Published in final edited form as:

Nat Methods. 2010 August ; 7(8): 615–618. doi:10.1038/nmeth.1478.

Genomewide Chromatin Maps Derived from Limited Numbers of Hematopoietic Progenitors

Mazhar Adli^{1,2,3}, Jiang Zhu^{1,2,3}, and Bradley E. Bernstein^{1,2,3}

¹ Howard Hughes Medical Institute, Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

² Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

³ Center for Systems Biology and Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts, USA

Abstract

Current methods for whole genome mapping of protein-DNA interactions, performed by coupling chromatin immunoprecipitation with high-throughput sequencing (ChIP-Seq), require large amounts of starting materials which precludes their application to rare cell types. Here, we combine a high-sensitivity ChIP assay with a novel library preparation procedure to map histone modifications in as few as 10,000 cells. We apply the technique to characterize murine hematopoietic progenitors and thereby gain insight into their developmental program.

ChIP-Seq has become a standard tool for whole genome mapping of histone modifications, transcription factors and chromatin-associated proteins¹. However, the technique is plagued by inefficiencies at ChIP and sequencing steps that translate into a requirement for large amounts of starting materials, typically on the order of millions or tens of millions of cells. ChIP assays yield small amounts of DNA whose availability for downstream assays is further reduced by DNA damage introduced during fixation and fragmentation. Although locus-specific analysis has been achieved for small cell numbers by coupling ChIP with PCR or promoter arrays^{2–5}, such assays do not achieve the comprehensiveness afforded by DNA sequencing approaches. Sequencing, however, requires relatively large amounts of DNA, in part because library preparation involves an inefficient ligation step that requires double-stranded DNA with intact ends. Direct sequencing of DNA without need for ligation or amplification⁶ can address some of these inefficiencies and interrogate small quantities of immunoprecipitated DNA. However, neither standard nor direct ChIP-Seq approaches have so far been successfully applied to small cell numbers. We therefore sought to develop a ChIP-Seq assay for limited numbers of cells using the Illumina Genome Analyzer, the dominant short read sequencing platform.

Correspondences to Bernstein.Bradley@mgh.harvard.edu.

Author's contributions

M.A. and B.E.B. designed the method, M.A. performed the experiments. M.A., J.Z. and B.E.B. analyzed the data and wrote the manuscript.

Competing Interest Statement

The authors declare competing financial interests: M.A. and B.E.B. have filed a patent application describing the methods presented here.

Accession codes

ata sets can be accessed from (http://www.broadinstitute.org/cgi-bin/seq_platform/chipseq/shared_portal/clone/Nano_ChIP_Seq.py) Gene Expression Omnibus (GEO) accession code for sequencing reads: GSE22075.

First, we optimized the ChIP assay for small samples. We identified reproducible conditions for shearing small amounts of chromatin (Supplementary Fig. 1; Methods) and minimized background due to non-specific pull-down by reducing sample volumes and titrating quantities of antibody and beads (Supplementary Fig. 1; Methods). These improvements enabled us to effectively enrich for histone H3 lysine 4 trimethylation (K4me3) in a ChIP performed with 10,000 mouse embryonic stem (ES) cells, as judged by quantitative PCR (qPCR).

However, the modified ChIP procedure yielded scarce quantities of DNA, below the detection limit of fluorometry and thus remained incompatible with standard ChIP-Seq protocols which require several nanograms of ChIP DNA. We estimated that a K4me3 ChIP performed on 10,000 cells yielded ~10–50 picograms of DNA, and sought to develop a library preparation procedure compatible with these amounts. We initially pursued a strategy that coupled a random primer-based amplification procedure widely used in microarray studies⁷ with standard Illumina library preparation. However, when we sequenced the resulting libraries and integrated the reads into density maps, we observed many false positive peaks, a conspicuous absence of signal over GC-rich regions and high numbers of unalignable reads that likely reflect primer artifacts.

We therefore implemented a series of strategies to overcome these problems (Fig. 1a). First, we tested modified random primers that carry bulky chemical groups or were designed to form secondary structures that prevent self-annealing. By using a random primer hairpin structure and implementing an exonuclease digestion step after priming, we were able to amplify small quantities of ChIP DNA while minimizing non-specific product (Fig. 1a, Supplementary Fig. 2; Online Methods). Second, we identified additives, cycling conditions and a specific polymerase enzyme that enabled faithful amplification of ChIP DNA, maintaining representation of GC-rich sequences (Fig. 1b; Online Methods). Finally, we digested the amplified DNA at a BciVI site introduced near the ends of the ChIP fragments to yield double-stranded products with 3' A overhangs that could be ligated directly to Illumina adapters for sequencing.

We validated this protocol with qPCR on the sequencing libraries because the scarce starting ChIP samples could not themselves be evaluated. This procedure quantifies short amplicons corresponding to positive- and negative-control genomic sites whose chromatin states are relatively invariant across cell types (Fig. 1b, Supplementary Fig. 1; Online Methods).

We deep sequenced a library prepared from an H3K4me3 ChIP performed on 10,000 ES cells. We obtained roughly 10 million 36 base reads. Initial alignments indicated that the first 9 bases of the sequencing reads had higher mismatch rates, possibly due to imperfect hybridization of random primers. Accordingly, subsequent alignments were performed using bases 10 to 36 of each read. Seven of the 10 million sequenced reads could be aligned to the mouse reference genome, as is typical for ChIP-Seq experiments. Aligned reads were processed into a density profile using standard procedures⁸. Visual analysis of the ChIP-Seq map generated from 10,000 ES cells suggested good concordance to a standard H3K4me3 ChIP-Seq map generated from roughly 10 million ES cells (Fig. 2a), with striking peaks at a majority of GC-rich promoters. A more quantitative analysis revealed roughly 93% overlap between a set of 11,193 H3K4me3-enriched promoters identified in the low cell number dataset and a set of 12,079 enriched promoters identified from the standard dataset (Fig. 2b). Considering the standard ChIP-Seq dataset as a gold standard, we estimate that the low cell number ChIP-Seq dataset has a sensitivity of ~80% at a specificity of ~90% (Fig. 2c; Online Methods).

Next, we profiled chromatin in hematopoietic stem and progenitor cells isolated from mouse bone marrow and enriched for the immunophenotype Lineage⁻, Sca-1⁺ and c-kit⁺ (LSK) (Supplementary Fig. 3; Online Methods). Using roughly 20,000 cells per assay, we profiled

H3K4me3 and two other H3 modifications, lysine 27 trimethylation (H3K27me3) and lysine 36 trimethylation (H3K36me3). These experiments yielded between 13 and 19 million reads, of which roughly 75% could be aligned to the mouse reference genome. Several lines of evidence suggest that the resulting chromatin maps accurately reflect the true patterns of histone modification in the LSK population. H3K4me3 and H3K36me3 localize to gene promoters and gene bodies, respectively, consistent with known roles for these marks in transcriptional initiation and elongation⁹. Integration of published gene expression profiles for LSKs¹⁰ confirmed expected positive (H3K4me3, H3K36me3) and negative (H3K27me3) correlations between the modifications and the transcriptional status of corresponding genes (Fig. 2d). Finally, we confirmed the reproducibility of the method by profiling all three modifications in a second LSK population (Fig. 2e,f).

The LSK chromatin profiles revealed several interesting findings of potential relevance to hematopoietic development. First, we observed a relatively small number of large H3K4me3 regions or ‘domains’ (>10 kb) that are distinct from the sharp peaks observed at most promoters (Fig. 3a). In ES cells, the largest H3K4me3 domains correspond to pluripotency genes such as *Oct4*⁸. In the hematopoietic progenitors, the largest H3K4me3 domains coincide with known hematopoietic regulators, including *HoxB4*, *HoxA7*, *HoxA9*, *Runx1*, *Meis1*, *Irf2* and *Tal1* (Fig. 3a-c, Supplementary Table 1).

To investigate the significance of the H3K4me3 domains, we examined chromatin maps for ES cells⁸. Roughly a third of loci with H3K4me3 domains in LSKs carry both H3K4me3 and H3K27me3 in ES cells. Such ‘bivalent domains’ have been correlated to developmental loci that are silent but poised for activation at later development stages^{11,12}. Reasoning that this subset of loci could illuminate genes with critical functions in hematopoietic progenitors, we collated genes marked by bivalent domains in ES cells and H3K4me3 domains in LSKs. The resulting list includes 40 genes, 30 of which encode transcription factors or other developmental regulators with previously described functions in hematopoiesis (Supplementary Table 2).

Although initially described in ES cells, bivalent domains are also present in multipotent cells, including hematopoietic progenitors^{8,13,14}. The LSK chromatin maps contain roughly 1700 promoters with detectable H3K27me3 and H3K4me3. A majority of these promoters are also bivalent in ES cells, but they have variable patterns in CD4⁺ T-cells, a differentiated progeny of LSKs¹⁵ (Fig. 3d, Supplementary Fig. 4).

Whereas essentially all promoters with H3K27me3 in ES cells also have high K4me3, K27me3-marked promoters in LSKs have a wide range of K4me3 levels. We used this opportunity to explore the significance of the bivalent state in an *in vivo* setting, specifically asking whether the level of H3K4me3 predicts the likelihood that a given promoter will be activated in differentiated progeny. We integrated our chromatin data with a compendium of gene expression profiles for differentiated hematopoietic cells^{15,16}. We found that the amount of H3K4me3 at a given H3K27me3-marked promoter in LSKs significantly correlated with the number of differentiated cell types in which the corresponding transcript could be detected (Fig. 3e). Accordingly, master regulators of hematopoietic lineages, such as *Pax5* and *Ebfl*, have high levels of H3K27me3 and H3K4me3 in LSKs (Fig. 3d· Supplementary Fig. 5). In contrast, master regulators that function in non-hematopoietic lineages, such as *MyoD* and *Sox3*, are enriched for H3K27me3, but not for H3K4me3 (Fig. 3d· Supplementary Fig. 5). Thus, the presence of bivalent chromatin in hematopoietic progenitors is associated with an increased likelihood of transcriptional induction during differentiation. Further study is needed to determine the causality of these chromatin structures and their relationship to transcriptional priming and to the developmental potential of progenitor cells¹⁷. This Chip-seq approach should be generally applicable for characterizing chromatin landscapes in biologically- and clinically-important cell models that have been inaccessible due to inadequate sample size.

Going forward, further reductions in cell numbers may be achievable by careful optimization of ChIP conditions. The library preparation scheme and accompanying validation procedure may also be useful for profiling transcription factors from relatively smaller samples. However, this application is likely to prove challenging due to variability among factors and affinity reagents and the exceedingly small DNA yields associated with such experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dave Flowers and Irwin Bernstein for expertise, reagents and assistance in isolation of hematopoietic cells. We also thank Noam Shoreh and Tarjei Mikkelsen for computational assistance, and Birgit Knoechel, Eric Mendenhall, Alon Goren and Andrew Chi for constructive discussions and critical reading of the manuscript. B.E.B. is a Charles E. Culpeper Medical Scholar and an Early Career Scientist of the Howard Hughes Medical Institute. This research was supported by funds from the Starr Cancer Consortium, the National Human Genome Research Institute, the NIH Roadmap for Epigenomics and the National Heart, Lung and Blood Institute.

References

1. Park PJ. *Nat Rev Genet* 2009;10:669–80. [PubMed: 19736561]
2. Acevedo LG, Iniguez AL, Holster HL, et al. *Biotechniques* 2007;43:791–797. [PubMed: 18251256]
3. O'Neill LP, VerMilyea MD, Turner BM. *Nat Genet* 2006;38:835–841. [PubMed: 16767102]
4. Attema JL, Papanthanasios P, Forsberg EC, et al. *Proc Natl Acad Sci U S A* 2007;104:12371–12376. [PubMed: 17640913]
5. Dahl JA, Collas P. *Nucleic Acids Res* 2008;36:e15. [PubMed: 18202078]
6. Goren A, Ozsolak F, Shoreh N, et al. *Nat Methods* 2009;7:47–49. [PubMed: 19946276]
7. Lieb JD, Liu X, Botstein D, et al. *Nat Genet* 2001;28:327–343. [PubMed: 11455386]
8. Mikkelsen TS, Ku M, Jaffe DB, et al. *Nature* 2007;448:553–560. [PubMed: 17603471]
9. Li B, Carey M, Workman JL. *Cell* 2007;128:707–719. [PubMed: 17320508]
10. Tothova Z, Kolipara R, Huntly BJ, et al. *Cell* 2007;128:325–339. [PubMed: 17254970]
11. Azuara V, Perry P, Sauer S, et al. *Nat Cell Biol* 2006;8:532–538. [PubMed: 16570078]
12. Bernstein BE, Mikkelsen TS, Xie X, et al. *Cell* 2006;125:315–326. [PubMed: 16630819]
13. Cui K, Zang C, Roh TY, et al. *Cell Stem Cell* 2009;4:80–93. [PubMed: 19128795]
14. Weishaupt H, Sigvardsson M, Attema JL. *Blood* 2009;115:247–256. [PubMed: 19887676]
15. Wei G, Wei L, Zhu J, et al. *Immunity* 2009;30:155–167. [PubMed: 19144320]
16. Chambers SM, Boles NC, Lin KY, et al. *Cell Stem Cell* 2007;1:578–591. [PubMed: 18371395]
17. Bottardi S, Ghiam AF, Bergeron F, et al. *Cell Cycle* 2007;6:1035–1039. [PubMed: 17457053]

Figure 1. Preparation of sequencing library from low amounts of ChIP DNA

(a) Schematic of ChIP-Seq library preparation for low cell numbers. Universal PCR sequence (orange), restriction site (gray) and random 9-mer (blue) are indicated. (b) qPCR data for H3K4me3 ChIP-Seq libraries from indicated cell numbers. Bar graph shows relative enrichments at four gene promoters with H3K4me3 in ES cells. Data are normalized against an intergenic control region on chromosome 18. Error bars indicate s.e.m. based on 3 experimental replicates

Figure 2. Validation of low cell number ChIP-Seq maps

(a) K4me3 H3K4me3 in ES cells. Data are normalized against an intergenic control region on chromosome 18. Error bars indicate s.e.m. based on 3 experimental replicates. Profiles are compared across a 50 kb region on the second chromosome. (b) Venn diagram shows overlap between H3K4me3-marked promoters. (c) Receiver operating characteristic (ROC) curve shows true-positive and false-positive rates for H3K4me3 low cell number ChIP-Seq data, evaluated for 1 kb windows across the genome, using the standard H3K4me3 ChIP-Seq dataset as a gold standard. Dotted red line indicates $X = Y$. (also see Supplementary Figure 6 & 7). (d) Chromatin maps from 20,000 LSK cells in biological replicate. Contour plots depict correlations between promoter (H3K4me3, H3K27me3) or gene body (H3K36me3) methylation levels and mRNA expression in LSKs. Each datapoint corresponds to a single gene. (e) Contour plots depict correlations between LSK replicates across gene promoters as well as gene bodies. (f) H3K4me3, H3K27me3 and H3K36me3 signal tracks for the independent LSK replicates are shown for the HoxA cluster and the *Flt3* locus. The Y-axis of each profile represents the estimated number of ChIP fragments at each position normalized by the total number of reads in a given dataset. RefSeq gene annotations are also shown.

Figure 3. Chromatin domains at developmental regulators in hematopoietic progenitors
(a) LSKs have large H3K4me3 domains that mark hematopoietic regulators and punctate peaks at most other promoters. **(b)** Histogram with the size distribution of H3K4me3 intervals associated with gene promoters in LSKs; the high end of the size distribution is enriched for key hematopoietic regulators. **(c,d)** ChIP-Seq signal tracks for ES cells and LSKs across *Zfp608* and *Zeb2* (c) and *Lef1* and *Sox3* (d). **(e)** Heat map of genes with H3K27me3 and variable levels of H3K4me3 in LSKs, depicts their expression levels in different hematopoietic cell types (blue indicates low expression; red indicates high expression). The number of cell types in which a gene is expressed is plotted vertically in the right panel.