Published in final edited form as:
Cell Rep. 2014 January 16; 6(1): 231–244. doi:10.1016/j.celrep.2013.11.044.

Hydroxymethylation at Gene Regulatory Regions Directs Stem/Early Progenitor Cell Commitment during Erythropoiesis

Jozef Madzo[#1], **Hui Liu**[#1], **Alexis Rodriguez**[#2], **Aparna Vasanthakumar**[1], **Sriram Sundaravel**[1], **Donne Bennett D. Caces**[1], **Timothy J. Looney**[3,4], **Li Zhang**[3,4], **Janet B. Lepore**[1], **Trisha Macrae**[1], **Robert Duszynski**[1], **Alan H. Shih**[8], **Chun-Xiao Song**[5,6], **Miao Yu**[5,6], **Yiting Yu**[7], **Robert Grossman**[2], **Brigitte Raumann**[2], **Amit Verma**[7], **Chuan He**[5,6], **Ross L. Levine**[8], **Don Lavelle**[9,10], **Bruce T. Lahn**[3,4], **Amittha Wickrema**[1,*], and **Lucy A. Godley**[1,*]

[1]Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, IL 60637, USA

[2]Center for Research Informatics, The University of Chicago, Chicago, IL 60637, USA

[3]Department of Human Genetics, The University of Chicago, Chicago, IL 60637, USA

[4]Howard Hughes Medical Institute, Chevy Chase, MD 20815-6789, USA

[5]Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA

[6]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA

[7]Albert Einstein College of Medicine, Bronx, NY 10461, USA

[8]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

[9]Department of Medicine, University of Illinois, Chicago, Chicago, IL 60612, USA

[10]Jesse Brown VA Medical Center, Chicago, IL 60612, USA

[#] These authors contributed equally to this work.

SUMMARY

Hematopoietic stem cell differentiation involves the silencing of self-renewal genes and induction of a specific transcriptional program. Identification of multiple covalent cytosine modifications raises the question of how these derivatized bases influence stem cell commitment. Using a replicative primary human hematopoietic stem/progenitor cell differentiation system, we demonstrate dynamic changes of 5-hydroxymethylcytosine (5-hmC) during stem cell commitment and differentiation to the erythroid line-age. Genomic loci that maintain or gain 5-hmC density throughout erythroid differentiation contain binding sites for erythroid transcription factors and

[*]Correspondence: awickrem@medicine.bsd.uchicago.edu (A.W.), lgodley@medicine.bsd.uchicago.edu (L.A.G.) http://dx.doi.org/10.1016/j.celrep.2013.11.044.

ACCESSION NUMBERS

The data corresponding to the input, 5-hmC-enriched, and RNA-seq tracks were deposited into the NCBI Gene Expression Omnibus under accession number GSE40243.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.11.044.

several factors not previously recognized as erythroid-specific factors. The functional importance of 5-hmC was demonstrated by impaired erythroid differentiation, with augmentation of myeloid potential, and disrupted 5-hmC patterning in leukemia patient-derived CD34+ stem/early progenitor cells with TET methylcytosine dioxygenase 2 (*TET2*) mutations. Thus, chemical conjugation and affinity purification of 5-hmC-enriched sequences followed by sequencing serve as resources for deciphering functional implications for gene expression during stem cell commitment and differentiation along a particular lineage.

INTRODUCTION

The propensity to generate multilineage hematopoietic cells from nascent uncommitted blood stem cells defines the complexity of the bone marrow system and serves as a paradigm for stem cell differentiation. Accumulated evidence has demonstrated the ability of a single hematopoietic cell to lose multilineage potential and commit to a specific blood cell type, a complex process involving extrinsic and intrinsic signals heavily influenced by the stem cell microenvironment (Ogawa, 1993; Smith et al., 1991; Spradling et al., 2001). Lineage commitment by stem cells is characterized by initiating a specific transcriptional program while simultaneously silencing large numbers of genes that maintain the self-renewal characteristics of the stem cell compartment.

Although accumulated data underscore the importance of extrinsic factors in lineage commitment, very little is known about the epigenetic changes that accompany lineage commitment and differentiation. Ten-Eleven-Translocation (TET) family members are dioxygenases that catalyze the conversion of 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC) and other covalent cytosine modifications, which have the potential to provide additional complexity to overall gene regulation (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Controversy exists as to whether Tet1 is required for pluripotency, with some studies showing that *Tet1* knockdown leads to spontaneous differentiation of mouse embryonic stem cells (ESCs) (Ito et al., 2010) and others failing to demonstrate compromised self-renewal capacity (Dawlaty et al., 2011; Koh et al., 2011; Williams et al., 2011). Genome-wide mapping of 5-hmC in mouse ESCs demonstrated an association with active chromatin marks as well as the enrichment of 5-hmC at transcriptional start sites and within enhancer regions, suggesting that 5-hmC plays a role in transcriptional regulation (Pastor et al., 2011; Stroud et al., 2011; Williams et al., 2011; Wu et al., 2011; Wu and Zhang, 2011b). Many studies have demonstrated a role for 5-hmC and the Tet enzymes in DNA methylation reprogramming in the mammalian zygote (Ficz et al., 2011; Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). However, whether 5-hmC functions as an intermediate in active or passive demethylation pathways, confers its own epigenetic function, or both, is not defined. Furthermore, many studies have examined 5-mC versus 5-hmC in cells in a single differentiation state and, therefore, have not been able to test how 5-hmC distribution changes during the course of differentiation. Until recently, most studies employed approaches that cannot distinguish 5-mC from 5-hmC, making it difficult to map dynamic changes precisely in the epigenetic landscape during stem cell commitment to a particular lineage. The current study provides comprehensive analysis of 5-hmC changes in a dynamic fashion during human stem/early progenitor cell commitment to the erythroid lineage and during subsequent differentiation, providing a valuable resource for understanding the relationship between epigenetic modifications and transcription factor (TF) binding as well as the generation of molecular hypotheses regarding stem cell commitment.

RESULTS

Global Levels of 5-hmC Change Dramatically during Erythropoiesis

To decipher the precise role(s) of 5-hmC in stem cell commitment, we chose a well-defined erythroid commitment and differentiation model (Kang et al., 2008; Tamez et al., 2009; Uddin et al., 2004), in which primary human hematopoietic stem/early progenitor cells differentiate during 17 days of in vitro culture in a replicative, synchronous, and orderly progression through all of the known erythroid intermediates (Figures 1A, S1A, and S1B). The day 0 starting cell population was highly enriched for stem/early progenitor cells (74.8% ± 6.8% CD34+/CD90+) and was devoid of cells expressing myeloid or lymphoid markers (Figure S1A; Table S1). Our culture conditions were permissive for erythroid-lineage commitment by day 3 as confirmed by expression of *EPOR*, *GATA1*, and *HBB*, all highly representative genes for erythroid cells (Figure S1C). These features were critical to our ability to utilize this model to assess the dynamic changes in 5-hmC marks and gene expression that accompany hematopoietic stem cell commitment to definitive erythropoiesis.

First, we used mass spectrometry to determine the global levels of 5-hmC (Figures 1B and S1D) and 5-mC (Figure 1C) at differentiation stages that correspond to defined erythroid intermediates, occurring at days 0, 3, 7, 10, 13, and 17 (stem/early progenitors, BFU-E, basophilic, polychromatic, orthochromatic, and reticulocytes, respectively). Total 5-hmC levels increased during stem/early progenitor cell commitment to the erythroid lineage, followed by a dramatic decrease throughout subsequent differentiation (Figures 1B and S1A–S1D). The effect was observed consistently in four independent biological replicates. In contrast, global 5-mC content decreased modestly throughout stem/early progenitor cell commitment and subsequent differentiation (Figure 1C), consistent with previous observations within murine erythropoiesis (Bock et al., 2012; Shearstone et al., 2011). We further confirmed these findings in populations of erythroid progenitors isolated directly from bone marrows of baboons (Figures 1D and 1E). *TET2* mRNA levels, measured by real-time PCR and confirmed by RNA sequencing (RNA-seq), were greatest in the day 0 CD34+ stem/early progenitors and exhibited a dramatic decrease thereafter, although maintained detectable levels until day 7 (Figure 1F). Expression levels of *TET1* and *TET3* were negligible at every time point (Figure 1F). The levels of TET2 protein were highest on day 3, the time period when CD34+ cells commitment to the erythroid line-age occurs (Figure 1G).

Locus-Specific Distribution of 5-hmC Undergoes Dynamic Changes during Erythropoiesis

Using chemical conjugation and affinity purification of 5-hmC-enriched sequences followed by next-generation sequencing (hMe-Seal) (Song et al., 2011), we determined the sites of dynamic changes in 5-hmC density across the entire genome on days 0, 3, 7, and 10, which allowed us to focus on the steps of stem/early progenitor cell commitment and subsequent differentiation into erythroblasts. We performed hMe-Seal on three independent samples derived from unique human donors and found that there was a high degree of correlation among them, with r^2 values ≥ 0.96 (Figure S2A). Simultaneously with these measurements, we performed RNA-seq (Table S2), on two of these individuals (the same donors whose samples yielded biological replicates #1 and #2 from the hMe-Seal data), enabling us to correlate 5-hmC dynamics with gene expression changes. Again, we found a high degree of correlation on days 0 and 10 between the biological replicates, with r^2 values ≥0.8 (Figure S2B).

We first quantified the total number of 5-hmC peaks that were gained or lost in the intervals days 0–3, days 3–7, and days 7–10, with a minimum fold change of two (Figure S2C). Strikingly, we observed much more loss of 5-hmC peaks across all intervals, with the

maximum loss occurring between days 0 and 3 (Figure S2C, white bars). 5-hmC peak gain occurred as erythroid commitment and differentiation began, peaking between days 3 and 7 (Figure S2C, black bars). We performed a similar analysis with gain and loss of gene expression and found that in parallel with the loss of 5-hmC peaks, initial stem/early progenitor cell commitment is associated most with a loss in gene expression (Figure S2D, white bars). Transcriptional upregulation was most pronounced as erythroid commitment and differentiation occurred between days 0 and 3 as well as from days 3 to 7 (Figure S2D, black bars). We observe the expression of many genes decreasing as differentiation proceeds, with the augmentation in gene expression of only a relatively small number of lineage-specific genes. This observation is in agreement with existing ideas of how stem cells display promiscuous gene expression of multiple lineages (Krause, 2002).

Next, we correlated the association between gains/losses of 5-hmC with particular gene expression patterns. Most notably, the greatest numbers of 5-hmC peaks were lost in genes whose expression levels fall immediately at day 3, many of which correspond to genes required for stem cell function, followed by genes whose expression levels decrease continuously throughout differentiation (Figure S2E). We also examined the patterns of gene expression as a function of the pattern of gain/loss in 5-hmC peaks. Loss of gene expression most often was associated with an initial loss of hydroxymethylation at day 3, and upregulation of gene expression was most often seen with initial gain in 5-hmC peaks (Figure S2F).

Given the complexity of the dynamics seen in both 5-hmC and gene expression patterns, we focused next on classifying the genomic location of 5-hmC peaks that were gained or lost in genes whose expression consistently increased (Figure S2G) or decreased (Figure S2H) over the four time points. Consistency in expression was defined at a change of 2-fold between at least two time points, with no change in any other time interval. The level of 2-fold was chosen because it accurately reflected the known dynamics for key genes classically recognized in erythropoiesis. We observed enrichment of 5-hmC peaks in gene bodies for genes with increased expression compared to genes whose expression decreased.

Peaks with consistently decreased expression generally lost 5-hmC peaks in their gene bodies, with predominant loss in introns, and exhibited rare gain in 5-hmC (Figure S2G). In contrast, peaks consistently upregulated across all time points showed much less loss in 5-hmC density (Figure S2H).

Next, to determine if there was a functional association of 5-hmC peaks at specific regions with gene expression, we overlaid data from hMe-Seal and RNA-seq for genes critical for stem and progenitor function (Figure 2; Table S2), including *CD34* (Figure 2A), *CD38* (Figure 2C), *CD45* (Figure 2D), *CD90* (Figure 2E), and *CD133* (Figure 2F). 5-hmC peaks paralleled the expression of the genes, with 5-hmC peaks appearing throughout the gene body at day 0, but disappearing quickly as stem/early progenitor cell commitment to the erythroid line-age occurred and differentiation proceeded. To determine the fate of the 5-hmC bases that disappear with differentiation, we employed single-base resolution using the Tet-assisted bisulfite sequencing (TAB-seq) method to distinguish unmodified cyto-sine, 5-mC, and 5-hmC. We chose two representative regions near or within *CD34* and observed that 5-hmC is present at high levels on day 0 but that these particular CpG residues were replaced by 5-mC by day 10 (Figure 2B). This is representative of 5-hmC loss with DNA replication in a gene that loses expression with differentiation. We picked the region 4 kb upstream of CD34 to measure cytosine modifications across all time points and found a significant increase in 5-hmC levels on day 3, followed by a highly significant decrease on day 7 (Figure S3A), indicating that the site underwent conversion to unmodified cytosine and immediate remethylation.

In contrast, genes that are induced during erythroid differentiation showed a more variable and dynamic pattern in 5-hmC distribution, as shown by the genes within the *hemoglobin* (*HB*) cluster, which expresses the β, δ, γ, and ε chains (Figure 3A). Different regions throughout the promoters and gene bodies gained and/or lost 5-hmC, indicating a level of dynamic change heretofore unappreciated in studies of cells at one particular stage of differentiation. BCL11A, a molecular regulator involved in HB switching (Lettre et al., 2008; Sankaran et al., 2008; Uda et al., 2008), binds to an intergenic region 5′ from the transcription start site of *HBD* (HB δ) as well as to the HPFH (hereditary persistence of fetal HB) and locus control (LCR) regions. We observed an increase in 5-hmC density at these regions as the cells underwent erythroid differentiation, which further implicates the importance of 5-hmC in the control of gene expression within the *HB* gene cluster (Figure 3A). To determine the dynamics of 5-hmC changes within *HBB* (Figure 3B), we used single-base resolution TAB-seq to examine the changes in cytosine modification for two 5-hmC peaks with opposite changes during differentiation, one peak that increased as differentiation proceeded, and one that was lost with differentiation. For the peak that gains 5-hmC density with differentiation (Figure 3C, left panel), we observed that this region was highly methylated when the gene was not expressed but gained extensive hydroxymethylation with gene expression and differentiation. This gain represents regions where 5-hmC density confers an epigenetic function unto itself because density increases even with DNA replication. This suggests that the base confers an important cell function because 5-hmC gain requires the action of two enzymes, DNMT and TET, in order to accumulate with DNA replication. In contrast, the region that lost 5-hmC density with differentiation (Figure 3C, right panel) showed decrease in 5-hmC density from 14% (day 0) to 0% at day 10, with substantial loss of 5-mC, indicating extensive demethylation resulting in mostly unmodified cytosines at this region. Additional examples from the overlay of data from RNA-seq and hMe-Seal for genes that represent various functional categories and play an important role during erythropoiesis are summarized in Figure S3B.

Loci that Gain 5-hmC during Differentiation Correspond to Gene Regulatory Regions

Given these significant shifts in 5-hmC peak locations, we next chose to focus on regions that gained 5-hmC density throughout differentiation. We reasoned that because cells are actively dividing in the first 10 days of culture, much of the loss in hydroxymethylation that we observe could be accounted for by passive demethylation due to DNA replication (Williams and Wang, 2012). Furthermore, we hypothesized that the DNA loci at which the 5-hmC base was most likely to confer important epigenetic function would be found in regions that demonstrated maintenance and, more so, gain of 5-hmC because both DNA methyltransferase and TET2 enzyme functions would be required at those positions to maintain or gain 5-hmC marks in this replicative system.

Therefore, we analyzed the DNA loci that gained 5-hmC density across any two time points and correlated the known gene annotations at those sites (Figure 4A). We found that gain in 5-hmC was most often found in gene bodies and TF binding sites. However, when we normalized the gene annotations to their respective base pair lengths, we found that only CpG islands failed to gain 5-hmC in proportion to their length, in contrast to all other genomic regions (Figure 4B). Furthermore, gains in 5-hmC were observed mainly in the early stages of lineage commitment and erythroid differentiation (up to day 7) and were not seen from days 7 to 10 (Figure 4B). When we analyzed the 5-hmC binding profile near transcription start sites of known annotated genes, we observed redistribution of 5-hmC density as erythroid differentiation progressed from days 0 to 10 (Figure 4C).

To understand the functional significance of the DNA loci that maintained or gained 5-hmC during erythropoiesis, we analyzed the sequence motifs present in these regions, specifically

focusing on the presence of known TF binding site motifs and correlating with their expression (Table 1). The highest number of hits for significantly enriched motifs included motifs for three TFs known to be critical in erythropoiesis: GATA1, GATA2, and KLF1. The expression of *GATA1* and *KLF1* is dramatically increased during erythroid differentiation, whereas *GATA2* levels are greatly reduced during the same time period (Table 1).

To test for experimental evidence of TF binding to the 5-hmC-enriched sites, we extracted the publicly available data from studies that have measured binding of these particular TFs during erythropoiesis by chromatin immunoprecipitation sequencing (ChIP-seq) (Fujiwara et al., 2009; Kang et al., 2012; Tallack et al., 2010; Xu et al., 2012) and analyzed overlap with the regions that gained 5-hmC density during erythroid differentiation (Figure 4D). We found that between 15% and 34% of the ChIP-seq peaks corresponded to DNA loci that gained 5-hmC density, indicating an overlap of TF binding sites with 5-hmC enrichment at a significant level ($p < 10^{-5}$). We found the presence of a CpG dinucleotide in close proximity to the TF binding sites of GATA1, GATA2, and KLF1. When we analyzed the distance from the TF binding motif to each CpG within the 5-hmC-enriched sequences, we observed co-occurrence between the TF binding motif and the CpG contained within it, suggesting colocalization of 5-hmC with TF binding (Figures S4A–S4C). We also examined the co-occurrence of 5-hmC binding with TF binding during erythroid differentiation at the *HB* gene cluster. Every peak corresponding to the binding of GATA1, GATA2, and/or KLF1 measured by ChIP-seq coincided with a region with 5-hmC gain, suggesting that 5-hmC marks enhancer/TF binding sites needed for stem/early progenitor cell differentiation (Figure S4D). There was also enrichment of GATA1 and GATA2 in the region surrounding the LCR, which is a distal regulatory element that controls gene expression within the *HBB* cluster (Bauer et al., 2012) (Figure S4D). Overlap of 5-hmC gains with GATA1 binding sites was enriched both for sites at which GATA1 functions as a transcriptional activator as well as at those at which it represses transcription (Figures S4E and S4F). The enrichment of 5-hmC at these sites was maintained within the *HB* cluster region in baboon erythroid progenitor cells, again indicating species conservation and likely preserved function (Figures S4G and S4H).

Next, we analyzed whether the genomic regions that gained 5-hmC during erythropoiesis correlated with the positions of modified histones of known function as measured by Xu and colleagues in an independent human ex vivo erythroid differentiation system (Xu et al., 2012). There were higher levels of normalized overlap between loci that gained 5-hmC and activating histone marks, such as H3K9 acetylation, H3K4me1, and H3K4me3, as erythroid differentiation proceeded (Figure 5A).

We recently published an analysis of how modified cytosine bases change throughout erythroid differentiation using the HELP assay, which does not distinguish between 5-mC and 5-hmC (Yu et al., 2013). By comparing our hMe-Seal data to these HELP data, we were able to define common peaks. We found that as differentiation progressed, the percentage of 5-hmC peaks that overlapped with the “peaks” obtained from the HELP assay diminished, suggesting that as erythroid cells develop, an increasing fraction of modified cytosines are due to actual 5-mC bases (Figure 5B).

To test the functional significance of reduced 5-hmC density on hematopoietic differentiation, we contrasted the potential of CD34+ cells derived from normal healthy donors to primary chronic myelomonocytic leukemia (CMML) samples that were either *TET2* wild-type ($TET2^{WT}$) or mutant ($TET2^{MUT}$) to differentiate down the myeloid and erythroid lineages. We chose CMML as a model because patients commonly demonstrate anemia and myeloproliferation, and 30%–50% of cases carry *TET2* mutations (Abdel-

Wahab et al., 2009; Jankowska et al., 2009; Tefferi et al., 2009; Yamazaki et al., 2012). CD34+ cells were selected from these patient samples and cultured under conditions to promote erythroid and myeloid differentiation for 14 days. Analysis on day 14 showed that CD34+ cells obtained from patients with *TET2*^{MUT} CMML were unable to differentiate properly into erythroid cells but were fully capable of generating myeloid cells (Figures 6A and 6B). CD34+ cells from patients with *TET2*^{WT} CMML were mixed in their hematopoietic differentiation capacity: two of the samples were able to differentiate equally well into myeloid and erythroid cells, whereas a third sample differentiated preferentially down the myeloid lineage (Figures 6A and 6B). Of note, the three *TET2*^{WT} samples were all *IDH1/IDH2*^{WT} in genotype. Because TET2 catalyzes the conversion of 5-mC to 5-hmC, we measured the total 5-hmC and 5-mC content of the erythroid cells at day 10. Patients with *TET2*^{MUT} had significantly lower 5-hmC than both nonmalignant and CMML *TET2*^{WT} samples (Figure 6C) as well as higher 5-mC levels (Figure 6D). In vitro erythroid differentiation was performed using murine cells with germline *Tet2* knockout alleles (Moran-Crusio et al., 2011), and mature erythroid (CD71+/Ter119+) cells had significantly lower 5-hmC and increased 5-mC content in comparison with *Tet2*^{wt} mice (Figures S5A and S5B).

In order to interrogate the genome-wide distribution of 5-hmC with *TET2* deficiency, we compared the 5-hmC patterns as measured by hMe-Seal in one *TET2*^{MUT} CMML sample at day 10 of erythroid differentiation to our established data from healthy donors. We observed 36% less coverage of the genome by 5-hmC in the *TET2*^{MUT} CMML sample (Figure 6E, inset), consistent with our data obtained from mass spectrometry (Figure 6C). Notably, there was significant redistribution of the 5-hmC density, with loss of 5-hmC peaks from intronic regions and gain in intergenic regions (Figure 6E). One striking example of this redistribution is seen at the *GATA1* locus, where there is gain of 5-hmC density in the CMML *TET2*^{MUT} sample, which is not seen with normal erythroid differentiation (Figure 6F). It is important to note from this example that the *TET2* mutation does not simply result in loss of 5-hmC density but allows for redistribution of the base, with focal regional gains.

DISCUSSION

To date, 5-hmC has been studied in a variety of contexts, including the early zygote and in ESCs and their differentiated counterparts. The majority of the work in ESCs has shown that 5-hmC is an essential component of stem cell function, with potential roles in transcriptional regulation (Dawlaty et al., 2011; Ficz et al., 2011; Gu et al., 2011; Iqbal et al., 2011; Ito et al., 2010; Koh et al., 2011; Pastor et al., 2011; Stroud et al., 2011; Williams et al., 2011; Wossidlo et al., 2011; Wu et al., 2011; Wu and Zhang, 2011a, 2011b; Xu et al., 2011). These studies left open the question whether this DNA modification is prevalent during adult stem cell commitment and whether 5-hmC played any role within the differentiated cells themselves.

Our work studies 5-hmC in human stem/early progenitor CD34+ cells as well as in several defined erythroid developmental stages, allowing us to follow the progression of 5-hmC dynamically from stem/early progenitor cell commitment throughout cellular differentiation along a single lineage. By combining 5-hmC profiling with RNA-seq, we were able to correlate alterations in 5-hmC with gene expression. As has been observed in previous studies (Ficz et al., 2011; Gu et al., 2011; Iqbal et al., 2011; Pastor et al., 2011; Stroud et al., 2011; Williams et al., 2011; Wossidlo et al., 2011; Wu et al., 2011; Wu and Zhang, 2011a, 2011b), expression of genes important in stem cell function is highly hydroxymethylated when expressed, like *CD34*. Using TAB-seq, we demonstrated that transcriptional repression of *CD34* was associated with loss of 5-hmC and gain of 5-mC at particular CpG residues. Like prior work in mouse ESCs (Bock et al., 2012; Shearstone et al., 2011), we

also observe that genes important for stem cell function are downregulated by DNA methylation in differentiating cells, accompanied by a large overall decrease in the amount of 5-hmC in the latter stages of differentiation.

Notably, we observe that particular DNA loci gain 5-hmC as differentiation proceeds. Because cells undergo DNA replication in our in vitro system, this argues that there is an active process in place to maintain and, in some cases, augment 5-hmC levels at these loci, implying that this epigenetic base has function at those sites. We observed a strong correlation between these regions and TF binding, supporting that 5-hmC is a positive regulator of TF function and implicating several TFs not previously known to be important in hematopoiesis. Taken together, these results suggest that 5-hmC acts both as an intermediate in demethylation to reactivate genes silenced in the primitive stem/early progenitor cell state as well as an activating mark that is distributed widely across expressed genes conferring stem-like function as well as at particular TF binding motifs.

Prior studies in ESCs have shown that 5-hmC is associated with the activating histone marks H3K4me1, H3K4me3, and H3K27ac (Pastor et al., 2011; Stroud et al., 2011; Williams et al., 2011). Similar to this work, we also observe a strong correlation between the presence of the activating histone marks H3K4me1–H3K4me3, H3K9ac, and H3K27ac, and regions that gain 5-hmC, suggesting that key mechanisms of gene transcription are conserved during embryonic as well as adult cell development. Thus, we predict that gain of 5-hmC in these locations acts in conjunction with activating histone marks to open condensed chromatin and allow TF binding and facilitate the function of these proteins.

Our genome-wide hMe-Seal and RNA-seq data will serve as valuable resources that will generate hypotheses for investigators interested in deciphering the molecular mechanisms that drive stem/early progenitor cell commitment with particular relevance to those studying hematopoiesis. The utility of this model is in identifying genomic regions that preserve or gain 5-hmC density as stem/early progenitor cells differentiate down particular lineages, and this will be a powerful strategy to decipher DNA loci that control gene expression. Along with well-known erythroid-specific TFs GATA1, GATA2, and KLF1, our analyses identified TFs that have been reported previously but not studied extensively as principal erythroid differentiation factors, including USF1, ATF3, and ATF4. USF1 (upstream TF 1), a basic-helix-loop-helix (bHLH) TF that binds to E box motifs (5′-CACGTG-3′), has been shown to work sequentially and cooperatively with other bHLH proteins such as Myc during erythroid differentiation (Anantharaman et al., 2011). USF1 has also been shown to recruit SET1 and NURF to chromatin insulator sequences to retain active chromatin structure in erythrocytes (Li et al., 2011b). *ATF3/ATF4* encode members of the cAMP-responsive element binding (CREB) protein family of TFs. ATF3, normally involved in cellular stress response, has been shown to play a role in platelet-derived growth factor (PDGF)-mediated signaling of erythropoietin activation (Xue et al., 2012). ATF4 has been shown to play a role in definitive erythropoiesis, with Atf4 knockout mice failing to proliferate and contribute to overall erythropoiesis (Masuoka and Townes, 2002). The data presented can be regarded as a starting point in the identification of previously uncharacterized factors that may be involved in erythroid-lineage differentiation and will serve as a unique resource for further experimentation and analysis of epigenetic modification in lineage-specific differentiation.

The functional importance of 5-hmC for hematopoietic differentiation is demonstrated by the preferential myeloid differentiation potential of *TET2*-deficient cells, both in human CMML samples as well as in *Tet2*-deficient mice, as has been observed previously (Ko et al., 2011; Li et al., 2011a; Moran-Crusio et al., 2011; Quivoron et al., 2011). *TET2* deficiency results in overall lower 5-hmC levels and skewing of 5-hmC distribution, not simply loss of 5-hmC density. This inherent block in differentiation implies that particular

therapeutic strategies may need to differ between $TET2^{WT}$ and $TET2^{MUT}$ myeloid diseases, especially if the treatments rely on induction of normal cell differentiation. Our work establishes hydroxymethylation as an important epigenetic signal that promotes commitment and differentiation of adult hematopoietic stem/early progenitor cells.

Our work further emphasizes that to interpret data correctly regarding DNA methylation status, it is essential to employ techniques that can distinguish among the known covalent cytosine modifications (Madzo et al., 2013). If techniques such as sodium bisulfite sequencing are used, then complete interpretation as to whether a particular cytosine residue is methylated or hydroxymethylated is not possible. Such techniques may give the impression that there is no change in methylation status but will miss dramatic shifts in modification. Pairing these techniques with a technique such as TAB-seq, which can differentiate 5-hmC from 5-mC, can help address the functional role that each individual base plays in transcriptional regulation and expression. The ability to distinguish among the covalent cytosine modifications will be essential to appreciate the dynamic nature of their distribution throughout differentiation.

EXPERIMENTAL PROCEDURES

Detailed procedures are provided in the Supplemental Experimental Procedures.

Hematopoietic Stem/Progenitor Cell Culture and Erythroid Differentiation

All human material was collected under studies approved by institutional review boards. Human CD34+ cells were purified from mobilized peripheral blood and cultured according to previously published methods (Tamez et al., 2009; Uddin et al., 2004). CD34+ cells from CMML patient samples were selected using the EasySep CD34 Positive Selection Kit (STEMCELL Technologies), and cells were placed into one of two culture conditions: the established erythroid differentiation conditions as detailed in the aforementioned publications, or myeloid conditions (i.e., IMDM with 15% FCS and 15% human serum, 10 ng/ml SCF, 20 ng/ml IL-3, 100 ng/ml FLT3L, and 10 ng/ml G-CSF). Cultures were analyzed after 7 and 14 days by Wright-Giemsa staining in the case of myeloid cells, and hematoxylin-benzidine staining for erythroid progenitors to allow visualization of hemoglobinization. Mouse stem and progenitor cells were cultured according to previously established protocols (Dev et al., 2010). All $Tet2^{-/-}$ mice were housed in a pathogen-free barrier facility at the University of Chicago, maintained under International Animal Care and Use Committee guidelines and an approved protocol.

Analysis of Global DNA Methylation and Hydroxymethylation

DNA hydrolysis was performed, and 5-mC and 5-hmC detected using a Zorbax XDB-C18 2.1 × 50 mm column (1.8 μm particle size) attached to an Agilent 1200 Series liquid chromatography machine in tandem with the Agilent 6410 Triple Quad Mass Spectrometer.

5-hmC-Labeling Reaction

Sonicated genomic DNA was labeled with chemically modified uridine diphosphoglucose glucose (UDP-6-N3-Glu) with standard methods (Song et al., 2011). The DNA samples were then purified by MinElute Reaction Cleanup Kit (Qiagen) and quantified using a Nanodrop UV spectroscope (Thermo Fisher Scientific).

5-hmC-Sequencing Data Mapping and Analysis

The 5-hmC reads for days 0, 3, 7, and 10 were mapped to the reference genome hg19 using BWA. Peak calling for each sample was performed using MACS.

RNA-Seq Data Mapping and Analysis

The RNA-seq reads for days 0, 3, 7, and 10 were mapped to the reference genome hg19 using TopHat. Gene expression analysis was performed with Cufflinks.

Bisulfite and TAB-Seq

Identification of 5-mC or 5-hmC with single-base resolution was performed as described (Yu et al., 2012).

Identification of TF Binding Site Motifs

All 5-hmC-peaks that gained in expression were extracted to perform a TF binding site analysis using HOMER with default parameters for motif identification.

5-hmC Overlap with Known TF and Histone Binding

To analyze the overlap between DNA regions that gain 5-hmC with known histone modifications, we used the data set generated by Stuart Orkin's group (Xu et al., 2012). We overlapped the 5-hmC gained peaks with histone modifica tions and compared them by permuting the 5-hmC peaks 1,000 times. From the actual versus randomized overlap, we calculated the normalized enrichment with the pybedtools randomstats function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge all of the members of the Wickrema and Godley laboratories for helpful discussions as well as constructive comments on this work from Emery Bresnick, Ursula Storb, and Vincent Marchesi. We also thank UIC MMPF members Jerry White and Bryan Zahakaylo for their assistance in running the LC-MS. This work was supported in part by NIH grants CA129831 (to L.A.G.) and CA129831-03S1 (to C.H., B.T.L., and L.A.G.), HL116336 (to A.W. and A. Verma), F32-DK092030 (to A. Vasanthakumar), and the Giving Tree Foundation (to A.W.).

REFERENCES

Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, Malinge S, Yao J, Kilpivaara O, Bhat R, et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood. 2009; 114:144–147. [PubMed: 19420352]

Anantharaman A, Lin IJ, Barrow J, Liang SY, Masannat J, Strouboulis J, Huang S, Bungert J. Role of helix-loop-helix proteins during differentiation of erythroid cells. Mol. Cell. Biol. 2011; 31:1332–1343. [PubMed: 21282467]

Bauer DE, Kamran SC, Orkin SH. Reawakening fetal hemoglobin: prospects for new therapies for the β-globin disorders. Blood. 2012; 120:2945–2953. [PubMed: 22904296]

Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol. Cell. 2012; 47:633–647. [PubMed: 22841485]

Dale, R. Metaseq 0.5. 2013. http://pythonhosted.org/metaseq

Dawlaty MM, Ganz K, Powell BE, Hu YC, Markoulaki S, Cheng AW, Gao Q, Kim J, Choi SW, Page DC, Jaenisch R. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell. 2011; 9:166–175. [PubMed: 21816367]

Dev A, Fang J, Sathyanarayana P, Pradeep A, Emerson C, Wojchowski DM. During EPO or anemia challenge, erythroid progenitor cells transit through a selectively expandable proerythroblast pool. Blood. 2010; 116:5334–5346. [PubMed: 20810925]

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011; 473:398–402. [PubMed: 21460836]

Fujiwara T, O'Geen H, Keles S, Blahnik K, Linnemann AK, Kang YA, Choi K, Farnham PJ, Bresnick EH. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Mol. Cell. 2009; 36:667–681. [PubMed: 19941826]

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011; 477:606–610. [PubMed: 21892189]

Iqbal K, Jin SG, Pfeifer GP, Szabó PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc. Natl. Acad. Sci. USA. 2011; 108:3642–3647. [PubMed: 21321204]

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010; 466:1129–1133. [PubMed: 20639862]

Jankowska AM, Szpurka H, Tiu RV, Makishima H, Afable M, Huh J, O'Keefe CL, Ganetzky R, McDevitt MA, Maciejewski JP. Loss of heterozygosity 4q24 and TET2 mutations associated with myelodys-plastic/myeloproliferative neoplasms. Blood. 2009; 113:6403–6410. [PubMed: 19372255]

Kang JA, Zhou Y, Weis TL, Liu H, Ulaszek J, Satgurunathan N, Zhou L, van Besien K, Crispino J, Verma A, et al. Osteopontin regulates actin cytoskeleton and contributes to cell proliferation in primary erythroblasts. J. Biol. Chem. 2008; 283:6997–7006. [PubMed: 18174176]

Kang YA, Sanalkumar R, O'Geen H, Linnemann AK, Chang CJ, Bouhassira EE, Farnham PJ, Keles S, Bresnick EH. Autophagy driven by a master regulator of hematopoiesis. Mol. Cell. Biol. 2012; 32:226–239. [PubMed: 22025678]

Ko M, Bandukwala HS, An J, Lamperti ED, Thompson EC, Hastie R, Tsangaratou A, Rajewsky K, Koralov SB, Rao A. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc. Natl. Acad. Sci. USA. 2011; 108:14566–14571. [PubMed: 21873190]

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell. 2011; 8:200–213. [PubMed: 21295276]

Krause DS. Regulation of hematopoietic stem cell fate. Oncogene. 2002; 21:3262–3269. [PubMed: 12032767]

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009; 324:929–930. [PubMed: 19372393]

Lettre G, Sankaran VG, Bezerra MA, Araújo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, Orkin SH. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc. Natl. Acad. Sci. USA. 2008; 105:11869–11874. [PubMed: 18667698]

Li Z, Cai X, Cai CL, Wang J, Zhang W, Petersen BE, Yang FC, Xu M. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood. 2011a; 118:4509–4518. [PubMed: 21803851]

Li X, Wang S, Li Y, Deng C, Steiner LA, Xiao H, Wu C, Bungert J, Gallagher PG, Felsenfeld G, et al. Chromatin boundaries require functional collaboration between the hSET1 and NURF complexes. Blood. 2011b; 118:1386–1394. [PubMed: 21653943]

Madzo J, Vasanthakumar A, Godley LA. Perturbations of 5-hydroxymethylcytosine patterning in hematologic malignancies. Semin. Hematol. 2013; 50:61–69. [PubMed: 23507484]

Masuoka HC, Townes TM. Targeted disruption of the activating transcription factor 4 gene results in severe fetal anemia in mice. Blood. 2002; 99:736–745. [PubMed: 11806972]

Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, Figueroa ME, Vasanthakumar A, Patel J, Zhao X, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell. 2011; 20:11–24. [PubMed: 21723200]

Ogawa M. Differentiation and proliferation of hematopoietic stem cells. Blood. 1993; 81:2844–2853. [PubMed: 8499622]

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011; 473:394–397. [PubMed: 21552279]

Quivoron C, Couronné L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, Do Cruzeiro M, Delhommeau F, Arnulf B, Stern MH, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell. 2011; 20:25–38. [PubMed: 21723201]

Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science. 2008; 322:1839–1842. [PubMed: 19056937]

Shearstone JR, Pop R, Bock C, Boyle P, Meissner A, Socolovsky M. Global DNA demethylation during mouse erythropoiesis in vivo. Science. 2011; 334:799–802. [PubMed: 22076376]

Smith LG, Weissman IL, Heimfeld S. Clonal analysis of hematopoietic stem-cell differentiation in vivo. Proc. Natl. Acad. Sci. USA. 1991; 88:2788–2792. [PubMed: 1672767]

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. 2011; 29:68–72. [PubMed: 21151123]

Spradling A, Drummond-Barbosa D, Kai T. Stem cells find their niche. Nature. 2001; 414:98–104. [PubMed: 11689954]

Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 2011; 12:R54. [PubMed: 21689397]

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; 324:930–935. [PubMed: 19372391]

Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, Perkins AC. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Res. 2010; 20:1052–1063. [PubMed: 20508144]

Tamez PA, Liu H, Fernandez-Pol S, Haldar K, Wickrema A. Stage-specific susceptibility of human erythroblasts to Plasmodium falciparum malaria infection. Blood. 2009; 114:3652–3655. [PubMed: 19706885]

Tefferi A, Lim KH, Abdel-Wahab O, Lasho TL, Patel J, Patnaik MM, Hanson CA, Pardanani A, Gilliland DG, Levine RL. Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. Leukemia. 2009; 23:1343–1345. [PubMed: 19295549]

Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc. Natl. Acad. Sci. USA. 2008; 105:1620–1625. [PubMed: 18245381]

Uddin S, Ah-Kang J, Ulaszek J, Mahmud D, Wickrema A. Differentiation stage-specific activation of p38 mitogen-activated protein kinase isoforms in primary human erythroid cells. Proc. Natl. Acad. Sci. USA. 2004; 101:147–152. [PubMed: 14694199]

Williams RT, Wang Y. A density functional theory study on the kinetics and thermodynamics of N-glycosidic bond cleavage in 5-substituted 2^0-deoxycytidines. Biochemistry. 2012; 51:6458–6462. [PubMed: 22809372]

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011; 473:343–348. [PubMed: 21490601]

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011; 2:241. [PubMed: 21407207]

Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev. 2011a; 25:2436–2452. [PubMed: 22156206]

Wu H, Zhang Y. Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle. 2011b; 10:2428–2436. [PubMed: 21750410]

Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. 2011; 25:679–684. [PubMed: 21460036]

Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Wang P, Xiao MT, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer Cell. 2011; 19:17–30. [PubMed: 21251613]

Xu J, Shao Z, Glass K, Bauer DE, Pinello L, Van Handel B, Hou S, Stamatoyannopoulos JA, Mikkola HK, Yuan GC, Orkin SH. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev. Cell. 2012; 23:796–811. [PubMed: 23041383]

Xue Y, Lim S, Yang Y, Wang Z, Jensen LD, Hedlund EM, Andersson P, Sasahara M, Larsson O, Galter D, et al. PDGF-BB modulates hematopoiesis and tumor angiogenesis by inducing erythropoietin production in stromal cells. Nat. Med. 2012; 18:100–110. [PubMed: 22138754]

Yamazaki J, Taby R, Vasanthakumar A, Macrae T, Ostler KR, Shen L, Kantarjian HM, Estecio MR, Jelinek J, Godley LA, Issa JP. Effects of TET2 mutations on DNA methylation in chronic myelomonocytic leukemia. Epigenetics. 2012; 7:201–207. [PubMed: 22395470]

Yu M, Hon GC, Szulwach KE, Song CX, Jin P, Ren B, He C. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat. Protoc. 2012; 7:2159–2170. [PubMed: 23196972]

Yu Y, Mo Y, Ebenezer D, Bhattacharyya S, Liu H, Sundaravel S, Giricz O, Wontakal S, Cartier J, Caces B, et al. High resolution methylome analysis reveals widespread functional hypomethylation during adult human erythropoiesis. J. Biol. Chem. 2013; 288:8805–8814. [PubMed: 23306203]

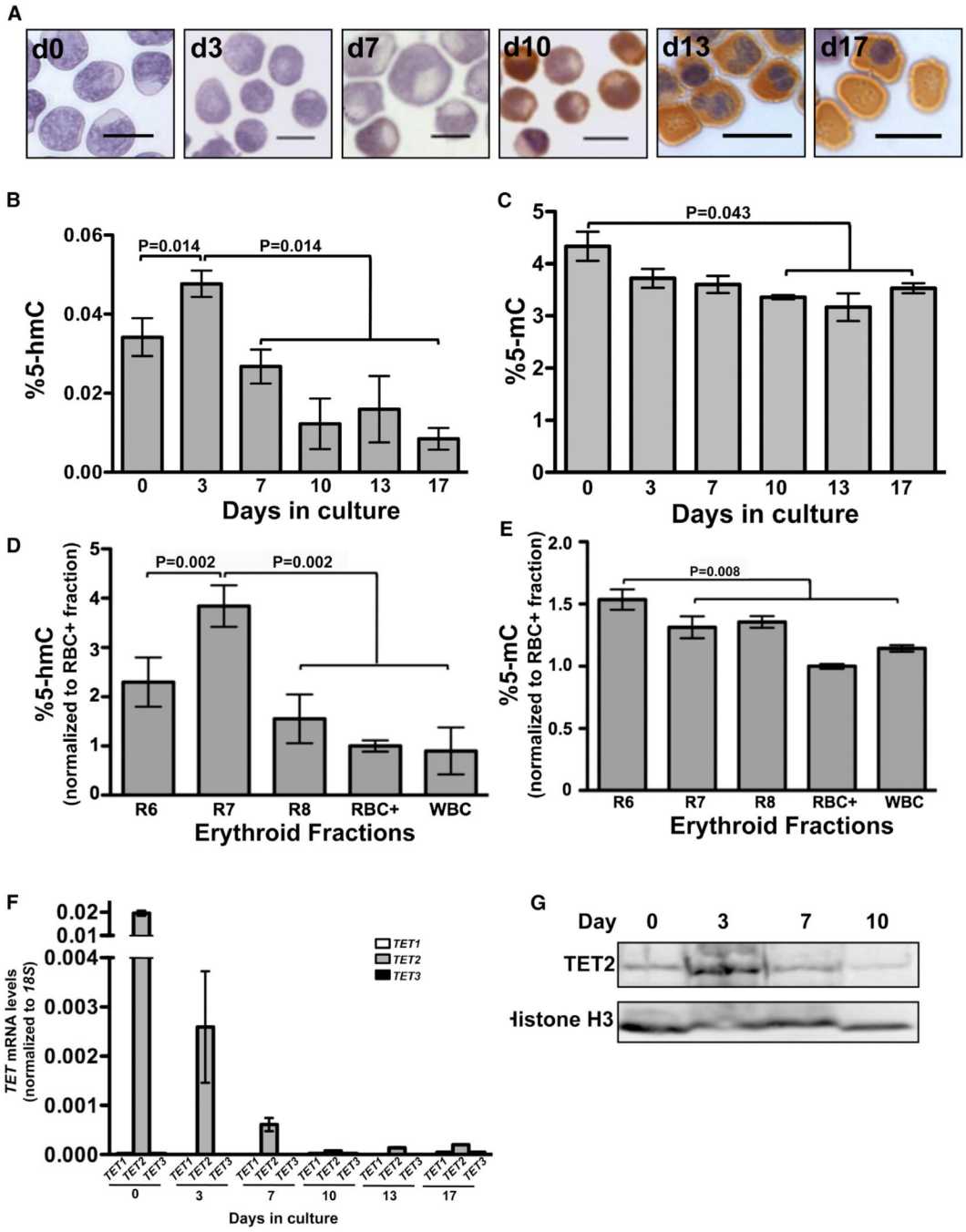

Figure 1. Dynamic Changes in 5-hmC and 5-mC Levels during Erythroid Differentiation

(A) Photomicrographs of hematoxylin and benzidine-stained cells cytospun onto slides on the days indicated. Changes in cellular morphology as well as acquisition of HB (brown) during the differentiation program are seen in the micrographs.

(B and C) Total 5-hmC (B) and 5-mC (C) content in cells at days 0, 3, 7, 10, 13, and 17 during in vitro erythroid differentiation, measured by LC-MS. The average (± SD) of two independent biological replicates obtained from two independent donor CD34+ cells is shown. Repeated measures ANOVA was used to determine statistical significance.

(D) Total 5-hmC content in the erythroid fractions isolated from baboon bone marrow. “R6” is the CD117+ CD36 bRBC fraction of cells, which form granulocytic, mixed, and large

BFUe colonies in methylcellulose. This population contains both early progenitors and later nonerythroid progenitors. “R7” is the CD117+CD36+ fraction, which forms CFUe and late BFUe in methylcellulose (10%–20% erythroid colony-forming cells). “R8” is the CD117–CD36+ bRBC– fraction, which does not form colonies in methylcellulose. “RBC+” is the fraction of erythroid precursors that do not form colonies and that were purified using Miltenyi Biotec columns with the baboon-specific red blood cell (RBC) antibody (BD Biosciences). All numbers were normalized to RBC+ values. White blood cell (WBC) was used as a control. Data are represented as the mean of two independent biological replicates ± SD.

(E) Total 5-mC content in the erythroid fractions isolated from baboon bone marrow, as described in (D). Data are represented as the mean of two independent biological replicates ± SD.

(F) qRT-PCR for *TET1* (white bars), *TET2* (gray bars), and *TET3* (black bars) expression normalized to *18S*, at days 0, 3, 7, 10, 13, and 17 during in vitro erythroid differentiation. The average (± SD) of two independent biological replicates is shown.

(G) Representative western blot showing expression of TET2 protein in nuclear extracts from cells at days 0, 3, 7, and 10. The blot was stripped and reprobed for histone H3, the loading control.

See also Figure S1 and Table S1.

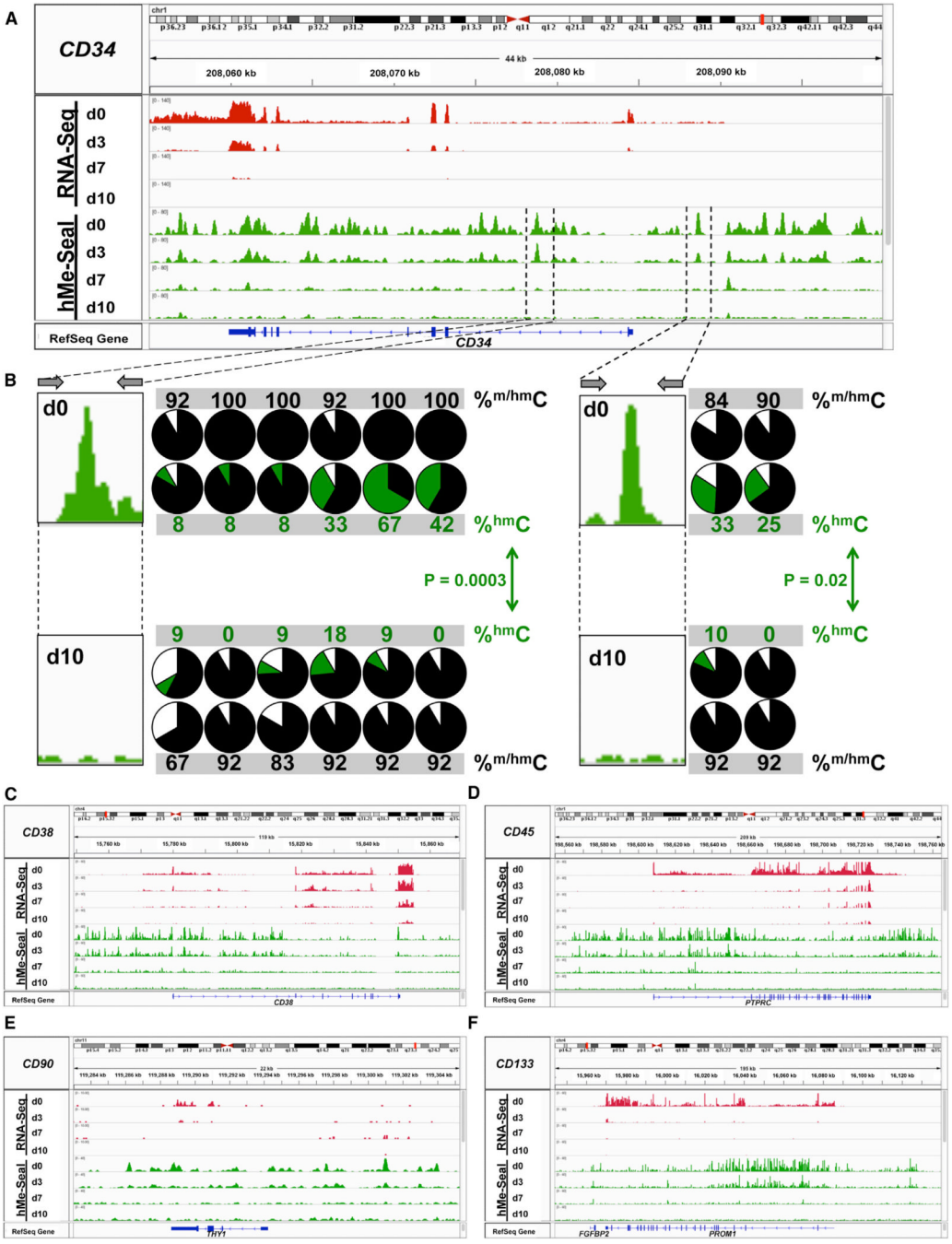

Figure 2. Dynamic Changes of 5-hmC Peaks and RNA-Seq in Regions Surrounding Stem and Progenitor Cell Genes

(A) RNA-seq (red) and hMe-Seal (green) tracks in the *CD34* gene.

(B) Pie charts depict results of bisulfite sequencing (outer rows) and TAB-seq (inner rows) showing quantitation of modified cytosines over regions of the *CD34* gene on day 0 (top) and day 10 (bottom). For bisulfite sequencing, the percentage of unprotected cytosines is shown in white and of protected cytosines in black; for TAB-seq, the percentage of unmodified cytosines is shown in white, of 5-mC in black, and of 5-hmC in green. p values were calculated by using the χ^2 test to compare 5-hmC percentages at days 0 and 10.

(C–F) RNA-seq (red) and hMe-Seal (green) tracks in regions surrounding *CD38* (C), *CD45* (D), *CD90* (E), and *CD133* (F).
See also Figures S2 and S3.

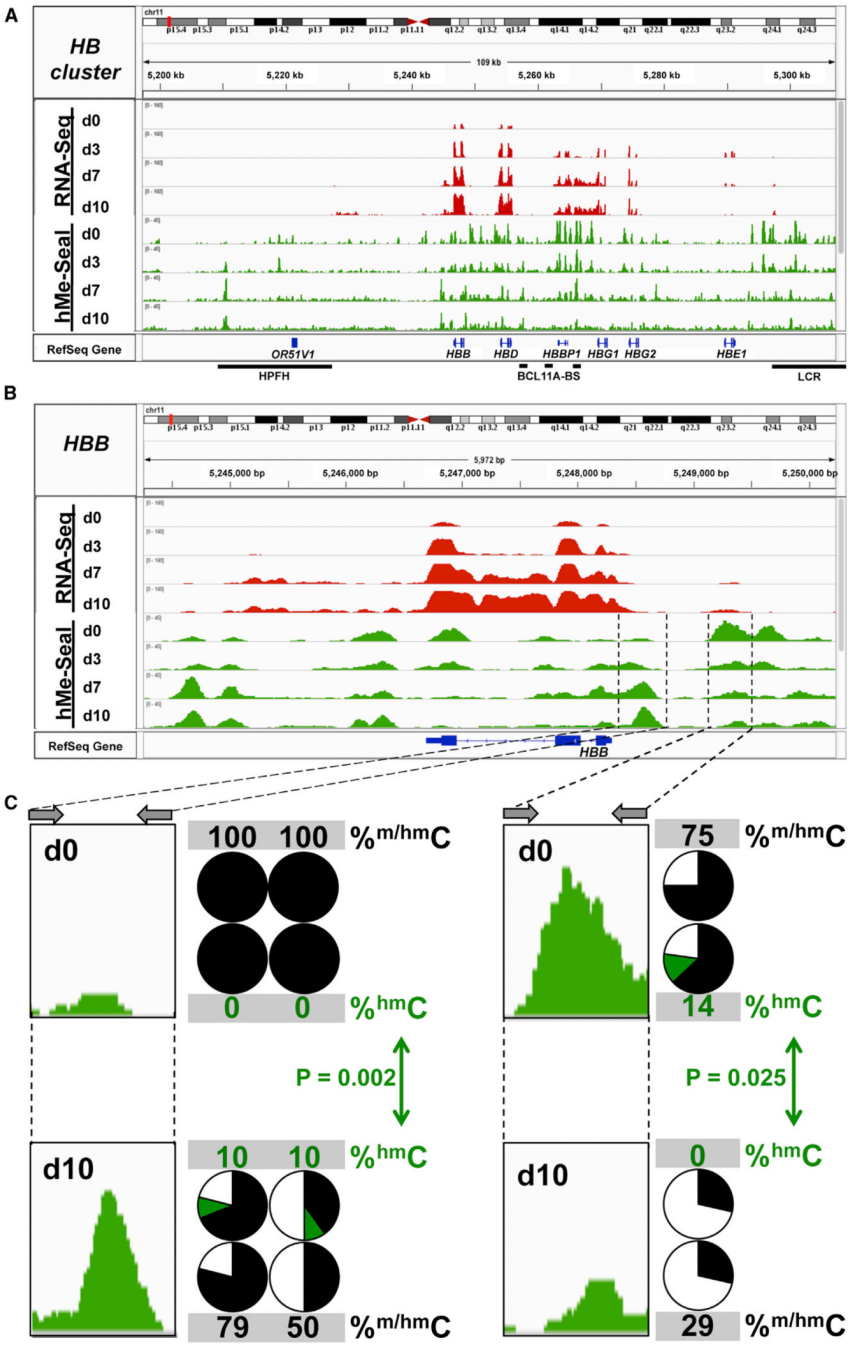

Figure 3. Dynamic Changes of 5-hmC Peaks and RNA-Seq in Regions Surrounding the HB Gene Cluster

(A) RNA-seq (red) and hMe-Seal (green) tracks in the region surrounding the HB cluster. Black bars indicate known regulatory regions: HPFH, BCL11A-BS (BCL11A binding site), and LCR.

(B) Results in (A) magnified to show the *HBB* gene.

(C) Pie charts depict results of bisulfite sequencing (outer rows) and TAB-seq (inner rows) showing quantitation of modified cytosines over regions of *HBB* on day 0 (top) and day 10 (bottom). For bisulfite sequencing, the percentage of unprotected cytosines is shown in white and of protected cytosines in black; for TAB-seq, the percentage of unmodified

cytosines is shown in white, of 5-mC in black, and of 5-hmC in green. p values were calculated by using the χ^2 test to compare 5-hmC percentages at days 0 and 10. See also Figures S2 and S3.

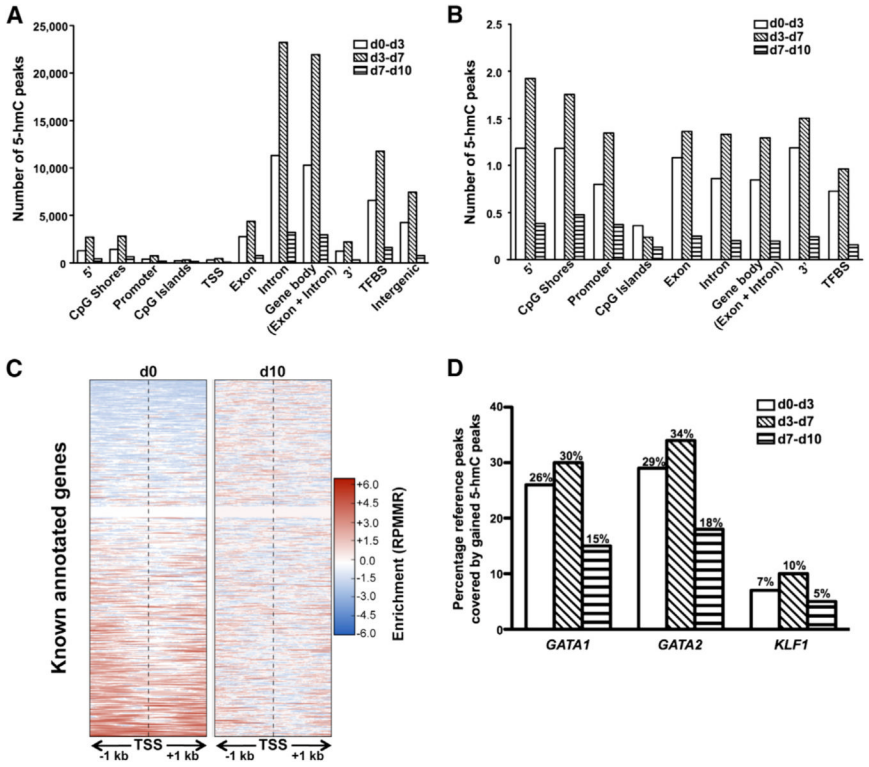

Figure 4. Functional Significance of 5-hmC Gains

(A) Total number of gained 5-hmC peaks between days 0 and 3 (days 0–3, white), 3 and 7 (days 3–7, hatched), and 7 and 10 (days 7–10, striped) with respect to relative position to annotated gene elements.

(B) Relative number of 5-hmC peaks from (A) normalized to the base pair lengths of each element.

(C) Comparison of the normalized 5-hmC reads surrounding the TSS, indicated by a vertical dotted line, ±1,000 bp for all known annotated genes at day 0 (left) versus day 10 (right), plotted by the Metaseq suite (Dale, 2013).

(D) Percentage of the overlap among publicly available GATA1, GATA2, and KLF1 ChIP-seq data (Fujiwara et al., 2009; Kang et al., 2012; Tallack et al., 2010) and 5-hmC gained peaks over days 0–3 (white), 3–7 (hatched), and 7–10 (striped) time points with significance ($p \leq 10^{-5}$).

See also Figure S4.

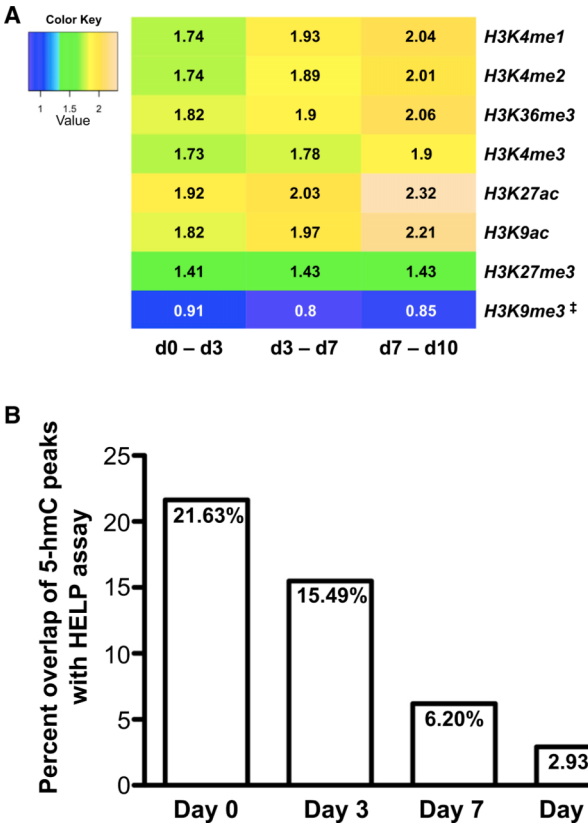

Figure 5. Overlap of Histone Marks as well as DNA Methylation Dynamics with 5-hmC Peaks that Were Gained between Days 0 and 3, 3 and 7, and 7 and 10

(A) Overlap of the distribution of particular histone modifications (deposited by Xu et al., 2012) and gain of 5-hmC density throughout erythroid differentiation. Scores were calculated as the ratio of actual overlap relative to a normal distribution generated by random permutations (yellow, high normalized enrichment; green, medium normalized enrichment; and blue, low normalized enrichment). All normalized enrichment scores except H3K9me3 reached significance of at least 10^{-5}. ‡, levels of significance for H3K9me3 versus 5-hmC enrichment were days 0–3 ($p = 0.066$), 3–7 ($p < 10^{-5}$), and 7–10 ($p = 0.075$). (B) Percentage of overlap between 5-hmC-specific peaks identified during erythroid differentiation and modified cytosines identified using the HELP assay (Yu et al., 2013).

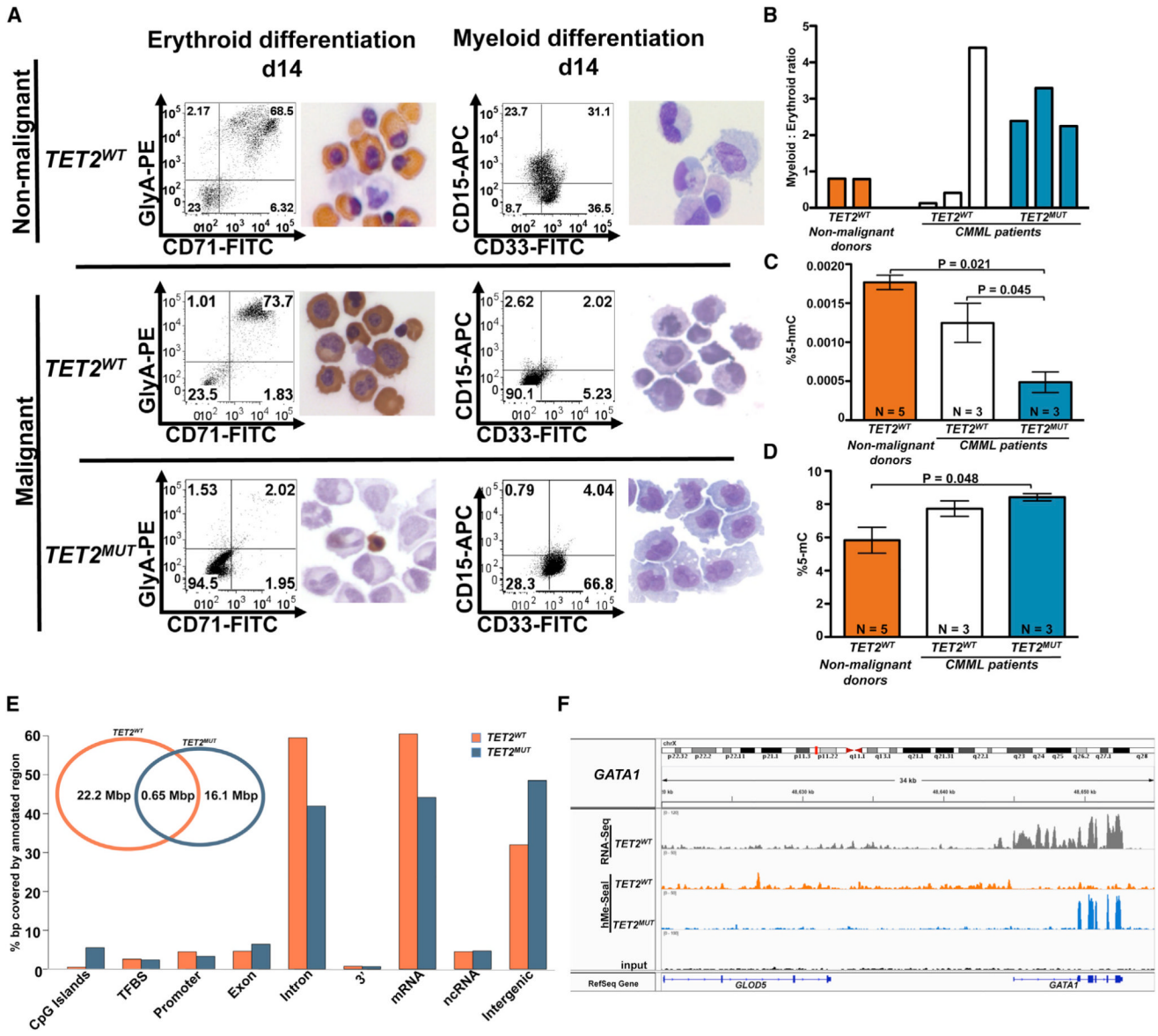

Figure 6. *TET2* Mutations and 5-hmC Deficiency Impede Erythroid Differentiation and Enhance Myeloid Development

(A) Bone marrow-derived CD34+ cells from healthy nonmalignant $TET2^{WT}$ donors (top panel) or from patients with CMML with $TET2^{WT}$ (middle panel) or $TET2^{mut}$ (bottom panel) were differentiated down the erythroid (left) or myeloid lineage (right). Differentiation was determined by FACS at day 14. Erythroid differentiation was determined by expression of Glycophorin A and CD71, and myeloid differentiation by expression of CD15 and CD33. Photomicrographs of cytospins prepared at each time point are shown to the right of each FACS plot.

(B) Differential differentiation potential was quantified by calculating a myeloid:erythroid (M:E) ratio for each sample. Myeloid expression was defined by the CD15+ and CD15+CD33+ populations, and erythroid differentiation was defined using the Glycophorin A+ CD71+ population. Samples from healthy nonmalignant $TET2^{wt}$ cells are shown in orange, those from patients with $TET2^{wt}$ CMML are shown in white, and those from patients with $TET2^{mut}$ CMML are shown in blue.

(C) Total 5-hmC levels measured at day 10 of in vitro erythroid differentiation, as determined by LC/MS. Data are represented as mean ± SEM.
(D) Total 5-mC levels measured in these same samples, as determined by LC/MS. Data are represented as mean ± SEM.
(E) Overlap in 5-hmC reads in nonmalignant CD34+ cells (orange) versus CD34+ cells derived from a patient with *TET2*-mutated CMML (blue) at day 10 after in vitro erythroid differentiation. Inset is a Venn diagram showing the number of base pairs covered by 5-hmC peaks in the respective samples.
(F) RNA-seq (top track, gray) and hMe-Seal tracks (for nonmalignant CD34+ cells, middle track, orange; versus patient with *TET2*-mutated CMML, bottom track, blue) in the region surrounding *GATA1*.
See also Figure S5.

Table 1

Transcription Factor Binding Sites within Regions that Gain 5-hmC during Erythroid Differentiation

Transcription Factor	3inding Site Sequence[a]	p Values			Transcripts per Cell			
		Days 0–3	Days 3–7	Days 7–10	Day 0	Day 3	Day 7	Day 10
GATA2	BBCTTATCTS	1×10^{-369}	$1 \times 10^{-1,742}$	1×10^{-69}	98.8	73.8	28.4	14.5
GATA1	SAGATAAGRV	1×10^{-368}	$1 \times 10^{-1,652}$	1×10^{-68}	9.2	100.2	127.7	126.0
KLF1	NWGGGTGTGGCY	1×10^{-36}	1×10^{-244}	1×10^{-19}	5.7	80.4	267.1	381.2
NF1	CYTGGCABNSSTGCCA	1×10^{-5}	1×10^{-154}	1×10^{-10}	6.7	2.7	3.0	2.3
STAT5A	RTTTCTNAGAAA	1×10^{-36}	1×10^{-52}	$< 1 \times 10^{-5}$	56.3	60.1	46.9	6.8
STAT1	NATTTCCNGGAAAT	1×10^{-29}	1×10^{-39}	$< 1 \times 10^{-5}$	12.6	4.8	5.5	7.1
NF-E2	GATGACTCAGCA	1×10^{-16}	1×10^{-18}	$< 1 \times 10^{-5}$	91.2	117.4	178.3	492.2
HIF1A	TACGTGCV	1×10^{-15}	1×10^{-15}	$< 1 \times 10^{-5}$	57.9	9.8	8.9	16.7
EGR1	TCCGCCCACGCA	1×10^{-14}	1×10^{-6}	$< 1 \times 10^{-5}$	71.9	12.4	30.6	63.0
MYB	GGCVGTTR	1×10^{-10}	1×10^{-25}	$< 1 \times 10^{-5}$	136.1	91.8	70.9	21.6
STAT3	CTTCCGGGAA	1×10^{-9}	1×10^{-11}	$< 1 \times 10^{-5}$	42.5	26.7	19.3	10.0
Myc	VCCACGTG	1×10^{-7}	1×10^{-35}	$< 1 \times 10^{-5}$	27.7	57.0	71.3	29.1
Max	RCCACGTGGYYN	1×10^{-7}	1×10^{-33}	$< 1 \times 10^{-5}$	36.7	21.4	19.1	32.0
USF1	SGTCACGTGR	1×10^{-6}	1×10^{-40}	$< 1 \times 10^{-5}$	34.9	21.5	20.8	25.2
ATF3	SGGTCACGTGAC	1×10^{-6}	1×10^{-27}	$< 1 \times 10^{-5}$	40.7	0.38	0.87	12.6
CEBP	ATTGCGCAAC	1×10^{-342}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	6.4	11.7	18.5	82.7
	DRTGTTGCAA	1×10^{-147}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$				
c-Jun	VTGACTCATC	1×10^{-112}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	515.4	1.2	0.86	3.8
RUNX1	AAACCACARM	1×10^{-97}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	41.2	26.4	12.7	6.7
RUNX2	NWAACCACADNN	1×10^{-75}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	8.2	0.6	0.2	0.2
	GCTGTGGTTW	1×10^{-68}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$				
ATF4	GATTGCATCA	1×10^{-66}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	176.2	208.5	196.9	286.78
SPI1	AGAGGAAGTG	1×10^{-23}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	53.5	68.1	12.2	4.0
ERG	ACAGGAAGTG	1×10^{-11}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	34.5	4.9	0.63	0.17
STAT6	TTCCKNAGAA	1×10^{-7}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	54.6	17	17.4	16.3
Sp1	GGCCCCGCCCCC	$< 1 \times 10^{-5}$	1×10^{-13}	$< 1 \times 10^{-5}$	21.6	11.7	15.0	29.2
ZNF711	AGGCCTAG	$< 1 \times 10^{-5}$	1×10^{-9}	$< 1 \times 10^{-5}$	32.4	5.2	2.8	1.5

The list is limited to motifs with an enrichment of $p > 1 \times 10^{-5}$ on the basis of HOMER software ChIP-seq analysis and a putative transcription factor mRNA expression of more than five transcripts per cell based on whole-transcriptome sequencing at days 0, 3, 7, or 10 of erythroid differentiation.

[a] Nucleic acid nomenclature. A, adenine; T, thymine; G, guanine; C, cytosine; R, guanine or adenine; Y, thymine or cytosine; M, adenine or cytosine; K, guanine or thymine; S, guanine or cytosine; W, adenine or thymine; H, adenine, cytosine, or thymine; B, guanine, thymine, or cytosine; V, guanine, cytosine, or adenine; D guanine, thymine, or adenine; N, any nucleic acid.