 °C, 55 °C, 70°C water bath
- Heat block: 100 °C
- Glass Dounce homogenizer (2-ml size, Fisher Cat# K885300)
- QIAquick PCR purification kit (Qiagen)
- QIAquick Gel Extraction kit (Qiagen)
- Magnetic Stand (Ambion, Cat# am10055)
- Nutator (VMR)
- Misonix Sonicator 4000 with microtip, Part# 418 (Qsonica, LLC)
- Applied Biosystems 7500 Real-Time PCR system or equivalent
- Agilent Bioanalyzer

Protocol Steps

Crosslink and sonicate chromatin complexes—

1. Obtain cells expressing the biotinylated transcription factor of interest.

   Typically, 4×10^7 cells (2 × 15 cm dishes at near confluence; will vary by cell type) will yield good signal to noise ratio for bioChIP-seq. This number will need to be modified depending on the transcription factor, its expression level, and the cells being studied. Using less cells is possible but may be accompanied by decreased signal to noise ratio.

   In addition to cells expressing the epitope-tagged protein and BirA, include a negative control expressing BirA only.

2. Remove dishes from incubator and place on the bench. Add formaldehyde to the media to a final concentration of 1% (v/v). Incubate at room temperature for 7 minutes.

   The indicated conditions represent a general starting point. Crosslinking conditions can vary significantly and both concentration and time of formaldehyde treatment may need individual optimization. Excessive crosslinking increases background.

3. Neutralize formaldehyde by adding glycine to a final concentration of 125 mM (1 ml 2.5M glycine per 20 ml media). Incubate at room temperature for 5 minutes, shaking plates occasionally.

4. Pour media off completely and rinse cells with cold PBS three times. Do not allow the cells to become dry. It is important to remove all fixative by washing well.

5. Add 5 ml cold PBS to each 15 cm dish. Use a cell lifter to scrape the cells off the dish. Transfer to a 50 ml centrifuge tube on ice. Add an additional 5 ml cold PBS, scrape one more time, and pool in the same 50 ml tube.

6. Centrifuge the cells at 3000 × *g* for 10 minutes at 4 °C.

7. Aspirate as much of the cold PBS as possible.

   Pause point: At this point, cells can be snap frozen in liquid nitrogen and stored at −80 °C for several months.

8. Using a P-1000 pipet, resuspend the cell pellet in 2 ml Hypotonic Buffer per ~4 × 10^7 cells. Leave tubes in ice for 15 minutes to permit hypotonic lysis. If cells clumps are seen, homogenize the suspension with five strokes in a pre-chilled dounce homogenizer. The need for homogenization depends on the cell type.

9. Transfer 1 ml cell suspension to each of two pre-chilled microfuge tubes. Centrifuge 13000 × *g* for 1 minute at 4 °C. Aspirate as much of the supernatant as possible.

10. Resuspend the pellet in each microfuge tube in 0.5 ml ChIP Dilution Buffer. Pool suspensions from the two microfuge tubes. Sonicate under conditions determined in pilot experiments to yield fragments with average length of ~150–200 bp (see Support Protocol).

    Practical aspects of how to sonicate the sample depends on the specific sonicator setup. We use a Misonix 4000 sonicator, with the sonicator tip at a fixed height in a sound reducing chamber, located in a cold room. The sample is placed in a 15 ml conical polypropylene centrifuge tube with the top half removed. This cut centrifuge tube is placed in an ice water bath on a height-adjustable sample stand. The sample is then carefully positioned so that the tip is 3 mm from the bottom of the tube and the probe is not contacting the side of the tube.

11. After sonication, transfer cell lysates to microfuge tubes and spin at 20,000 × *g* for 15 min at 4 °C to pellet the cell debris.

12. The supernatant is the sheared chromatin. Transfer it to 2 new microfuge tubes. The sheared chromatin should be half-transparent and slightly yellow.

13. Estimate the sheared chromatin concentration. Pipet 1 µl of supernatant into 99 µl TE and measure the DNA concentration using a Nanodrop. Use this value to equalize the amount of DNA used for streptavidin pulldown.

14. Save a 25 µl aliquot of supernatant as the input sample by storing it at −20 °C. This will be reverse cross-linked at step 23.

Pull down biotinylated protein-DNA complexes on streptavidin—

15. Prepare 50 µl Protein A-magnetic beads and 50 µl streptavidin-magnetic beads for each ChIP reaction from ~4 × 10^7 cells. Wash the beads three times with PBS/1% BSA, then block the streptavidin beads by incubating with 500 µl PBS/1% BSA for at at least one hour in the cold room. This step can be done while samples are being sonicated.

16. Preclear the sheared chromatin: Add the Protein A beads prepared in step 15 to each sample and incubate for 1 hour in the cold room. This step removes proteins that bind non-specifically to the bead matrix.

17. Briefly spin (allow the microfuge to reach 6000 rpm then stop) the sheared chromatin samples containing Protein A beads and place on magnetic stand for 30 seconds. Transfer the supernatant containing the sheared chromatin to a new siliconized, pre-cooled microfuge tube. Discard the Protein A beads.

18. Briefly spin the blocked streptavidin beads prepared in step 15. Place on magnetic stand for 30 seconds. Decant the supernatant and resuspend the beads in 50 µl PBS/1% BSA. Add these beads to the sheared chromatin samples and incubate in the cold room for 1 hour to overnight.

19. Briefly centrifuge the samples and place on magnetic stand for 30 seconds. Pipet out the supernatant as much as possible using a P-10 pipet. Removing as much wash solution as possible helps to reduce background.

20. Sequentially incubate with 1 ml wash solution indicated below at room temperature for 5 minutes. After each incubation, briefly centrifuge, place on magnetic stand, and pipet out as much supernatant as possible, as in step 19. Proceed with next wash as indicated:

    a. SDS Wash Buffer

    b. SDS Wash Buffer

    c. High Salt Buffer

    d. LiCl Buffer

    e. TE Buffer

21. Add 1 ml TE Buffer. Resuspend beads and transfer to a new siliconized microfuge tube. Incubate 5 minutes, briefly centrifuge, place on magnetic stand, and pipet out as much supernatant as possible. This second TE wash in a new tube helps to reduce nonspecific background.

Elute and purify bioChIP DNA—

22. Pellet beads and place on magnetic stand. Remove supernatant. To remove as much wash buffer as possible, briefly centrifuge again and replace on magnetic stand. Pipet off as much residual supernatant as possible.

23. To reverse crosslinks, add 100 µl of SDS ChIP elution buffer to the pelleted beads and place in 70 °C waterbath overnight. At the same time, thaw the saved input chromatin aliquot from step 14, add 175 µl of SDS ChIP elution buffer, and reverse crosslink along with pulldown samples.

24. Remove samples from 70 °C water bath and cool to room temperature. Briefly spin, replace on magnetic stand, and decant the supernatant into a new siliconized microfuge tube. Add 100 µl TE buffer to the beads, vortex vigorously for 10 seconds and combine the supernatant with the first eluate. Add 2 µl Proteinase K (20 mg/ml) to each of the bioChIP and input chromatin samples, and incubate for 2 hours at 55 °C.

25. Add 2 µl RNase A (10 mg/ml) for 1 hour at 37 °C.

26. Purify the bioChIPed DNA using a QIAquick PCR purification kit following the manufacturer’s protocol. Eluted the bioChIPed DNA in 50 µl of the kit’s EB

buffer (10 mM Tris pH 8.0) into a siliconized microfuge tube. This is the purified bioChIP DNA. At this point the bioChIP material can be validated by qPCR (see Support Protocol) or used to build a high throughput sequencing library (see Support Protocol for construction of an Illumina GA2 ChIP-seq library).

Note: at this point, DNA concentration and yield will be too low to be quantified by Nanodrop. However, it is enough for real-time quantitative PCR and ChIP-seq library preparation. If storing the bioChIPed DNA rather than proceeding immediately to construction of the ChIP-seq library, it is critical to store the DNA in siliconized microfuge tubes to avoid loss of DNA.

Expression of in vivo biotinylated transcription factors

The method used to express biotinylated proteins of interest depend on the cell line being studied. Potential methods include development of stable cell lines, transient transfection, and transduction with appropriate viral vectors. Similarly, the design of the expression construct will depend on the factor being expressed. For example, some transcription factors may tolerate placement of the epitope tag at one terminus but not the other.

We used two constructs to permit titratable expression of biotinylated proteins of interest (Figure 1). The first, pCMV-rtTA-IRES-BirA, uses the CMV promoter to express the reverse tet activator protein (rtTA) and the E. coli protein BirA. The second construct expresses a tagged factor from a doxycycline (DOX)-regulated promoter (TRE).

We used the Gateway cloning system (Invitrogen) to facilitate placing the gene of interest into the TRE plasmid. Briefly, the flag-bio tag was cloned into a gateway entry cassette in the plasmid pENTR-flbio. An ORF encoding the gene of interest (without the stop codon) is then PCR amplified and cloned between unique restriction sites in this plasmid, yielding a C-terminally tagged ORF. The Gateway LR Clonase is then used to transfer the tagged gene of interest into pTRE-gateway to yield the TRE-driven expression plasmid (generically called pTRE-GOI^{flbio}). pCMV-rtTA-IRES-BirA and pTRE-GOI^{flbio} can be introduced into cells as dictated by the model system. For example, the constructs can be introduced by stable transfection, or by lentiviral or adenoviral expression. Pilot experiments are then performed to optimize the level of expression by adjusting the dosage of doxycycline.

The pCMV-rtTA-IRES-BirA, pENTR-flbio, and pTRE-gateway constructs are available through Addgene.

Validation by bioChIP-qPCR

Quantitative PCR (qPCR) is an effective method to analyze DNA pulled down by bioChIP. bioChIP-qPCR can be used to validate a pull-down sample or a bioChIP library prior to submission for high throughput sequencing. bioChIP-qPCR can also be used to define the false positive and false negative rate for a set of peak calls obtained by bioChIP-seq. This support protocol does not provide detailed information about how to set up and run a qPCR experiment, but provides information about the application of qPCR to bioChIP.

For bioChIP-qPCR, take 1 μl of each bioChIPed DNA that results from the main bioChIP protocol as qPCR template. Negative control samples are the input chromatin after crosslink reversal, and bioChIP of the sample expressing BirA but not a *bio* epitope tagged protein. Primers should be designed to the genomic region of interest. A pair of primers to a region of the genome not anticipated to be pulled down in the bioChIP reaction should be used as an internal control to monitor the extent of non-specific pulldown; we used a region in the

first intron of the murine *Actb* gene (foward, 5'-CGTATTAGGTCCATCTTGAGAGTACACAGTATT-3'; reverse, 5'-GCCATTGAGGCGTGATCGTAGC-3'). This region is refer to as Actb-1st intron in the next steps. Because the average fragment length in bioChIP-seq libraries is 100–300 bp (200–400 bp minus the size of the ligated linker, which are 92 bp), we recommend designing primers so that amplicons are less than 100 bp. We used Primer3 with optimal Tm 60 ± 1 °C and primer length 20 ± 2 bp.

To determine the fold enrichment of each peak, perform the following calculations:

$$\text{ChIPed DNA}=\frac{2^{-\text{Cl (ChIPed peak)}}}{2^{-\text{Ct (ChIPed neg ctrl)}}}$$

$$\text{Input DNA}=\frac{2^{-\text{Cl (Input peak)}}}{2^{-\text{Ct (Input neg ctrl)}}}$$

$$\text{BirA DNA}=\frac{2^{-\text{Ct (BirA peak)}}}{2^{-\text{Cl (BirA neg ctrl)}}}$$

$$\text{Fold Enrichment}=\frac{\text{ChIPed DNA} - \text{Input DNA}}{\text{BirA DNA} - \text{Input DNA}}$$

Support Protocol 1: Optimization of Sonication Conditions

Proper fragmentation of chromatin is essential to achieving good results (see critical parameters below). On most sonicators, the variables that can be adjusted are amplitude, pulse on time, rest time, and cycle number. Specific values that we provide are for the Misonix 4000 with microtip, and will need to be determined empirically depending on the sonicator setup. Input cell number, sample volume, and sonication buffer components all affect sonicator settings, and therefore pilot experiments should use parameters that reflect those that will be used in actual experiments.

Amplitude is the percent of maximal power. Too low power will lead to long sonication times and incomplete fragmentation. Excessive power leads to foaming. We use 70% amplitude. Pulse on time is the time per cycle during which sonication occurs. Excessive pulse on time causes local heating, which damages chromatin. We use a pulse on time of 15 seconds. Rest time is the time per cycle during which no sonication occurs. This is required for dispersal of heat. We use a rest time of 1 minute. Cycle number is the number of sonication cycles. After establishing the other settings, cycle number is the main variable that we adjust depending on different cell types. We typically use between 20–32 cycles.

All sonication is performed in the cold room in an ice water bath. The sonicator tip needs to be positioned carefully as the sonication conditions depend on the volume being sonicated and the position of the tip within the sample. Also, the tip can be damaged if it touches the sample tube.

To optimize sonication conditions, pilot experiments should be performed on the cell type of interest.

1. Proceed through steps 1–11 of the main protocol, preparing 3 chromatin samples.
2. Sonicate the chromatin samples, keeping the amplitude, pulse on, and pulse off times constant while varying the number of cycles.
3. Remove a 25 µl aliquot (analogous to the input sample) and follow steps 23–26 of the main protocol to analyze the size distribution of the sonicated chromatin.
4. Repeat steps 2 and 3, adjusting the amplitude and cycle number to achieve an average fragment size of ~100–300 bp. An example of two different sonication conditions is shown in Figure 2.

Support Protocol 2: Conversion of bioChIP-seq DNA to a library for sequencing on the Illumina Genome Analyzer 2

ChIP-seq library preparation is perfomed according to the Illumina ChIP-seq preparation kit manual, but with a few modification. Use siliconized microfuge tubes throughout this procedure.

Materials List

- ChIP-seq library preparation kit (Illumina)
- QIAquick PCR Purification Kit (Qiagen)
- QIAquick Gel Extraction Kit (Qiagen)
- MinElute PCR purification kit (Qiagen)
- Agarose
- Ethidium bromide
- 100 bp DNA ladder (New England Biolabs)
- Agarose gel electrophoresis and documentation setup
- Long wavelength UV transilluminator or Dark Reader (Clare Chemical)
- Razor blades
- Quant-it PicoGreen dsDNA Reagent (Invitrogen)
- Siliconized microfuge tubes
- Thermal cycler
- 37 °C heat block

1. Prepare 1 U/µl of Klenow DNA polymerase by 1:5 fold dilution of original polymerase in 1x DNA ligase buffer.
2. Prepare the following reaction mix and incubate in a thermal cycler for 30 minutes at 20 °C.

bioChIP DNA	40 µl
T4 DNA ligase buffer with 10mM ATP	5 µl
dNTP mix	2 µl
T4 DNA polymerase	1 µl
diluted Klenow DNA polymerase	1 µl

T4 PNK	1 µl

3. Purify on one QIAquick PCR Purification Kit column, following the manufacturer's instructions. Elute in 34 µl of the kit's EB buffer (10 mM Tris pH 8).

4. Prepare the following mix and incubate for 30 minutes at 37 °C

DNA sample (from last step)	34 µl
Klenow buffer	5 µl
dATP	10 µl
Klenow exo$^-$ (3' to 5' exo minus)	1 µl

5. Purify on one QIAquick PCR Purification Kit column, following the manufacturer's instructions. Elute in 34 µl of the kit's EB buffer.

6. Dilute the Adapter oligo mix 1:20 with water to adjust for the smaller quantity of DNA. Prepare the following reaction mix and incubate for 15 minutes at room temperature.

DNA sample	10 µl
T4 DNA ligase buffer	15 µl
Diluted adapter oligo mix	1 µl
T4 DNA ligase	4 µl

7. Purify on one MinElute PCR Purification Kit column, following the manufacturer's instructions. Elute in 10 µl of the kit's EB buffer. Then add another 10 µl EB buffer and elute a second time. Combine the two eluates.

8. Prepare a 100 ml, 2% agarose gel with 1×TAE buffer. Just before pouring the gel, add ethidium bromide (EtBr) to 400 ng/ml.

9. Add 2 µl of 10x DNA loading buffer and an additional 2 µl 80% glycerol to the DNA from the purified ligation reaction.

10. Load the gel. In one lane, load 500 ng of 100 bp DNA ladder. Load the entire sample in another lane, leaving at least one empty lane between the ladder and the sample.

11. Run gel slowly for best separation from free adaptors (e.g. 80 V for 4 hours). Free adapter will compete with ChIP DNA for PCR amplification and can give rise to spurious PCR products that consume sequencing capacity.

12. View the gel with a minimum of UV exposure. For example, use a long wavelength UV transilluminator or a Dark Reader transilluminator. Short wavelength UV damage impairs the efficiency that ChIP DNA can be PCR amplified and incorporated into the sequencing library.

13. Using a clean razor blade, excise a gel slice from the sample lane between ~200–400 bp (corresponding to the peak chromatin size plus 92 bp of ligated adapters). No band will be evident due to low sample concentration (see Figure 3). Photograph the gel before and after cutting out the gel slice.

14. Cut a slice of the same size from an empty well on the same gel and take this sample through gel purification and PCR. No visible PCR product should be present.

15. Recover DNA from the gel slices using the Qiagen Gel Extraction Kit. Elute DNA in 36 µl EB buffer.

16. Prepare the following PCR reaction mix:

DNA	36 µl
5x Phusion buffer	10 µl
dNTP mix (1.5 µl) PCR primer 1.1	1.5 µl
Illumina PCR primers 1.1	1 µl
Illumina PCR primers 2.1	1 µl
Phusion polymerase	0.5 µl

Note: To save the reagent cost, we found that the PCR primers can be diluted 2-fold with water without significantly influencing the bioChIP-seq library. The total volume should be 50 µl.

17. Amplify using the following PCR protocol:

30 seconds at 98°C

18 cycles of 10 seconds at 98 °C, 30 seconds at 65 °C, 30 seconds at 72 °C

5 minutes at 72 °C

Hold at 4 °C

18. Purify on one MinElute PCR Purification Kit column, following the manufacturer's instructions. Elute in 15 µl of the kit's EB buffer.

19. Purify again on a 2% agarose gel, following steps 8–14 of this protocol. This time a band should be present that corresponds to the ChIP-seq library (Figure 3). Purify DNA from the gel slice using the MinElute Gel Extraction Kit. Elute in 10 µl of the kit's EB buffer. Add another 10 µl and repeat. Pool the two eluates.

Note: This second gel purification step is essential to remove primer dimers and excess oligonucleotides. Running the gel at low voltage for several hours will enhance separation of the desired product away from contaminants.

20. The DNA is ready for quantitation, quality control, and sequencing with the Illumina GA2. Accurate measurement of library concentration is essential for properly loading the Illumina sequencer. Picogreen measurement of library concentration is a convenient and sensitive quantitation method. Real time quantitative PCR against a standard reference sample is another method. Analyzing the size distribution of the library on a Bioanalyzer is also recommended prior to submitting the sample to the sequencing core.

Note: We recommend storing the bioChIP-seq library in siliconized microfuge tubes and sequencing the sample as soon as possible, preferably within 2–3 weeks of library preparation.

Reagents and Solutions

Hypotonic Buffer

20 mM HEPES pH 7.5

10 mM KCl

1 mM EDTA

0.1 mM activated Na_3VO_4

0.2% (vol/vol) Nonidet P40 (NP-40)

10% (vol/vol) glycerol

Store at 4 °C for several months

At time of use, add fresh DTT to 1 mM, PMSF to 1 mM, and concentrated protease inhibitor cocktail (1:1,000).

ChIP Dilution Buffer

20 mM TrisCl (pH8.0)

2 mM EDTA

150 mM NaCl

0.1% SDS

1% Triton X-100

Store at room temperature for several months

At time of use, add concentrated protease inhibitor cocktail (1:50).

SDS Wash Buffer

2% SDS, prepared from 20% SDS stock solution by dilution in nuclease-free water.

Store at room temperature for several months

High Salt Buffer

50 mM HEPES (pH 7.5)

500 mM NaCl

1 mM EDTA

0.1% sodium deoxycholate

1% Triton X-100

Store at room temperature for several months

LiCl Buffer

10 mM Tris-HCl (pH 8.1)

250 mM LiCl

1 mM EDTA

0.5% (vol/vol) NP-40

0.5% (vol/vol) sodium deoxycholate

Store at room temperature for several months

TE buffer

10 mM Tris-HCl (pH 7.5)

1 mM EDTA

Store at room temperature for several months

SDS elution buffer

50 mM Tris-HCl (pH 8.1)

10 mM EDTA

1% (vol/vol) SDS

Store at room temperature for several months

Concentrated Protease Inhibitor Cocktail

1 Roche Protease Inhibitor Cocktail (Cat No. 11697498001) tablet dissolved in 1ml dH_20.

Store at −20 °C for 1 month

Commentary

Background Information

Chromatin immunoprecipitation has become the method of choice to identify transcription factor chromatin occupany in vivo, and genome-wide application of this approach by hybridization-based (ChIP-chip) and high throughput sequencing-based (ChIP-seq) technologies has fueled rapid advances in the understanding of transcriptional regulation. While hybridization-based approaches continue to be useful for targeted analyses, sequencing based approaches are both less expensive and have greater dynamic range. Therefore, ChIP-seq has become the method of choice for unbiased genome-wide studies.

One barrier to application of ChIP approaches is availability of “ChIP-grade” antibodies that efficiently and specifically pull-down the protein of interest even after treatment with formaldehyde. This limitation can be circumvented by placing an epitope tag on the protein of interest, which allows different factors to be pulled down by a uniform method, facilitating comparisons between factors. The *bio* tag is a 22 amino acid sequence that is specifically biotinylated by the E. coli enzyme BirA. In BirA-expressing cells, *bio*-tagged proteins become biotinylated, allowing them to be pulled down with extraordinary affinity onto immobilized streptavidin. de Boer, Strouboulis, and colleagues used the bio tag for single step affinity purification of transcription factors from mammalian cells (de Boer et al., 2003), and Orkin and colleagues applied the technique to genome-wide location analysis (Kim et al., 2008).

Critical Parameters

Expression of tagged protein—The strategy for expressing the biotinylated protein of interest depends on the gene transfer technology that is optimal for the model system under study. Important factors to consider are expression of the correct protein isoform, the fraction of cells that express the biotinylated protein (i.e., gene transfer efficiency), and the level of expression in expressing cells. Obviously, when a gene can be expressed as multiple different isoforms as a result of alternative promoter usage or alternative splicing, it is desirable to select the isoform relevant to the system under study. Obtaining high gene transfer efficiency improves signal to noise and decreases the number of input cells required, since only expressing cells contribute to the ChIP signal but all cells contribute to noise. Within expressing cells, it is desirable to achieve expression at near endogenous levels, as marked overexpression may lead to occupancy of non-physiological binding sites. Using a titratable expression system, such as that based on the reverse tet-activator, facilitates optimization of expression level.

Chromatin preparation—To fix dynamic protein-DNA complexes under the conditions being studied investigation, it is desirable to crosslinking proteins to DNA with a minimum of intervening manipulations. While in some cases FACS sorting may be required to isolate the population of interest, it is possible that chromatin occupancy measured after long FACS procedures may not be representative of the condition of interest. If cells need to be frozen prior to proceeding with bioChIP, certainly it is desirable to fix the cells prior to freezing. The extent of crosslinking is also an important parameter. Under crosslinking reduces signal, while over crosslinking can impair chromatin fragmentation and increase non-specific background.

Adequate fragmentation of chromatin to a mean fragment size of 150–300 bp is crucial for success of bioChIP-seq. Insufficient fragmentation leads to inefficient use of input material. More importantly, it reduces resolution of bioChIP-seq peaks and increases background. Sonication conditions vary by target cells and growth conditions. Therefore, pilot experiments should be performed to empirically optimize sonication conditions prior to proceeding with preparation of bioChIP-seq samples.

Controls—Two essential controls for bioChIP-seq experiments are input chromatin and chromatin pull down in the presence of BirA but the absence of *bio*-tagged protein. Input chromatin consists of sonicated chromatin prior to pull-down. This sample establishes the background distribution of chromatin fragments, which is not uniform because of non-random chromatin fragmentation by sonication. The BirA sample controls for background caused by endogenously biotinylated proteins and non-specific protein binding to streptavidin beads.

High Throughput Sequencing—36 bp single end or paired end sequencing can be used for bioChIP-seq. Paired end sequencing improves mappability of reads and directly yields the pulled down genomic fragment. With single end sequencing, a greater fraction of reads cannot be uniquely mapped. In addition, the other end of the genomic fragment must be inferred from the average length of the material incorporated into the sequencing library. This potentially reduces the precision with which bound DNA peaks can be identified. On the other hand, paired end sequencing is more expensive and many “peak caller” software packages do not make use of the paired end sequencing data.

The depth of sequencing required for adequate detection of transcription factor binding sites depends to some extent on the number of expected binding sites and how “adequate detection” is defined. Initially it was expected that a saturation point would be reached, beyond which additional sequencing would not yield substantially more binding sites. However, simulations have demonstrated that this is not the case when statistical thresholds are used (Park, 2009), perhaps because increased sequencing depth permits more sensitive detection of low affinity binding sites. However, the biological significance of such low affinity binding sites is unclear. Empirically, we found that 10–20 million mappable reads, corresponding to two to three lanes on an Illumina GA2, provided a good balance between cost and sensitivity.

Analysis—The millions of sequence reads generated by a bioChIP-seq experiment require computational analysis to identify enriched regions. In brief, sequence reads are aligned to the genome, using short read aligners such as Bowtie, Eland, Maq, and Novoalign. Then a “peak caller” scans the genome to identify areas enriched for aligned reads (Park, 2009). Different peak caller implementations are available, each with different strengths and weaknesses. A number of these were recently reviewed (Pepke et al., 2009).

For bioChIP-seq data, we have used CisGenome (Ji et al., 2008) and Sole-Search (Blahnik et al., 2010). We compared the experimental bioChIP sample to input. To control for background binding or endogenously biotinylated proteins, we used compared the BirA sample to input. Peaks from the BirA vs Input comparison are then subtracted from the bioChIP vs Input comparison. In our experience there has been little overlap between peaks called in the experimental and negative control comparisons.

Anticipated Results

Depending on the tagged transcription factor and the peak calling parameters, bioChIP-seq generally yields between 1,000–50,000 transcription factor binding sites. Important properties to experimentally define for this transcription factor binding site list are the false positive and false negative rates. The false negative rate is the frequency that "gold standard" transcription factor binding sites established by prior experimentation fail to be identified in the observed list of transcription factor binding sites. The "gold standard" list should be validated in the specific experimental system used for bioChIP-seq through bioChIP-qPCR. Conversely, the false positive rate is the frequency that transcription factor binding sites identified by bioChIP-seq fail to be validated by bioChIP-qPCR. We typically observed false positive rates of 0–25%; if the false positive rate is unacceptably high, then more stringent peak calling parameters may be necessary. On the other hand, unacceptably high false negative rates may be due to many factors, such as excessively stringent peak calling parameters, inadequate sequencing depth, or low signal or high background in the chromatin pulldown. Other properties of bioChIP-seq peaks that might validate the data are identification of the expected transcription factor binding motif among the bioChIP-seq peaks, and evolutionary conservation of the bioChIP-seq peaks.

Time Considerations

Chromatin preparation: 0.5 days

BioChIP binding and washing: 2 hours to overnight

Reverse-cross-linking and purification of ChIPed DNA: 5 hours

ChIP-seq library preparation: 1 day

Literature Cited

Beckett D, Kovaleva E, Schatz PJ. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci. 1999; 8:921–929. [PubMed: 10211839]

Blahnik KR, Dou L, O'Geen H, McPhillips T, Xu X, Cao AR, Iyengar S, Nicolet CM, Ludascher B, Korf I, Farnham PJ. Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. Nucleic Acids Res. 2010; 38:e13. [PubMed: 19906703] Describes a bioinformatic tool for read alignment, peak calling, and peak annotation.

de Boer E, Rodriguez P, Bonte E, Krijgsveld J, Katsantoni E, Heck A, Grosveld F, Strouboulis J. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A. 2003; 100:7480–7485. [PubMed: 12802011] Application the bio epitope tag for one step purification of mammalian proteins.

Farnham PJ. Insights from genomic profiling of transcription factors. Nat Rev Genet. 2009; 10:605–616. [PubMed: 19668247]

Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol. 2008; 26:1293–1300. [PubMed: 18978777]

Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008; 132:1049–1061. [PubMed: 18358816] Application of in vivo transcription factor biotinylation to genome-wide location analysis by bioChIP-chip.

Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009; 10:669–680. [PubMed: 19736561]

Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. Nat Methods. 2009; 6:S22–S32. [PubMed: 19844228]

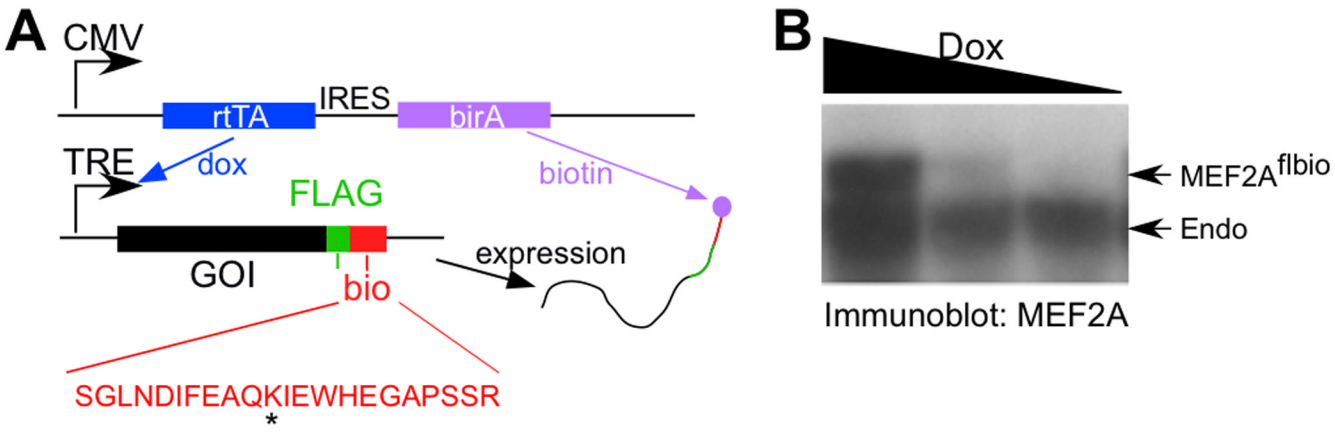

Figure 1. System for expression of biotinylated transcription factors

A. Two vector system for titratable expression of TFs with C-terminal FLAG and bio peptides. BirA recognizes and biotinylates the *bio* sequence on the central lysine residue (asterisk). **B.** Titration of tagged TF expression by adjusting the concentration of Dox. Tagged TFs are slightly larger than endogenous the corresponding endogenous protein due to the epitope tags.

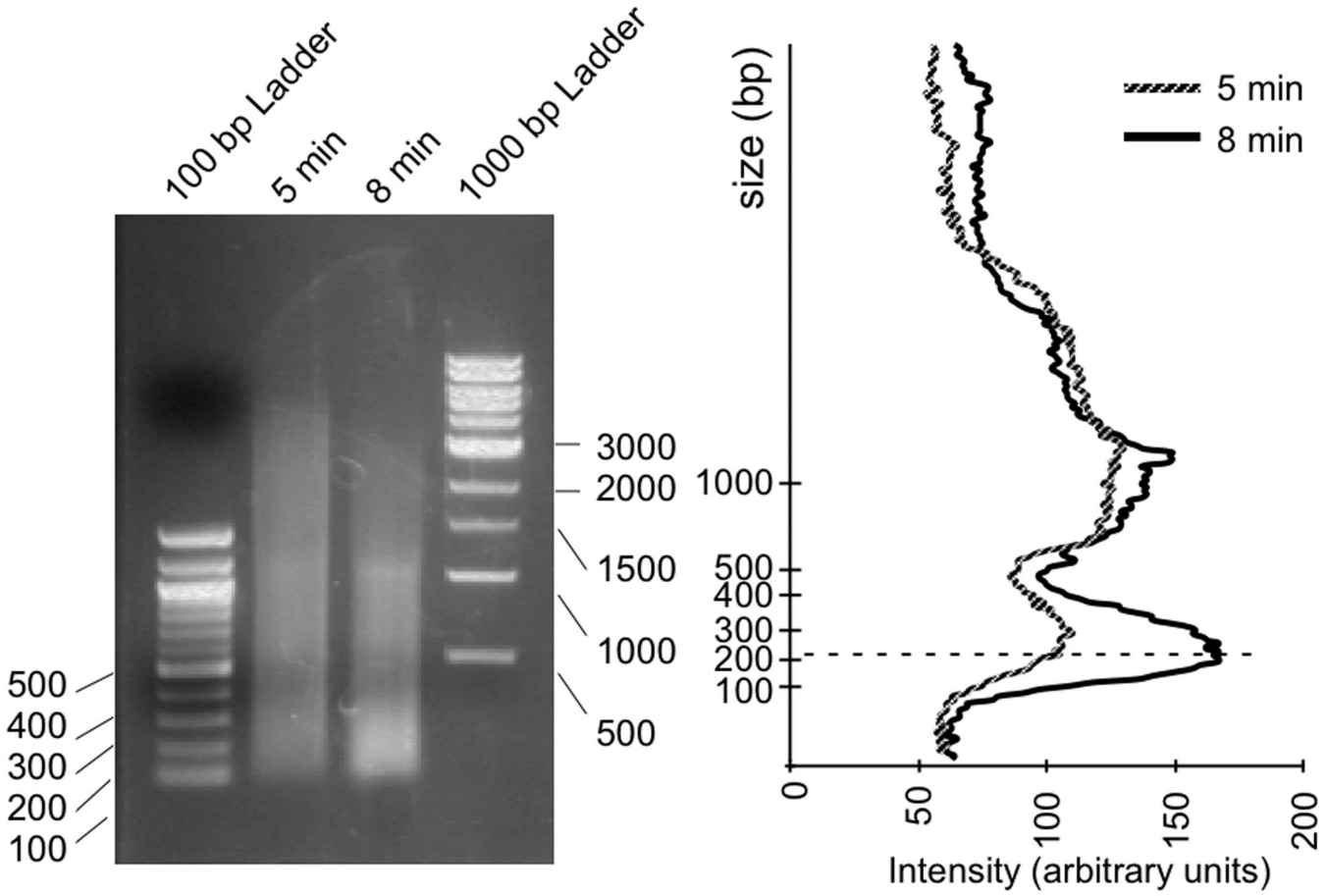

Figure 2. Optimization of Sonication Conditions

2 ml of nuclear extract was sonicated for 5 or 8 minutes using a Misonix 4000 with the amplitude of 70, cycled 15 sec on and 1 min off. Plot on right shows size versus intensity profile of sonicated chromatin. 8 minutes led to the desired size peak of ~200 bp.

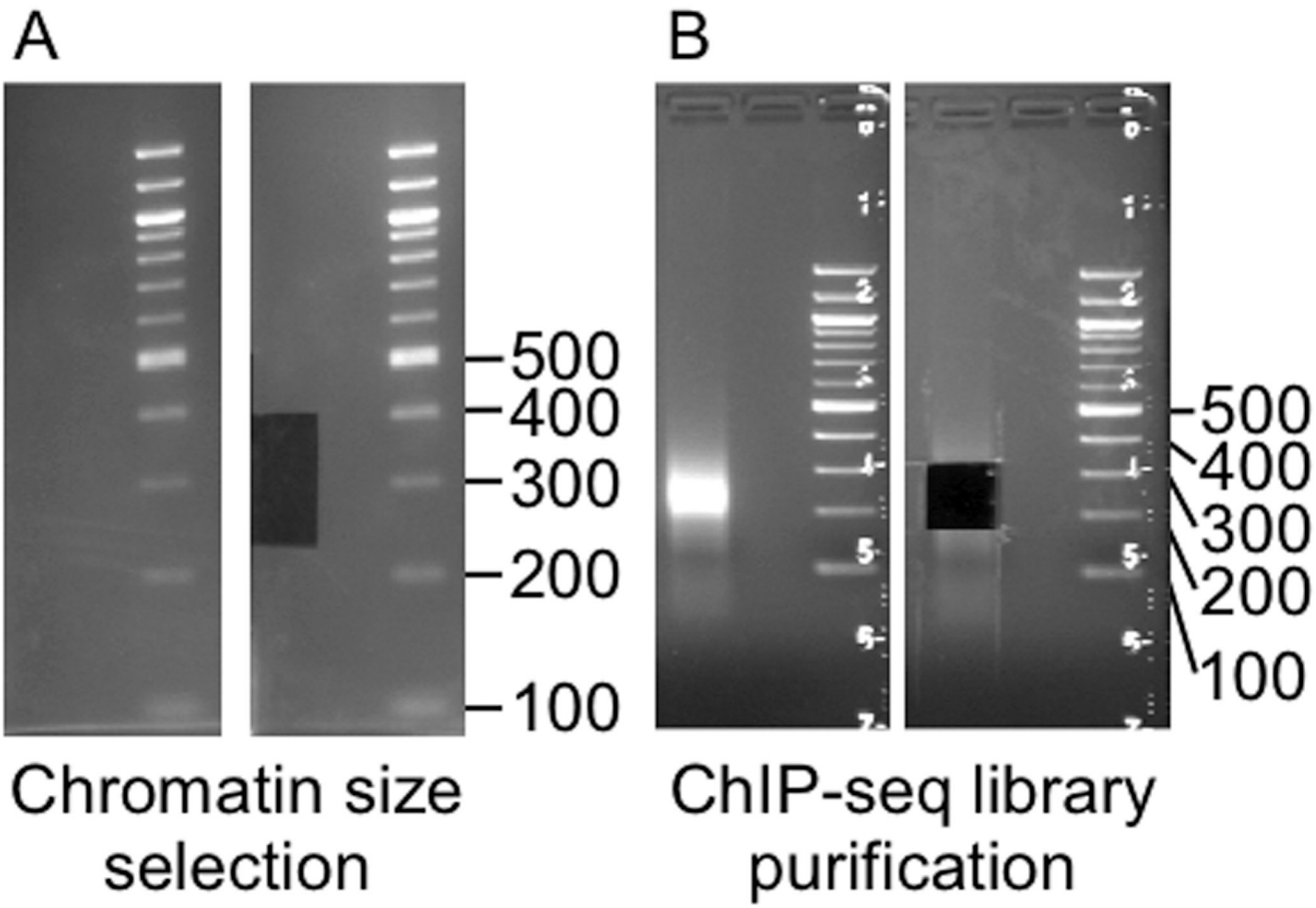

Figure 3. Gel size selection of fragmented chromatin and ChIP-seq library