Application Note

Small cell number ChIP-seq using ThruPLEX® DNA-seq Kit as a tool for epigenetic profiling

Yanina Bogliotti[1], Kamran Shazand[2], John Paul Jerome[2], Brian Walsh[2], and Pablo Ross[1]

1. University of California, Davis, California, United States

2. Rubicon Genomics, Inc., Ann Arbor, Michigan, United States

Introduction

DNA-binding proteins, such as polymerases, transcription factors and chromatin modifiers, are critical for a variety of essential cellular processes from DNA replication and repair to control of gene expression and chromatin structure. Individual DNA-binding proteins may bind to specific genomic loci, such as replication origins or promoters, or may bind DNA with varying sequence dependence, such as single stranded DNA binding proteins, helicases, or polymerases (1-3). Of primary concern to many researchers is the role that the location and modification of histones and nucleosomes play in gene regulation (4). Tri-methylation of histone 3 at lysine 4 has been shown to promote increased transcription of bound loci, while tri-methylation of histone 3 at lysine 27 results in decreased transcription (3). Chromatin immunoprecipitation (ChIP) has become an invaluable tool for studying the interactions between DNA-binding proteins and their genomic targets. Recent developments in next-generation sequencing (NGS) technologies have allowed scientists to more readily determine the target sequences of DNA-binding proteins on a genome-wide scale using ChIP coupled to high throughput DNA sequencing (ChIP-seq).

The general workflow for a ChIP-seq experiment requires crosslinking proteins to DNA using a reversible cross-linker such as formaldehyde, cleaving the DNA by sonication or enzymatically, precipitating the protein-DNA complex of interest using specific antibodies coupled to magnetic beads, reversing the crosslink, preparing the library with the released DNA, and finally, high throughput sequencing (Figure 1). By its very nature, the amount of DNA recovered from ChIP is very low, as only regions bound by a single protein or complex are selected. Moreover, ChIP from limited number of cells pose an extra challenge. For those samples, the implementation of the ThruPLEX® technology, which is designed for sample inputs as low as 50 pg, is key.

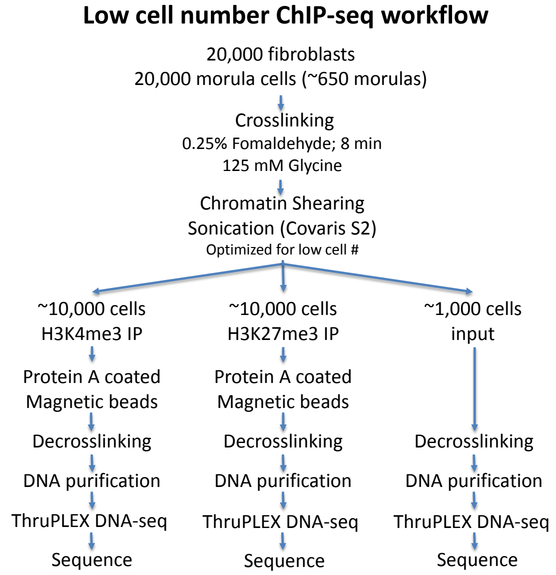


Figure 1. ***Low cell number ChIP-seq workflow.***
General workflow for low cell number ChIP-seq. Approximately 20,000 fibroblasts and 20,000 morula cells were each subjected to formaldehyde crosslinking, quenching by glycine addition, and chromatin shearing by sonication. Following shearing, chromatin from fibroblasts or morula cells is divided into subpopulations for both immunoprecipitations, or 10% input controls, followed by decrosslinking, DNA purification, library preparation with ThruPLEX DNA-seq, and sequencing.

Methods

Cells and Embryos

Bovine fibroblasts were grown to a confluence of 80%-90%. Cell media was aspirated and cells were washed twice with PBS and harvested using Trypsin-EDTA (1x). Cell pellets (20 million and 20,000 cells) were flash frozen and stored at -80°C until use.

Morula embryos were obtained by in vitro fertilization (IVF) of in vitro matured (IVM) oocytes aspirated from slaughterhouse-derived cow ovaries (Cargill, Fresno, CA) as previously described (5). Morulas were collected 120 hours post fertilization, washed 3 times in SOF-HEPES, and zona pellucida depleted using 5% of pronase. Blastomeres from morula embryos were washed 5 times in PBS/PVA 1 mg/ml and snap frozen in ~10 µl of PBS/PVA and stored at -80°C until pooling and use.

Pellets containing 20 million fibroblasts or 20,000 cells (from either fibroblast or embryonic cells) were thawed on ice and resuspended in PBS and crosslinked with 1% or 0.25% formaldehyde respectively for 8 min at room temperature followed by quenching with 125 mM glycine for 5 minutes at room temperature in a rotating wheel.

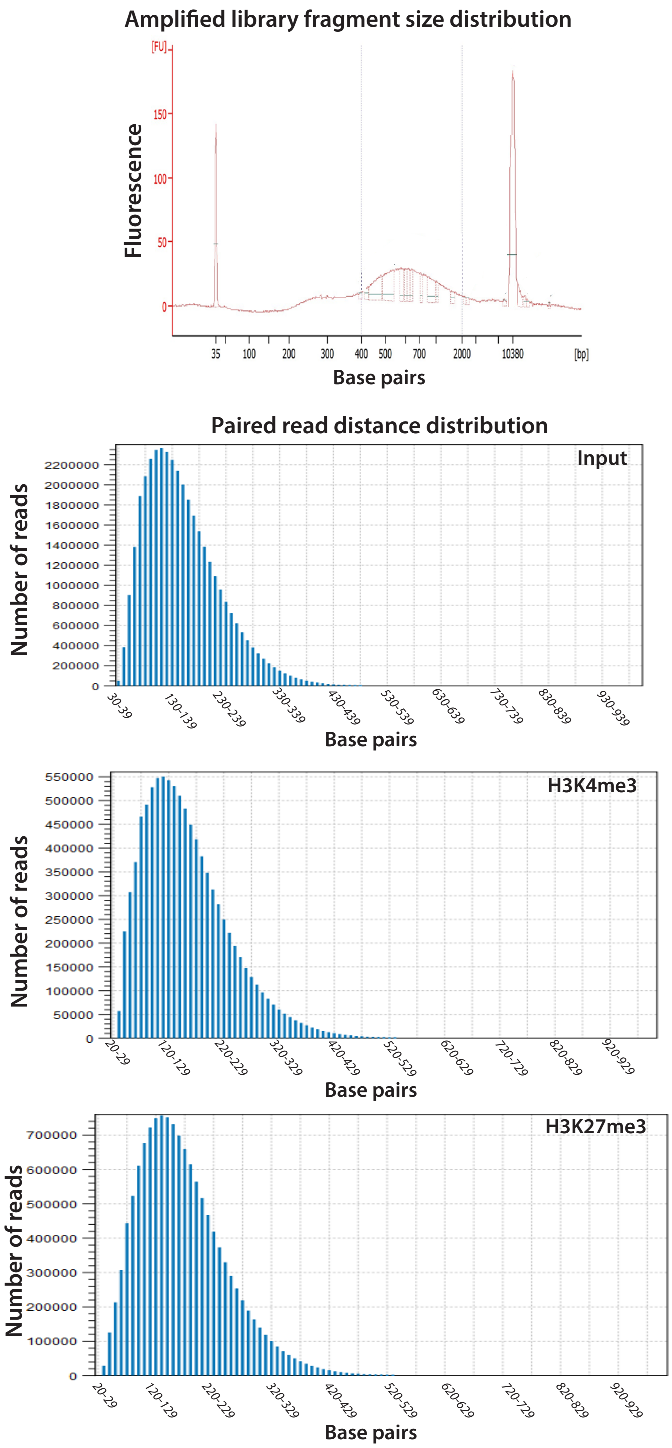


Figure 2. ***Size distributions of library fragments and mapped reads from low cell number fibroblasts.***

Top: Bioanalyzer® trace showing library fragment size distribution post-amplification. Bottom: Distribution of distance between paired reads, in nucleotides, for input DNA and ChIP libraries.

Chromatin immunoprecipitation

High-cell number ChIP (10 million fibroblasts) was done following a protocol from Dahl and Collas (6). Sonication was done by 7 cycles 30 seconds on / 30 seconds off at 30% power in a Labsonic® M sonicator in 20 million cells. Low cell-number ChIP (10,000 cells) was done following the True MicroChIP kit protocol with minor modifications (Diagenode C01010130). Crosslinked cells (20,000 cells) were lysed for 5 min on ice and sonicated using a Covaris® S2 sonicator for 12 minutes at duty 5%, intensity 3, and bursts 200. In both cases, 10% of the sonicated material was separated and used as the input control for library preparation and sequencing. The remaining material was equally divided for the H3K4/K27me3 precipitations. The immunoprecipitation was performed using H3K4me3 (provided in the Diagenode True MicroChIP Kit) and H3K27me3 (Millipore ABE44) antibodies.

Library Preparation

Libraries were prepared from either H3K27me3 or H3K4me3 ChIPed chromatin and 10% input chromatin. Low cell number libraries were prepared using the ThruPLEX DNA-seq kit with dual indexes, and 16 cycles during the final amplification step. Libraries from the high cell number ChIP and its input were prepared using the Illumina® TruSeq® DNA sample preparation kit according to the manufacturer's instructions.

The quality and size of the libraries were verified running a HS DNA chip on an Agilent Bioanalyzer after library preparation. The size of fragments generated by ThruPLEX DNA-seq amplification was found to be between 500-1000 bp (Figure 2).

Sequencing

Low cell number ChIP and 10% input libraries were pooled and loaded onto an Illumina NextSeq® 500 flow cell and sequencing was performed as a 2 x 75 bp paired-end run. Sequencing generated ~20M read pairs for each ChIP sample, and 49M - 61M read pairs for each input (Figure 3). High cell number ChIP and their input libraries were sequenced on an Illumina HiSeq® 2000 in a 50 bp single-end run. The high cell number sequencing run resulted in ~35M read pairs from the input and ~25M read pairs from both ChIP samples (Figure 3). Paired read distance (the distance on the chromosome between where either end of the read maps) distributions were very similar across the low cell number input and ChIP samples (Figure 2).

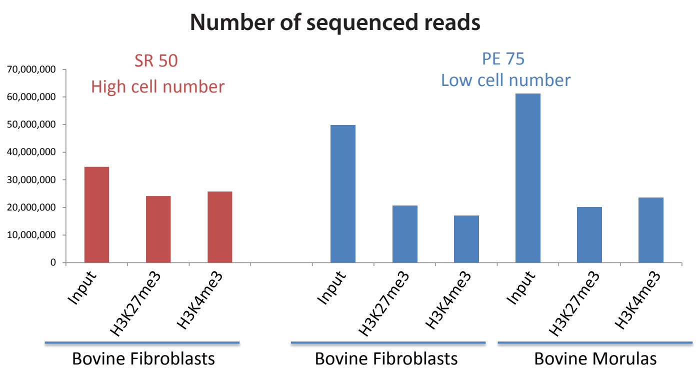
Data Analysis

The ChIP-seq reads were checked for quality using FastQC and then aligned to the annotated bovine genome (UMD 3.1 assembly) using Bowtie2 (7). Peak calling was done using MACS2, and the peaks were visualized using integrative genomics viewer (IGV) (8, 9).

Called peaks were further analyzed using PeakAnalyzer to find peak associations with genes, transcript start sites, and other genomic features. Gene ontology was done using DAVID (the Database for Annotation, Visualization, and Integrated Discovery) (10, 11).

Results & Discussion

The low yield from chromatin immunoprecipitation, especially for the low-cell number condition, make it challenging to prepare libraries for high throughput sequencing. Despite these difficulties, ChIP-seq has become a dominant technique for the investigation of site specific protein-DNA interactions and transcriptional regulation (12). Here, ChIP-seq was performed on ~10,000 bovine fibroblast cells and blastomeres from morula embryos using antibodies against H3K4me3 or H3K27me3 and the ThruPLEX DNA-seq kit for library preparation. The libraries generated by ThruPLEX from small cell numbers were similar to those generated from 10^7 cells by Illumina TruSeq. The number of sequenced reads from the 10,000 cell ChIP (Figure 3, Low cell number) was very similar to the results seen in a ChIP-seq experiment previously performed with 10^7 fibroblasts using the Illumina TruSeq kit (Figure 3, High cell number). The low cell number ChIP-seq samples and inputs generated a low number of clonal reads given the small amount of DNA used for starting material (Figure 3).



	Low cell number		
	# reads	# duplicate reads	% clonal reads
Fibroblasts Input	49,817,987	673,228	1%
Fibroblasts H3K27me3	20,697,494	1,443,766	7%
Fibroblasts H3K4me3	17,055,886	2,828,338	17%
Morula Input	61,191,932	1,217,478	2%
Morula H3K27me3	20,183,380	2,092,028	10%
Morula H3K4me3	23,669,174	1,931,430	8%

Figure 3. ***Number of sequenced reads per condition.***
Low cell number ChIP-seq samples and inputs were sequenced to similar depths as high cell number ChIP-seq samples and low cell number ChIP-seq libraries had low percentages of PCR duplicates. Each ChIP and sequencing experiment was performed as a single technical replicate.

ThruPLEX amplification of ChIP and input DNA generated libraries of similar quality using 1000-fold less cells than the TruSeq prepared libraries, with 95-96% of input reads aligning to the reference bovine genome, and nearly identical input peaks called genome-wide (Figure 4).

Even though the low cell number ChIP showed a lower percentage of total reads aligning to the genome as a result of the extremely low amount of starting material (Figure 4), the number of genes associated with these marks (after peak calling) in fibroblasts were very similar (Figure 5). As expected, fewer genes were detected from the ChIP experiment in bovine morulas due to limited methylation levels of histones during this stage of embryonic development (13). The results of the low cell number ChIP-seq experiment correlate extremely well with the literature and previous RNA-seq results, as seen in Figure 6, with a high percentage of H3K4me3-bound genes being expressed, while the majority of H3K27me3-bound genes have very low transcription rates (3).

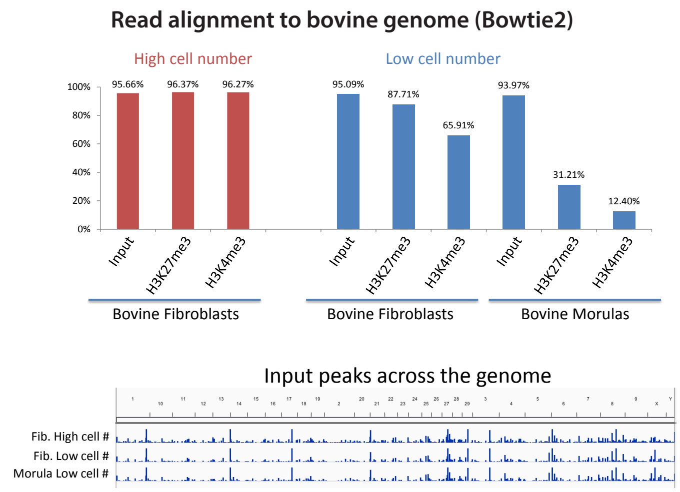


Figure 4. ***Read alignment to bovine genome.***
Percentages of reads aligning to the bovine genome, as calculated using the Bowtie2 package, are shown for high cell number input and ChIP samples, as well as low cell number input and ChIP samples. Below, the peaks identified genomewide from input DNA are nearly identical when compared between high cell number library preparation using TruSeq or low cell number library preparation using ThruPLEX DNA-seq.

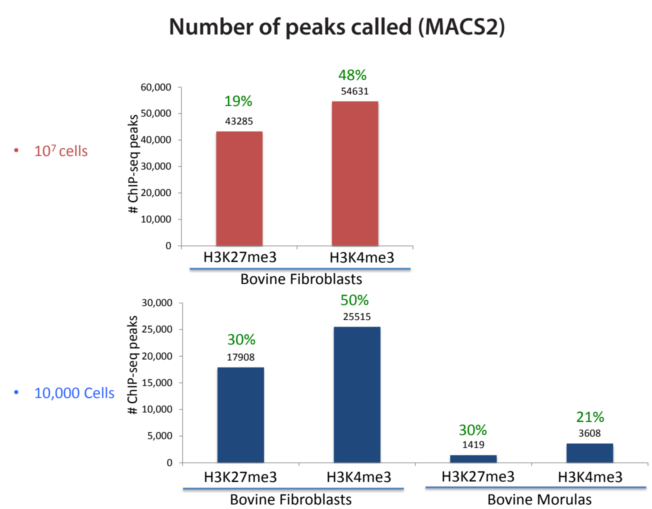


Figure 5. ***ChIP-seq peaks called to the bovine genome.***
Above, in black: the number of peaks called to the bovine genome are shown, per ChIP-seq experiment. Above, in green: the percentage of peaks called that specifically overlap genes.

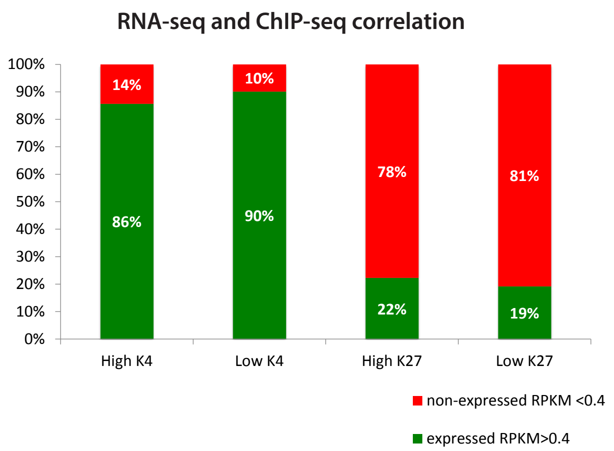


Figure 6. ***Correlation between genes with ChIP-seq peak and transcription level.***

The percentage of genes with peaks identified in both high and low cell number ChIP-seq and their correlation with transcription level was analyzed. Genes immunoprecipitated with H3K4me3 show high levels of gene expression in prior RNA-seq results. Genes immunoprecipitated with H3K27me3 show reduced transcription.

Conclusion

The primary limitations of preparing sequencing libraries with small amounts of ChIPed DNA can be overcome by using the ThruPLEX DNA-seq kit for library preparation. The ThruPLEX chemistry robustly amplifies ChIP DNA from very low numbers of bovine fibroblasts and embryonic cells (~10,000). As seen here, low cell number ChIP-seq results from both H3K4me3- and H3K27me3-precipitated DNA correlate extremely well with the results from high cell number ChIP-seq, and yield low numbers of duplicate reads. The performance of the ThruPLEX DNA-seq kit is complemented by its fast and simple, single-tube, three-step protocol. Library preparation with ThruPLEX DNA-seq overcomes the significant technical challenges associated with amplifying ChIP DNA and stands out as an important tool of the ChIP-seq workflow (14).

Trademarks

References

1. Farnham PJ. Insights from genomic profiling of transcription factors. Nat Rev Genet. 2009;10(9):605-16. doi: 10.1038/nrg2636. PubMed PMID: 19668247; PubMed Central PMCID: PMC2846386.
2. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007;316(5830):1497-502. doi: 10.1126/science.1141319. PubMed PMID: 17540862.
3. Martin C, Zhang Y. The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol. 2005;6(11):838-49. doi: 10.1038/nrm1761. PubMed PMID: 16261189.
4. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37. doi: 10.1016/j.cell.2007.05.009. PubMed PMID: 17512414.
5. Ross PJ, Rodriguez RM, Iager AE, Beyhan Z, Wang K, Ragina NP, *et al.* Activation of bovine somatic cell nuclear transfer embryos by PLCZ cRNA injection. Reproduction. 2009;137(3):427-37. doi: 10.1530/REP-08-0419. PubMed PMID: 19074500.
6. Dahl JA, Collas P. MicroChIP--a rapid micro chromatin immunoprecipitation assay for small cell samples and biopsies. Nucleic Acids Res. 2008;36(3):e15. doi: 10.1093/nar/gkm1158. PubMed PMID: 18202078; PubMed Central PMCID: PMC2241883.
7. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PubMed Central PMCID: PMC3322381.
8. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, *et al*. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. doi: 10.1186/gb-2008-9-9-r137. PubMed PMID: 18798982; PubMed Central PMCID: PMC2592715.
9. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, *et al.* Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6. doi: 10.1038/nbt.1754. PubMed PMID: 21221095; PubMed Central PMCID: PMC3346182.
10. Salmon-Divon M, Dvinge H, Tammoja K, Bertone P. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics. 2010;11:415. doi: 10.1186/1471-2105-11-415. PubMed PMID: 20691053; PubMed Central PMCID: PMC2923140.
11. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003;4(5):P3. PubMed PMID: 12734009.
12. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009;10(10):669-80. doi: 10.1038/nrg2641. PubMed PMID: 19736561; PubMed Central PMCID: PMC3191340.
13. Canovas S, Cibelli JB, Ross PJ. Jumonji domain-containing protein 3 regulates histone 3 lysine 27 methylation during bovine preimplantation development. Proc Natl Acad Sci U S A. 2012;109(7):2400-5. doi: 10.1073/pnas.1119112109. PubMed PMID: 22308433; PubMed Central PMCID: PMC3289350.
14. Cao Z, Chen C, He B, Tan K, Lu C. A microfluidic device for epigenomic profiling using 100 cells. Nat Methods. 2015. doi: 10.1038/nmeth.3488. PubMed PMID: 26214128.