Chromatin immunoprecipitation and deep sequencing (ChIP-Seq) protocol

A. Wills, R, Gupta, E. Chuong and J. Baker

Chromatin preparation materials and equipment

- Methanol-free formaldehyde (Sigma F8775)
- 0.125M Glycine
- 1X PBS
- Magnetic particle concentrator (Invitrogen 123.21D)
- Dynabeads protein A (Invitrogen 10001D)
- Dynabeads protein G (Invitrogen 10003D)
- Antibody (diluted according to manufacturer instructions)
- PBS+5% BSA: 250mg BSA (Sigma A9647-50G) in 50ml PBS (Gibco 10010-023). Store at 4° and discard after one week.
- RIPA buffer (4°C, 1.25 ml per set of 50 embryos): 50 mM Tris-HCl, pH 7.4, 1% Igepal CA-630 (NP-40) (Sigma I3021), 0.25% Na-Deoxycholate, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.5 mM DTT, 5 mM Na-Butyrate
- Complete, Mini Protease Inhibitor Cocktail Tablets (Roche 11836153001), 1 tablet per 10mL RIPA buffer.
- Rotator or Nutator, at 4**°**C
- Sonicator
- Refrigerated centrifuge

Detailed chromatin preparation procedure

Day 1:

- Culture embryos to desired stage.
- Fix 1 hour in 1% formaldehyde/PBS
- Wash 5 minutes in 0.125M glycine
- Wash 3 times, quickly, in PBS. Divide embryos into batches of up to 100 *X. laevis* or 250 *X. tropicalis*.
- Remove all PBS and freeze up to 6 months at -80**°**C, or proceed directly to Day 2 steps.
- Prepare Dynabeads with antibody (see section 3)

Day 2:

- Thaw crosslinked embryos on ice, 10-15 minutes.
- Add 600µl cold RIPA+Protease Inhibitor to each sample.
- Break embryos by pipetting with a P1000 tip or by gentle disruption with a plastic pestle, until embryos are broken into small fragments and the solution is gray in color. The embryos need not be completely homogenized yet.
- Centrifuge at 14,000XG for 10 minutes using a refrigerated centrifuge.

- Decant supernatant, and wipe the walls of the tube with a kimwipe to remove traces of yolk.
- Add 100µl cold RIPA, and homogenize thoroughly, taking care to avoid bubbles, until no visible fragments of embryo remain and homogenate is a uniform gray.
- Add 550µl cold RIPA.
- Chill embryo samples on wet ice for 20-30 minutes, until ready to sonicate. Gently resuspend each sample by pipetting immediately before sonication.
- Sonicate using empirically determined conditions. We recommend trying 2, 5, or 10 cycles at 30%, 50% or 100% intensity to start.
- Centrifuge at 14,000XG for 10 minutes using a refrigerated centrifuge.
- Transfer 60µl of the resulting lysate to new tubes for input controls.
- Transfer supernatant to pre-clear beads, and then to antibody-conjugated beads (see section 3).

Immunoprecipitation materials

- Low salt buffer: 0.1% SDS, 1% TritonX-100, 2mM EDTA, 20mM Tris-HCL pH8.0, 150mM NaCl (Store at 4°C)
- High salt buffer: 0.1% SDS, 1% TritonX-100, 2mM EDTA, 20mM Tris-HCL pH8.0, 500mM NaCl (Store at 4°C)
- LiCl salt buffer: 0.25M LiCl (L-8895 Sigma), 1% IGEPAL CA630 (Sigma I-3021), 1% deoxicholate acid, 1mM EDTA, 10mM Tris-HCL pH8.0 (Store at 4°C)
- TE buffer: 10mM Tris-HCl, 1mM EDTA, pH 8.0 (Store at 4°C)
- TES buffer: 50mM Tris-HCl pH8.0, 10mM EDTA, 1% SDS (store at room temperature)
- 5M NaCl
- Phenol/Chloroform/Isoamyl Alcohol (Thermo Fisher BP1752)
- Chloroform (Thermo Fisher BP1145)
- Sodium Acetate buffer solution (Sigma S-7899)
- Glycogen (Fermentas cat# R0561 Fermentas)
- Qiagen PCR Purification Kit (Qiagen 28004)
- RNase A (Roche 10109169001)
- 70% and 100% Ethanol

Detailed immunoprecipitation procedure

Day 1 (the day before homogenization and sonication):

- Resuspend protein G or protein A Dynabeads by vortexing. Use 50µl dynabeads for each experiment, including mock pulldown, if performing. All dynabeads can be aliquoted in one volume and washed together.

- Place tubes in bead separator, wait until beads have migrated against magnet and sample is clear.
- Wash beads twice with 1ml of PBS+5% BSA, vortexing beads well and then returning to magnet for each wash.
- Resuspend beads in PBS+5% BSA, using 200µl for each experiment. Vortex beads well and aliquot into individual tubes for each experiment. Bring volume up to 1ml (+800µl) with PBS+5% BSA.
- Add primary antibody according to manufacturer-recommended dilutions.
- Place tubes on a nutator or rotator at 4°C overnight.

Day 2 (same day as sonication):

- Prepare 20µl of washed Dynabeads per sample as above, resuspending in 1 mL PBS+5% BSA. These are the "pre-clear" beads
- Place against magnet to pellet pre-clear beads, discard supernatant, and replace with the supernatant of sonicated, centrifuged embryo lysate.
- Incubate pre-clear beads and lysate with rotation at 4°C for one hour.
- Use magnet to pellet ANTIBODY-conjugated beads. Remove and discard supernatant. Use magnet to pellet PRE-CLEAR beads, and transfer supernatant (lysate) from pre-clear beads onto antibody- conjugated beads. Discard pre-clear beads.
- Incubate antibody- conjugated beads plus lysate 4°C overnight, with rotation.

Day 3:

- Use magnet to pellet antibody-conjugated beads, discard supernatant. Perform washes at 4°C, using 1 ml of solution for each wash, and using the magnet to pellet beads between each wash:
- Low salt solution (2 washes, 5 minutes each)
- High salt solution (2 washes, 5 minutes each)
- LiCl buffer (2 washes, 5 minutes each)
- TE (2 washes, 5 minutes each)
- Remove all TE, and replace with 200µl TES. Return to room temperature.
- Vortex beads very thoroughly, let settle, and vortex again.
- Incubate beads at 65 degrees for 15 minutes, vortexing every 5 minutes.
- Vortex thoroughly once more, pellet beads using magnet, and transfer supernatant carefully to new, labeled microfuge tube.
- Remove input samples from freezer and thaw.
- Add 16µl 5M NaCl to input samples and immunoprecipitated samples. Cap lids tightly and incubate at least 5 hours or overnight at 65°C. *Optional: for higher-quality DNA, include 7µl proteinase K/Glycogen solution.

Day 4:

- Add 1 volume (approx. 200µl) Phenol/Chloroform to samples. Vortex until milky, and centrifuge at 12,000XG for 5-7 minutes.

- Transfer supernatant to new, clean, labeled microfuge tube.
- Repeat extraction with 1 volume (200µl) Chloroform; vortex, spin and transfer supernatant as above.
- Add 1/10 vol (20µl) 3M Sodium acetate, 2.5 vol (500µl) 100% Ethanol, and 1µl glycogen. Precipitate at least 5 hours or overnight at -20°C.
- Centrifuge at full speed for 15 minutes, taking care to note the orientation of the tubes. Small pellets should be visible in input samples, but may not be visible in IP samples.
- Carefully remove supernatant and wash pellets with 500µl 70% ethanol.
- Centrifuge at full speed for 1 minute, taking care to note the orientation of the tubes and position of pellets. Carefully remove all traces of supernatant.
- Resuspend in 15µl of nuclease-free water and quantify yield as described below. Alternatively, if the resulting DNA will be used for qPCR, perform the following additional steps:
- Incubate in 100ul RNAse A/TE for 1 hour at 37°C.
- Purify using a Qiagen Minelute reaction purification kit, according to kit instructions, using 15µl as the final elution volume
- Quantify yield using nanodrop. If yield is low (<50ng/µl), use a high-sensitivity method such as Qbit to accurately quantitate yield.

Library preparation materials

All inclusive kits:

- Genomic DNA prep kit from Illumina (requires Qiagen Minelute reaction cleanup reagents and columns as well)
- TruSEQ kit from Illumina

Piecemeal:

- T4 ligase and buffer. Invitrogen 15224-017
- T4 DNA Polymerase. NEB M0203L
- 10mM dNTPs. Invitrogen 18427-013
- Klenow DNA Polymerase
- T4 PNK
- Minelute reaction cleanup kit. Qiagen 28204
- dATP. Invitrogen 10216-018
- Klenow Fragment (3'-5' Exo). NEB M0212L
- HPLC-purified Adapters (consult Illumina for sequences)
- E. coli DNA ligase. Invitrogen 18052-019
- 2% eGel. Invitrogen G6610-02
  –or-
- NuSieve GTG agarose. Lonza 50080

- 50bp ladder. Invitrogen 10416-014
- Phusion high-fidelity 2X master mix polymerase kit. NEB M0532S
- HPLC-purified primers (consult Illumina for sequences)

Library prep procedure

Day 1:

- Aliquot enough ChIP DNA to reach 1µg (minimum) or 5µg (ideal). Bring up DNA samples to 30ul with nuclease-free water.
- Repair ends by preparing the following reaction mix, using PCR tubes:

ChIP DNA	30µl
Nuclease free water	45µl
T4 ligase buffer	10µl
dNTPs	4µl
T4 DNA polymerase	5µl
Klenow DNA Polymerase	1µl
T4 PNK	5µl
Total	70µl

- Incubate at 20°C for 30 minutes
- Clean up the reaction using Qiagen Minelute reaction clean up kit, as follows (or follow manufacturer instructions):
  -To the 70µl reaction mix, add 300µl buffer ERC, mix.
  -Pipet into Minelute column inside collection tube.
  -Centrifuge at full speed one minute.
  -Discard flow-through.
  -Add 750ul Buffer PE to the top of column, centrifuge 1 minute.
  -Discard flow through, and centrifuge again 1 minute to remove traces of ethanol.
  -Transfer Minelute column to new, clean, labeled microfuge tube.
  -Add 32µl buffer nuclease-free water to the middle of each column, let sit one minute.
  -Centrifuge 1 minute at full speed.
- Transfer samples to PCR tubes.
- To each sample (32µl), add:

Klenow Buffer	5µl
dATP	10µl
Klenow exonuclease (3' to 5' exo)	3µl
Total	50µl

- Incubate at 37°C 30'.
- Clean up reaction using Minelute column as above, eluting in 18µl final volume.

- Transfer samples to PCR tubes.
- Ligate adapters by preparing the following reaction mix:

End-repaired DNA	18µl
DNA ligase buffer	25µl
Adapter oligo mix	2µl
DNA ligase	5µl
Total:	50µl

- React at 20°C for 15 minutes.
- Gel-purify samples using either a 2% e-Gel or 3% Nuseive Agarose gel (see above).
  For eGel:
  -Load samples into upper wells, leaving a space between each sample to avoid contamination. Choose one lane to run a 50bp ladder.
  -Load 20ul water into any empty top wells, and all bottom wells.
  -Select "2%" mode. Run eGel for 13 minutes.
  -Use UV illumination to check position of ladder. DNA in samples should be visible as a smear. Add 10µl water to each bottom well.
  -Run gel slowly, checking frequently, until bottom wells are positioned between the 200 and 250 bp band. Isolate these bands from each sample. If the gel runs too long, change mode to "reverse e-Gel" and run in reverse until the desired band comes into the bottom wells.
  -If desired, add 20µl additional water to bottom wells, and run gel until 300bp band is positioned in bottom well. Isolate these bands from samples.

- For Agarose:
  -Prepare 3% gel using NuSeive GTG agarose and TAE
  -Combine samples with 10X loading dye. Also prepare a 1:10 dilution of 50bp ladder with loading dye.
  -Load samples, leaving an empty well between each sample to avoid contamination.
  -Run gel according to usual methods. Check by UV visualization; the DNA should be visible as a smear in wells containing samples.
  -Use a clean scalpel blade to isolate the 200-250bp region.
  -Purify DNA from gel fragment using Minelute gel purification kit, following manufacturer instructions, and eluting in 20µl final volume.

- Amplify the library by preparing the following PCR reaction:

Gel-purified DNA	4µl
Phusion DNA polymerase 2X Master Mix	25µl
Primer 1.1	1µl
Primer 2.1	1µl

Water	19µl
Total	50µl

- Perform the following PCR reaction, using 15 cycles if the initial amount of ChIP DNA was 1µg, or 12 cycles if the initial amount of DNA was 2µg or more.

  30 seconds at 98°C
  [10 seconds at 98°C
  30 seconds at 65°C
  30 seconds at 72°C] X 12-15 cycles
  5 minutes at 72°C

- Purify the resulting PCR products using a Minelute PCR cleanup kit, following manufacturer instructions, eluting in 20µl final volume.
- Check DNA concentration using a high-sensitivity method, such as Qbit or Bioanalyzer. If concentrations are too low (below 3ng/µl), repeat PCR and combine and concentrate reactions.

6. Analysis

6.1 Sequencing platform considerations

A standard ChIP-SEQ analysis requires a sequenced ChIP library as well as a background library, which is usually input chromatin or IGG control ChIP. The most common sequencing platforms at the time of this writing are the Illumina GAIIx or HiSeq machines. Both platforms are capable of producing different read lengths, ranging from 36 to 100bp, as well as single or paired-end reads. For most ChIP applications, single-end 36 bp reads are sufficient. The primary advantage of longer and paired ends is increased mapability in unique regions, which may be beneficial if investigating the chromatin of repetitive genomic DNA. The major difference between these platforms is read throughput: the GAIIx routinely produces 30-50 million reads, while the HiSeq platform routinely generates 100-200 million reads per lane. For most ChIP experiments, 20-30 million reads is sufficient to produce a high quality library, although this depends on large part on the protein being investigated. For example, TF binding sites cover small punctuate sites (<300 bp) throughout the genome, and require less read coverage than histone modifications that cover broad (1 kb+)

regions of the genome. While HiSEQ lanes are more costly, multiple samples can be “bar-coded” or indexed on the HiSEQ platform (consult Illumina for sequences of indexing primers). This way, multiple ChIP libraries can be sequenced on the same lane, making the effective cost far lower than the GAII without sacrificing read depth, since the HiSEQ platform yields much higher read count. We recommend anywhere from 4-6 samples be sequenced per HiSeq lane. Read indexing is usually performed by using Illumina adapter indexes or custom barcodes, but researchers should consult their sequencing facility.

6.2 Alignment

Because both RNA-SEQ and ChIP-SEQ libraries are double stranded DNA, they are handled by the sequencing platform in the same way. Sequencing centers usually generate a compressed “FASTQ” file, which includes all read sequences accompanied by quality score. Upon request they will usually align the reads to a user-specified genome. If not, there are multiple short-read aligners to accomplish this task. In Figure 2 we outline a general pipeline for sequence analysis, beginning with alignment. We use BWA (http://bio-bwa.sourceforge.net/), although Bowtie (http://bowtie-bio.sourceforge.net/index.shtml) and SOAP (http://soap.genomics.org.cn/) are other popular aligners. For all of these programs, the original genome fasta file must be downloaded from the UCSC genome browser or Xenbase, and this file must be first be indexed by the aligner. Next, a single command is usually used to align the FASTQ file to the genome index, and the result is an alignment file in SAM format. The SAM format is widely adopted as the standard alignment format, and is often stored in compressed form as a BAM file. Importantly, the SAM output from an aligner may include redundant or unmapped reads, and these should be discarded. Programs such as Samtools (a basic toolkit for SAM manipulation)(http://samtools.sourceforge.net/) or Picard (which is more powerful but slightly more difficult to use)(http://picard.sourceforge.net/) should be used to “clean” SAM files prior to ChIP analysis, ensuring only uniquely mapping reads are utilized.

6.3 Peak calling and peak inspection

Once a BAM file is generated for both ChIP and background samples, they can be used as input into a ChIP analysis program. Many programs exist for this purpose, and we recommend trying several as new and improved algorithms are constantly being developed. A major caveat when analyzing frog data is once again the scaffold arrangement. These gaps in the sequence break certain programs like QuEST (http://mendel.stanford.edu/SidowLab/downloads/quest/), but we found that the MACS program (http://liulab.dfci.harvard.edu/MACS/) works well with scaffolds when using the genome size set to 1.5e9. Most ChIP analysis programs will create a BED file that lists the coordinates of predicted ChIP peaks, as well as a WIG or BDG file that contains continuous ChIP coverage information over the genome.

With all programs, the user should manually inspect predicted ChIP peaks with the raw reads, allowing for both evaluation of the quality of the data and for insight into threshold setting. The most reliable way to manual inspect the peaks is to upload the BAM file and any BED/WIG files as tracks in genome browser (e.g. using “add track” at the UCSC Genome Browser or Xenbase). At each predicted ChIP peak, the BAM track should display a smooth “pileup” of both positive and negative reads, with the width of the pileup dictated by the type of protein being detected. Typically transcription factor peaks will have a narrower peak (Figure 3A) and histone modifications a broader peak (Figure 3B). If there are <10 reads in most peaks, or if reads form a vertical pile rather than a smooth pileup (Figure 3C), or if reads are all coming from a single strand, then the settings should be adjusted to be more stringent (e.g. set a lower FDR or P value threshold). Conversely, if a region exhibits a characteristic pileup of reads but the program did not predict that region, the settings may be too stringent. If the majority of high-scoring peaks do not pass visual inspection, then the initial ChIP library was likely of poor quality. Another good metric of quality is to determine the % of uniquely mapped reads that fall within predicted peaks. This can be

calculated by using BedTools (http://code.google.com/p/bedtools/) to count the number of reads from the BAM file that overlap the predicted peaks in the BED file (described below). A low fraction (< 10%) indicates that the library is noisy, and the ChIP may need to be repeated. Finally, researchers should also be aware of PCR "bottlenecking" where many reads pile in sparse vertical stacks rather than smooth piles across the genome (Fig 3C). This is indicative of either poor ChIP quality or low starting ChIP material when building the sequencing library, and results from PCR over-amplification. Overall, there are several specific features indicative of high quality ChIP libraries and we recommend that labs new to ChIP obtain a library of known quality for comparison.

The number of peaks predicted for a library can range anywhere from a few hundred to tens of thousands. We rarely see libraries with over 100,000 predicted peaks. Some peak-caller programs such as MACS can also be used to determine whether adequate coverage has been obtained. For example, if a ChIP library was returned with 30 million reads, would resequencing the library to attain a total of 60 million reads identify more significant peaks? By randomly selecting a subset of reads, and re-running the ChIP program, one can determine whether adding more reads would be beneficial. MACS performs this automatically using the "--diag" mode, but essentially this mode predicts peaks using progressively larger subsets of the total reads (e.g. 20, 40, 60, 80, and 100%). If the number of peak predictions continue to increase substantially as the number of reads used increases, particularly from 80 to 100%, then more sequencing may be effective. However, if using 80% of reads identifies a similar number of peaks as 100%, then more reads are unnecessary. For some libraries, such as those with poor overall ChIP enrichment or if the protein of interest is widely bound in the genome (e.g. H3K9me3), attaining full coverage may be prohibitive. We usually find that peaks identified with further sequencing are usually those with low enrichment, so the most strongly enriched peaks will normally be found within 20-30 million reads.

6.4 Analysis tools suggestions

Once the user is assured of the quality of ChIP predictions, data analysis may proceed. The most standard analyses involve associating ChIP peaks with nearby genes, or determining the association of ChIP peaks across different libraries. In general, these analyses are performed using BED files. Annotated gene files, from Refseq or the *Xenopus* Gene Collection (XGC) can be downloaded in BED format from the UCSC Genome Table Browser. A number of programs exist to manipulate and compare BED files (determining overlap, adjacent elements, etc.). For most users we recommend using the online Galaxy suite of tools (https://main.g2.bx.psu.edu/), which offers a web interface for uploading and manipulating BED files. Downloadable programs such as BEDTools offer similar functionality with more customizability for those comfortable with command-line usage. Another common analysis is to predict sequence motifs enriched within a ChIP dataset. Many programs exist for this usage, but one of the most widely used is the MEME suite (http://meme.nbcr.net/meme/), which is capable of both *de novo* motif discovery (using Meme) and identifying instances of known motifs (using Fimo). Though a ChIP dataset may contain many thousands of peak predictions, we recommend using just the top 2000 highest-scoring peaks. These programs take as input the DNA sequence of all peaks in FASTA format, which can be generated using BEDTools or Galaxy.

Most analyses are readily performed on modern desktop computers. The most intensive step is often the alignment of reads to the genome, which can take several hours to a full day for moderately sized (20+ million reads) libraries. Most alignment tools, including BWA (using the –t option), can utilize multiple CPUs which are featured on most modern desktop computers. Peak calling may also take several hours, but most downstream analysis of BED files is relatively computationally unintensive. A major consideration is that sequence files and intermediate processed files are often several gigabytes each, and it is important to invest in sufficient hard drive storage space and backup solutions. If an adequately powerful machine is not readily available, remote servers such as

Galaxy may provide a temporary solution, and many institutions now offer bioinformatics core servers that will usually have commonly used alignment software installed.

Example commands used for ChIP-Seq analysis
Using software: BWA 0.5.9, samtools 0.1.10, and MACS 1.4

To align reads to a genome "xenTro3.fa" downloaded from UCSC:

```
# creates a searchable index file for the genome
bwa index –a bwtsw xenTro3.fa

# aligns "reads.fastq" short-read file
# optional -t N where N is number of processors to use, ex) -t 4

bwa aln -q 10 xenTro3.fa reads.fastq >reads.sai
bwa samse xenTro3.fa reads.sai reads.fastq  >output.bwa.sam

# Converts sam file to bam file, and only retains reads with quality scores of 1 or greater (scores of 0 indicate multiply mapped reads)
samtools view -S -b -q 1 output.bwa.sam >output.bwa.uniq.bam

# Calls peaks given uniquely aligned ChIP and control libraries ("chipReads.bam" and "inputReads.bam").
macs14 -t chipReads.bam -c  inputReads.bam -g 1.5e9 bam -n MyChIP
```