Genome-Wide Location Analysis by Pull Down of In Vivo Biotinylated Transcription Factors

Aibin He[1,2] and William T. Pu[1,2]

[1]Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts
[2]Harvard Stem Cell Institute, Harvard University, Boston, Massachusetts

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. ChIP 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). *Curr. Protoc. Mol. Biol.* 92:21.20.1-21.20.15. © 2010 by John Wiley & Sons, Inc.

Keywords: ChIP-seq • transcription factor • biotinylation • genome-wide location analysis

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 cross-linked chromatin (ChIP). The critical reagent for ChIP is the immunoprecipitating antibody. However, antibodies suitable for ChIP are not available for many transcription factors. In addition, an antibody may exhibit nonspecific binding that is idiosyncratic with respect 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 identifying the transcription factor binding locations genome-wide (Kim et al., 2008).

The Basic Protocol in this unit describes how to perform bioChIP-seq starting from cells expressing the biotinylated transcription factor. This protocol is similar in principle to the ChIP protocols presented in *UNITS 21.3, 21.9, & 21.13*, 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

Current Protocols in Molecular Biology 21.20.1-21.20.15, October 2010
Published online October 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/0471142727.mb2120s92

than by PCR or microarray hybridization. The output from the Basic Protocol is purified DNA enriched for fragments bound by the transcription factor. Steps are provided at the end of the Basic Protocol for validation of bioChIP-seq results by bioChIP followed by quantitative PCR (bioChIP-qPCR). Support Protocol 2 describes how to convert the DNA from the Basic Protocol into a library suitable for sequencing on an Illumina Genome Analyzer 2.

STRATEGIC PLANNING

Expression of in vivo biotinylated transcription factors

The method used to express biotinylated proteins of interest will 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.

The authors used two constructs to permit titratable expression of biotinylated proteins of interest (Fig. 21.20.1). 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).

The Gateway cloning system (Invitrogen) was used 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.

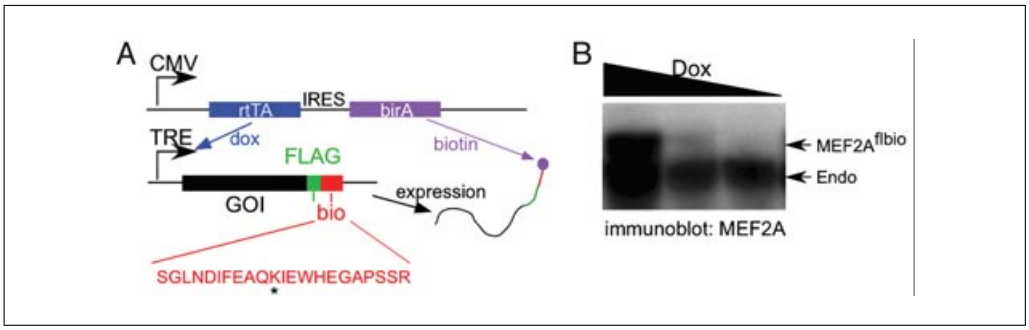

Figure 21.20.1 System for expression of biotinylated transcription factors. (**A**) Two-vector system for titratable expression of transcription factors with C-terminal FLAG and *bio* peptides. BirA recognizes and biotinylates the *bio* sequence on the central lysine residue (asterisk). (**B**) Titration of tagged transcription factor expression by adjusting the concentration of Dox. Tagged transcription factors are slightly larger than the corresponding endogenous protein due to the epitope tags. For the color version of this figure go to *http://www.current protocols.com/protocol/mb2120.*

GENOME-WIDE LOCATION ANALYSIS OF IN VIVO BIOTINYLATED TRANSCRIPTION FACTORS BY STREPTAVIDIN PULL-DOWN FOLLOWED BY HIGH-THROUGHPUT SEQUENCING (bioChIP-seq)

BASIC PROTOCOL

This protocol uses adherent cultured cells that express the transcription factor of interest. Protein-DNA complexes are stabilized by cross-linking 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 cross-link reversal.

Materials

Cell type of interest, grown under appropriate cell culture conditions in appropriate medium, and expressing BirA and the *bio* tagged transcription factor (see Strategic Planning)
Negative control: cell type of interest expressing BirA only
37% formaldehyde (Fisher Scientific, cat. no. F79-500)
2.5 M glycine (American Bioanalytic, cat. no. AB00730)
Phosphate-buffered saline (PBS; *APPENDIX 2*), cold
Hypotonic buffer (see recipe)
ChIP dilution buffer (see recipe)
TE Buffer (see recipe)
PBS (*APPENDIX 2*) containing 1% (w/v) BSA
Protein A magnetic beads (Invitrogen)
M-280 streptavidin magnetic beads (Invitrogen)
SDS wash buffer (see recipe)
High-salt buffer (see recipe)
LiCl buffer (see recipe)
SDS ChIP elution buffer (see recipe)
20 mg/ml proteinase K, DNase-free
10 mg/ml RNase A, DNase-free
QIAquick PCR purification kit (Qiagen)
Quant-It PicoGreen dsDNA DNA reagent (Invitrogen), for qPCR (see *UNIT 15.8*)

15-cm tissue culture dishes
Cell culture incubator
Cell lifter
15- and 50-ml conical centrifuge tubes
Tissue culture centrifuge
Glass Dounce homogenizer (2-ml size, Fisher, cat. no. K885300)
1.7-ml microcentrifuge tubes, prechilled
Refrigerated microcentrifuge
Misonix Sonicator 4000 with microtip (part no. 418, Qsonica, *http://nano.sonicator.com/*)
Nanodrop spectrophotometer
Siliconized nonstick microcentrifuge tubes, precooled
Magnetic stand (Ambion, cat. no. am10055)
37°C, 55°C, and 70°C water baths
Applied Biosystems 7500 Real-Time PCR system or equivalent (also see *UNIT 15.8*)

Additional reagents and equipment for optimizing sonication conditions for bioChIP-seq (Support Protocol 1) and qPCR (*UNIT 15.8*)

Cross-link and sonicate chromatin complexes

1. Culture cells expressing the biotinylated transcription factor of interest in 15-cm dishes. In addition to cells expressing the epitope-tagged protein and BirA, include a negative control expressing BirA only.

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 fewer cells is possible but may be accompanied by decreased signal to noise ratio.

2. Remove dishes from incubator and place on the bench. Add 37% formaldehyde to the medium to a final concentration of 1% (v/v), to cross-link DNA and bound proteins. Incubate at room temperature for 7 min.

   The indicated conditions represent a general starting point. Cross-linking conditions can vary significantly, and both concentration and time of formaldehyde treatment may need individual optimization. Excessive cross-linking increases background.

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

4. Decant medium 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 thorough washing.

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 to the dish, scrape one more time, and pool in the same 50-ml tube.

6. Centrifuge the cells 10 min at $3000 \times g$, 4°C.

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

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

8. Using a pipettor with a 1000-µl pipet, resuspend the cell pellet in 2 ml hypotonic buffer per $\sim 4 \times 10^7$ cells. Leave tubes in ice for 15 min to permit hypotonic lysis. If cells clumps form, homogenize the suspension with five strokes in a prechilled Dounce homogenizer.

   The need for homogenization depends on the cell type.

9. Transfer 1 ml cell suspension to each of two prechilled microcentrifuge tubes. Centrifuge 1 min $13{,}000 \times g$, 4°C. Aspirate as much of the supernatant as possible.

10. Resuspend the pellet in each microcentrifuge tube in 0.5 ml ChIP dilution buffer. Pool suspensions from the two microcentrifuge tubes. Sonicate under conditions determined in pilot experiments to yield fragments with average length of ∼150 to 200 bp (see Support Protocol 1).

    Practical aspects of how to sonicate the sample depends on the specific sonicator setup. The authors 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 does not contact the side of the tube.

11. After sonication, transfer cell lysates to microcentrifuge tubes and centrifuge 15 min at $20{,}000 \times g$, 4°C, to pellet the cell debris.

12. Transfer the supernatant (containing the sheared chromatin) to two new microcentrifuge tubes.

    The solution should be semi-transparent and slightly yellow.

13. Estimate the sheared chromatin by pipetting 1 µl of supernatant into 99 µl TE buffer and measure the DNA concentration using a Nanodrop spectrophotometer. Use this value to equalize the amount of DNA used for streptavidin pull-down.

14. Save a 25-µl aliquot of supernatant as the input sample and store at −20°C.

    This sample 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 least 1 hr in the 4°C cold room.

    This step can be performed while samples are being sonicated (in step 10).

16. Preclear the sheared chromatin: add the Protein A beads prepared in step 15 to each sample and incubate for 1 hr in the cold room.

    This step removes proteins that bind nonspecifically to the bead matrix.

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

18. Briefly spin the blocked streptavidin beads prepared in step 15. Place on magnetic stand for 30 sec. 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 hr to overnight.

19. Briefly centrifuge the samples and place on magnetic stand for 30 sec. Remove as much of the supernatant as possible using a pipettor with a 10-µl tip.

    Removing as much wash solution as possible helps to reduce background.

20. Sequentially incubate at room temperature with 1 ml of each wash solution indicated below for 5 min. After each incubation, briefly centrifuge, place on magnetic stand, and remove as much supernatant as possible by pipet, as described 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 microcentrifuge tube. Incubate 5 min, briefly centrifuge, place on magnetic stand, and remove as much supernatant as possible by pipet.

    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. Remove as much residual supernatant as possible using a pipet.

23. To reverse cross-links, add 100 µl of SDS ChIP elution buffer to the pelleted beads and place in a 70°C water bath overnight. At the same time, thaw the saved input chromatin aliquot from step 14, add 175 µl of SDS ChIP elution buffer, and reverse cross-link along with the pull-down samples.

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

25. Add 2 µl of 10 mg/ml RNase A for 1 hr at 37°C.

26. Purify the bioChIPed DNA using a QIAquick PCR purification kit following the manufacturer's protocol. Elute the bioChIPed DNA in 50 µl of the kit's EB buffer (10 mM Tris·Cl, pH 8.0) into a siliconized microcentrifuge tube.

    This is the purified bioChIP DNA. At this point the bioChIP material can be validated by qPCR (steps 27 to 29) or used to build a high-throughput sequencing library (see Support Protocol 2 for construction of an Illumina GA2 ChIP-seq library).

    At this point, DNA concentration and yield will be too low to be quantified using a Nanodrop spectrophotometer. 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 microcentrifuge tubes to avoid loss of DNA.

Validate bioChIP DNA 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 pulled-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. The following steps do not provide detailed information about how to set up and run a qPCR experiment, but provide information about the application of qPCR to bioChIP. *UNIT 15.8* provides more detailed information on performing qPCR.

27. For bioChIP-qPCR, use 1 µl of each bioChIPed DNA that results from the main bioChIP protocol as qPCR template. Prepare negative control samples from (1) the input chromatin after cross-link reversal, and (2) bioChIP of the sample expressing BirA but not a *bio* epitope–tagged protein.

28. Design primers for 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 nonspecific pull-down. As an example, the authors used a region in the first intron of the murine Actb gene:

    Forward: 5′-CGTATTAGGTCCATCTTGAGAGTACACAGTATT-3′

    Reverse: 5′-GCCATTGAGGCGTGATCGTAGC-3′.

    This region is referred to as Actb-1st intron. Because the average fragment length in bioChIP-seq libraries is 100 to 300 bp (200 to 400 bp minus the size of the ligated linker, which is 92 bp), it is recommend that the primers be designed so that amplicons are less than 100 bp. In this example, the authors used a primer with optimal T_m of $60^o \pm 1°C$ and length 20 ± 2 bp.

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

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

$$\text{input DNA} = \frac{2^{-Ct}\left(\text{input peak}\right)}{2^{-Ct}\left(\text{input neg ctrl}\right)}$$

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

$$\text{fold enrichment} = \frac{\text{ChIPed DNA} - \text{input DNA}}{\text{BirA DNA} - \text{input DNA}}$$

SUPPORT PROTOCOL 1

OPTIMIZATION OF SONICATION CONDITIONS FOR bioChIP-seq

Proper fragmentation of chromatin is essential to achieving good results (see Critical Parameters). On most sonicators, the variables that can be adjusted are amplitude, pulse-on time, rest time, and cycle number. Specific values are provided here for the Misonix 4000 with microtip, but individual settings 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 conditions that reflect those that will be used in actual experiments.

Amplitude is the percent of maximal power. Insufficient power will lead to long sonication times and incomplete fragmentation. Excessive power leads to foaming. The authors 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. The authors use a pulse-on time of 15 sec. 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 min. Cycle number is the number of sonication cycles. After establishing the other settings, cycle number is the main variable that requires adjustment depending on different cell types, but is typically between 20 and 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.

For materials, see Basic Protocol.

1. Proceed through steps 1 to 11 of the Basic Protocol, preparing three 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 to 26 of the Basic Protocol to analyze the size distribution of the sonicated chromatin.

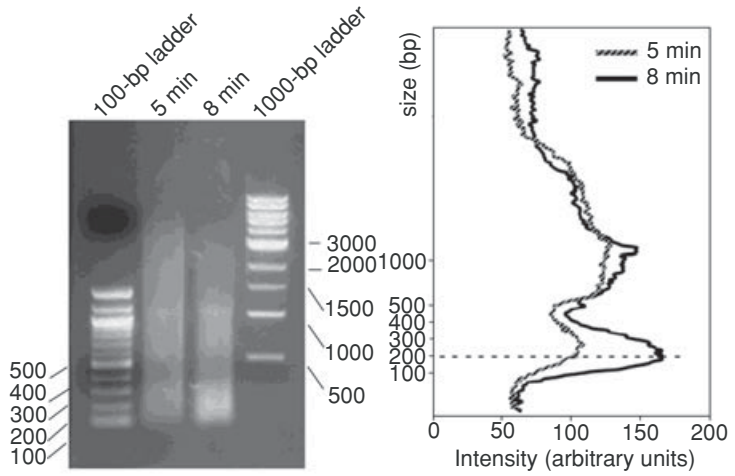

Figure 21.20.2 Optimization of sonication conditions. A 2-ml quantity of nuclear extract was sonicated for 5 or 8 min using a Misonix 4000 with the amplitude of 70, cycles 15 sec on and 1 min off. The plot on the right shows size versus intensity profile of sonicated chromatin. A sonication time of 8 min led to the desired size peak of ~200 bp.

4. Repeat steps 2 and 3, adjusting the amplitude and cycle number to achieve an average fragment size of ~100 to 300 bp.

   An example of two different sonication conditions is shown in Figure 21.20.2.

SUPPORT PROTOCOL 2

CONVERSION OF bioChIP-seq DNA TO A LIBRARY FOR SEQUENCING ON THE ILLUMINA GENOME ANALYZER 2

ChIP-seq library preparation is performed according to the Illumina ChIP-seq preparation kit manual, but with a few modifications as described below. Use siliconized microcentrifuge tubes throughout this procedure.

Materials

bioChIP DNA samples (from Basic Protocol)
10× T4 DNA ligase buffer (from Illumina kit) containing 10 mM ATP
QIAquick PCR Purification Kit (Qiagen)
ChIP-seq library preparation kit (Illumina) including:
- Adapter oligo mix
- T4 DNA ligase
- Klenow DNA polymerase
- T4 polynucleotide kinase (PNK)
- 1 mM dATP
- T4 DNA ligase buffer
- 5× Phusion buffer
- 10 mM dNTP mix
- Klenow buffer
- Klenow exo$^-$ (3′ to 5′ exo minus)
- Illumina PCR primers 1.1
- Illumina PCR primers 2.1
- Phusion DNA polymerase

Agarose
1× TAE buffer (*APPENDIX 2*)
Ethidium bromide (EtBr)

10× DNA loading buffer (*UNIT 2.5A*)
80% (v/v) glycerol
100-bp DNA ladder (New England Biolabs)
QIAquick Gel Extraction Kit (Qiagen)
MinElute PCR purification kit (Qiagen)
Ethidium bromide

Siliconized microcentrifuge tubes
Thermal cycler
Long-wavelength UV transilluminator or Dark Reader (Clare Chemical)
Razor blades

Additional reagents and equipment for agarose gel electrophoresis (*UNIT 2.5A*)

1. Prepare 1 U/µl Klenow DNA polymerase by 1:5 fold dilution of original polymerase in T4 DNA ligase buffer.

   Both the polymerase and the buffer are components of the Illumina kit.

2. Prepare the following reaction mix and incubate in a thermal cycler for 30 min at 20°C.

   40 µl bioChIP DNA
   5 µl T4 DNA ligase buffer with 10 mM ATP
   2 µl 10 mM dNTP mix
   1 µl T4 DNA polymerase
   1 µl diluted Klenow DNA polymerase (from step 1)
   1 µl T4 polynucleotide kinase.

3. Purify DNA 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·Cl, pH 8).

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

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

5. Purify DNA 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 min at room temperature.

   10 µl DNA sample from step 5
   15 µl T4 DNA ligase buffer
   1 µl diluted adapter oligo mix (see above)
   4 µl T4 DNA ligase.

7. Purify DNA 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 (*UNIT 2.5A*). Just before pouring the gel, add ethidium bromide (EtBr) to 400 ng/ml.

9. Add 2 µl of 10× DNA loading buffer and an additional 2 µl of 80% glycerol to the DNA from the purified ligation reaction (step 7).

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 (*UNIT 2.5A*) for best separation from free adapters (e.g., 80 V for 4 hr).

    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 of ChIP DNA PCR amplification and the extent of incorporation into the sequencing library.

13. Using a clean razor blade, excise a gel slice from the sample lane between ~200 and 400 bp (corresponding to the peak chromatin size plus 92 bp of ligated adapters). Photograph the gel before and after cutting out the gel slice.

    No band will be evident due to low sample concentration (see Fig. 21.20.3).

14. Cut a slice of the same size from an empty well on the same gel and process this sample through gel purification and PCR.

    No visible PCR product should be present.

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

16. Prepare the following PCR reaction mix:

    36 µl DNA (from step 15)
    10 µl 5× Phusion buffer
    1.5 µl 10 mM dNTP mix
    1.5 µl PCR primer 1.1
    1 µl Illumina PCR primers 1.1
    1 µl Illumina PCR primers 2.1
    0.5 µl Phusion polymerase.

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

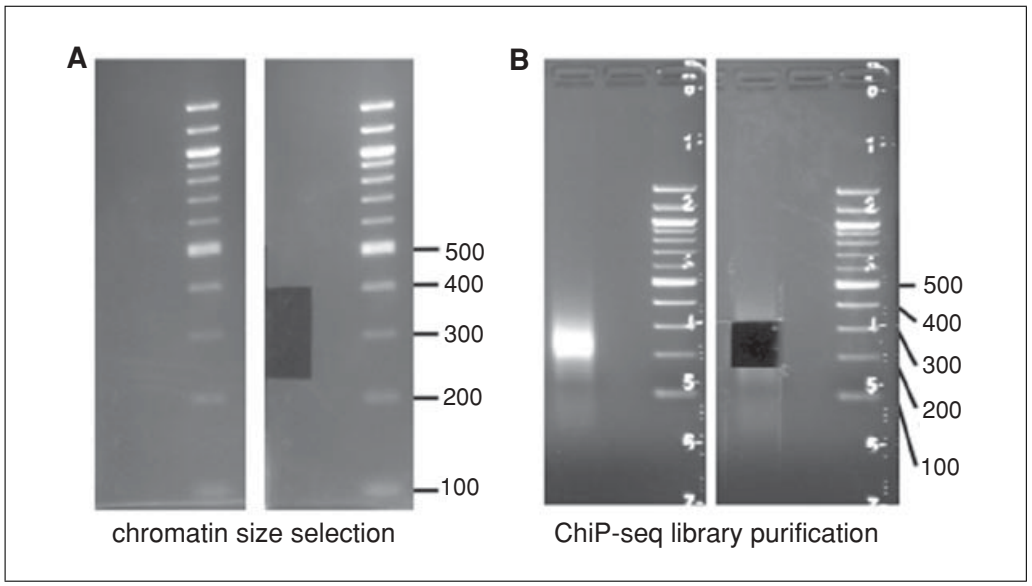

Figure 21.20.3 Gel size selection of fragmented chromatin and ChIP-seq library.

17. Amplify using the following PCR protocol:

1 cycle:	30 sec	98°C	(initial denaturation)
18 cycles:	10 sec	98°C	(denaturation)
	30 sec	65°C	(annealing)
	30 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
	Indefinitely	4°C	(hold).

18. Purify PCR products 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 to 14 of this protocol; this time a band should be present that corresponds to the ChIP-seq library (Fig. 21.20.3). 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.

    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 from contaminants.

20. Perform 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 (see UNIT 15.8). Analyzing the size distribution of the library on a Bioanalyzer is also recommended prior to submitting the sample to the sequencing core.

    Store the bioChIP-seq library in siliconized microcentrifuge tubes and sequence the sample as soon as possible, preferably within 2 to 3 weeks of library preparation.

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps. For common stock solutions, see ***APPENDIX 2****; for suppliers, see* ***APPENDIX 4****.*

ChIP dilution buffer

20 mM Tris·Cl, pH 8.0 (*APPENDIX 2*)
2 mM disodium EDTA
150 mM NaCl
0.1% (w/v) SDS
1% (v/v) Triton X-100
Store at room temperature for up to several months

At time of use, add 1 part concentrated protease inhibitor cocktail (see recipe) per 50 parts buffer.

Concentrated protease inhibitor cocktail

Dissolve one Roche Protease Inhibitor Cocktail tablet (cat. no. 11697498001) in 1 ml distilled water. Store up to 1 month at −20°C.

High-salt buffer

50 mM HEPES, pH 7.5
500 mM NaCl
1 mM disodium EDTA
0.1% (w/v) sodium deoxycholate
1% (v/v) Triton X-100
Store at room temperature for up to several months

Hypotonic buffer

20 mM HEPES pH 7.5
10 mM KCl
1 mM EDTA
0.1 mM activated Na_3VO_4 (to activate the sodium orthovanadate, boil a 200 mM stock solution for 10 min, then cool to room temperature and add the appropriate amount to the buffer mixture)
0.2% (v/v) Nonidet P-40 (NP-40)
10% (v/v) glycerol
Store at 4°C with the above components up to several months
At time of use, add:
Fresh DTT to 1 mM
PMSF to 1 mM
1 part concentrated protease inhibitor cocktail per 1000 parts buffer

LiCl buffer

10 mM Tris·Cl, pH 8.1 (*APPENDIX 2*)
250 mM LiCl
1 mM disodium EDTA
0.5% (v/v) Nonidet P-40 (NP-40)
0.5% (w/v) sodium deoxycholate
Store at room temperature for up to several months

SDS ChIP elution buffer

50 mM Tris·Cl, pH 8.1 (*APPENDIX 2*)
10 mM disodium EDTA
1% (w/v) SDS
Store at room temperature for up to several months

SDS wash buffer

Prepared from 20% SDS stock solution (*APPENDIX 2*) by dilution in nuclease-free water. Store at room temperature for up to several months

TE buffer

10 mM Tris·Cl, pH 7.5 (*APPENDIX 2*)
1 mM disodium EDTA
Store at room temperature for up to several months

COMMENTARY

Background Information

Chromatin immunoprecipitation (ChIP) has become the method of choice to identify transcription factor chromatin occupancy in vivo. Genome-wide application of this approach by hybridization-based (ChIP-chip; *UNIT 21.13*) and high-throughput sequencing-based (ChIP-seq; *UNIT 21.19*) 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 (*UNIT 21.19*) 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 the signal-to-noise ratio 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 nonphysiological 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, it is desirable to cross-link proteins to DNA with a minimum of intervening manipulations. While in some cases fluorescence-activated cell sorting (FACS) 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 cross-linking is also an important parameter. Under-cross-linking reduces signal, while over-cross-linking can impair chromatin fragmentation and increase nonspecific background.

Adequate fragmentation of chromatin to a mean fragment size of 150 to 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, so 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 nonrandom chromatin fragmentation by sonication. The BirA sample controls for background caused by endogenously biotinylated proteins and nonspecific 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 (*UNIT 21.15*) 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 to 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 compared the BirA sample to Input. Peaks from the BirA versus Input comparison are then subtracted from the bioChIP versus 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 and 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 at which "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% to 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 pull-down. 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 hr to overnight.

Reverse-cross-linking and purification of ChIP-ed DNA: 5 hr.

ChIP-seq library preparation: 1 day.

Literature Cited

Beckett, D., Kovaleva, E., and Schatz, P.J. 1999. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* 8:921-929.

Blahnik, K.R., Dou, L., O'Geen, H., McPhillips, T., Xu, X., Cao, A.R., Iyengar, S., Nicolet, C.M., Ludascher, B., Korf, I., and Farnham, P.J. 2010. Sole-Search: An integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic Acids Res*. 38:e13.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. 2003. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* 100:7480-7485.

Farnham, P.J. 2009. Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* 10:605-616.

Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., and Wong, W.H. 2008. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.* 26:1293-1300.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132:1049-1061.

Park, P.J. 2009. ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10:669-680.

Pepke, S., Wold, B., and Mortazavi, A. 2009. Computation for ChIP-seq and RNA-seq studies. *Nat. Methods* 6:S22-S32.

Key References

de Boer et al., 2003. See above.

Application of the bio epitope tag for one step purification of mammalian proteins.

Blahnik et al., 2010. See above.

Describes a bioinformatic tool for read alignment, peak calling, and peak annotation.

Kim et al., 2008. See above.
Application of in vivo transcription factor biotinylation to genome-wide location analysis by bioChIP-chip.

Internet Resources

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

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