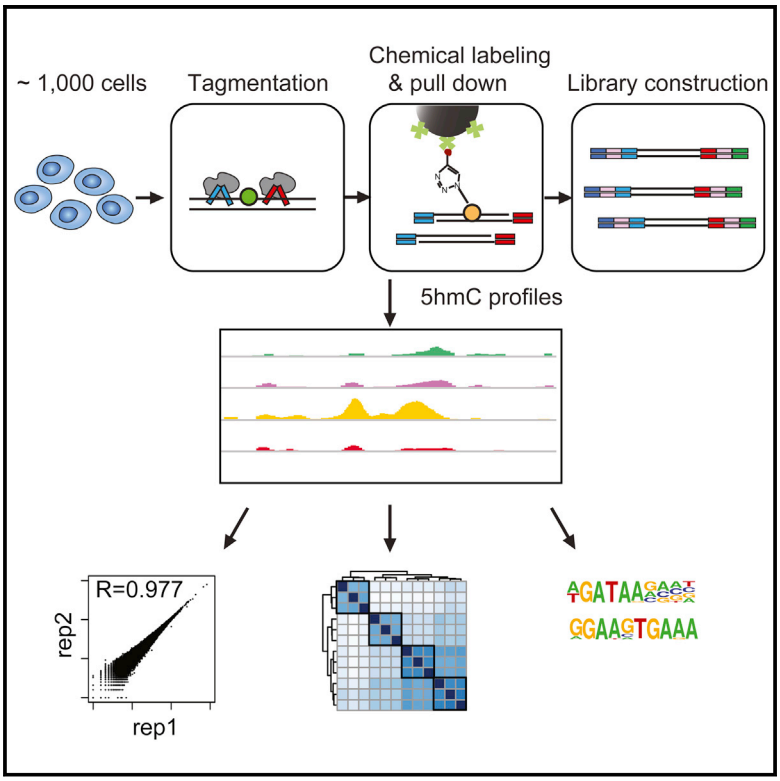
Technology

Molecular Cell

A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations

Graphical Abstract



Highlights

- Chemical labeling enables highly sensitive profiling of 5hmC using limited input DNA
- Genome-wide 5hmC distributions obtained for hematopoietic stem cell differentiation
- 5hmC changes correlate with differential gene expression in an AML model
- Redistributions of 5hmC in the Tet2-deficient leukemia stem cells

Authors

Dali Han, Xingyu Lu, Alan H. Shih, ..., Ari M. Melnick, Ross L. Levine, Chuan He

Correspondence

leviner@mskcc.org (R.L.L.), chuanhe@uchicago.edu (C.H.)

In Brief

Han et al. developed a highly sensitive and robust 5hmC sequencing approach using genomic DNA isolated from ~1,000 cells. Using their approach, 5hmC maps across different stages of mouse hematopoietic differentiation were obtained and insights into roles of 5hmC in tumor-initiating cells in an AML mouse model were gained.

Accession Numbers

GSE77967

Han et al., 2016, Molecular Cell *63*, 1–9
August 18, 2016
http://dx.doi.org/10.1016/j.molcel.2016.06.028

CellPress

Please cite this article in press as: Han et al., A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations, Molecular Cell (2016), http://dx.doi.org/10.1016/j.molcel.2016.06.028

Molecular Cell
Technology

A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations

Dali Han,[1,8] Xingyu Lu,[1,8] Alan H. Shih,[2,3,8] Ji Nie,[1,4] Qiancheng You,[1] Meng Michelle Xu,[5] Ari M. Melnick,[6,7] Ross L. Levine,[2,3,*] and Chuan He[1,*]
[1]Department of Chemistry, Department of Biochemistry and Molecular Biology, and Institute for Biophysical Dynamics, Howard Hughes Medical Institute, University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA
[2]Leukemia Service, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[3]Human Oncology and Pathogenesis Program and Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[4]Beijing National Laboratory for Molecular Sciences (BNLMS), MOE Key Laboratory of Bioorganic Chemistry and Molecular Engineering, College of Chemistry, Peking University, Beijing 100871, China
[5]Department of Radiation and Cellular Oncology, University of Chicago, Chicago, IL 60637, USA
[6]Department of Medicine/Hematology-Oncology and Department of Pharmacology, Weill Cornell Medical College, New York, NY 10065, USA
[7]Weill Cornell Medical College, Cornell University, 413 East 69th Street, BB-1462, New York, NY 10021, USA
[8]Co-first author
*Correspondence: leviner@mskcc.org (R.L.L.), chuanhe@uchicago.edu (C.H.)
http://dx.doi.org/10.1016/j.molcel.2016.06.028

SUMMARY

We present a highly sensitive and selective chemical labeling and capture approach for genome-wide profiling of 5-hydroxylmethylcytosine (5hmC) using DNA isolated from ~1,000 cells (nano-hmC-Seal). Using this technology, we assessed 5hmC occupancy and dynamics across different stages of hematopoietic differentiation. Nano-hmC-Seal profiling of purified *Tet2*-mutant acute myeloid leukemia (AML) murine stem cells allowed us to identify leukemia-specific, differentially hydroxymethylated regions that harbor known and candidate disease-specific target genes with differential 5hmC peaks compared to normal stem cells. The change of 5hmC patterns in AML strongly correlates with differential gene expression, demonstrating the importance of dynamic alterations of 5hmC in regulating transcription in AML. Together, covalent 5hmC labeling offers an effective approach to study and detect DNA methylation dynamics in in vivo disease models and in limited clinical samples.

INTRODUCTION

5-hydroxymethylcytosine (5hmC), the oxidative product of 5-methylcytosine (5mC) catalyzed by ten-eleven translocation (TET) enzymes, is found in various mammalian tissues and cell types. Emerging evidence indicates that 5hmC is not only an intermediate of DNA demethylation, but also acts as a potential epigenetic mediator, which modulates a spectrum of biological processes and human diseases (Hackett et al., 2013; Lu et al., 2015b; Sun et al., 2014; Topalian et al., 2012). The recent development of high-throughput, genome-wide sequencing technologies has enabled genome-wide mapping of 5hmC in mammalian systems (Booth et al., 2012; Cui et al., 2014; Pastor et al., 2011; Song et al., 2011, 2012; Sun et al., 2015; Yu et al., 2012). While applications of these methods have provided key information about the distribution of 5hmC and its functional insights, the need for a large amount of cells to obtain sufficient genomic DNA starting material for 5hmC localization precludes their use with rare cell populations including normal and malignant stem cells, homogeneous neuronal cells, and clinical isolates including needle biopsies, circulating tumor cells, and cell-free DNA. Therefore, new approaches are needed to allow for the detection of 5hmC in rare cell populations.

Here, we introduce a sensitive and robust 5hmC sequencing approach which allows genome-wide profiling of 5hmC based on a previously invented selective chemical labeling (Song et al., 2011) using a limited amount of genomic DNA that can be readily isolated from ~1,000 cells (nano-hmC-Seal). To demonstrate the advantage and utility of this strategy, we have applied this approach to compare 5hmC profiles between hematopoietic stem cell (HSC) and progenitor cell populations. We found that 5hmC is enriched in the gene body of highly expressed genes and the level of 5hmC positively correlates with histone modifications that mark active transcription. Moreover, we observed that the differentiation of murine HSCs to progenitor cells is strongly associated with dynamic alterations in 5hmC patterns with lineage-specific enhancers marked by pronounced 5hmC peaks. We further applied this technology to profile leukemia stem cells from a murine model of *Tet2*-mutant acute myeloid leukemia (AML)

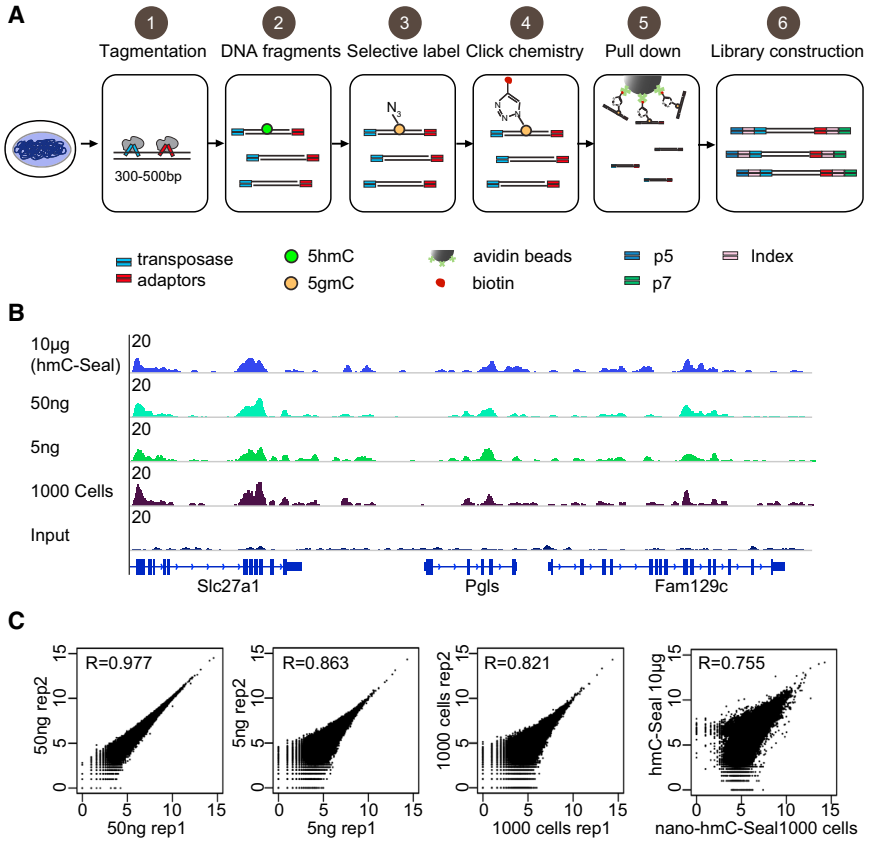


Figure 1. Nano-hmC-Seal to Generate Genome-wide 5hmC Maps from Ultra-Low DNA Starting Materials

(A) Schematic overview of the nano-hmC-Seal approach.

(B) Genome browser views of 5hmC signals detected in a 30 kb region from libraries generated with 5 ng–10 μg of starting genomic DNA from mESCs. In blue, the 5hmC profile obtained using regular hmC-Seal with 10 μg genome DNA (top). Approximately 5–6 ng of genomic DNA could be isolated from 1,000 mESCs (1,000 cells).

(C) Scatterplots showing correlation between nano-hmC-Seal replicates with Pearson correlation (r) displayed. Each dot represents a 5hmC-enriched peak. The read counts were transformed to log2 base. From left to right: correlation between replicate libraries prepared from 50 ng and 5 ng mESC genomic DNA, and genomic DNA isolated from 1,000 mESCs, respectively, and between libraries using 1,000 cells with regular hmC-Seal using 10 μg genomic DNA.

See also Figure S1 and Table S1.

and obtained high-quality maps of 5hmC in tumor-initiating cells.

DESIGN

We modified an engineered Tn5 transposase-based library construction strategy (Adey et al., 2010; Buenrostro et al., 2013; Schmidl et al., 2015), which allows fragmentation of genomic DNA into 300–500-bp fragments and appends sequencing-compatible adaptors in a single tagmentation step (Adey et al., 2010). Next, to enrich 5hmC-containing DNA fragments, we took advantage of the selective 5hmC chemical labeling (hmC-Seal) strategy, which has been previously developed for efficient, unbiased, and genome-wide labeling and covalent pull down of 5hmC (Song et al., 2011, 2012). Specifically, sequencing adaptors are incorporated through the transposase-catalyzed DNA tagmentation (Adey et al., 2010). The T4 bacteriophage enzyme β-glucosyltransferase (βGT) is then employed to transfer an engineered glucose moiety containing an azide-group to 5hmC in duplex DNA, yielding β-6-azide-glucosyl-5-hydroxymethyl-cytosine (N3-5-gmC), as reported previously (Song et al., 2011). Then a biotin tag is installed onto the azide group by using Huisgen cycloaddition (click) chemistry. Finally, 5hmC-containing DNA fragments with biotin tags are efficiently captured from the random DNA fragments pool by avidin beads. A regular PCR amplification reaction generates the library, which is then subjected to high-throughput sequencing (Figure 1A).

RESULTS

Generation of Nano-hmC-Seal Libraries with Ultra-Low Starting Material

We tested the nano-hmC-Seal approach with 50 ng and 5 ng genomic DNA isolated from mouse embryonic stem cells (mESCs) or directly starting from 1,000 mESCs in replicates (Table S1). We found that nano-hmC-Seal enrichment profiles are similar to results obtained from the regular hmC-Seal profiling starting with 10 μg mESC DNA (Figures 1B and S1A–S1C). The 5hmC profiles were highly reproducible between replicates (R = 0.979 for 50 ng, 0.863 for 5 ng, and 0.821 for 1,000-cell samples) (Figure 1C). The pairwise correlation of 5hmC signals in 2,000-bp tiling regions across the genome is 0.98 (Pearson correlation coefficient) when comparing the two replicates using 50 ng DNA, while regular hmC-Seal libraries showed a correlation of 0.88 (Figure S1C). The 5hmC signals from 5 ng and 1,000-cell libraries correlated well with those obtained from the regular hmC-Seal libraries (Pearson's r ranging from 0.76 to 0.82) (Figure S1C), albeit with slightly lower levels of correlations likely due to reduced input materials used (Figures 1C and S1C). Furthermore, the fractions of reads in high density 5hmC clusters (see Experimental Procedures) were calculated as a measure of specific enrichment. All of the libraries generated by nano-hmC-Seal had similar 5hmC enrichment levels compared to the regular hmC-Seal libraries (Figure S1D). We next employed the PreSeq (Daley and Smith, 2013) package to extrapolate and estimate library complexity (Figure S1E). The results showed that libraries using 50 ng DNA have similar complexity as regular hmC-Seal libraries. Although libraries built from 5 ng or 1,000 cells displayed lower complexity, deeper sequencing of these libraries could still generate enough unique reads for downstream analysis. The

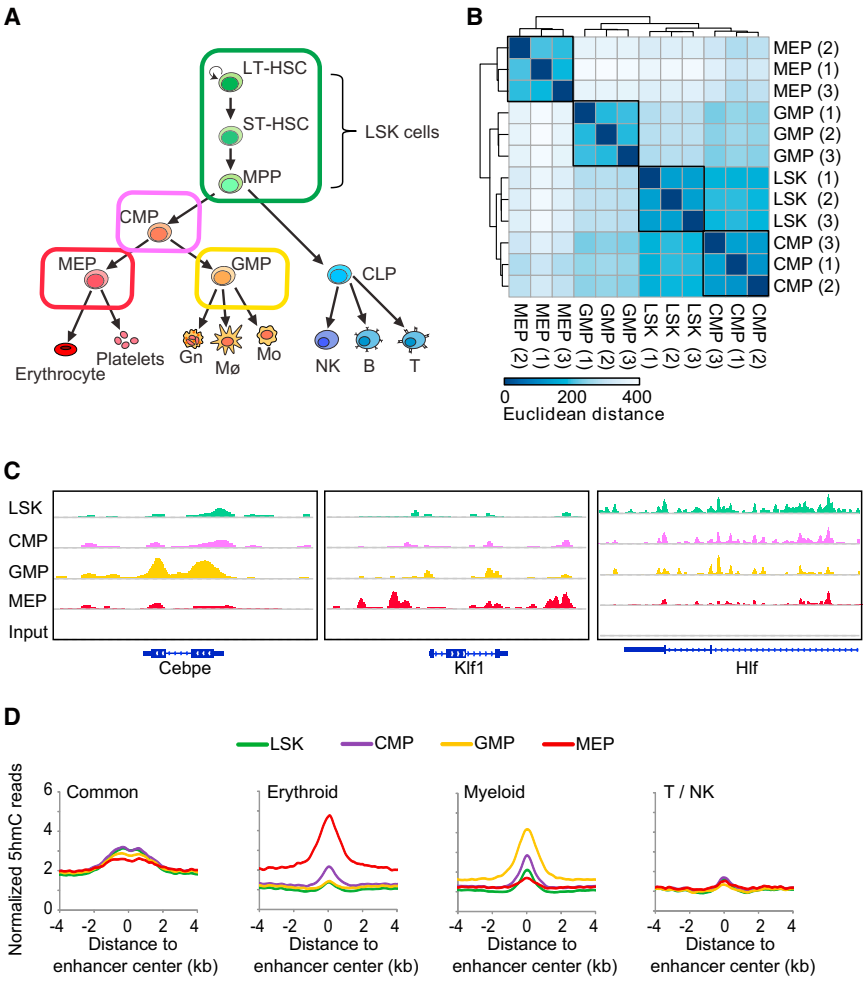


Figure 2. Nano-hmC-Seal Provides Dynamic 5hmC Profiles during Early Hematopoiesis

(A) Schematic of the hematopoietic differentiation stages. The cell types investigated in this study are outlined. The cell surface phenotypes were LSK (lin^- Sca^+ $cKit^+$), MPP (lin^- Sca^+ $cKit^+$ $CD48^+$ $CD150^-$), CMP (lin^- Sca^- $cKit^+$ $CD34^+$ $CD16/32^-$), MEP (lin^- Sca^- $cKit^+$ $CD34^-$ $CD16/32^-$), and GMP (lin^- Sca^- $cKit^+$ $CD34^+$ $CD16/32^+$).

(B) Relationship of 5hmC profiles in four different types of hematopoiesis stem and progenitor cells. Hierarchical clustering applied to the matrix of sample-to-sample distance based on rlog-transformed read counts in 304,069 detected 5hmC-enriched peaks is shown.

(C) Representative examples of 5hmC profile in several loci (from left to right: *Cebpe*, *Klf1*, and *Hlf*).

(D) Distribution of 5hmC signals at lineage-specific enhancers (from left to right: common enhancers, erythroid enhancer, myeloid enhancers, and T/NK cells enhancers). The lineage-specific enhancers were determined based on K-means analysis of H3K4me1 signals (see Figure S4). The genomic locations of enhancers were previously defined (Lara-Astiaso et al., 2014).

See also Figures S2–S4.

5hmC peaks detected from libraries using 50 ng DNA have similar number and quality compared to peaks obtained from regular hmC-Seal libraries; the quality starts to decrease with libraries employing 5 ng DNA or from 1,000 cells, but still yields a large number of 5hmC-enriched peaks for analysis (Figure S1F).

To compare our data sets with the "gold-standard" base-resolution 5hmC maps generated by using TAB-seq (Yu et al., 2012), 5hmC sites detected by TAB-seq were divided into three sets with low (10%–25%), medium (25%–40%), and high (>40%) 5hmC percentage. 5hmC signals obtained using nano-hmC-Seal around these sites positively correlate with the 5hmC abundance determined by TAB-seq (Figure S1G). In addition, a strong enrichment of nano-hmC-Seal signals was also observed at the 5hmC-containging CpG islands (CGIs) detected by oxRRBS (Booth et al., 2012) (Figure S1H). We therefore conclude that nano-hmC-Seal provides a reliable approach to the genome-wide detection of 5hmC using limited genomic DNA materials.

Nano-hmC-Seal Reveals Dynamic Hydroxymethylation Localization at Enhancer Sites during Early Hematopoietic Differentiation

Having validated nano-hmC-Seal with limited starting genomic DNA, we next applied this approach to the study of hematopoiesis, a process in which epigenetic regulation plays a central role (Rice et al., 2007; Shih et al., 2012). In order to profile the dynamics of 5hmC during hematopoietic cell differentiation (Figure 2A), we performed nano-hmC-Seal on cells isolated from mouse bone marrow by fluorescence-activated cell sorting (FACS). We analyzed flow sorted HSCs (LSK), common myeloid progenitors (CMP), granulocyte-macrophage progenitors (GMP), and megakaryocyte erythroid progenitors (MEP), with approximately 5,000 cells isolated from each purified population. Each cell type was sequenced in triplicate to monitor biological variability. These libraries yielded a total of 226.62 million reads (Table S1), of which, 92.5% could be aligned to the mouse reference genome (see Experimental Procedures). A total of 323,854 5hmC peaks in the genome were identified. The replicates displayed high similarity of 5hmC signal density in these peaks, thus demonstrating high reproducibility of our method (Figure 2B). Of note, the 5hmC pattern of LSK bears more resemblance to the 5hmC pattern of CMP, and the 5hmC pattern of GMP cells are closer to that of CMP than LSK, whereas the 5hmC pattern of MEP is distinct from the other three lineages (Figure 2B), indicating that the transition from CMP to MEP is associated with significant changes in 5hmC localization. Furthermore, at the differentially methylated regions (DMRs), which distinguish between GMP and MEPs obtained using RRBS (Bock et al., 2012), the 5hmC density showed an inverse comparative pattern in comparison with 5mC changes (Figures S2A–S2C).

Our results represent comprehensive genome-wide 5hmC maps obtained during in vivo differentiation of HSCs to

committed progenitors. We next examined differential 5hmC localizations at genes encoding master transcriptional factors known to be expressed or silenced during HSC differentiation (Lara-Astiaso et al., 2014; Moignard et al., 2013). For example, 5hmC was observed at the highest level across the gene body of the *Cebpe* gene in GMP; *Cebpe* encodes a bZIP transcription factor responsible for the lineage determination of GMP cells (Lara-Astiaso et al., 2014; Lekstrom-Himes and Xanthopoulos, 1999) (Figure 2C). In contrast, *Klf1*, a gene encoding a transcription factor critical for erythropoiesis, showed an increased 5hmC deposition at genic-proximal regions in MEP. Moreover, we observed a stepwise loss of 5hmC across the intragenic regions of *Hlf*, a gene essential for maintaining HSC function (Gazit et al., 2013).

To assess the regulatory role of 5hmC during differentiation, we compared 5hmC levels for each gene with previously reported chromatin immunoprecipitation sequencing (ChIP-seq) signals of histone modifications (Lara-Astiaso et al., 2014). The distributions of 5hmC and histone modifications were plotted ±2 kb around annotated genes and sorted based on their expression levels (Figure S2D). As expected, 5hmC is enriched in the gene-body of highly expressed genes. In all cell types, there is a positive correlation of gene-body 5hmC with histone modifications that mark gene activation, especially for H3K4me1 (Figures S2D and S2E). We next probed the relationship between 5hmC distribution and chromatin accessibility detected by assay for transposase-accessible chromatin (ATAC)-seq (Buenrostro et al., 2013; Lara-Astiaso et al., 2014). We divided ATAC-seq peaks into four groups based on signal intensity and found that 5hmC is depleted in the center of ATAC-seq peaks with high chromatin accessibility, but enriched in ATAC-seq peaks with lower signal intensity in all cell stages (Figure S3A). The distribution of H3K4me1 peaks showed similar patterns with depletion in the center of ATAC-seq peaks, but not for other active histone makers (Figure S3A). These observations are consistent with previous findings that 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), the further oxidized intermediates in active demethylation, mark open chromatin sites (Lu et al., 2015a), whereas 5hmC tends to mark less active or "poised" chromatin elements (Lu et al., 2015a; Pastor et al., 2011; Yu et al., 2012) and enriches around, but not on the transcription factor binding sites (Yu et al., 2012).

Next, we studied the potential association between changes in 5hmC sites and in histone modifications during differentiation (Figure S3B). We identified 21,791 regions that display differential 5hmC peaks across different differentiation stages. Clustering of these regions by their dynamic 5hmC profiles revealed four clusters. 5hmC sites in cluster I do not show noticeable trends of 5hmC changes, nor correlation patterns to reported histone modification differences. Gene set enrichment analysis with the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) showed that genes associated to these 5hmC sites had enriched for genes with a role in immune system development and related to abnormalities in hematopoietic cell number and immune cell physiology (Figure S3C). H3K4me1 and H3K4me2 signals show similar correlation patterns with respect to differential 5hmC in clusters II–IV (Figure S3B). However, cluster III comprises peaks with the highest 5hmC level in LSK cells, which gradually reduces during differentiation. GREAT annotation of these peaks showed association with genes highly enriched for leukocyte differentiation. Cluster II comprises sites with increased 5hmC signals during differentiation from CMP to MEP cells, and cluster IV contains sites with gradually increasing 5hmC signals upon differentiation from LSK to GMP cells, but with slightly reduced 5hmC signals in MEP cells. Cluster II sites are linked to genes essential for erythrocyte differentiation and cluster IV sites are related to genes whose knockout phenotypes exhibit defects in immune tolerance and acute inflammation (Figure S3C). Both clusters contain regions that are pre-marked with relatively low levels of 5hmC in the stem cell stage; however, the 5hmC level increases upon differentiation to GMP or MEP cells, which is accompanied by the appearances of H3K4me1 and H3K4me2 marks that were not observed at the stem cell stage. These observations may suggest that 5hmC precede H3K4me1/2 at these regions in the chromatin activation process during HSC differentiation. In general gene-body 5hmC changes correlate with histone mark changes.

Previous studies have shown that enhancer dynamics during hematopoiesis are critical for the differential access of transcriptional factors that drive lineage specification (Lara-Astiaso et al., 2014). Furthermore, studies in mESCs have suggested that the establishment and maintenance of enhancer regions primarily depend on TET-mediated active demethylation, and that the activity of these enhancers correlates with the abundance of 5hmC (Hon et al., 2014; Stadler et al., 2011). Analysis of genome-wide data obtained by nano-hmC-Seal allowed us to observe that 85.6% (41,450/48,415) of predicated enhancers (Lara-Astiaso et al., 2014), defined by enriched H3K4me1/2 histone modification signatures, are co-localized with 5hmC peaks in hematopoietic cells. We therefore hypothesized that dynamic changes in lineage specific enhancers during hematopoiesis (particularly those marked by H3K4me1; Figure S4A) are associated with corresponding changes in 5hmC. Indeed, 5hmC signatures at common enhancers are remarkably similar across all cell types, whereas 5hmC signatures were gradually established in the myeloid specific enhancers during myeloid transition from LSK to GMP cells, but not in MEP cells (Figure 2D). Moreover, 5hmC is highly enriched in erythroid specific enhancers in MEP cells, indicating a strong correlation of 5hmC distribution with lineage commitment. Further analyses verified that 5hmC signals detected in these cells are not enriched at enhancers specific for T or NK cells (Figure 2D).

Nano-hmC-Seal Analysis of Murine Leukemia Stem Cells with *Tet2* Loss and *Flt3^{ITD}* Mutation

After mapping 5hmC dynamics during normal hematopoiesis, we next chose to focus our attention on AML, a disease marked by recurrent mutations in epigenetic regulators. We therefore applied nano-hmC-Seal to dissect how *Tet2* loss, combined with other known AML disease alleles, could potentially dysregulate 5hmC and contribute to leukemogenesis. To this end, we generated global 5hmC maps of MPP and GMP isolated from wild-type mice and T2F3 mice (a murine AML model harboring *Tet2* and *Flt3^{ITD}* mutations; Shih et al., 2015). Of note, previous studies have shown that MPP, but not GMP, from *Tet2* and

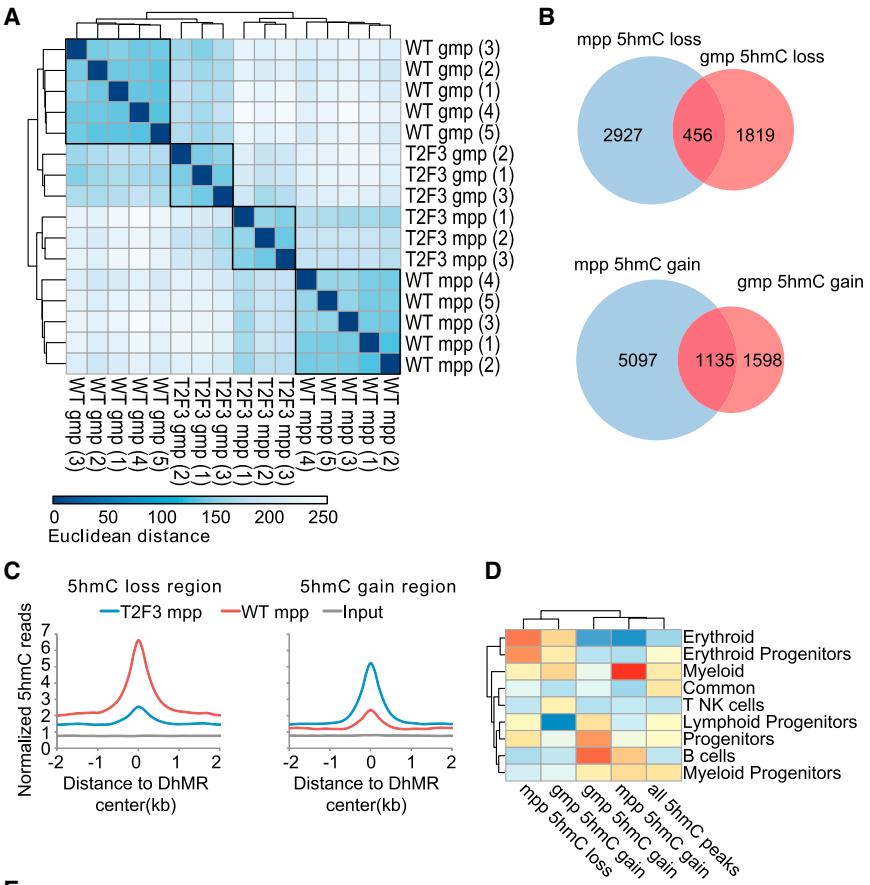


Figure 3. Nano-hmC-Seal Reveals 5hmC Redistribution in a Murine AML Model

(A) Comparison of 5hmC profiles in bone marrow MPP and GMP cells from WT and *Tet2*$^{-/-}$;*Flt3*^{ITD} mice. Hierarchical clustering applied to the matrix of sample-to-sample distance based on rlog-transformed read counts in 272,087 detected 5hmC-enriched peaks is shown.

(B) Venn diagram showing the overlap of detected DhMRs between MPP and GMP cells.

(C) Distribution of 5hmC signals at DhMRs (left: 5hmC loss and right: 5hmC gain) in MPP cells.

(D) Relative enrichment of the genomic overlap between DhMRs and lineage-specific enhancers. Only DhMRs with at least 2-fold 5hmC change, adjust $p < 0.1$, and mean of normalized counts >20 were used for analysis. All 5hmC peaks were used as control set.

(E) De novo motif analysis by HOMER at DhMRs (left: 5hmC loss and right: 5hmC gain) in MPP cells.

See also Figure S4.

Flt3^{ITD} mice have leukemia stem cell (LSC) potential with the ability to serially transplant in vivo (Shih et al., 2015). Unsupervised hierarchical clustering of a 5hmC-enriched region demonstrated clear separation of leukemic T2F3 samples from wild-type (WT) samples in both cell types (Figure 3A). To pinpoint specific loci that display differential 5hmC profiles between leukemic samples and WT samples, we proceeded to identify and characterize differentially hydroxymethylated regions (DhMRs). A total of 9,204 DhMRs were found in MPP and 5,008 DhMRs in GMP (Figures 3B and 3C).

We compared DhMRs in MPP with those identified in GMP and found that the majority (>80%) of these DhMRs are not shared between the two cell types, suggesting distinct 5hmC deposition and maintenance in these two cell types isolated from this AML model. Regions with 5hmC-loss and 5hmC-gain in MPP and GMP were examined for enrichment in lineage specific enhancers. Regions with increased 5hmC in MPP were found preferentially overlapping with the myeloid specific enhancers, while losses of 5hmC in MPP were enriched for the erythroid-related enhancers (Figure 3D). These results suggest that the alterations of 5hmC deposition may favor an aberrant differentiation toward myeloid cells at the expense of erythroid cells, a conclusion consistent with previous observations of increased GMP and reduced MEPs in T2F3 mice (Shih et al., 2015). To better understand the regulatory sequence codes that are associated with 5hmC changes, we performed de novo motif analysis within these DhMR regions. The most enriched recognition motifs in loci with reduced 5hmC matched the binding sites of known regulators implicated in AML pathogenesis, including GATA and MYB (Figure 3E). In contrast, we detected a known PU.1 motif (AGAGGAAGTG, $p = 1 \times 10^{-106}$) in the regions that gained 5hmC (Figure S4B), supporting the notion that reciprocal antagonism between PU.1 and the GATA family is critical for the myeloid lineage commitment in normal and malignant hematopoiesis (Walsh et al., 2002). Interestingly, we detected a PU.1:IRF composite motif in the regions that gained 5hmC (Figure 3E).

We also observed an accumulation of 5hmC at the edge of DNA methylation canyons, which is a recently reported epigenetic feature of AML driven by mutations in epigenetic regulators (Jeong et al., 2014). Although no significant variations of 5hmC were noticed within the canyon, a decrease of 5hmC outside the canyon boundary region was observed in AML samples compared to the WT (Figure S4C), suggesting that the disruption of 5hmC distribution in the leukemia model may affect the maintenance of methylation canyons.

To further confirm the functional influence of the observed 5hmC alternation, we performed RNA-seq of mRNA obtained from the same cells. As expected, the alternation of 5hmC positively correlated with gene expression (Figure 4A). In MPP, we

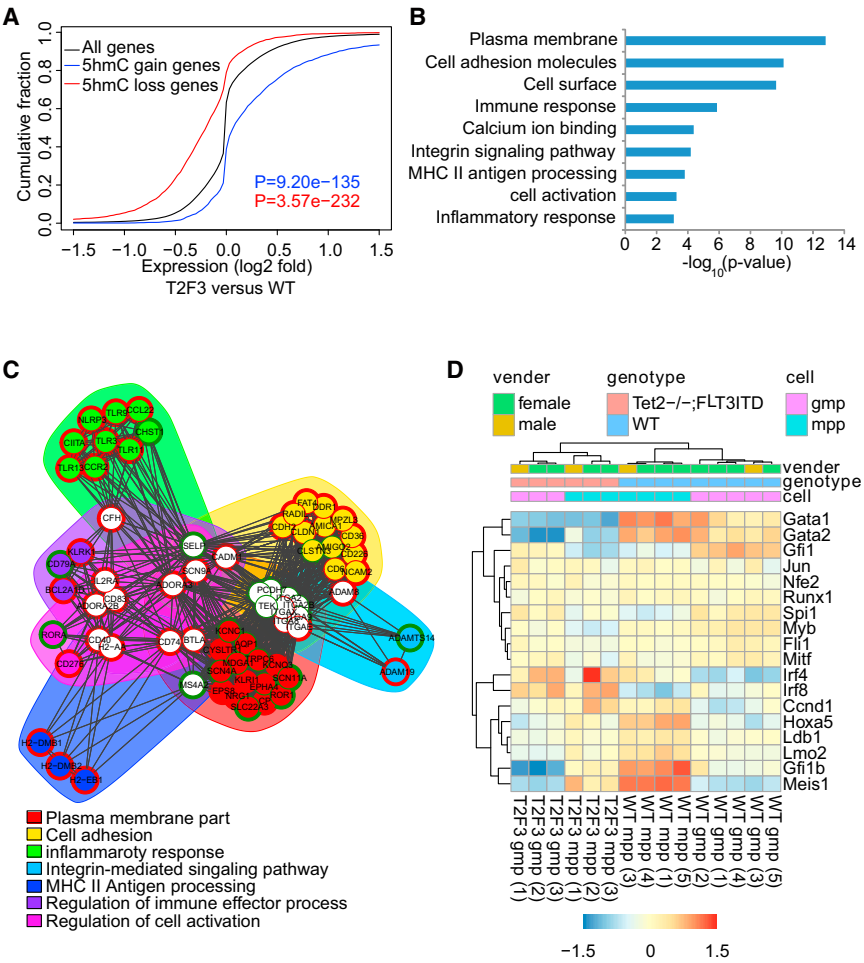


Figure 4. Alternations of 5hmC in Gene Body Correlate with Gene Expression Changes in AML Model

(A) Cumulative distribution of 5hmC gain genes (blue) and 5hmC loss genes (red) correlate with expression changes in MPP cells from *Tet2*$^{-/-}$;*Flt3*^{ITD} (T2F3) mice versus WT mice. All genes were used as control. p values were calculated versus all genes (two-sided Wilcoxon rank-sum test).

(B) GO enrichment analysis of 366 differentially expressed genes (|fold change| >2 and adjust $p < 0.1$) associated with significant 5hmC changes (adjust $p < 0.1$) in MPP cells.

(C) Functional gene networks derived from GO enrichment analysis. The border color of the nodes denotes the upregulation (red) or downregulation (green) of genes in MPP cells from *Tet2*$^{-/-}$;*Flt3*^{ITD} versus WT mice. The genes in the same terms were linked to each other. The genes within the same clusters are surrounded by a common background color.

(D) Heatmap of RNA expression to compare gene expression of hematopoietic transcriptional regulators in different samples. The genes and samples were clustered by Euclidean distance using centered rlog-transformed expression counts.

identified 366 differentially expressed genes that were associated with significant 5hmC changes. We performed gene ontology analysis (GO) and found that these genes are enriched in functions such as cell adhesion, inflammatory response, and regulation of immune effector process (Figures 4B and 4C). Consistent with the motif analysis, we also confirmed the downregulation of *Gata1*/*Gata2* and upregulation of *Irf8* (Figure 4D). Together, these results highlight the utility of nano-hmC-Seal as a sensitive tool to map 5hmC distribution and dynamic methylation/demethylation in rare cell populations.

DISCUSSION

Several methods have been developed to map 5hmC genome wide. Bisulfite-based approaches such as TAB-seq (Yu et al., 2012) and oxBS-seq (Booth et al., 2012) provide the most comprehensive and quantitative information because they are able to detect 5hmC at single-base resolution. However, these methods can be prohibitively expensive due to the requirement for high sequencing coverage. The oxidation and bisulfite treatments also lead to substantial degradation of genomic DNA, which limits their utility to investigate systems with limited input materials. By contrast, enrichment-based profiling methods require lower sequencing costs and can be routinely applied to 5hmC studies. However, antibody-based enrichment still requires microgram levels of input genomic DNA and can introduce sequence biases. In addition, it cannot be used to profile biological processes or clinical samples where target cells or input DNA exist in rare quantities and often are difficult-to-obtain. We show here that nano-hmC-Seal, an approach based on selective chemical labeling and capture, provides a highly sensitive and robust method to profile 5hmC with a few nanograms of genomic DNA (~1,000 cells). We have demonstrated the utility of this method in the study of HSC differentiation from mouse models. Importantly, we show that the covalent labeling and pull-down technology could be widely applied to profile 5hmC in sorted homogenous neurons and stem cells, clinical biopsy samples, circulating tumor cells, and cell-free DNA. We anticipate expansion of this approach into the study of dynamics of 5mC with limited input DNA materials when combined with Tet-mediated covalent labeling of 5mC (TAmC-seq) (Zhang et al., 2013).

Because 5hmC serves as a potential indicator of active DNA demethylation, a crucial mechanism commonly associated with transcriptional activation, the distribution patterns of 5hmC not only provide a global view of gene activation, but also uncover insight into the use of lineage/disease-specific enhancers and promoters. As such, differential 5hmC patterns reflect the underlying gene expression patterns and the binding activities of transcription factors at enhancer regions, especially for pioneer factors. Our studies using nano-hmC-Seal in normal

HSC and progenitor cells and in AML LSCs allowed us to define multiple specific loci with dynamic changes in 5hmC patterns that are enriched in enhancers and gene bodies with a known or putative role in normal and malignant hematopoiesis and which strongly correlate with differential gene expression. Therefore, integrated states of 5hmC at enhancer and gene-body regions in patients can potentially unravel the molecular mechanisms that drive transformation from normal stem/progenitor cells to different malignancies.

Our results reveal new insights of 5hmC dynamics during HSC differentiation: (1) gene-associated 5hmC is positively correlated with active histone modifications, in particular with H3K4me1/2 changes. The genes proximal to these regions are annotated to blood cell functions including leukocyte/erythrocyte differentiation and immunological disorders and (2) 5hmC peaks tend to locate in genomic regions that are marked by less strong ATAC-seq signal intensity and are depleted in the center of ATAC-seq peaks that may mark high chromatin accessibility. We identified regions with gain and loss of the Tet2-deficient LSCs, indicating that the inactive mutation of Tet2 does not simply lead to loss of 5hmC, but instead drives a redistribution of this modification (Madzo et al., 2014). In addition, we identified PU.1 motifs, a key factor required for the generation of myeloid and lymphoid cells (Walsh et al., 2002), as significantly enriched at sites with increased 5hmC. PU.1 is known to drive different normal and pathogenic hematopoietic functions by co-binding with various transcription factors including c-Jun, C/EBPA, GATA1/2, and IRF4/8 (Gupta et al., 2009). The identification of a significant PU.1:IRF ($p = 1 \times 10^{-321}$) motif discovered by de novo motif analysis in these loci implies a specific PU.1 regulatory mechanism associated with active demethylation and 5hmC distribution. It will be of interest to investigate whether the PU.1/IRF complex has a functional role in leukemogenesis. These observations also open an interesting question about how the binding of PU.1, which was reported to physically interact with TET2 (de la Rica et al., 2013), contributes to 5hmC redistribution induced by inactivation of TET2. One possible explanation is that PU.1 might also interact with other Tet proteins, thus leading to a unique genomic binding pattern in the absence of TET2.

Our 5hmC profiling further showed that regions with loss of 5hmC in AML stem cells exhibit significant enrichment of binding motifs for GATA factors, which are known to be mutated and altered by translocations and differential expression in AML (Gröschel et al., 2014). This result underscores our previous observation that LSCs have reduced expression of *Gata2* in the same AML mice models, and that restoration of *Gata2* expression abrogates leukemogenesis in *Tet2*/*Flt3*-mutant AML (Shih et al., 2015). Together, these findings suggest a multi-layered epigenetic regulation of GATA factors by repressing transcription initiation and silencing of downstream binding sites. These studies demonstrate the utility of nano-hmC-Seal or similar approaches based on covalent 5hmC labeling to probe differential 5hmC in normal and pathologic disease states with limited samples and provide a robust platform to assess 5hmC localizations in tumor models and clinical samples in order to identify target loci and biomarkers with biologic and therapeutic relevance.

Limitations

This manuscript presents a robust and highly sensitive approach (nano-hmC-Seal) for genome-wide profiling of 5hmC using rare cell populations and limited input DNA. This sequencing method is based on affinity purification that precludes locating 5hmC at base resolution and measuring its absolute abundance at the modification site; however, base-resolution methods could not be used to study rare cell populations in their current forms. In addition, the library complexity for nano-hmC-Seal reduces when the DNA starting material is very limited (i.e., at picogram level), which may affect the sensitivity to detect weak 5hmC signals. To address this limitation, we expect that our method could be further adapted to employ an indexing-first strategy, which would enable 5hmC profiling analysis from a lower number of cells by allowing multiple barcoded samples to be pooled for subsequent enrichment, purification, and library preparation steps.

EXPERIMENTAL PROCEDURES

Nano-hmC-Seal Protocol

Genomic DNA is extracted from cells using Quick-gDNA MicroPrep (Zymo) according to the manufacture's instruction. Tagmentation reactions were performed in a 50 μl solution with 1× tagmentation buffer from Nextera DNA Sample Preparation Kit (FC-121-1031). The input DNA ranged from 5 ng to 50 ng. Reactions were performed and purified according to the manufacturer's instruction. In brief, the fragmentation reaction was performed in 50 μl solutions containing 25 μl 2× tagmentation buffer, gDNA, and 0.5-5 ×I Tagmentase, at 55°C for 5 min. The fragmented DNA was eluted into 17.5 μl ddH_2O. The glucosylation reactions were performed in a 20 μl solution containing 50 mM HEPES buffer (pH 8.0), 25 mM MgCl2, purified DNA, 100 μM N3-UDP-Glc, and 1 μM βGT, at 37°C for 1 hr. After the reaction, 2 μl DBCO-PEG4-DBCO (Click Chemistry Tools, 20 mM stock in DMSO) was added to the reaction mixture, and the mixture was incubated at 37°C for 2 hr. Next, the modified DNA was purified by a Micro Bio-Spin 30 Column (Bio-Rad). The purified DNA was incubated with 5 μl C1 Streptavidin beads (Life Technologies) in 2× buffer (1× buffer: 5 mM Tris [pH 7.5], 0.5 mM EDTA, and 1 M NaCl) for 15 min according to the manufacture's instruction. The beads were subsequently undergone six 5 min washes with 1× buffer. All binding and washing were done at room temperature with gentle rotation. The captured DNA fragments were amplified with 12–17 cycles of PCR amplification using the enzyme mix supplied in the Nextera kit. The PCR products were purified using AMPure XP beads according to the manufacture's instruction. Separate input libraries were made by direct PCR from fragmented DNA without labeling and capture according to the manufacture's instruction in the Nextera kit. DNA concentration of each library was measured with a Qubit fluorometer (Life Technologies). Sequencing was performed on the HiSeq instrument.

Cell Culture

mESCs were cultured in feeder-free gelatin-coated plates in Dulbecco's modified Eagle medium (DMEM) (Invitrogen Cat. No. 11995) supplemented with 15% fetal bovine serum (FBS) (Gibco), 2 mM L-glutamine (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 1× nonessential amino acids (Gibco), 1,000 units/ml LIF (Millipore Cat. No. ESG1107), 1× pen/strep (Gibco), 3 mM CHIR99021 (Stemgent), and 1 mM PD0325901 (Stemgent). Half of the medium was changed every day for 5 days before harvesting mEBs by sedimentation.

Isolation of Hematopoietic Progenitor Cells

Mouse strains, antibody staining, and FACS protocol were as previously described (Shih et al., 2015). Briefly, leg and arm bones were isolated from mice and bone marrow was isolated by centrifugation at 6,000 rcf for 1 min. Cells were then lysed in ACK lysis buffer. Stem cell enrichment was performed using the Progenitor Cell Enrichment Kit (STEMCELL Technologies).

Antibodies used for flow cytometry were as follows: (anti-mouse) Gr1 (Ly6G), B220 (RA3-682), CD34 (RAM34), CD16/32 (93) Sca-1 (D7), cKit (2B8), Mac-1 (CD11b) (M1/70), NK1.1 (PK136), Ter119 (Ter119,553673), CD3 (145-2C11), CD150 (TC-15-12F2.2), and CD48 (HM48-1) from BioLegend and eBioscience. The "lineage cocktail" included CD3, Gr-1, Mac-1 (CD11b), NK1.1, B220, and Terr-119. FACS was performed on a BD FACSAria sorter. DNA and RNA were isolated using the AllPrep DNA/RNA Mini Kit from QIAGEN.

All animal procedures were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at Memorial Sloan Kettering Cancer Center.

RNA-Seq and Analysis

RNA-seq libraries were constructed by TrueSeq Stranded mRNA Sample Preparation Kit (Illumina). The reaction was performed according to the manufacture's instruction.

Sequencing reads were aligned to mm9 genome by STAR (Dobin et al., 2013). Count matrices were generated by summarizeOverlaps functionality from R package GenomicAlignments (Lawrence et al., 2013). Differential expression analysis was performed by R package DESeq2 (Love et al., 2014).

Data Processing and Analysis

Illumina reads were post-processed and aligned to the mouse mm9 assembly using the bowtie program with default parameters. To visualize sequencing signals in the genome browser, we generated bedGraph files with HOMER (Heinz et al., 2010). deepTools (Ramírez et al., 2014) software was used to plot the heatmap of signal distribution, perform K-means clustering, and calculate genome-wide correlations (2,000-bp tiling regions). The identification of 5hmC-enriched regions (peaks) in each sample was performed using MACS (Feng et al., 2012). Peaks that were detected by all replicates were considered as high confident peaks. Peaks from all samples were combined into one unified catalog for each study separately by "mergePeak" functionality from HOMER (Heinz et al., 2010). Potential library complexity was determined by using the PreSeq (Daley and Smith, 2013) package with extrapolate function. To show the correlation between Nano-hmC-Seal and standard hmC-Seal in mESC (or between replicates), tag counts in merged high-confident peaks were calculated by HOMER. The $\log_2$ transform values were then used to generate scatterplots and calculate the Pearson correlation. The data set of standard hmC-Seal using bulk mESCs was obtained from a previously published study (Song et al., 2013). The fractions of reads in high density 5hmC clusters were calculated as a measure of specific enrichment. High density 5hmC cluster were regions that contained at least five TAB-Seq detected 5hmCs, each within 200 bp of each other. To detect 5hmC-containing CGIs from an ox-RRBS data set, R package methylKit (Akalin et al., 2012) was used for analysis.

To assess overall difference between samples, we calculated the Euclidean distance based on rlog-transformed 5hmC signals and visualized the distance in a heatmap figure by using an R package "pheatmap". To detect the genotype-specific DhMR or genes, the Bioconductor DESeq2 packages was used for analysis. Functional annotation of DhMRs was obtained with GREAT (McLean et al., 2010). To detect DMR from an RRBS data set, R package methylKit (Akalin et al., 2012) was used for analysis. The HOMER software was used to perform de novo motif analysis at ±500bp around DhMR. GO term analyses were performed by DAVID (Huang et al., 2009) and visualized as functional gene network by FGNet (Aibar et al., 2015).

Definition of Lineage Specific Enhancer Subgroups

The genomic locations of 48,415 enhancers identified across 16 stages of hematopoietic differentiation were previously defined (Lara-Astiaso et al., 2014). To be consistent with previous studies, these enhancers were further grouped into nine major clusters based on log-transformed H3K4me1 level using K-means cluster, as described before (Lara-Astiaso et al., 2014). R package pheatmap were used to generate Figure 4D.

External Data

TAB-Seq data sets were derived from Yu et al. (2012). The mESC ox-RRBS data sets were derived from Booth et al. (2012). RRBS data sets for MEP and GMP cells were derived from Bock et al. (2012). ATAC-seq and histone modification ChIP-seq of H3K4me1/2/3 and H3K27ac data sets were derived from Lara-Astiaso et al. (2014).

ACCESSION NUMBERS

The accession number for the sequencing data reported in this article is NCBI GEO: GSE77967.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.06.028.

AUTHOR CONTRIBUTIONS

X.L., D.H., and C.H. conceived the idea. C.H., D.H., X.L., A.M.M., and R.L.L. designed the experiments; X.L., A.H.S., J.N., D.H., and Q.Y. performed experiments; D.H. performed bioinformatics analysis; R.L.L. and C.H. supervised the project; D.H. and C.H. wrote the paper with inputs from X.L., A.H.S., M.M.X., A.M.M., and R.L.L. All authors discussed the results and commented on the manuscript.

ACKNOWLEDGMENTS

This work was supported by the NIH (R01HG006827 to C.H., U54CA193419 to C.H. and R.L.L., and 5R01CA173636-03 to A.M.M. and R.L.L.), the Howard Hughes Medical Institute (to C.H.), the Gabrielle's Angel's Foundation (to A.M.M., C.H., and R.L.L.), a Starr Cancer Consortium grant (to R.L.L. and A.M.M.), and a Leukemia Lymphoma Society (LLS) SCOR grant (7006-13 to A.M.M.). A.H.S. is a Special Fellow of the LLS, and R.L.L. is a Scholar of the LLS. We thank S.F. Reichard (University of Chicago) for editing the manuscript. We thank the University of Chicago Genomic Facility for providing technical support. The University of Chicago has filed for patent protection on the original hmC-Seal technology. C.H. was one of the inventors. X.L. is a shareholder of a company that has licensed the patent.

Received: March 14, 2016
Revised: May 24, 2016
Accepted: June 20, 2016
Published: July 28, 2016

REFERENCES

Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., and Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. *11*, R119.

Aibar, S., Fontanillo, C., Droste, C., and De Las Rivas, J. (2015). Functional gene networks: R/Bioc package to generate and analyse gene networks derived from functional enrichment and clustering. Bioinformatics *31*, 1686–1688.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. *13*, R87.

Bock, C., Beerman, I., Lien, W.H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol. Cell *47*, 633–647.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science *336*, 934–937.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cui, L., Chung, T.H., Tan, D., Sun, X., and Jia, X.Y. (2014). JBP1-seq: a fast and efficient method for genome-wide profiling of 5hmC. Genomics *104*, 368–375.

Daley, T., and Smith, A.D. (2013). Predicting the molecular complexity of sequencing libraries. Nat. Methods *10*, 325–327.

de la Rica, L., Rodríguez-Ubreva, J., García, M., Islam, A.B., Urquiza, J.M., Hernando, H., Christensen, J., Helin, K., Gómez-Vaquero, C., and Ballestar, E. (2013). PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol. *14*, R99.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740.

Gazit, R., Garrison, B.S., Rao, T.N., Shay, T., Costello, J., Ericson, J., Kim, F., Collins, J.J., Regev, A., Wagers, A.J., and Rossi, D.J.; Immunological Genome Project Consortium (2013). Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. Stem Cell Reports *1*, 266–280.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A., Erpelinck, C., van der Velden, V.H., Havermans, M., Avellino, R., van Lom, K., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell *157*, 369–381.

Gupta, P., Gurudutta, G.U., Saluja, D., and Tripathi, R.P. (2009). PU.1 and partners: regulation of haematopoietic stem cell fate in normal and malignant haematopoiesis. J. Cell. Mol. Med. *13*, 4349–4363.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448–452.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hon, G.C., Song, C.X., Du, T., Jin, F., Selvaraj, S., Lee, A.Y., Yen, C.A., Ye, Z., Mao, S.Q., Wang, B.A., et al. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol. Cell *56*, 286–297.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G.A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R., et al. (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. Nat. Genet. *46*, 17–23.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943–949.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118.

Lekstrom-Himes, J., and Xanthopoulos, K.G. (1999). CCAAT/enhancer binding protein epsilon is critical for effective neutrophil-mediated response to inflammatory challenge. Blood *93*, 3096–3105.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, X., Han, D., Zhao, B.S., Song, C.X., Zhang, L.S., Doré, L.C., and He, C. (2015a). Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. Cell Res. *25*, 386–389.

Lu, X., Zhao, B.S., and He, C. (2015b). TET family proteins: oxidation activity, interacting molecules, and functions in diseases. Chem. Rev. *115*, 2225–2239.

Madzo, J., Liu, H., Rodriguez, A., Vasanthakumar, A., Sundaravel, S., Caces, D.B., Looney, T.J., Zhang, L., Lepore, J.B., Macrae, T., et al. (2014). Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis. Cell Rep. *6*, 231–244.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Moignard, V., Macaulay, I.C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F.J., Kinston, S., Joshi, A., Hannah, R., Theis, F.J., et al. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat. Cell Biol. *15*, 363–372.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394–397.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Rice, K.L., Hormaeche, I., and Licht, J.D. (2007). Epigenetic regulation of normal and malignant hematopoiesis. Oncogene *26*, 6697–6714.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods *12*, 963–965.

Shih, A.H., Abdel-Wahab, O., Patel, J.P., and Levine, R.L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. Nat. Rev. Cancer *12*, 599–612.

Shih, A.H., Jiang, Y., Meydan, C., Shank, K., Pandey, S., Barreyro, L., Antony-Debre, I., Viale, A., Socci, N., Sun, Y., et al. (2015). Mutational cooperativity linked to combinatorial epigenetic gain of function in acute myeloid leukemia. Cancer Cell *27*, 502–515.

Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. *29*, 68–72.

Song, C.X., Yi, C., and He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. Nat. Biotechnol. *30*, 1107–1116.

Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678–691.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Sun, W., Zang, L., Shu, Q., and Li, X. (2014). From development to diseases: the role of 5hmC in brain. Genomics *104*, 347–351.

Sun, Z., Dai, N., Borgaro, J.G., Quimby, A., Sun, D., Corrêa, I.R., Jr., Zheng, Y., Zhu, Z., and Guan, S. (2015). A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. Mol. Cell *57*, 750–761.

Topalian, S.L., Hodi, F.S., Brahmer, J.R., Gettinger, S.N., Smith, D.C., McDermott, D.F., Powderly, J.D., Carvajal, R.D., Sosman, J.A., Atkins, M.B., et al. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N. Engl. J. Med. *366*, 2443–2454.

Walsh, J.C., DeKoter, R.P., Lee, H.J., Smith, E.D., Lancki, D.W., Gurish, M.F., Friend, D.S., Stevens, R.L., Anastasi, J., and Singh, H. (2002). Cooperative and antagonistic interplay between PU.1 and GATA-2 in the specification of myeloid cell fates. Immunity *17*, 665–676.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368–1380.

Zhang, L., Szulwach, K.E., Hon, G.C., Song, C.X., Park, B., Yu, M., Lu, X., Dai, Q., Wang, X., Street, C.R., et al. (2013). Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. Nat. Commun. *4*, 1517.