ChIP-seq analysis of protein–DNA interactions on the Ion Proton™ System

Key findings

- Optimized ChIP-seq protocol for limited cell populations—The protocol combines the MAGnify™ Chromatin Immunoprecipitation System and sample multiplexing via barcoded libraries using the Ion Proton™ System
- Comparable ChIP-Seq results using 1 million or 20 million cells—Cell sorted populations of 1, 3, 5, or 20 million CD11b+ lineage dendritic cells (DC) were analyzed for genomic binding sites of the transcription factor Irf4, demonstrating that results for 1 million cells are comparable to those for 20 million cells, enabling epigenomic analysis of limited cell populations.
- Highly concordant results to ENCODE project data—Following the optimized ChIP-Seq protocol, a cell sorted population of 1 million control cell line K 562 cells was analyzed for genomic binding sites occupied by a transcription factor (CTCF), histone activation (H3K27ac) and histone silencing (H3K27me3), with highly concordant results compared to ENCODE project data.

Introduction

Chromatin immunoprecipitation sequencing (ChIP-seq) is the evolution of an earlier microarray method (ChIP-chip) to enrich and map binding sites for transcription factors, chromatin-modifying complexes, histone modifications, and other chromatin-associated proteins. Treating cells with a fixative, such as formaldehyde, crosslinks the DNA and associated proteins that compose chromatin; following cell lysis and chromatin shearing, antibody binding (immunoprecipitation) of the protein of interest is used to enrich for the bound DNA (Figure 1).

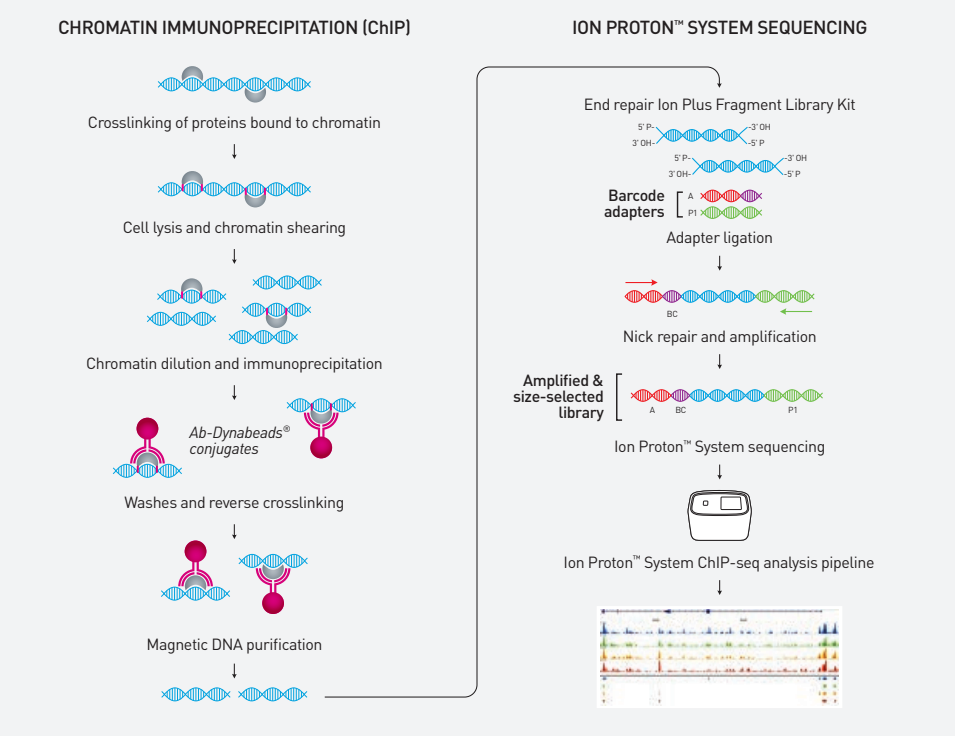

Figure 1. The Ion Proton™ System ChIP-seq workflow.

life technologies

Massively parallel sequencing of the enriched DNA provides an unbiased, genome-wide view of the binding sites for a diversity of chromatin-associated proteins in any cell type or organism with known genomic sequence. Traditional ChIP-seq workflows require relatively large numbers of cells as starting material, hindering efforts to examine chromatin dynamics in the small populations of specific cell types that are relevant to specific diseases. A robust ChIP-seq methodology compatible with small populations of flow cytometry–sorted, highly purified cells is required to accurately define epigenetic and chromatin dynamics regulating cell functions that when disregulated may contribute to disease.

Analyze transcriptional networks and the dynamic epigenome

Epigenetics can be broadly described as those molecular processes that modify the output of genomic loci without changing the original DNA sequence [1]. Epigenetic modifications can shape the life of a cell and an organism in dynamic ways through structural alterations of chromosomal regions that can modify activity states, such as the regulation of gene loci activity and transcription. Of key interest in the study of the epigenome are changes that influence chromatin structure and accessibility such as DNA methylation and histone modifications. Histone proteins package DNA into nucleosomes (a key structural component of chromatin) with a large number of modifications now described, such as trimethylation of histone H3 on lysine 27 (H3K27me3) associated with transcriptional repression and acetylation of histone H3 on lysine 27 (H3K27ac) correlated with transcriptional activation and enhancer regions [2].

Global, unbiased methods, such as ChIP-seq, have been crucial to revealing the epigenome and were employed by the ENCODE (Encyclopedia of DNA Elements) project in an attempt to systematically map the regions of transcription, transcription factor binding, chromatin structure, and histone modification in a number of different human cell types [3].

Integrated genomic analysis

ChIP-seq may be used for the identification of candidate regulatory regions by mapping the binding sites of transcription factors and components of the RNA polymerase complex. These enriched sequence regions can be further subjected to motif analysis for putative binding sites for the assayed factor or protein complex as well as the discovery of both canonical and distinct secondary motifs. Putative regulatory regions identified by ChIP-seq can be assessed for enrichment by a combination of factors or contexts including: association with specific modified histones; integration with chromatin patterns; nucleosome positioning; DNA methylation; and GC-content. Further, ChIP-seq data can be integrated with other genomics-based data such RNA-seq to holistically investigate chromatin patterns with the output of the transcriptome. Integrating ChIP-seq with gene expression data can potentially elucidate a functional relationship between chromatin dynamics and changes in transcription. For example, microarray expression profiling was used in conjunction with ChIP-seq analysis (performed on the Ion PGM™ System) to identify mouse Pitx2 target genes, with many targets corresponding to previously identified genes in human atrial fibrillation genome-wide association studies (GWAS) [4].

ChIP-seq challenges

The isolation of DNA by ChIP can yield relatively low amounts of starting material that can span a wide size range for next-generation sequencing (NGS) library construction. This can be particularly acute when ChIP is performed on rare or limited populations of cells. The research utility of semiconductor-based ChIP-seq analysis using sub-nanogram amounts of DNA has been previously demonstrated for the Ion PGM™ System [5]. ChIP was performed using antibodies against a histone mark, histone 3 trimethylated on lysine 4 (H3K4me3), and the carboxyl-terminus domain of RNA polymerase II, to characterize genomic sites in lipopolysaccharide-stimulated mouse DCs. Using the equivalent of 10 million cells, the results obtained with the Ion PGM™ sequencer were compared to data from the Illumina HiSeq® 2000 System with excellent agreement between results and highly correlated enrichment scores. Successful libraries with reasonable enrichment score correlation were achieved from immunoprecipitated DNA equivalent to ~20,000 cells. These preliminary data suggest that the longer read lengths possible through semiconductor-based sequencing might be advantageous for ChIP-seq applications and suggests that the scalability of Ion Torrent™ instruments may enable the optimization of ChIP-seq for limited starting populations of specific cell types.

Results

Sequence results and library complexity assessment

The optimized ChIP-seq workflow illustrated in Figure 1 was followed; for more detail see the "Optimized ChIP-seq workflow for the Ion Proton™ System" in the Supplementary information section. Multiplexing four samples on an Ion PI™ v2 Chip, replicate runs resulted in 89.4 and 89.7 million reads (after quality filtering and trimming) with an excellent barcode balance of 19.4–25.8 million reads per sample aligned to the human genome (hg19) using the TMAP alignment tool with default parameters (Table1A, Supplementary information). The mean read length ranged from 124–132 bp, with the fraction of unique reads for CTCF, H3K27me3, and H3K27ac estimated to be 32% (6.9–7.1 million), 82% (19.1–21.1 million), and 82% (16.0–16.8 million), respectively.

Discovery of genomic regions interacting with a protein and experimental reproducibility can be influenced by library complexity. To assess library complexity, the nonredundant fraction was calculated. For the transcription factor sample, CTCF, the nonredundancy fraction was 0.49–0.56, while broad source factors (delineating large chromatin domains) H3K27me3 and H3K27ac had nonredundant read fraction of 0.87–0.89, indicating suitable complexity for target region discovery (Table1A, Supplementary information). The differential in nonredundancy rates could be an indication of the expected sequencing coverage partitioned over a relatively small percentage of the genome targeted by a sequence-specific DNA-binding protein, such as CTCF, compared with the greater percentage bound by genome-wide chromatin elements, such as histone modifications. As well, the nonredundancy differences suggest that sufficient sequencing depth was achieved at the level of four multiplexed samples per Ion PI™ v2 Chip.

Robust immunoenrichment

Peak calling was performed via Model-based Analysis for ChIP-Seq (MACS) v2.0.10 [6]. To improve the reliability of peak detection and eliminate artifacts, MACS removes redundant reads to retain a single read per genomic locus. Using a random sample of enriched regions, the fragment size for a particular enrichment is estimated and read positions are adjusted with subsequent calculation of peak enrichment relative to input control background. The MACS default p-value threshold was used to empirically calculate the false discovery rate (FDR), with a FDR threshold of 10^{-3} used to identify putative loci (Table 1B, Supplementary information).

The modal lengths of CTCF, H3K27me3, and H3K27ac loci were 165, 260, and 186 bp, respectively. Note that the locus size distribution for H3K27me3 was considerably more dispersed compared to peaks for H3K27ac and CTCF (Figure 2).

The ENCODE Consortium recommends performing ChIP-seq experiments twice to assess reproducibility [7]. Replicate analyses confirmed a Pearson correlation coefficient (R) was 0.89 for the two CTCF ChIP-seq experiments, demonstrating that the two datasets are highly correlated; run 1 was size-selected with the Pippin Prep™ instrument during library creation, whereas run 2 was size-selected with a double SPRI method (described in detail: see Supplementary information).

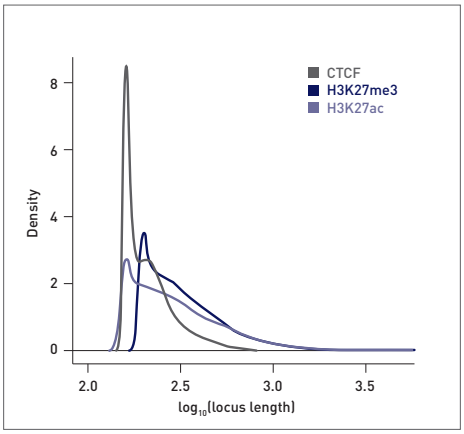

Figure 2. Locus length distribution for ChIP-seq peaks identified using the MAGnify™ Chromatin Immunoprecipitation System and the Ion Proton™ System. Shown is the distribution of peak size and density for CTCF (grey), H3K27me3 (blue), and H3K27ac (violet).

Strong correlation with ENCODE project data

Genomic binding sites for CTCF, H3K27ac, and H3K27me3 identified in K-562 cells using the MAGnify™ Chromatin Immunoprecipitation System and the Ion Proton™ System were compared to ENCODE project results for the same cell type (Table 1B, Supplementary information and Figure 3). A comparison between replicate semiconductor-based ChIP-seq results demonstrates consistent peak detection with ENCODE project results. For transcription factor CTCF, 52–69% of the peaks were concordant with an ENCODE dataset and 31–48% were unique to ChIP-seq data generated on the Ion Proton™ System. Binding sites for histone modifications H3K27ac and H3K27me3 were highly correlated with ENCODE project data with 71–73% and 89% of the peaks overlapping, respectively; with 27–29% of the H3K27ac peaks and only 11% of the H3K27me3 peaks unique to the semiconductor-based ChIP-seq data.

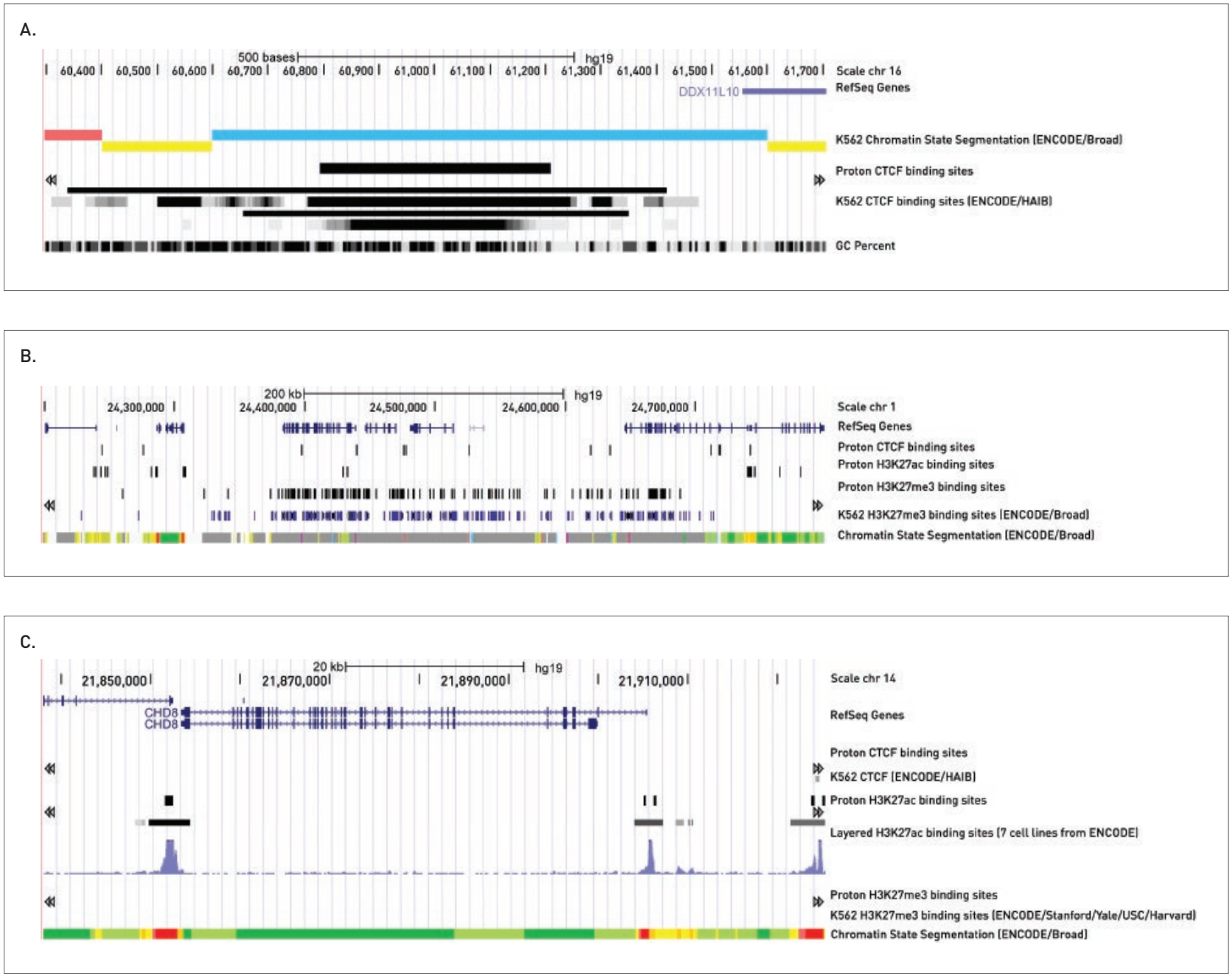

Figure 3. Genomic loci and concordance with ENCODE project results for ChIP-seq data from the Ion Proton™ System [10]. **(A)** Enrichment of CTCF ChIP-seq loci at an insulator on chromosome 16 (chr16: 60794–61210; depth: 116; $\log_{10}$ q-value: 196; fold-change: 64). **(B)** ChIP-seq loci for H3K27me3 are coincident with polycomb-repressed chromatin on chromosome 1. **(C)** ChIP-seq peaks for H3K27ac correspond to active promoters around the *CHD8* gene on chromosome 14.

The chromatin binding factor encoded by the human CTCF gene is an important component of epigenetic regulation that can act as transcriptional activator or repressor in a context-dependent fashion [8]. Further, CTCF can bind insulator elements that act to block interactions between enhancer and promoter elements [9]. ChIP-seq data from the Ion Proton™ System showed significant enrichment at insulator domains with ~41% of the identified loci overlapping with insulator elements and ~17% of the loci demonstrating enrichment for promoter/enhancer domains as determined by ENCODE project data (Figure 3A) [10].

Associated with transcriptional repression, ~74% enrichment in H3K27me3 loci was observed for genomic regions associated with closed/inactive chromatin while ~19% of the putative H3K27me3 loci were associated with the heterochromatin [10]. There was no correlation (R: –0.1) between peaks identified for H3K27me3 and H3K27ac, reassuringly, since acetylation of H3 histone at lysine 27 is associated with active genomic regions and enhancers (Figure 3B).

Peaks identified for H3K27ac showed ~55% and ~35% enrichment for enhancers and promoters, respectively [10]. Illustrative of the link between H3K27ac and transcriptional activation, the association with active promoters (red) and strong enhancers (orange) for the *CHD8* gene—encoding a chromatin remodeling factor—is shown (Figure 3C).

Motif analysis of CTCF genomic binding sites

GEM, a software tool for motif discovery, was used to discover a DNA-binding motif for CTCF from Ion Proton™ ChIP-seq data [11]. Enriched motifs (k-mers) were discovered by comparing frequencies between experimental and input control ChIP-seq results. Optimal GEM analysis (p-value (hgp) of 1e-446.1 and 449-fold enrichment) identified an overrepresented motif in 1347 sites or 5% of the peaks in CTCF run 1 with a positional weight matrix that matches the established CTCF consensus motif (Table 1B, Supplementary information and Figure 4). These results recapitulated the canonical CTCF motifs identified using MEME, a *de novo* motif discovery tool, that was applied across different cell types and vertebrate species (human, mouse, and chicken) [8].

Dendritic cells as a model for genomic programing

Dendritic cells (DCs) orchestrate a range of important functions during innate and adaptive immune responses [12]. A central DC function is the capture, processing, and presentation of antigens—via major histocompatibility complex (MHC) proteins—to T-cells. Located at the body surfaces, DCs interact with the environment to capture antigens. In response to specific stimuli such as microbial products, micro-environmental cues, and endogenous immune system signaling, DCs can differentiate from bone marrow stem cells along multiple lineage pathways. These lineage pathways are governed via groups of transcription factors that tightly regulate gene expression patterns specific to the immune function of each lineage [13]. For example, the transcription factor Irf4 is required for DC lineage development from bone marrow stem cells to CD11b+ DCs; a lineage implicated in inflammatory disease such as asthma [14]. Thus, the multiple DC lineage pathways provide an excellent model system for understanding the environmental regulation of transcriptional programs governing cellular development and the differentiation programs governing important immune cell activities that may contribute to inflammatory disease.

ChIP-seq with lower-input numbers of purified cells

Depending on the sample source, cell lines versus *in vivo*–derived cells, ChIP-seq experiments can be confounded by low-input DNA amounts. To improve the signal-to-noise ratio it is standard to recommend that at least 10^7 cells be used for immunoprecipitation [15]. However, working with flow-sorted, specific cell types from *in vivo* samples often limits availability on the order of 10^6 cells. For example, the genome-wide characterization of the CD11b+ DC lineage by ChIP-seq has been difficult due to the low amount of DNA that results from highly purified DC subpopulations. To address the challenges of limited sample input and to test the utility and robustness of the optimized ChIP-seq protocol, ChIP with antibodies to Irf4 was used to determine the genomic binding sites present using flow cytometry–sorted populations of 1, 3, 5, or 20 million mouse DCs of the CD11b+ lineage.

Highly correlated peak identification across a range of cell number inputs

Peaks associated with Irf4 binding were compared between cell-sorted populations of 1, 3, 5, and 20 million CD11b+ DCs; with peak comparisons restricted to significant predictions. To be considered statistically significant, a peak must pass a minimum threshold of $-\log_{10}$ q-value ≥20 in all samples in which the peak appears.

A Venn diagram illustrates excellent agreement with 61% (n = 10,632), out of a total of 17,417 significant peaks, that were in common across all cell inputs (Figure 5). When compared to 20 million cell input, ChIP-seq libraries made from 5 and 3 million cells observed a decline in total number of significant peaks but were highly correlated, with 93% (n = 11,092) and 94% (n = 10,704) in common, respectively. At the lowest input level of 1 million cells there was a decline to 80% (n = 11,263) of the identified regions that were

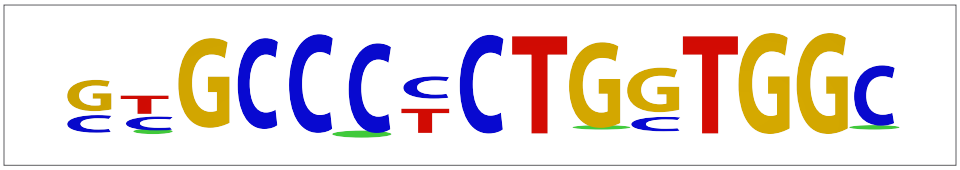

Figure 4. Motif analysis for CTCF consensus binding site in ChIP-seq peaks identified from K-562 cells. Optimal GEM motif analysis (p-value (hgp) of 1e-446.1 and 449-fold enrichment using kmin = 16, kmax = 23) identified 1347 peaks enriched, compared to 3 peaks from the input control, with a position weight matrix (PWM) of 15.21/25.32.

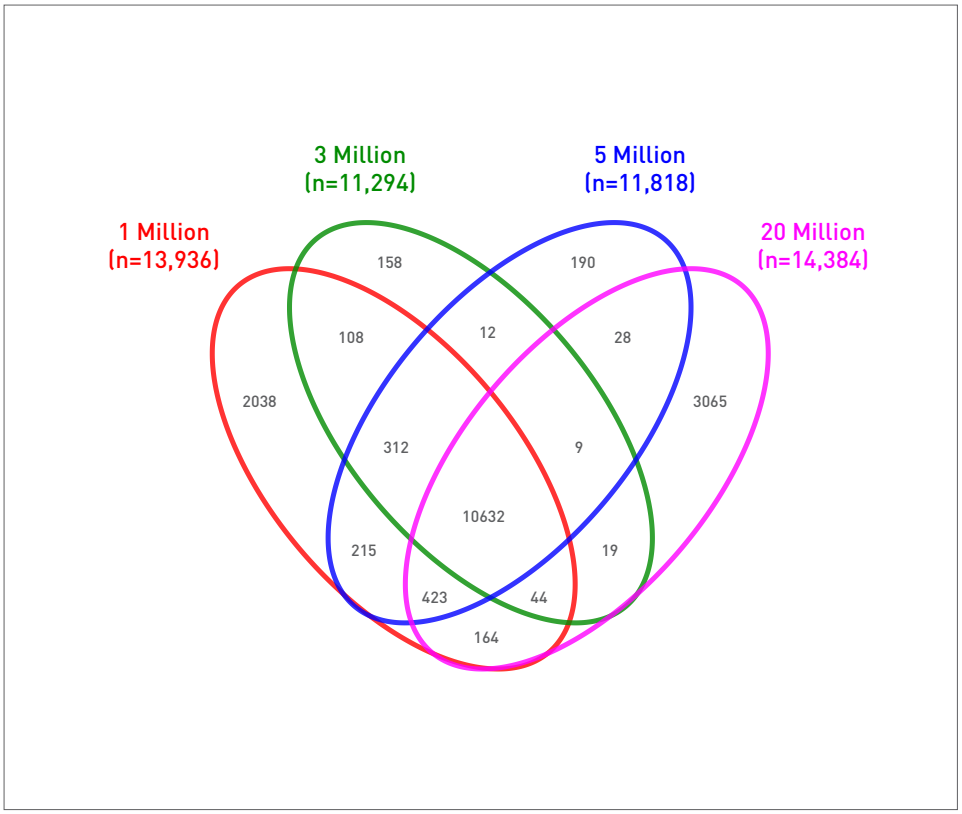

Figure 5. Venn diagram of the overlap in significant peaks with q-values of ≥20 (n = 17,417) identified by ChIP-seq from cell sorted populations of 1, 3, 5, and 20 million CD11b+ lineage mouse dendritic cells.

in common with 20 million cell input, due largely to an increase to 12% (n = 2038) of the peaks that were novel to the lowest cell input and not present in ChIP-seq data from the other input amounts.

Pathway analysis of Irf4 ChIP-seq peaks in CD11b+ DCs

Ion Proton™ semiconductor-based ChIP-seq results were compared to data from a recent publication examining Irf4 regulation of DC via ChIP-seq analysis using the Illumina HiSeq® System [16]. The peaks detected by Irf4 ChIP from CD11b+ DCs were ranked by significance, confirming the Irf4 gene targets highlighted by Lugt *et al*. Genes and biological pathways identified by significant peaks (q-value) include: *CD46*, a complement receptor that regulates DC function and adaptive immunity; *Mid1*, which encodes an E3 ligase implicated in innate immunity and asthma; *F13a*, a component of the coagulation pathway and known mediator of innate immunity and DC function; and *Zbtb46*, a transcription factor involved in DC differentiation and a known target of Irf4 (Figure 6). Correspondence between putative Irf4 peaks in *Zbtb46* and *F13a1* from the immunoprecipitation of 1, 3, 5, or 20 million mouse DCs of the CD11b+ lineage supports the ability to use the Ion Proton™ platform to accurately interrogate small populations of primary immune cells via ChIP-seq (Figure 7).

Conclusions

Specifically developed for research with precious samples such as primary cells, stem cells, and archived sample material, the MAGnify™ Chromatin Immunoprecipitation System, in combination with an optimized ChIP-seq protocol for the Ion Proton™ System, demonstrates high-quality data from limited cell populations of flow cytometry–sorted mouse CD11b+ DCs. Further, sample multiplexing of four barcoded libraries on a single Ion PI™ Chip, revealed highly concordant results when compared to ENCODE project data using the equivalent of 1 million K-562 cells when analyzed for genomic binding sites occupied by a transcription factor (CTCF), and histone modifications associated with activation (H3K27ac) and silencing (H3K27me3). Successful ChIP-seq analysis of low cell number inputs using the Ion Proton™ System indicates that the discovery of regions and associated with Irf4 binding are concordant with larger cell number inputs. Overlapping Irf4-associated peaks with significance were used to identify genes with potential biological significance for mouse CD11b+ DC function. The Ion Proton™ System, with scalable chips, rapid sequencing run times, and affordable instrument pricing, enables the implementation of a ChIP-seq solution that is highly reproducible, robust, and accessible for research studies with sample size variability.

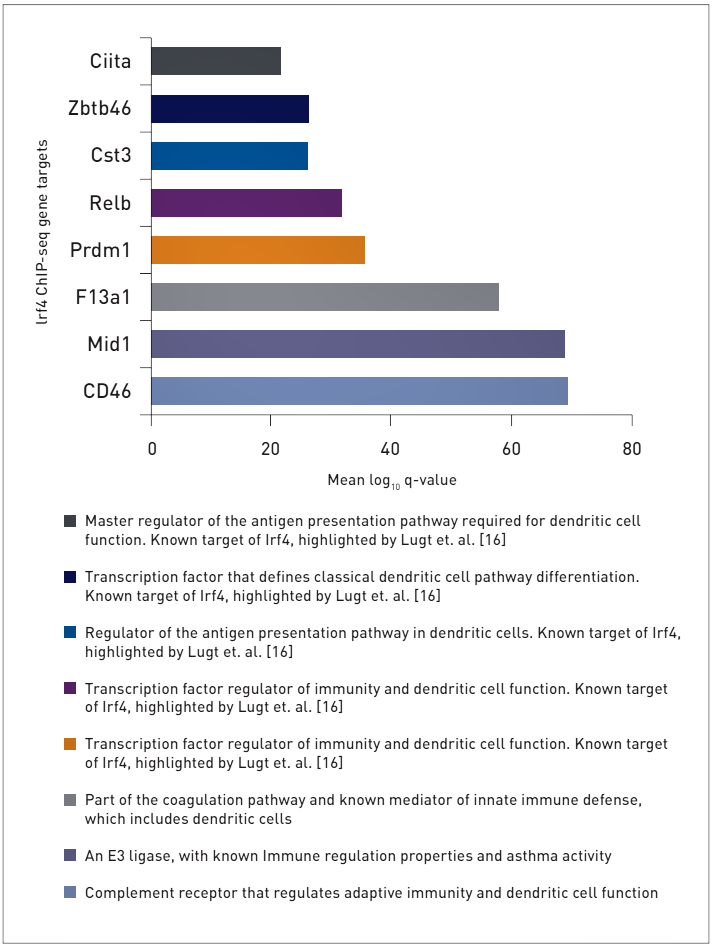

Figure 6. Genes and biological pathways identified by Irf4 ChIP-seq peaks. Relevant biological pathway genes with high-confidence Irf4 binding sites (q-value). Known Irf4 binding targets are noted.

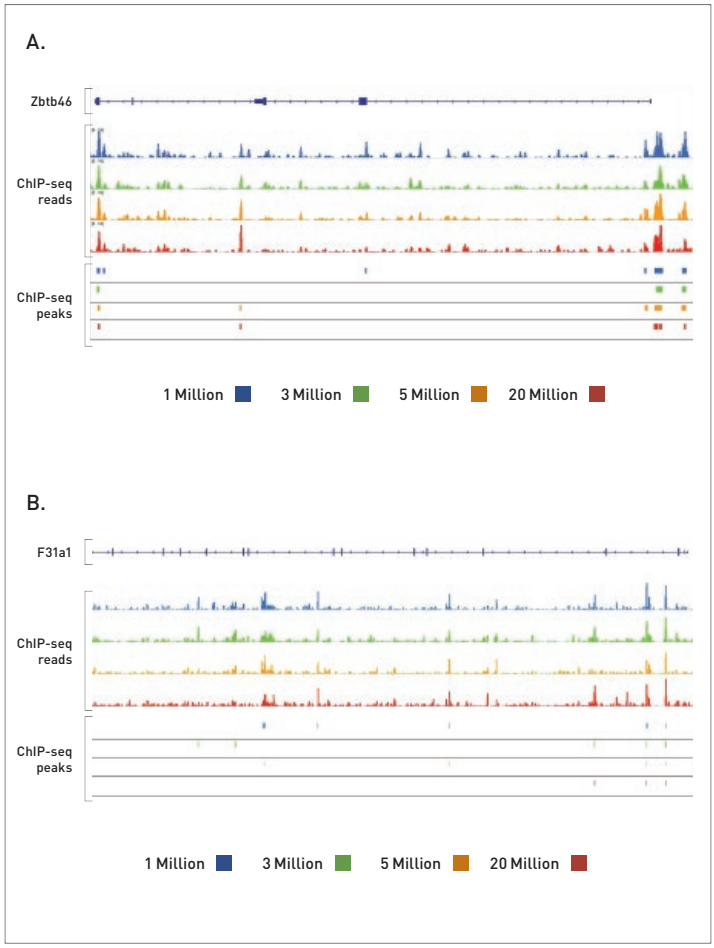

Figure 7. ChIP-seq peaks from the immunoprecipitation of 1, 3, 5, or 20 million mouse CD11b+ DCs for *Zbtb46* (A) and *F13a1* (B) genes.

Acknowledgements

Dr. Brian P. O'Connor, Dr. Sonia Leach, and the flow cytometry facility (Integrated Center for Genes, Environment & Health (CGEH) at National Jewish Health). Research in the O'Connor laboratory is supported by the Walter Scott Foundation.

References

1. Goldberg AD, Allis CD, and Bernstein E (2007) Epigenetics: A Landscape Takes Shape. *Cell* 128:635–638
2. Rada-Iglesias A, Bajpai R, Swigut T et al. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470:279–283
3. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57-74
4. Tao Y, Zhang M, Li L, et al. (2014) Pitx2, an Atrial Fibrillation Predisposition Gene, Directly Regulates Ion Transport and Intercalated Disc Genes. *Cir Cardiovasc Genet* 7:23–32
5. Cheng CS, Rai K, Garber M, et al. (2013) Semiconductor-based DNA sequencing of histone modification states. *Nat Commun* 4:2672
6. Feng J, Liu T, Qin B, Zhang Y, and Liu XS (2012) Identifying ChIP-seq enrichment using MACS. *Nat Protocols* 7:1728–1740. (https://github.com/taoliu/MACS/)
7. Landt SG, Marinov GK, Kundaje A, et al. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia *Genome Res* 22:1813–1831
8. Martin D, Pantoja C, Minan AF, et al. (2011) Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* 18:708–714
9. Wallace JA and Felsenfeld G (2007) We gather together: insulators and genome organization. *Curr Opin in Genet Dev* 17:400–407
10. ENCODE comparisons are presented against BroadChromHMM, K562 cell-line data for Broad Histone, Hudson Alpha (HAIB) CTCF, HAIB RR DNA methylation (genome.ucsc.edu)
11. Guo Y, Mahony S, and Gifford DK (2012) High Resolution Genome Wide Binding Event Finding and Motif Discovery Reveals Transcription Factor Spatial Binding Constraints. *PLoS Comput Biol* 8:e1002638
12. Steinman RM and Banchereau J (2007) Taking dendritic cells into medicine *Nature* 449:419–426
13. Satpathy AT, KC W, Albring JC et al. (2012) Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J Exp Med* 209:1135–1152
14. Satpathy AT, Wu X, Albring JC, and Murphy KM (2012) Re(de)fining the dendritic cell lineage *Nat Immunol* 13:1145–1154
15. Kidder BL, Hu G, and Zhao K (2011) ChIP-Seq: technical considerations for obtaining high-quality data *Nat Immunol* 12:918–922
16. Lugt BV, Khan AA, Hackney JA, et al. (2014) Transcriptional programming of dendritic cells for enhanced MHC class II antigen presentation. *Nat Immunol* 15:161–167

For qualified epigenetic antibodies for ChIP-seq, visit:
lifetechnologies.com/ionchip-seq

life technologies

A Thermo Fisher Scientific Brand

For Research Use Only. Not for use in diagnostic procedures. © 2014 Thermo Fisher Scientific Inc. All rights reserved. Pippin Prep is a trademark of Sage Science, Inc. BioRuptor is a trademark of Diagenode SA. Agencourt and AMPure are trademarks of Beckman Coulter, Inc. Bioanalyzer and Agilent are trademarks of Agilent Technologies, Inc. HiSeq is a trademark of Illumina, Inc. ChiP-IT is a trademark of Active Motif, Inc. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. **CO28266 1014**

Optimized ChIP-seq workflow for the Ion Proton™ System

The optimized ChIP-seq workflow is illustrated in Figure 1. It is recommended to quantify the chromatin, using the Qubit® HS kit or a similar method, and standardize the input amount prior to ChIP isolation. Similarly, to assess the completeness of sonication, it is advisable to run sonicated chromatin on a 2% agarose e-gel following the reversal of crosslinking to determine that the majority of the chromatin is <300 bp in size.

For optimization of library preparation steps, Ready-to-ChIP Chromatin from K-562 cells (Active Motif, catalog no. 53020)—previously grown, cross-linked, and pre-sonicated (via the ChIP-IT® Express protocol)—was sheared by sonication using Bioruptor® Pico to <300 bp in size and used for ChIP. Using the equivalent of 1 million cells, the MAGnify™ Chromatin Immunoprecipitation System was used to isolate and purify genomic regions interacting with CTCF (Cell Signaling Technology, catalog no. 2899), H3K27me3 (Abcam, catalog no. ab6002), and H3K27ac (Abcam, catalog no. ab4729).

An enhanced Ion Proton™ System library construction protocol was used, with full details found here: http://ioncommunity.iontorrent.com/docs/DOC-9003. To reduce sample variability it is recommended to use a single protocol across all samples, with standardization of the critical steps of size selection and amplification. In summary, the following was performed: (i) end repair of ≤10 ng of ChIP DNA fragments followed by purification using the Agencourt® AMPure® XP Kit; (ii) P1 adapters and A adapters with barcodes are ligated and fragments nick-repaired followed by purification; (iii) library amplification with 18 cycles, avoiding over amplification, and subsequent purification; (iv) size selection using Pippin Prep™ system Library or double solid-phase reversible immobilization (SPRI) (AMPure® XP Kit) size selection with ensuing size distribution analysis on the Bioanalyzer® instrument using the Agilent® High Sensitivity DNA Kit. Performing size selection later in library construction improves the sequence diversity of the ChIP fragments and decreases adapter dimer contamination, resulting in easier sample multiplexing, greater sequencing output, and simpler bioinformatics analysis. Assessing the quality of the ChIP-seq libraries prior to template preparation and sequencing is critical. Based on the estimated fragment size range from the Bioanalyzer® instrument, samples can be quantified and multiplexed.

Samples were multiplexed by normalizing libraries to the same molarity and combined to create a final pool. Template preparation and sequencing were performed with four barcoded samples per reaction. Multiplexing experimental and input control samples helps minimize run-to-run variation that can complicate bioinformatics analysis. Libraries were clonally amplified with the Ion PI™ Template OT2 200 Kit v3 and the Ion OneTouch™ 2 System for template generation and enrichment prior to chip loading. Sequencing was performed using the Ion PI™ Sequencing 200 Kit v3 and the Ion PI™ Chip v2.

life technologies

Table 1. ChIP-seq results on the Ion Proton™ System using the Ion PI™ Chip and the MAGnify™ Chromatin Immunoprecipitation System. ChIP was performed using antibodies against CTCF, H3K27me3, and H3K27ac to characterize genomic occupancy in K-562 cells. Sequencing was performed on the Ion Proton™ System using the Ion PI™ Chip v2 with 4 samples per chip. Primary data analysis was performed with Torrent Suite™ Software and peak calling was performed via Model-based Analysis for ChIP-Seq (MACS v2.0.10) using command-line with standard defaults (macs2 callpeak -t ChIP.bam -c Control.bam -f BAM -g mm -n OutputName_uniq -q 0.001) [6].

A. Sequencing statistics

ChIP	Ion Proton™ sequencing run	Total reads (million)	Percent of reads/ Ion PI™ Chip	Fraction of unique reads*	Non-redundant fraction**	Mean read length
CTCF	Run 1 (LUC-242)	22.1	25.1	0.321	0.49	124
CTCF	Run 2 (GOB-276)	21.0	24.0	0.327	0.56	124
H3K27me3	Run 1 (LUC-242)	23.2	26.3	0.823	0.87	131
H3K27me3	Run 2 (GOB-276)	25.8	29.5	0.817	0.89	132
H3K27ac	Run 1 (LUC-242)	20.1	22.8	0.834	0.87	131
H3K27ac	Run 2 (GOB-276)	19.4	22.1	0.823	0.89	132
Input	Run 1 (LUC-242)	22.7	25.8	0.855	0.87	125
Input	Run 2 (GOB-276)	21.3	24.3	0.841	0.88	126

* Internal plug-in calculates fraction of unique reads (mapped to chromosome 1 only for speed)
** 1-redundant fraction calculated by MACS [6]

B. Immunoenrichment and ENCODE project concordance for ChIP-seq using K-562 cells

ENCODE dataset	Ion Proton™ sequencing run	ChIP target	Total no. of sites*	Putative occupancy concordance with indicated ENCODE dataset			
				Sites with no overlap (No. of sites)	Sites with any overlap (No. of sites)	Sites with ≥ 50% overlap (No. of sites)	Sites with 100% overlap (No. of sites)
HAIB†	Run 1 (LUC-242)	CTCF	27,042	48% (12,949)	52% (14,092)	51% (13,852)	45% (12,199)
	Run 2 (GOB-276)	CTCF	17,538	31% (5,519)	69% (12,018)	68% (11,888)	60% (10,571)
Broad‡	Run 1 (LUC-242)	H3K27ac	60,686	29% (17,535)	71% (43,150)	68% (41,243)	63% (35,101)
	Run 2 (GOB-276)	H3K27ac	52,861	27% (14,144)	73% (38,716)	70% (37,105)	66% (35,101)
Broad§	Run 1 (LUC-242)	H3K27me3	56,749	11% (6,224)	89% (50,524)	88% (49,932)	87% (49,469)
	Run 2 (GOB-276)	H3K27me3	57,742	11% (5,969)	89% (48,772)	88% (48,123)	87% (47,632)

* False discovery rate (FDR) threshold of 10^{-3}
† wgEncodeHaibTfbsK562CtcfcPcr1xPkRep2
‡ wgEncodeBroadHistoneK562H3k27acStdPk
§ wgEncodeBroadHistoneK562H3k27me3StdPk

For qualified epigenetic antibodies for ChIP-seq, visit:
lifetechnologies.com/ionchip-seq

life technologies
A Thermo Fisher Scientific Brand

For Research Use Only. Not for use in diagnostic procedures. © 2014 Thermo Fisher Scientific Inc. All rights reserved. Pippin Prep is a trademark of Sage Science, Inc. BioRuptor is a trademark of Diagenode SA. Agencourt and AMPure are trademarks of Beckman Coulter, Inc. Bioanalyzer and Agilent are trademarks of Agilent Technologies, Inc. HiSeq is a trademark of Illumina, Inc. ChiP-IT is a trademark of Active Motif, Inc. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. **CO28266 1014**