TECH NOTE

Low cell number ChIP-seq using the SMARTer ThruPLEX DNA-seq kit as a tool for epigenetic profiling

Data kindly provided by: Yanina Bogliotti and Pablo Ross at University of California, Davis, California, 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 (Farnham 2009; Johnson et al. 2007; Martin and Zhang 2005). Of primary concern to many researchers is the role that the location and modification of histones and nucleosomes play in gene regulation (Barski et al. 2007). 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 (Martin and Zhang 2005). 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 a limited number of cells poses an extra challenge. For those samples, the implementation of SMARTer ThruPLEX technology, which is designed for sample inputs as low as 50 pg, is key.

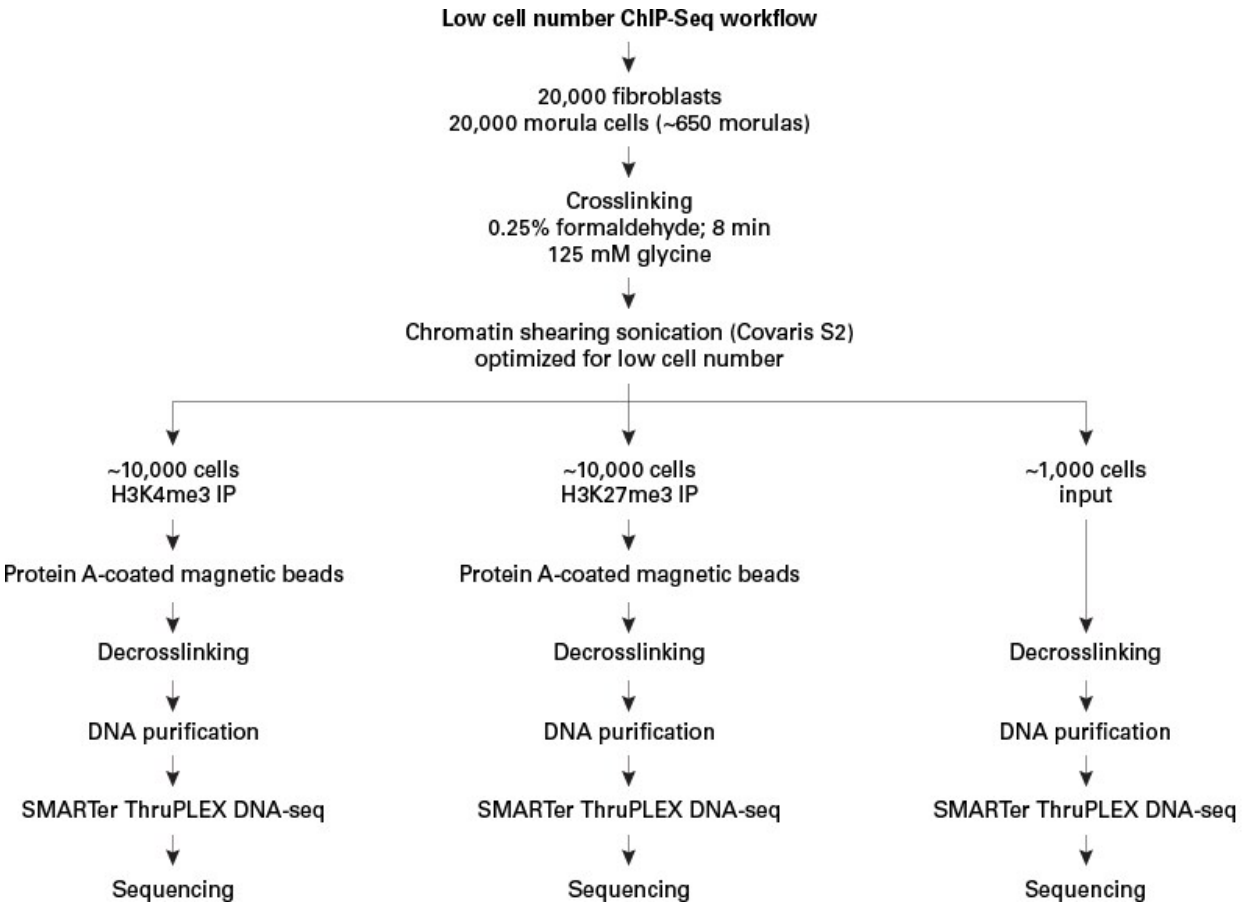
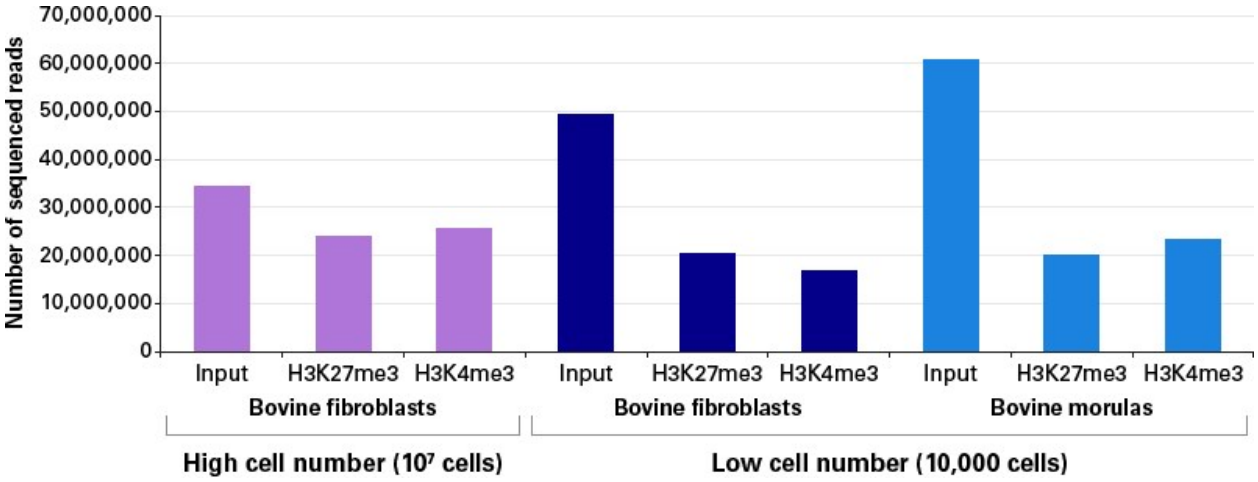


Figure 1. Low cell number ChIP-seq workflow. This is a general workflow for low cell number ChIP-seq. Approximately 20,000 fibroblasts and 20,000 morula cells were each subjected to formaldehyde crosslinking, quenching by glycine addition, and chromatin shearing by sonication. Following shearing, chromatin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

Results

The low yields from chromatin immunoprecipitation, especially for low cell number conditions, 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 (Park 2009). Here, ChIP-seq was performed on ~10,000 bovine fibroblast cells and blastomeres from morula embryos using antibodies against H3K4me3 or H3K27me3 and the SMARTer ThruPLEX DNA-seq kit for library preparation. The libraries generated by SMARTer ThruPLEX from low 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 2, 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 as starting material (Figure 3).

A



B

Number of sequenced reads			
	Low cell number		
	Number of reads	Number of duplicated 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 2. 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 (Panel A), and low cell number ChIP-seq libraries had low percentages of PCR duplicates (Panel B). Each ChIP and sequencing experiment was performed as a single technical replicate.

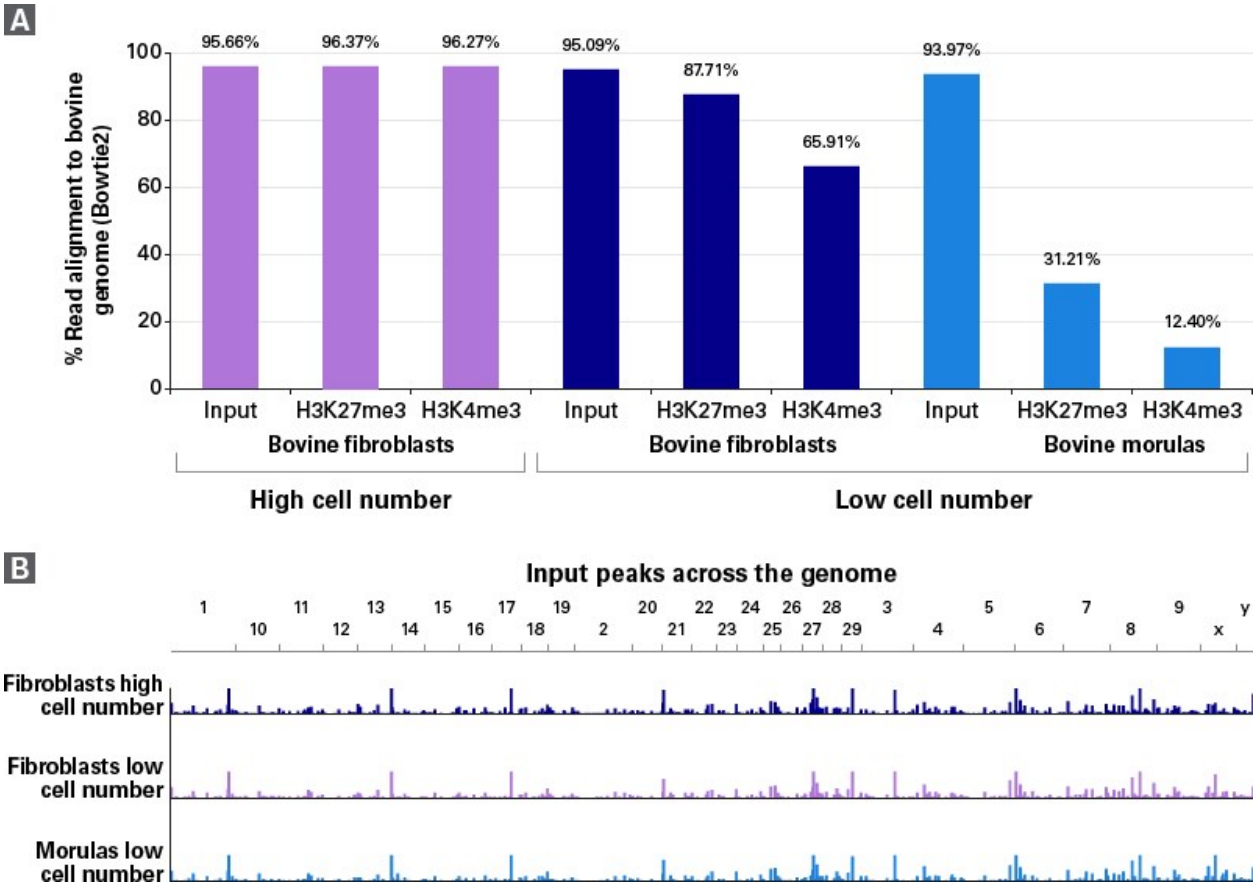


Figure 3. Read alignment to the bovine genome. Panel A. Percentages of reads aligning to the bovine genome, as calculated using the Bowtie2 package, are shown for high cell number input and ChIP samples, as well as low cell number input and ChIP samples. Panel B. The peaks identified genome-wide from input DNA are nearly identical when compared between high cell number library preparation using TruSeq and low cell number library preparation using SMARTer ThruPLEX DNA-seq.

SMARTer ThruPLEX amplification of ChIP and input DNA-generated libraries of similar quality using 1,000-fold less cells than the TruSeq prepared libraries results in 95–96% of input reads aligning to the reference bovine genome, and nearly identical input peaks called genome-wide (Figure 3).

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 3), the number of genes associated with these marks (after peak calling) in fibroblasts were very similar (Figure 4). 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 (Canovas, Cibelli, and Ross 2012). 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 5, with a high percentage of H3K4me3-bound genes being expressed, while the majority of H3K27me3-bound genes have very low transcription rates (Martin and Zhang 2005).

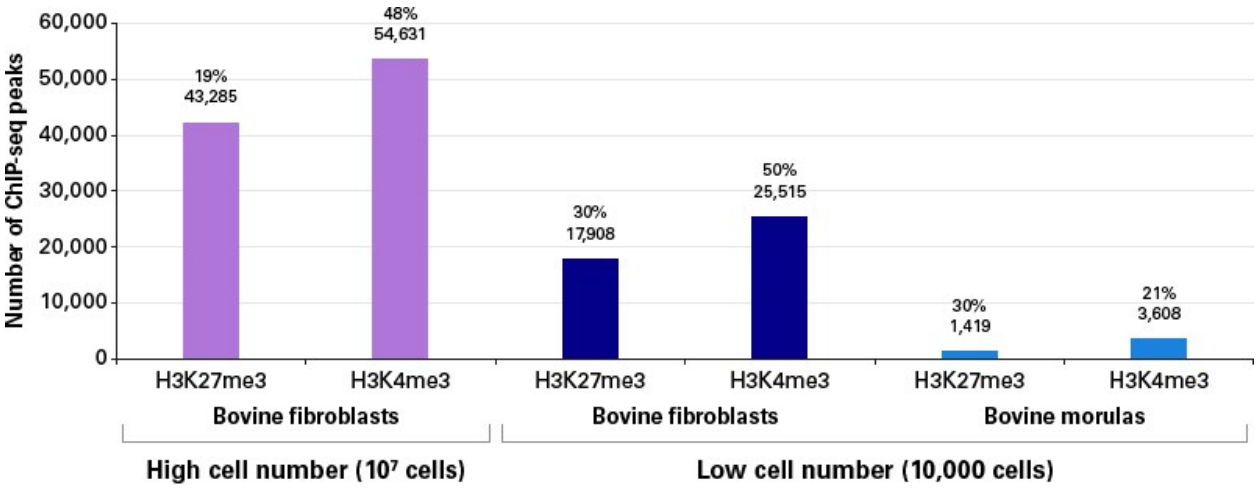


Figure 4. ChIP-seq peaks called to the bovine genome. Above, in black: the number of peaks called to the bovine genome are shown, per ChIP-seq experiment. Above, in red: the percentage of peaks called that specifically overlap genes.

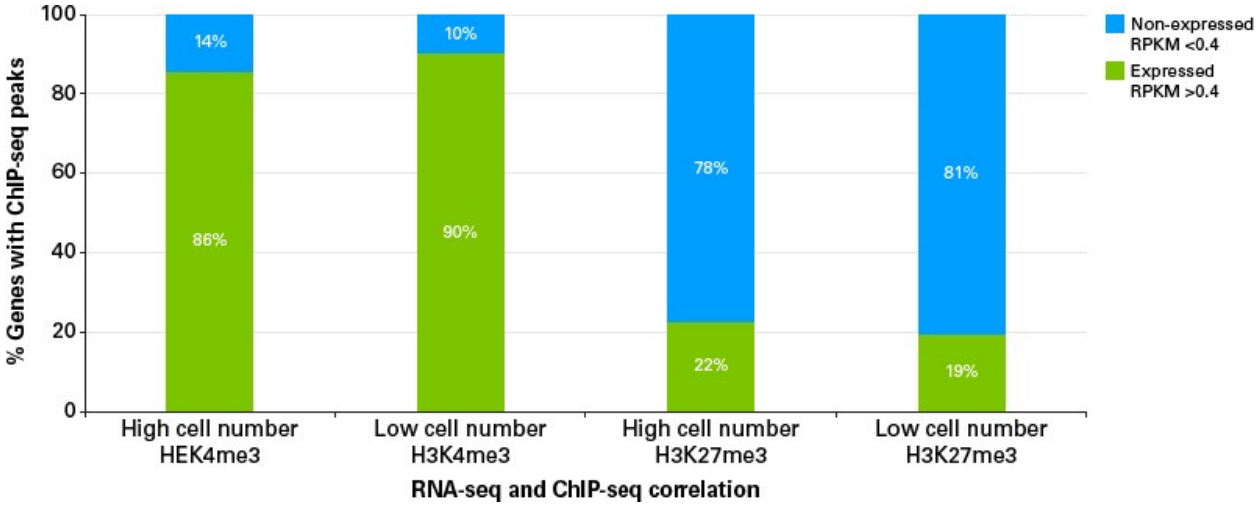


Figure 5. Correlation between genes with ChIP-seq peaks and transcription levels. The percentage of genes with peaks identified in both high and low cell number ChIP-seq and their correlation with transcription level was analyzed. Genes immunoprecipitated with H3K4me3 show high levels of gene expression 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 SMARTer ThruPLEX DNA-seq kit for library preparation. The SMARTer 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 SMARTer ThruPLEX DNA-seq kit is complemented by its fast and simple, single-tube, three-step protocol. Library preparation with SMARTer 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 (Cao et al. 2015).

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 (Ross et al. 2009). 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.

Chromatin immunoprecipitation

High cell number ChIP (10 million fibroblasts) was performed according to a protocol from Dahl and Collas (Dahl and Collas 2008). Sonication was carried out using 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 performed according to 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 SMARTer 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 SMARTer ThruPLEX DNA-seq amplification was found to be between 500–1,000 bp (Figure 6). 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.

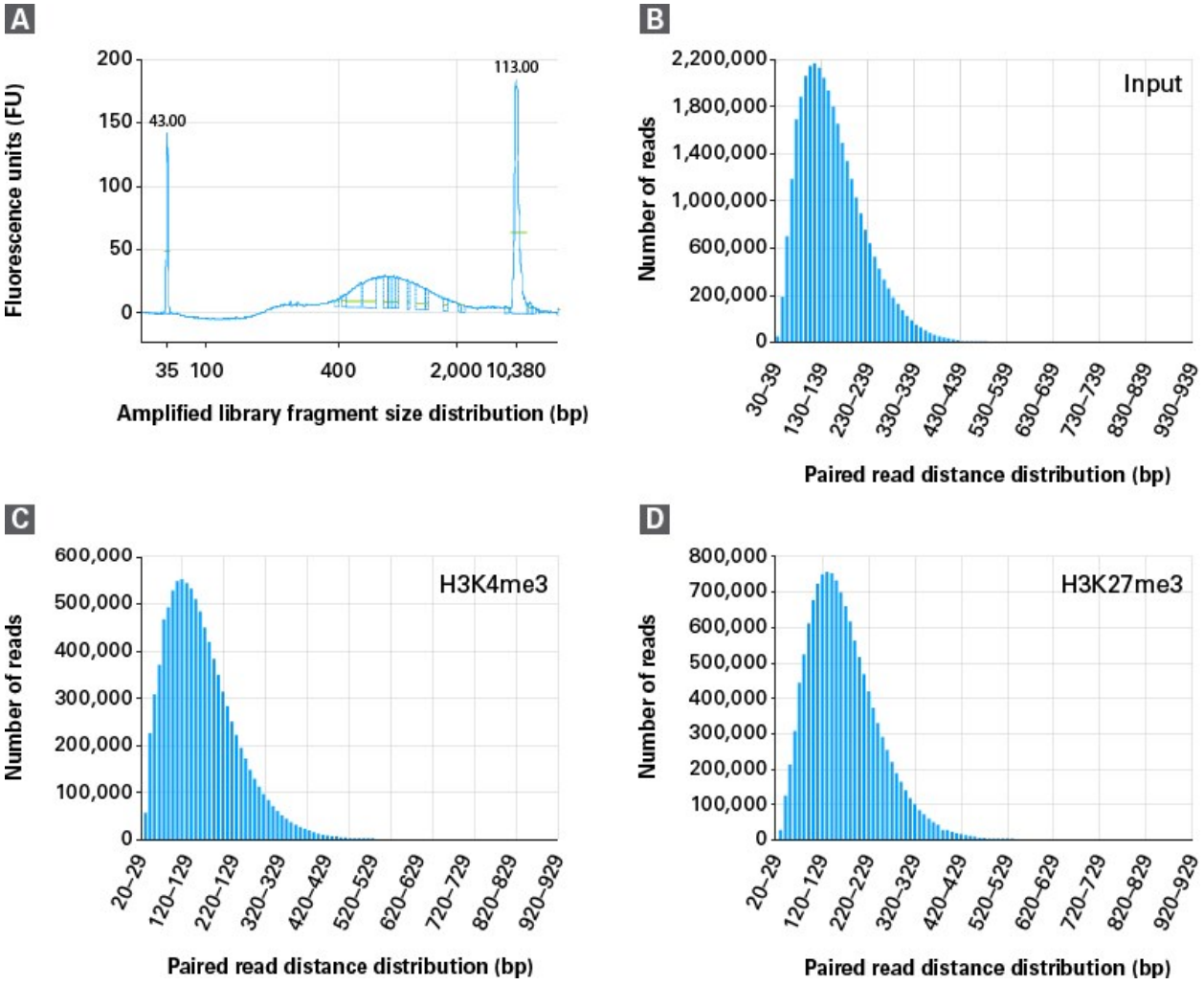


Figure 6. Size distributions of library fragments and mapped reads from low cell number fibroblasts. Panel A. Bioanalyzer trace showing library fragment size distribution post-amplification. Panels B–D. Distribution of distance between paired reads, in nucleotides, for input DNA and ChIP libraries.

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 2). 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 2). 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 6).

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 (Langmead and Salzberg 2012). Peak calling was done using MACS2, and the peaks were visualized using the integrative genomics viewer (IGV) (Zhang et al. 2008; Robinson et al. 2011). 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) (Salmon-Divon et al. 2010; Dennis et al. 2003).

References

Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–37 (2007).

Canovas, S., Cibelli, J. B. & Ross, P. J. Jumonji domain-containing protein 3 regulates histone 3 lysine 27 methylation during bovine preimplantation development. *Proc. Natl. Acad. Sci.* 109, 2400–2405 (2012).

Cao, Z., Chen, C., He, B., Tan, K. & Lu, C. A microfluidic device for epigenomic profiling using 100 cells. *Nat. Methods* 12, 959–962 (2015).

Dahl, J. A. & Collas, P. µChIP-a rapid micro chromatin immunoprecipitation assay for small cell samples and biopsies. *Nucleic Acids Res.* 36, e15 (2008).

Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4, P3 (2003).

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

Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* 316, 1497–1502 (2007).

Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012).

Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* 6, 838–849 (2005).

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

Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26 (2011).

Ross, P. J. *et al.* Activation of bovine somatic cell nuclear transfer embryos by PLCZ cRNA injection. *Reproduction* 137, 427–437 (2009).

Salmon-Divon, M., Dvinge, H., Tammoja, K. & Bertone, P. PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* 11, 415 (2010).

Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008).

Related Products

Cat. #	Product	Size	License	Quantity	Details
R400586	SMARTer® ThruPLEX® Tag-seq 96D Kit	96 Rxns			
R400585	ThruPLEX® Tag-seq 48S Kit	48 Rxns			
R400584	ThruPLEX® Tag-seq 6S (12) Kit	12 Rxns			

Add to Cart

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999