

UCSF

UC San Francisco Previously Published Works

Title

The Cell Type-Specific 5hmC Landscape and Dynamics of Healthy Human Hematopoiesis and TET2-Mutant Preleukemia.

Permalink

<https://escholarship.org/uc/item/6bv854nh>

Authors

Nakauchi, Yusuke

Azizi, Armon

Thomas, Daniel

et al.

Publication Date

2022-07-06

DOI

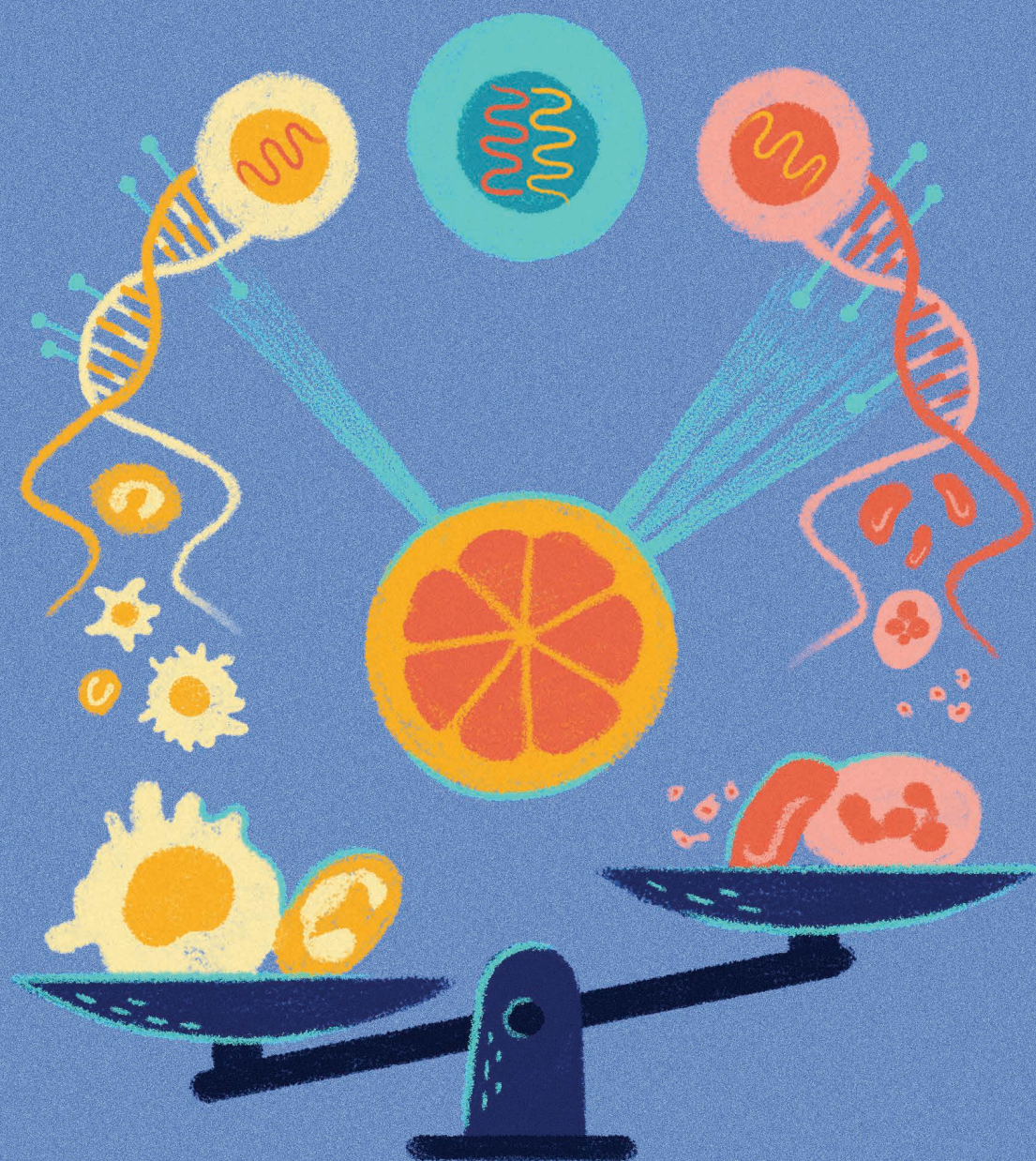
10.1158/2643-3230.BCD-21-0143

Peer reviewed

The Cell Type–Specific 5hmC Landscape and Dynamics of Healthy Human Hematopoiesis and *TET2*-Mutant Preleukemia



Yusuke Nakauchi^{1,2,3}, Armon Azizi^{1,2,3,4}, Daniel Thomas^{1,2,3,5}, M. Ryan Corces^{1,2,3,6,7,8}, Andreas Reinisch^{1,2,3,9}, Rajiv Sharma^{1,2,3}, David Cruz Hernandez^{1,2,3,10}, Thomas Köhnke^{1,2,3}, Daiki Karigane^{1,2,3}, Amy Fan^{1,2,3}, Daniel Martinez-Krams^{1,2,3}, Melissa Stafford^{1,2,3}, Satinder Kaur^{1,2,3}, Ritika Dutta^{1,2,3}, Paul Phan^{1,2,3}, Asiri Ediriwickrema^{1,2,3}, Erin McCarthy¹¹, Yuhong Ning¹¹, Tierney Phillips¹¹, Christopher K. Ellison¹¹, Gulfem D. Guler¹¹, Anna Bergamaschi¹¹, Chin-Jen Ku¹¹, Samuel Levy¹¹, and Ravindra Majeti^{1,2,3}



ABSTRACT

The conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) is a key step in DNA demethylation that is mediated by ten-eleven translocation (TET) enzymes, which require ascorbate/vitamin C. Here, we report the 5hmC landscape of normal hematopoiesis and identify cell type-specific 5hmC profiles associated with active transcription and chromatin accessibility of key hematopoietic regulators. We utilized CRISPR/Cas9 to model *TET2* loss-of-function mutations in primary human hematopoietic stem and progenitor cells (HSPC). Disrupted cells exhibited increased colonies in serial replating, defective erythroid/megakaryocytic differentiation, and *in vivo* competitive advantage and myeloid skewing coupled with reduction of 5hmC at erythroid-associated gene loci. Azacitidine and ascorbate restored 5hmC abundance and slowed or reverted the expansion of *TET2*-mutant clones *in vivo*. These results demonstrate the key role of 5hmC in normal hematopoiesis and *TET2*-mutant phenotypes and raise the possibility of utilizing these agents to further our understanding of preleukemia and clonal hematopoiesis.

SIGNIFICANCE: We show that 5-hydroxymethylation profiles are cell type-specific and associated with transcriptional abundance and chromatin accessibility across human hematopoiesis. *TET2* loss caused aberrant growth and differentiation phenotypes and disrupted 5hmC and transcriptional landscapes. Treatment of *TET2* KO HSPCs with ascorbate or azacitidine reverted 5hmC profiles and restored aberrant phenotypes.

INTRODUCTION

5-methylcytosine (5mC) plays a critical role in regulating gene expression in higher organisms. This reversible epigenetic mark is added to CpG dinucleotides through the enzymatic action of DNA methyltransferases. The ten-eleven translocation (TET) family of proteins (*TET1*, *TET2*, and *TET3*) are involved in the Fe²⁺ and ascorbate-dependent oxidation of 5-methylcytosine to 5-hydroxymethylcytosine (5hmC), a critical intermediate in the active demethylation pathway culminating in base excision repair (1, 2). Studies

in brain tissue have identified 5hmC accumulation at gene bodies independent of 5mC demethylation (3), suggesting a potential epigenetic regulatory role. Similar to 5mC, 5hmC is globally downregulated in cancer through a variety of mechanisms including loss-of-function mutations within genes such as *TET2*, or mutant isocitrate dehydrogenase production of (R)-2-hydroxyglutarate that inhibits *TET2* function (4).

Genomic studies have identified somatic mutations in *TET2* across hematologic malignancies, preleukemic stem cells, and clonal hematopoiesis, suggesting that regulation of 5hmC may be critical to stem cell and cancer phenotypes (5, 6). Indeed, mouse studies have shown that mutations in *TET2* promote hematopoietic stem cell self-renewal and myeloid-skewed differentiation (7). Furthermore, these mutations are associated with epigenetic changes underlying disease phenotypes, specifically the reduction of global 5hmC in *TET2*-mutated hematopoietic malignancy (8).

Clonal hematopoiesis (CH) is the aberrant expansion of blood cell clones derived from a single hematopoietic stem cell (HSC), and its incidence is associated with aging (9, 10). Large cohort sequencing studies have determined that *TET2* is one of the most frequently mutated genes in CH. Individuals with CH have an increased risk of developing hematologic malignancies, and presumably, these CH clones acquire additional mutations to evolve into cancer. Indeed, such preleukemic HSCs have been identified in patients with acute myeloid leukemia (AML), including clones with *TET2* mutations (11, 12).

Recent studies have charted the transcriptional, open chromatin, and methylation landscapes of normal human hematopoiesis (12, 13). These studies have shown that most epigenetic changes contributing to cellular phenotypes during hematopoietic differentiation occur in regions with low CpG levels (14). Furthermore, integrated analyses of DNA methylation and gene-expression data have revealed cell type-specific epigenetic programs, not only for mature blood

¹Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, California. ²Cancer Institute, Stanford University School of Medicine, Stanford, California. ³Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California. ⁴University of California Irvine School of Medicine, Irvine, California. ⁵South Australian Health and Medical Research Institute (SAHMRI), University of Adelaide, Adelaide, Australia. ⁶Gladstone Institute of Neurological Disease, San Francisco, California. ⁷Gladstone Institute of Data Science and Biotechnology, San Francisco, California. ⁸Department of Neurology, University of California San Francisco, San Francisco, California. ⁹Division of Hematology, Medical University of Graz, Graz, Austria. ¹⁰MRC Molecular Haematology Unit and Oxford Centre for Haematology, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ¹¹Bluestar Genomics Inc., San Mateo, California.

Note: Supplementary data for this article are available at Blood Cancer Discovery Online (<https://bloodcancerdiscov.aacrjournals.org/>).

Y. Nakauchi and A. Azizi contributed equally to this article.

Corresponding Author: Ravindra Majeti, Stanford University School of Medicine, Stanford Institute for Stem Cell Biology and Regenerative Medicine, Lokey Stem Cell Building, 265 Campus Drive, Stanford, CA 94305-5463. Phone: 650-721-6376; Fax: 650-736-2961; E-mail: rmajeti@stanford.edu

Blood Cancer Discov 2022;3:346-67

doi: 10.1158/2643-3230.BCD-21-0143

©2022 American Association for Cancer Research

cell types but also for individual hematopoietic stem and progenitor cell (HSPC) subpopulations (15). We recently assayed the chromatin profiles of hematopoietic cell types and identified distinct chromatin signatures across the hierarchy, as well as key regulators governing hematopoietic differentiation (12).

Initial investigations of DNA methylation in human cancers, including hematopoietic malignancy, revealed severely altered patterns of methylation such as promoter, CpG island (CGIs), and CGI-flanking CpG-shore hypermethylation, but also repetitive element hypomethylation, highlighting the importance of methylation in regulating cellular function and phenotypes (16). 5hmC appears to be found at demarcation borders of CpG islands and CpG canyons in normal tissues and across gene bodies of actively expressed genes and poised bivalent promoters, but has not been comprehensively studied in human cells (17). Evidence from both clinical specimens and experimental models suggests that TET2 is required to prevent aberrant 5mC and 5hmC changes at enhancer sites (7, 18, 19). This activity is attenuated by *TET2* mutations, leading to enhancer hypermethylation and to a critical deregulation of enhancer-associated gene-expression patterns. This places *TET2* mutations in line with other so-called landscaping mutations that predispose cells to malignant transformation (20).

A recent study has revealed that 5hmC may have multiple independent functions in the active demethylation process (21). 5hmC modifications are thought to be either stable epigenetic marks or derived from *de novo* methylated 5mC during early embryonic development (22). Additionally, 5hmC modification at gene bodies and enhancers is associated with transcriptional activation (23). Notably, 5hmC-modified loci serve as informative biomarkers for a number of human cancers including hematologic malignancy and other diseases (24–26). Despite the wealth of knowledge gained in recent years about 5hmC in healthy and malignant tissue, the contribution of *TET2* to HSC clonal growth advantage and increased genomic instability is poorly understood.

TET2 activity itself is subject to dynamic regulation. Several studies using mouse ESC and iPSCs have shown that ascorbate is able to augment the activity of wild-type *TET2* and dramatically increase 5hmC production (27). Recently, a role for ascorbate in maintaining physiologic levels of 5hmC and fully functional TET enzymatic activity in HSCs has been reported (28, 29). In addition, several pharmacologic agents have been linked with *TET2* activity. Notably,

DNA hypomethylating agents such as 5-azacitidine (azacitidine) and 5-aza-2'-deoxycytidine (decitabine) elicit a higher response rate in patients with *TET2* mutations (30), but it is not known whether these agents can restore global and genome-wide 5hmC levels.

Here, we explored these questions by investigating 5hmC in normal and *TET2*-mutant human hematopoiesis using a recently developed genome-wide 5hmC sequencing assay. We characterized the 5hmC landscape in healthy hematopoietic cell types and primary human CD34⁺ HSPCs engineered with disrupted *TET2* using CRISPR/Cas9 to mimic *TET2*-mutated preleukemia/CH. We also used these methods to investigate the effects of pharmacologic agents reported to be active against *TET2*-mutant cells including azacitidine and ascorbate and demonstrate that they revert aberrant 5hmC profiles resulting from *TET2* disruption.

RESULTS

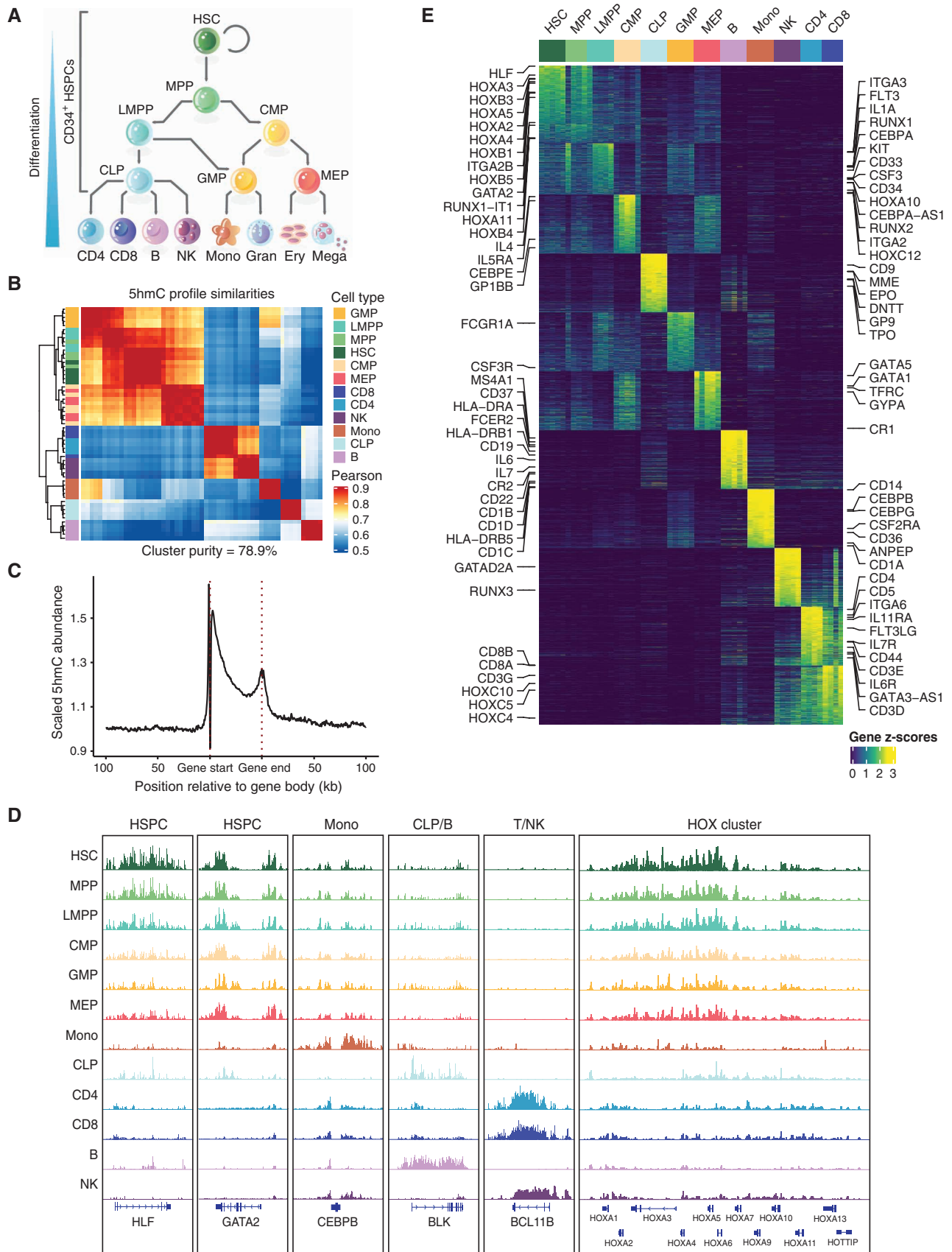
The 5hmC Landscape of the Normal Human Hematopoietic Hierarchy

To characterize the genome-wide distribution of 5hmC, we utilized 5hmC-seq, a recently developed 5hmC pulldown and sequencing assay. This approach biotinylates 5hmC with subsequent streptavidin bead enrichment to create DNA libraries with fragments containing 5hmC (Supplementary Fig. S1A; refs. 24, 31). We applied this methodology to 7 HSPC populations and 5 mature cell types fluorescence-activated cell sorting (FACS)-purified from the bone marrow (BM) and peripheral blood of 5 healthy human donors (Fig. 1A; Supplementary Fig. S1B–S1D).

Unsupervised clustering of 5hmC profiles grouped cell types with high specificity demonstrating that 5hmC profiles are cell type-specific (Fig. 1B). In previous studies (12), accessible distal chromatin elements were found to be more cell type-specific than promoters. However, 5hmC at both promoters and distal elements was cell type-specific, suggesting that 5hmC plays cell type-specific roles in both local regulation of transcription and distal epigenetic regulation (Supplementary Fig. S2A and S2B).

Prior literature has described the accumulation of 5hmC around gene bodies in other tissue types (31). Across the hematopoietic cell types, there was strong enrichment of the 5hmC signal within gene bodies, extending to almost 10 kb away from the transcription start site (TSS) and transcription termination site (TTS), suggesting potential interactions

Figure 1. The 5-hydroxymethylation landscape of the hematopoietic hierarchy. **A**, Schematic of the human hematopoietic hierarchy showing the 12 primary cell types analyzed in this work. Granulocytes, megakaryocytes, and erythroid cells were excluded. The cell types comprising CD34⁺ HSPCs are indicated. Colors used in this schematic are consistent throughout the figures. HSCs (Lin⁻ CD34⁺ CD38⁻ CD45RA⁻ CD90⁺ CD10⁻) multipotent progenitors (MPP; Lin⁻ CD34⁺ CD38⁻ CD45RA⁻ CD90⁺ CD10⁻), lymphoid-primed multipotent progenitors (LMPP; Lin⁻ CD34⁺ CD38⁻ CD45RA⁺ CD10⁻), common myeloid progenitors (CMP; Lin⁻ CD34⁺ CD38⁺ CD123⁺ CD45RA⁻ CD10⁻), MEPs (Lin⁻ CD34⁺ CD38⁺ CD123⁻ CD45RA⁻ CD10⁻), and GMPs (Lin⁻ CD34⁺ CD38⁺ CD123⁻ CD45RA⁺ CD10⁻), CD4⁺ T-cell (CD4; CD3⁺ CD4⁺), CD8⁺ T-cell (CD8; CD3⁺ CD8⁺), B cell (B; CD19⁺ CD20⁺), natural killer cell (NK; CD56⁺), monocyte (Mono; CD14⁺), granulocyte (Gran), erythroid (Ery), megakaryocyte (Mega). **B**, Unsupervised hierarchical clustering of 5hmC data from all replicates of 12 normal hematopoietic cell types. Values shown are Pearson correlation coefficients. Clustering was performed on all peaks using the log₂(CPM+1) of the 5hmC reads mapped to each peak. **C**, Normalized 5hmC read depth across all gene bodies averaged across all normal hematopoietic cell types. An overall signal is scaled using the 50-kb regions located 50–100 kb upstream and downstream of each gene. **D**, Normalized 5hmC profiles at developmentally important genes and regions across all normal hematopoietic cell types. Each profile represents the average across all biological replicates for each cell type. $N \geq 4$ for all cell types. **E**, Heatmap of 5hmC gene scores across all normal hematopoietic samples. The top 500 genes that were significantly 5-hydroxymethylated in each cell type relative to others are shown. Genes were selected if they had a normalized 5hmC abundance that was significantly higher than all other cell types (*t* test adjusted for multiple hypothesis testing, $P_{\text{adjusted}} \leq 0.05$). Genes associated with hematopoietic development are highlighted. $N \geq 4$ for all cell types.



with transcriptional machinery during active transcription (Fig. 1C).

We characterized cell type-specific 5hmC differences at gene bodies and regulatory loci across the hematopoietic hierarchy (Fig. 1D). *HLF* and *GATA2*, hallmark genes abundant in immature HSPCs, were highly 5-hydroxymethylated in those cells. Differentiated monocytes, B cells, and T cells also contained specific 5hmC signatures including *CEBPB*, *BLK*, and *BCL11B*, respectively, genes previously demonstrated to be important in the development of these cell types. The *HOX* locus, a region vitally important in hematopoietic development, showed vast reorganization and overall reduction of 5hmC as cells differentiated. To quantify gene-specific 5hmC enrichment, we developed a “gene score” that normalized the number of 5hmC reads mapped to each gene locus across samples. These gene scores recapitulated the immunophenotype and transcriptional profile of these cells, with the 5hmC profiles of *HOX* family genes enriched in HSCs and MPPs, *MS4A1* and *CD37* enriched in B cells, and *CD14* and *CD36* enriched in monocytes, among other enrichments (Fig. 1E). These results demonstrate cell-type specificity of 5hmC at gene loci across hematopoiesis.

5hmC Is Associated with Open Chromatin and Active Transcription at Gene Bodies and Enhancers

We previously reported the transcriptional and chromatin accessibility landscape of the hematopoietic hierarchy (12), and given the cell type and gene-specific 5hmC associations described here, we performed multiomic analyses to integrate these data sets across the 12 cell types. We observed that 5hmC abundance at genes relevant in hematopoietic differentiation was positively correlated with both chromatin accessibility and transcription level (Fig. 2A). We found that 5hmC is depleted within open chromatin sites and enriched at the 5' and 3' ends of these sites (Fig. 2A and B). There are exceptions, where the 5hmC signal was not always associated with a transcriptionally active gene body. For example, *CEBPB* showed distal 5-hydroxymethylation that tracked with chromatin accessibility and transcript abundance during monocyte differentiation.

We interrogated transcription start sites across cell types and observed that 5hmC abundance is enriched in the 5' and 3' flanks of accessible TSSs (Fig. 2B). Additionally, 5hmC abundance was positively associated with transcript abundance in multiple cell types with R values ranging from 0.38 to 0.47 (Fig. 2C; Supplementary Fig. S3A and S3B; Supplementary Tables S1 and S2). 5hmC abundance also showed a dynamic positive association with transcript change as cells differentiated from their progenitor to terminally differentiated states, suggesting concordant 5hmC and transcript dynamics during hematopoiesis. Furthermore, important hematopoietic regulators were among the genes with the highest association between 5hmC and transcript fold change (Fig. 2C).

To evaluate 5hmC at genomic sites distal to gene bodies, we identified a set of 5677 putative active enhancer regions from CD34⁺ HSPCs (12, 32). 5hmC abundance was closely correlated with ATAC-seq signal at enhancers across all cell types and demonstrated dynamic association

during differentiation (Fig. 2D; Supplementary Fig. S3C and S3D). Generally, chromatin accessibility was more associated with gene expression than 5hmC deposition, consistent with its role in transcriptional regulation (Supplementary Fig. S3E and S3F). Analysis of gene bodies and enhancers showed that enhancers contained significantly greater 5hmC per kb than gene bodies or background regions (Supplementary Fig. S3G). These data demonstrate that 5hmC abundance is correlated with chromatin accessibility and gene expression at both gene and enhancer loci, uncovering intricacies underlying the epigenetic regulation of blood differentiation.

5hmC Gene Deposition Patterns Display Unique Transcriptional Dynamics

5hmC tended to accumulate at terminal ends of gene bodies (centrally depleted) or was consistently present across the entire gene (centrally enriched) across cell types. We classified genes into these two groups across all healthy hematopoietic samples (Supplementary Fig. S4A).

To determine the biological relevance of these two 5hmC signatures, we compared 5hmC abundance and gene expression between the groups. There was no difference in overall 5hmC abundance; however, genes with a centrally enriched 5hmC signature had significantly greater expression across hematopoietic samples (Supplementary Fig. S4B). Additionally, 5hmC abundance was more correlated with transcript expression in the centrally enriched genes relative to the centrally depleted genes across all hematopoietic cell types (Supplementary Fig. S4C–S4E).

Generation of *TET2* Knockout Primary Human HSPCs

TET2 is a key mediator of 5hmC production in hematopoietic cells and is commonly mutated in hematologic malignancies, preleukemia, and CH (9–12, 33). To model human *TET2*-mutated preleukemia/CH and *TET2* mutation-driven leukemogenesis, we used CRISPR/Cas9 to disrupt the *TET2* gene in primary human HSPCs. Two complementary approaches were utilized (Supplementary Table S3). The first used homology-directed repair (HDR) to simultaneously disrupt *TET2* and tag the locus with a fluorescent reporter (HDR method (Supplementary Fig. S5A–S5E), whereas the second used two sgRNAs to introduce a deletion into the *TET2* gene (Hit and Run method; Supplementary Fig. S5F and S5G). The cells edited using these methods were measured by flow cytometry or TIDE analysis (34), respectively. As a control in all experiments, we used the same approaches to target the *adeno-associated virus integration site 1* (*AAVS1*) locus (Supplementary Fig. S5A), a well-validated “safe harbor” in the human genome (35).

TET2 Knockout Human HSPCs Exhibit Increased Serial Replating and Defective Erythroid Differentiation *In Vitro*

The phenotypic effects of *TET2* disruption in human HSPCs were first evaluated *in vitro* (Supplementary Fig. S6). Proliferation was assessed by culturing CD34⁺ HSPCs for 3 to 4 weeks under conditions that can promote HSPC retention. With both HDR and Hit and Run methods, *TET2*

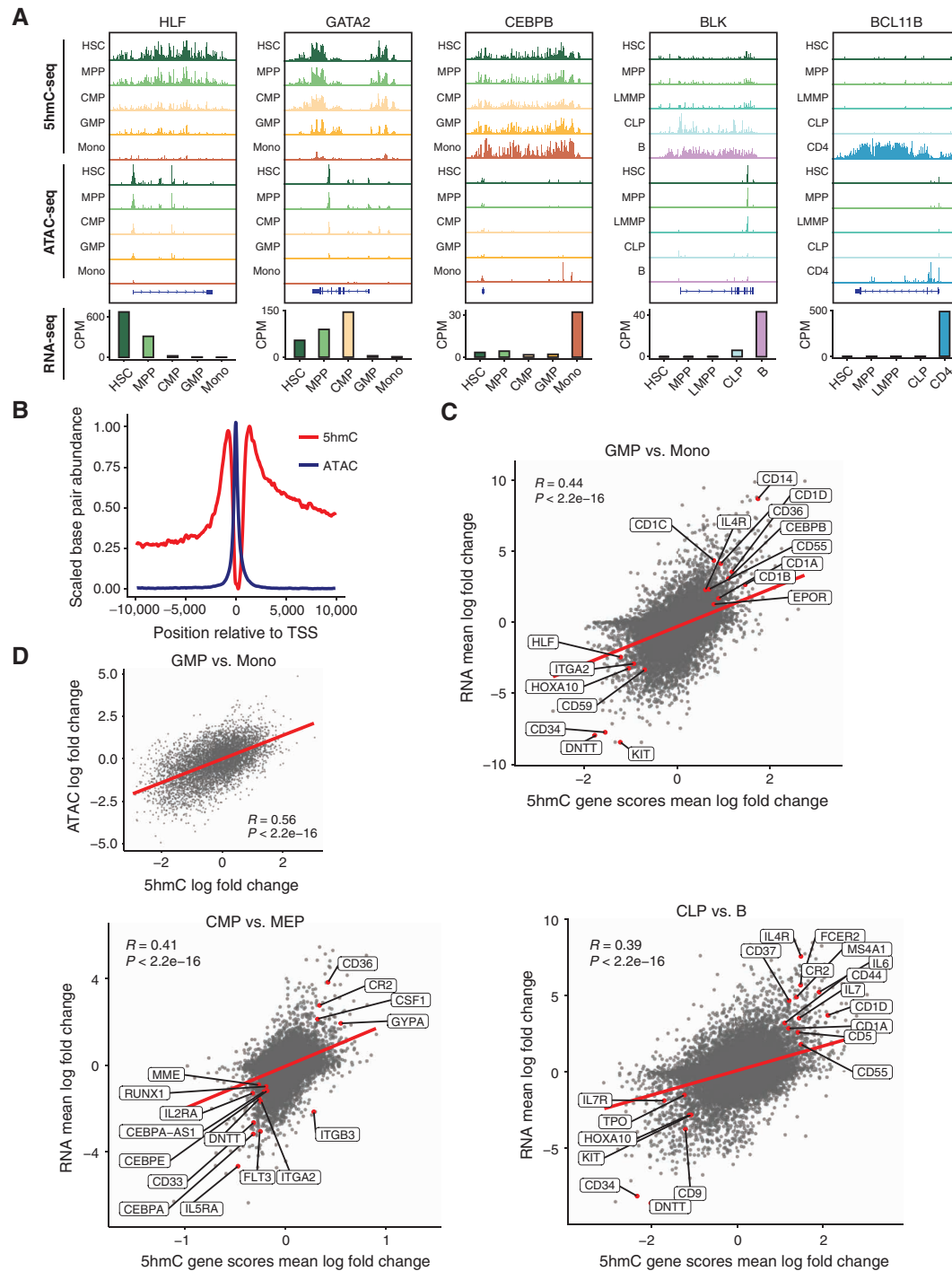


Figure 2. 5-Hydroxymethylation association with chromatin accessibility and transcriptome data. **A**, Normalized 5hmC profiles, ATAC-seq profiles, and transcript abundance at developmentally important genes in select differentiation trajectories. 5hmC and ATAC-seq profiles represent the average of biological replicates. RNA-seq values represent the average TPM across all biological replicates. *HLF*, *GATA2*, and *CEBPB* are shown for the cell types in the HSC to Monocyte differentiation trajectory (HSC, MPP, CMP, GMP, and Mono). *BLK* profiles are shown for the HSC to B-cell trajectory (HSC, MPP, LMPP, CLP, B). *BCL11B* profiles are shown for the HSC to CD4⁺ T-cell trajectory (HSC, MPP, LMPP, CLP, and CD4). $N \geq 4$ for all 5hmC experiments, $N \geq 4$ for all ATAC experiments, and $N \geq 5$ for all RNA-seq experiments. **B**, Normalized 5hmC abundance and ATAC-seq abundance across all TSSs across all healthy hematopoietic samples. Both 5hmC and ATAC-seq values were scaled such that the minimum normalized read depth in each plot was 0, and the maximum normalized read depth in each plot was 1. **C**, Scatter plot comparing the log₂ fold change of 5hmC gene scores (x-axis) and log₂ fold change of transcript abundance (y-axis) across all genes. Fold changes are calculated by comparing the normalized read depth in monocytes to GMPs (GMP vs. Mono), MEPs to CMPs (CMP vs. MEP), and B cells to CLPs (CLP vs. B). Developmentally relevant genes that showed high concordance between 5hmC change and transcript change during these differentiation transitions are highlighted. **D**, Scatter plot comparing the log₂ fold change of 5hmC read counts (x-axis) and log₂ fold change of Tn5 insertions (ATAC-seq, y-axis) across all enhancer regions. Fold change is calculated by comparing normalized signals between monocytes and GMPs.

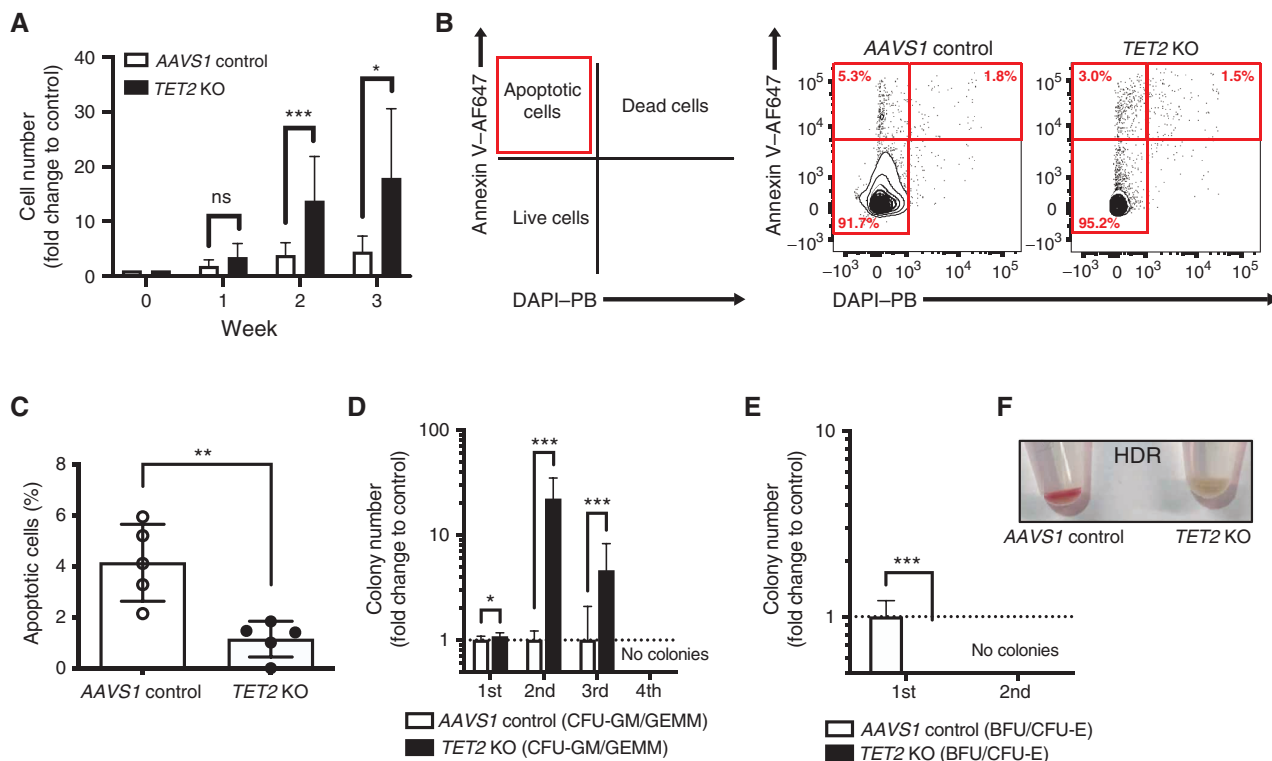


Figure 3. *TET2* KO increases serial replating of human HSPCs and blocks erythroid-megakaryocytic differentiation *in vitro*. **A**, Cell proliferation assay of AAVS1 control and *TET2* KO cells in the HSPC-expansion medium. Absolute HSPC counts were determined by flow cytometry every 7 days using counting beads. The results show fold changes normalized to control on day 0. AAVS1 control, $n = 10$; *TET2* KO, $n = 12$, all from different donors. **B**, Representative flow plots for analysis of apoptosis of AAVS1 control and *TET2* KO cells. Cells were cultured in an HSPC-expansion medium and stained with Annexin V and DAPI. Gating strategy is shown in the far-left panel. **C**, Statistical analysis of apoptotic cells within AAVS1 control or *TET2* KO cells. $N = 5$, all from different donors; unpaired *t* test. **D** and **E**, Methylcellulose colony-forming assay of AAVS1 control or *TET2* KO cells. The number of myeloid (g) and erythroid (h) colonies were scored by morphology 12 to 16 days after plating, and cells were replated up to four cycles. The results show fold changes in the number of each colony normalized to control. $N = 4$, all from different donors; paired *t* test. **F**, Representative cell pellets from AAVS1 control or *TET2* KO cells after first plating from **D** and **E**. (continued on following page)

disruption increased the absolute cell numbers after more than 2 weeks of culture and decreased apoptosis (Fig. 3A–C; Supplementary Fig. S7A). Progenitor function was assessed by serial replating in methylcellulose. No significant differences in the number of CFU-GM/GEMM colonies were observed after the first plating compared with AAVS1 control; however, upon replating, there was a distinct advantage for *TET2* knockout (KO) cells (Fig. 3D; Supplementary Fig. S7B). Moreover, we observed a statistically significant decrease in BFU/CFU-E erythroid colonies from *TET2* KO cells consistent with previous reports of *TET2* KO-mediated granulocyte–macrophage bias (ref. 36; (Fig. 3E and F; Supplementary Fig. S7C).

We investigated erythroid and megakaryocytic differentiation using liquid culture assays (Supplementary Fig. S6). *TET2* KO or AAVS1-targeted control HSPCs were placed into culture under erythroid or megakaryocytic differentiation-promoting conditions. The number of CD71⁺GPA⁺ mature erythroblasts at 10 days was found to be significantly decreased in *TET2* KO cells (Fig. 3G and H; Supplementary Fig. S7D). Similarly, *TET2* KO cells exhibited decreased production of CD41a⁺CD61⁺ megakaryocytes (Fig. 3I and J). Consistent with previous studies (7, 36, 37), these results

indicate that the disruption of *TET2* leads to a proliferative and reduced apoptosis advantage, enhanced progenitor replating, and a block in erythroid and megakaryocytic differentiation of human HSPCs.

TET2* KO Induces a Competitive Advantage, Myeloid Skewing, and Chronic Myelomonocytic Leukemia-like Cells *In Vivo

To explore the effect of *TET2* KO on human hematopoiesis *in vivo*, we transplanted Hit and Run *TET2* KO CD34⁺ HSPCs into NOD/SCID/IL2RG-null (NSG) mice (Supplementary Fig. S5F and S5G; Supplementary Table S3) and analyzed both primary and secondary transplants. BM aspirates were analyzed for engraftment on weeks 16 to 18 for primary transplants and weeks 12 and 36 for secondary transplants (Fig. 4A). Analysis of human CD45⁺ engraftment in the BM from mice transplanted with *TET2* KO cells showed that nearly all engrafted cells were *TET2* KO. Notably, there were no statistical differences in human BM chimerism levels in *TET2* KO HSPC-engrafted mice after primary (Fig. 4B) and secondary (Fig. 4C) transplantations compared with AAVS1-targeted control cells. Moreover, there were no statistically significant differences in CD34⁺CD38[−] immature HSPCs, spleen size, or

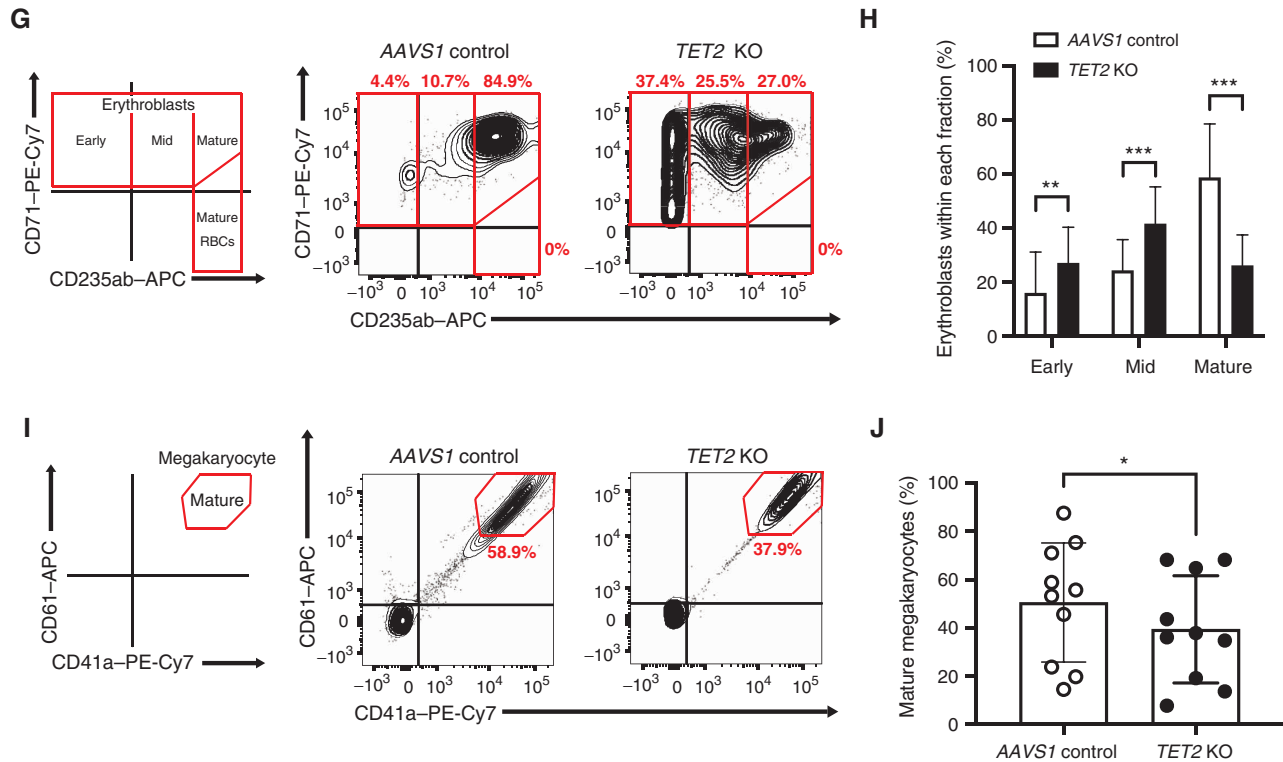


Figure 3. (Continued) **G**, Representative flow plots for the erythroid differentiation assay of AAVS1 control or TET2 KO cells. Cells were cultured in erythroid differentiation media for 10 days, and then stained with CD71 and CD235a (GPA) as erythroblast markers. The gating strategy is shown in the far-left panel. **H**, Statistical analysis of each erythroblast fraction within AAVS1 control or TET2 KO cells. $N = 14$, all from different donors; paired t test. **I**, Representative flow plots of the megakaryocytic differentiation assay of AAVS1 control and TET2 KO cells. Cells were cultured in megakaryocytic differentiation media for 10 days, and then stained with CD41a and CD61 as mature megakaryocyte markers. Gating strategy is shown in the far-left panel. **J**, Statistical analysis of the proportion of mature megakaryocytes from **I**. $N = 10$, all from different donors; paired t test. All the cells were edited by the biallelic HDR method and used for the experiments. ns = not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. Error bars, \pm SD.

complete blood count after 12, 36, and 36 weeks of transplantation, respectively (Supplementary Fig. S8A–S8D).

Lineage analysis of engrafted cells showed that mice transplanted with TET2 KO cells exhibited myeloid-skewed engraftment after 36 weeks of secondary transplantation compared with AAVS1 control and TET2 KO primary transplants (Fig. 4D–G). Furthermore, the proportion of TET2 KO cells, as measured by TIDE analysis, increased over time, indicating that TET2 KO cells had a competitive advantage (Fig. 4H; Supplementary Fig. S8E and S8F).

Chronic myelomonocytic leukemia (CMML) is a myeloproliferative overlap syndrome with a high frequency (>65%) of TET2 mutations (7, 38) and stable karyotype. CMML often presents with a characteristic increase in the fraction of CD14⁺CD16⁻ classic monocytes (>94.0% diagnostic of CMML; ref. 39). We performed a phenotypic analysis of CD33⁺ cells isolated from mice engrafted with TET2 KO cells (Supplementary Fig. S8G). We detected a high proportion of human CD14⁺CD16⁻ CMML-like cells in these mice, with a trend toward an increase compared with AAVS1 control (Supplementary Fig. S8H). This definition was based solely on the immunophenotype of the cells, and CBCs and spleen size were not evaluated in these mice. This suggests that TET2 KO HSPCs may give rise to a CMML-like phenotype over time in transplanted mice.

TET2 KO Affects Hematopoiesis in a Cell Context-Dependent Manner in Immature HSPCs

TET2 mutations are found in preleukemic/CH HSCs (11), suggesting that TET2 mutations act in a cell context-dependent manner. To investigate this, six human HSPC subpopulations from human cord blood were sorted by FACS, subjected to HDR TET2 KO, and cultured *in vitro* under progenitor-retaining or differentiation-promoting conditions (Supplementary Fig. S6). After 10 days of culture, CD34⁺ cell retention and an erythroid-megakaryocyte differentiation block were observed with TET2 KO HSCs and MPPs, but not with more committed progenitors (Supplementary Fig. S9A–S9D). Notably, these effects were not observed with TET2 KO in MEPs. These results indicate that the effect of TET2 KO is dependent on the cell type within which it occurs and that these effects are restricted to the most immature human HSPCs.

TET2 KO Results in Global and Gene-Specific Depletion of 5hmC Which Is Associated with Aberrant Transcriptional Output

We next assessed global 5hmC and transcriptional dynamics in TET2 KO-engineered cells (Supplementary Fig. S10).

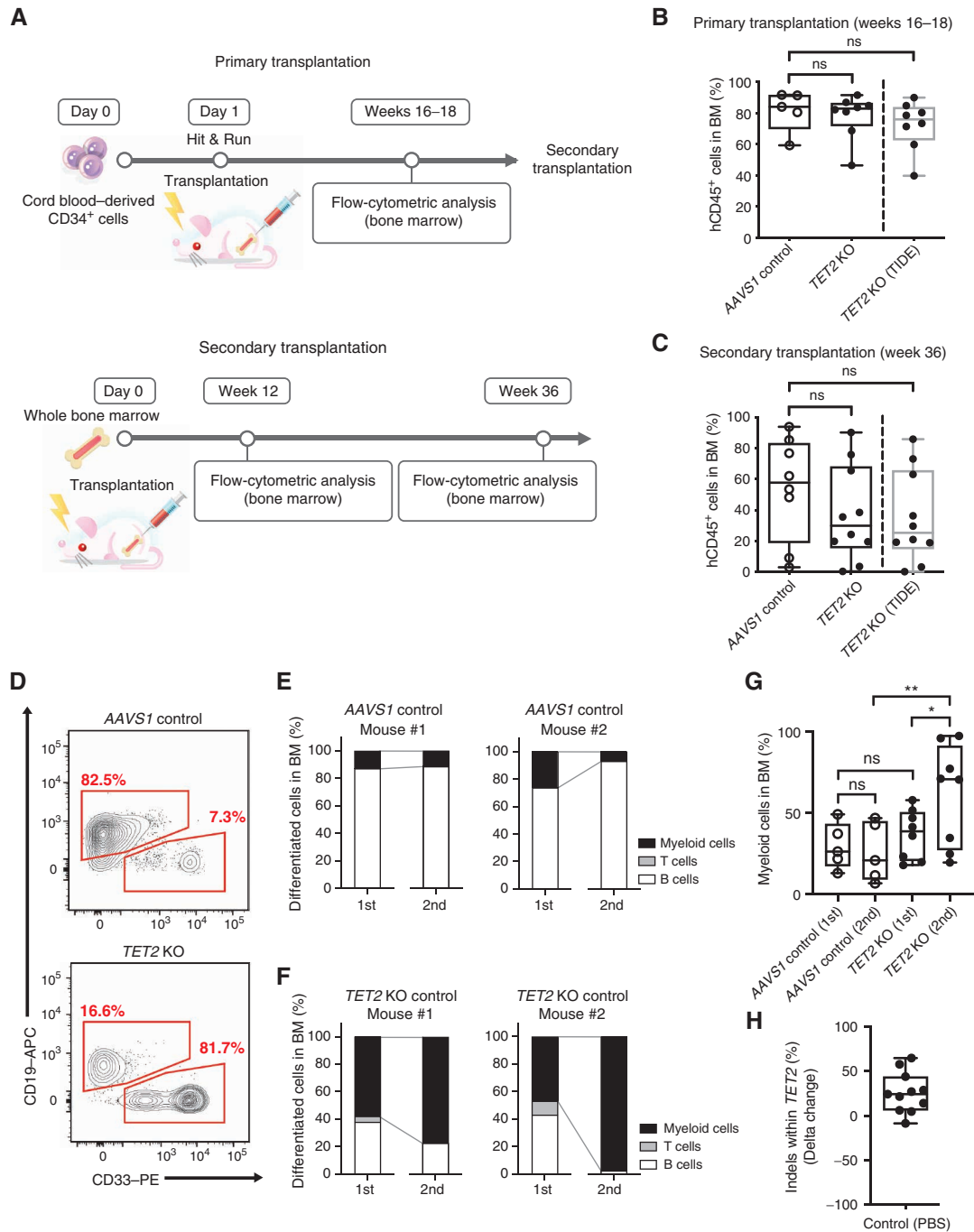


Figure 4. *TET2* KO promotes the myeloid skewing of human HSPCs *in vivo*. **A**, Experimental schematic and timeline of primary transplantation and secondary transplantation. Whole BM cells from primary transplantation NSG mice were transplanted intrafemorally into NSG secondary recipient mice. Flow-cytometric analyses were conducted on weeks 16–18 for primary transplantation mice and weeks 12 and 36 for secondary transplantation mice from BM aspirates. **B** and **C**, Summary of human engraftment chimerism levels in mouse BM after the primary (**B**) and secondary (**C**) transplantation. Dots represent human CD45⁺ cells in the BM of individual engrafted mice at the indicated time points after transplantation. Far-right bar [*TET2* KO (TIDE)] indicates the engraftment of human CD45⁺ cells edited at the *TET2* locus determined by TIDE analysis. Primary, AAVS1 control, *n* = 5; *TET2* KO, *n* = 8; secondary, AAVS1 control, *n* = 8; *TET2* KO, *n* = 10; unpaired *t* test. **D**, Representative flow plots of engrafted human CD19⁺ and CD33⁺ human cells from mice transplanted with AAVS1 control (top) or *TET2* KO (bottom) cells after 36 weeks of secondary transplantation. **E** and **F**, Proportion of engrafted human mature cells (CD19⁺ B cells, CD33⁺ myeloid cells, CD3⁺ T cells) in BM of two representative mice after 36 weeks of secondary transplantation. BMs were analyzed by flow cytometry from AAVS1 control (**E**) or *TET2* KO (**F**) transplanted mice. **G**, Statistical analysis of engrafted human CD33⁺ myeloid cells in BM from the transplanted mice. Mice engrafted >0.5% with human CD45⁺ cells, AAVS1 control, *n* = 5; *TET2* KO, *n* = 8; paired *t* test. **H**, Summary of the delta change of *TET2* KO cells in mouse BM. Mice were intrafemorally transplanted with a five-to-one ratio of AAVS1 control and *TET2* KO HSPCs, and 3 months later, these mice were treated with PBS (control). Before and after treatment, we analyzed the proportion of *TET2*-disrupted CD45⁺ cells in the mouse BM by TIDE analysis. *N* = 11. All the cells were edited by the Hit and Run method and used for the experiments. ns = not significant; *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001; ****, *P* < 0.0001. Error bars, ± SD.

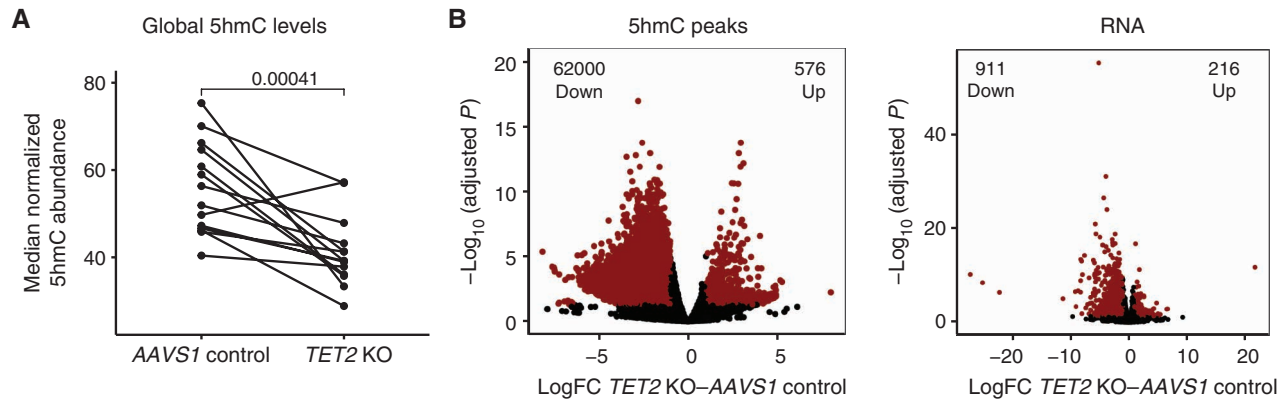


Figure 5. The 5hmC landscape of *TET2*-mutated HSPCs. **A**, Median 5hmC gene scores were determined after normalizing to the WGS background control in *AAVS1* control and *TET2* KO samples (see Methods for global 5hmC normalization). Each dot represents an individual donor. Samples are paired by their experimental condition and biological replicate. Pairs are connected by lines. Paired *t* test, $P = 0.00041$. **B**, left, volcano plot of differential 5hmC abundance in *TET2* KO samples compared with *AAVS1* control samples. Significant peaks (DESeq2 adjusted $P \leq 0.05$) with an absolute \log_2 fold change of at least 1 are highlighted in red and enumerated. DESeq2 was used to determine 5hmC fold change and significance for each peak (paired test corrected for multiple hypothesis testing, global 5hmC scaling factors applied in workflow). See Methods for full differential testing details. Right, volcano plot of RNA-seq data from *TET2* KO compared with *AAVS1* control samples. DESeq2 was utilized to calculate \log_2 fold changes and significance for each gene (paired test corrected for multiple hypothesis testing). Significant genes (adjusted $P \leq 0.05$) with absolute \log_2 fold change greater than 1 are highlighted and enumerated. (continued on next page)

Normalized 5hmC abundance was globally depleted across all *TET2* KO samples (Fig. 5A). Differential analysis of 5hmC abundance across all consensus peaks revealed that the majority of differential loci showed a decrease in 5hmC abundance and that a limited number of loci had increased 5hmC abundance (Fig. 5B, left). RNA sequencing (RNA-seq) on the same biological conditions revealed that the majority of differentially expressed transcripts were downregulated in *TET2* KO cells relative to *AAVS1* controls (Supplementary Fig. S10; Fig. 5B, right), implying a causal link between *TET2*-mediated 5hmC present at gene loci and activation of transcription.

A number of erythroid differentiation-associated genes including *GATA1*, *GYP A*, *HBA1*, *HBA2*, and *KLF1* were depleted of 5hmC and had decreased expression in *TET2* KO cells with *GATA1* and *KLF1* being among the most transcriptionally depleted genes upon *TET2* disruption (Fig. 5C and D). Evaluation of 5hmC fold change within consensus peaks and the associated transcript change of their nearest genes indicated that the majority of peak-gene pairs that were significantly differential upon *TET2* KO were both decreased in 5hmC and transcript abundance (Fig. 5E). The select few peak-gene pairs that displayed paradoxically elevated 5hmC but decreased expression may be attributable to imperfect peak-gene linkages and/or 5hmC associations with repressive elements. These data demonstrate that *TET2* disruption results in aberrant 5hmC and transcriptional signatures globally, as well as at key erythroid genes, linking epigenetic defects to the observed erythroid differentiation block in these cells.

We additionally measured 5hmC and 5mC as a percentage of total DNA in edited HSPCs over time. This analysis indicated that in *TET2* KO HSPCs, 5hmC became depleted in a time-dependent manner and that 5mC initially decreased in abundance but accumulated to levels surpassing control cells after 10 days (Supplementary Fig. S11A). Additionally, 5hmC profiles tended to be more maintained during steady-state

growth relative to perturbed growth as determined by comparing 5hmC profiles before and after serial replatings (Supplementary Fig. S11B).

To determine the degree to which 5hmC abundance in *TET2*-mutant hematopoietic cells could be due to 5hmC that is part of active demethylation, we compared differential 5hmC regions in *TET2* KO HSPCs to differentially 5mC-methylated regions (DMR) derived from *TET2*-mutant and *TET2* wild-type human AML samples. A small percentage of all 5hmC peaks (<1%) overlapped with DMRs and a significantly higher proportion of differential 5hmC peaks overlapped with DMRs (Supplementary Fig. S11C). This suggests that there is a low, but significant and biologically meaningful association, between 5hmC and 5mC changes in *TET2*-mutant cells, but that this does not fully explain overall 5hmC dynamics. Of these overlaps, hypermethylated regions contained less 5hmC relative to hypomethylated regions, suggesting an inverse association between 5hmC and 5mC change in *TET2*-mutant samples, consistent with the idea that *TET2* mutations interfere with active demethylation, leading to regions of hypermethylation and reduced 5hmC (Supplementary Fig. S11D).

Gene set enrichment analysis of RNA-seq data derived from *TET2* KO and *AAVS1* control cells showed that there were common patterns of pathway dysregulation. There was a depletion of transcriptional programs associated with mature hematopoietic cells, megakaryocyte and erythroid differentiation, and apoptosis in the *TET2* KO cells, consistent with the observed differentiation block and stemness phenotypes (Supplementary Fig. S11E). Most evaluated gene sets were downregulated, consistent with the overall depletion of transcripts in *TET2* KO HSPCs (Supplementary Fig. S11F). Overall, these findings link *TET2* deficiency to aberrant gene-expression programs through modulation of 5hmC at 5hmC loci and gene bodies, thereby promoting the blocked erythroid differentiation and increased stem cell activity observed in preleukemic/CH cells.

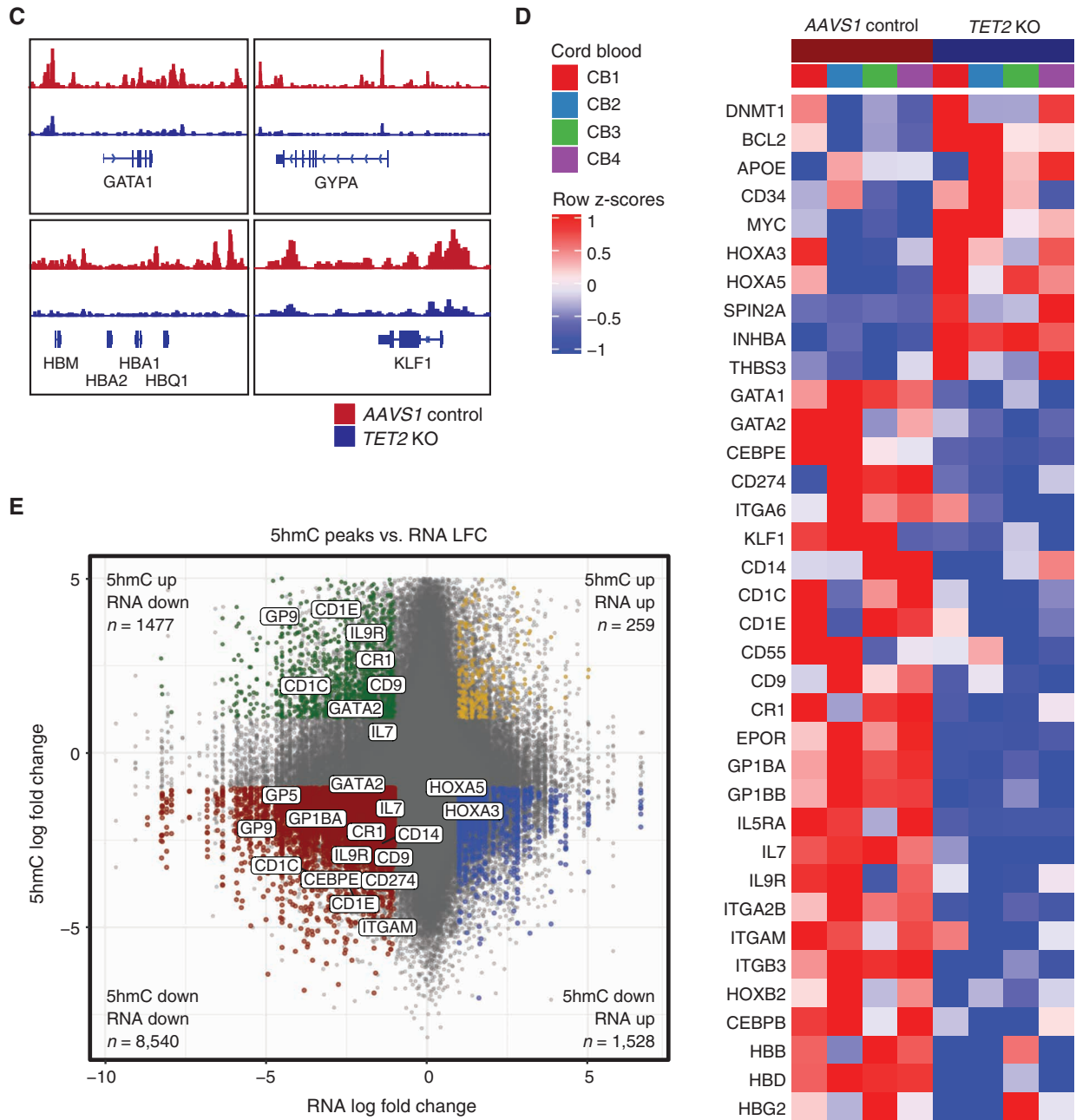


Figure 5. (Continued) **C**, Normalized 5hmC profiles at *AAVS1* control and *TET2* KO cells at select gene loci. *AAVS1* control and *TET2* KO scales are identical for each gene. Tracks are normalized based on 5hmC normalization factors, which are derived using the 5hmC enrichment within low-coverage regions compared with the input control sequencing sample. **D**, Heatmap of transcripts related to hematopoietic development that were determined to be significantly differentially expressed after multiple hypothesis correction between *AAVS1* control and *TET2* KO samples. Values represent gene z-scores of TPM values. **E**, Scatter plot of 5hmC peak \log_2 fold change and transcript \log_2 fold change when comparing *TET2* KO samples to *AAVS1* control samples. \log_2 fold changes for each gene (RNA-seq) and peak (5hmC) were derived from the DESeq output from **B**. Peaks were annotated using a nearest-gene approach and, therefore, there are multiple peaks per gene. Points were highlighted if the gene and associated 5hmC peak were significantly differential in both 5hmC and transcript. The number of significant peak-gene pairs in each quadrant are displayed. Significant genes related to hematopoietic development are highlighted. All the cells were edited by the biallelic HDR method. GFP and mCherry double-positive cells were sorted by FACS and used for each analysis.

Depleted Transcription Factor 5-Hydroxymethylation and Expression Are Associated with Aberrant Differentiation Phenotypes in *TET2* KO HSPCs

Following *TET2* KO in HSPCs, a select few transcription factors (TF) were depleted in both 5hmC and transcript

abundance relative to *AAVS1* control cells. These included *GATA1*, *GATA2*, *MAX*, and *PBX1* (Supplementary Fig. S12A). We hypothesized that changes in the 5hmC profile and expression of these TFs contribute to the erythroid differentiation block and myeloid skewing present in the *TET2* KO cells.

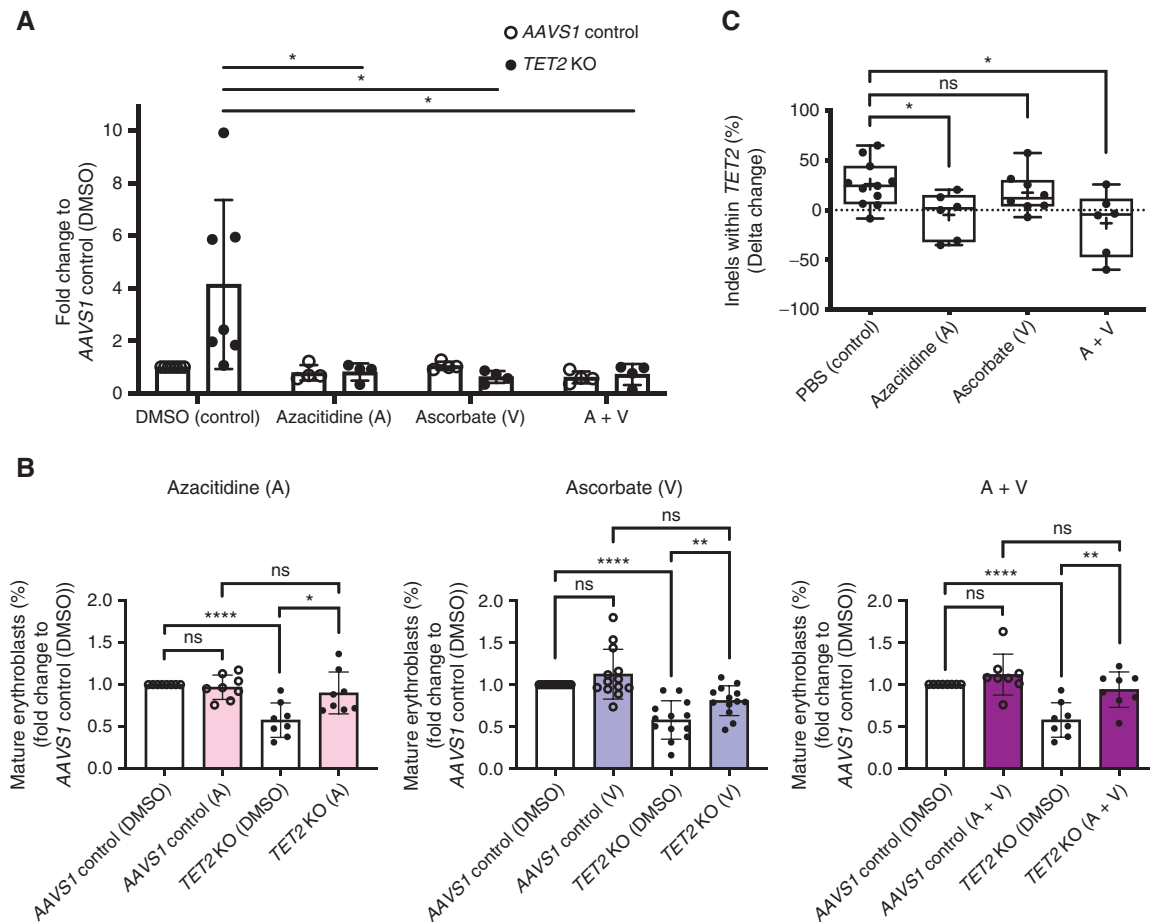


Figure 6. Azacitidine and ascorbate suppress the outgrowth of *TET2* KO cells *in vitro* and *in vivo*. **A**, *In vitro* cell proliferation assay of AAVS1 control and *TET2* KO cells with various indicated treatments (HDR method). Absolute HSPC counts after 10 days of culture in an HSPC-expansion medium were determined by flow cytometry using counting beads and normalized to AAVS1 control dimethyl sulfoxide (DMSO)-treated control. DMSO, $n = 6$; azacitidine (A), $n = 4$; ascorbate (V), $n = 4$; A + V, $n = 4$; all from different donors; paired *t* test. **B**, *In vitro* erythroid differentiation block rescue experiments with drug treatments (HDR method). Cells were cultured in erythroid differentiation media with the indicated treatments for 10 days and stained with CD71 and CD235a (GPA) as mature erythroblast markers. The percentage of mature erythroblasts was measured in AAVS1 control and *TET2* KO cell cultures, and normalized to DMSO control. Azacitidine (A), $n = 8$; ascorbate (V), $n = 13$; A + V, $n = 8$; all from different donors; one-way ANOVA. **C**, Summary of the delta change of indels within the *TET2* locus in bulk human CD45⁺ cells isolated from mouse BM before and after each treatment as determined by TIDE analysis. PBS, $n = 11$; azacitidine (A), $n = 6$; ascorbate (V), $n = 8$; A + V, $n = 6$; all from different donors; one-way ANOVA. The plus (+) symbol in each box depicts the average delta value. All the cells were edited by the Hit and Run method for *in vivo* and the biallelic HDR method for *in vitro* experiments. ns = not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. Error bars, \pm SD.

To test this hypothesis, we utilized our HDR editing methodology to simultaneously KO *TET2* and insert an overexpression cassette for one or multiple of these TFs. Notably, overexpression of *GATA1*, *GATA2*, *GATA1+GATA2*, or *PBX1* restored erythroid differentiation to levels significantly higher than the *TET2* KO HSPCs (Supplementary Fig. S12B and S12C). These results suggest that the 5hmC-associated decrease in expression of these TFs contributes to the erythroid differentiation block in *TET2* KO cells.

The *In Vivo* Competitive Advantage of *TET2* KO HSPCs Is Suppressed by Azacitidine and Ascorbate

Several pharmacologic agents have been reported to be active against *TET2*-mutant hematologic malignancies or have direct effects to increase *TET2* activity including azacitidine (30) and ascorbate (29), but have not been tested in

molecularly directed clinical trials of CH, CMML, or AML. Here, we examined the effects of these compounds on the engineered *TET2* KO HSPC phenotypes. First, we examined the effects on *in vitro* proliferation by culturing bulk CD34⁺ HSPCs edited by HDR in HSPC-expansion medium. When treated with either azacitidine (A), ascorbate (V), or both (A + V), *TET2* KO cell proliferation was significantly decreased compared with untreated *TET2* KO controls (Fig. 6A). At the doses used, the agents had little to no cytotoxicity to AAVS1 control cells. Next, we investigated the effect of these agents on erythroid differentiation and found that azacitidine, ascorbate, and the combination of these two agents (A + V) were able to reverse the erythroid differentiation block imparted by *TET2* deficiency (Fig. 6B).

To determine if these agents negated the competitive advantage of *TET2* KO cells *in vivo*, we conducted competition studies with azacitidine, ascorbate, or A + V compared

with control (Fig. 6C). NSG mice were transplanted intrafemorally with a five-to-one ratio of *AAVS1* control and *TET2* KO HSPCs, and 16 weeks later, these mice were treated with azacitidine (2.5 mg/kg/dose, i.p.), ascorbate (4 g/kg/dose, i.p.), A + V, or PBS control for 6 additional weeks. We used TIDE analysis to determine the proportion of *TET2* KO indels in human CD45⁺ cells isolated from the mouse BM before and after treatment (Fig. 6C). In control mice, the average percentage of *TET2* KO cells increased from 38.3% to 64.3% over the course of the 6-week treatment (average delta change of 26.0%), indicating that *TET2* KO cells had a competitive advantage over *AAVS1* control cells. Interestingly, azacitidine and A + V, but not ascorbate alone, reverted the expansion of *TET2* KO cells relative to PBS control mice (average delta change -4.8% with azacitidine and -13.2% with A + V).

5hmC Depletion Is Concentrated at Erythroid Loci but Myeloid Loci Are Spared in *TET2* KO Cells

Given the efficacy of ascorbate and azacitidine in reverting *TET2* KO phenotypes, we further evaluated 5hmC dynamics in untreated cells (Supplementary Fig. S13A, *in vitro*). We first identified transcriptional signatures for erythroid and myeloid cell types by generating gene sets with expression restricted to either HSCs, MPPs, CMPs, MEPs, erythrocytes, granulocyte-monocyte progenitors (GMPs), or monocytes, in addition to a set of randomly selected genes. Next, using 5hmC gene scores, we evaluated the overall change in 5hmC at these gene loci following *TET2* disruption. Erythroid-associated genes were significantly more depleted in 5hmC compared with randomly selected genes. Furthermore, 5hmC was spared at myeloid (GMP and monocyte) loci with an equal or greater abundance of 5hmC at these genes compared with randomly selected genes (Fig. 7A).

We evaluated changes in 5hmC at enhancer sites in *TET2* KO cells and performed TF motif enrichment analysis by identifying motifs with increased abundance of 5hmC relative to controls within these enhancers (Fig. 7B). Consistent with the depletion of 5hmC at erythroid loci and spared 5hmC at myeloid loci, 5hmC was depleted at enhancer motifs

corresponding to *GATA* family genes and increased at *CEBP* family genes in *TET2* KO cells relative to control.

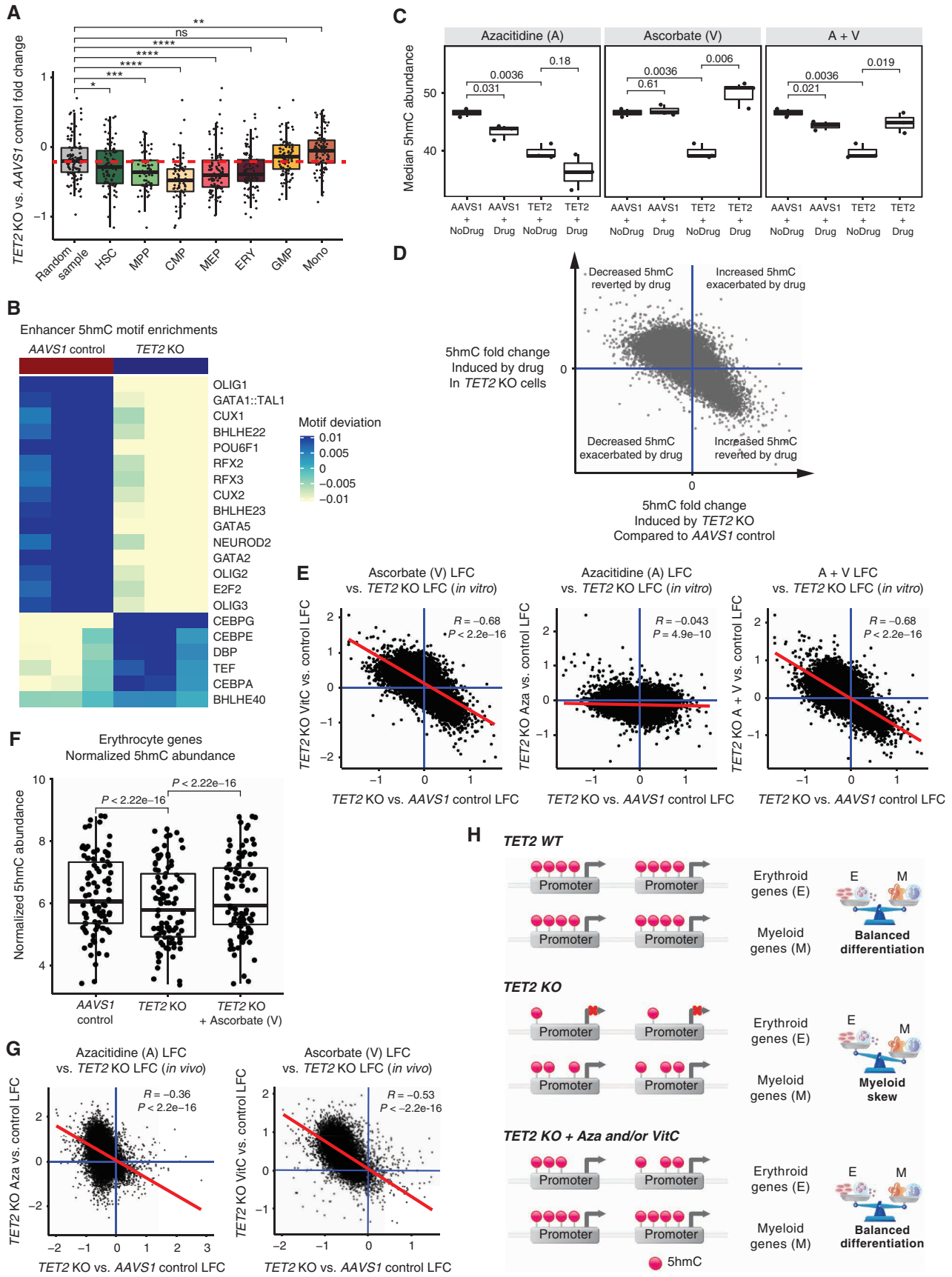
Global and Cell Type-Specific 5hmC Depletion in *TET2* KO Cells Is Reversed by Pharmacologic Agents

We next examined the effects of ascorbate and azacitidine on the genome-wide distribution of 5hmC. 5hmC-seq was performed on HSPCs either cultured *in vitro* or engrafted *in vivo* as described (Supplementary Fig. S13A). Globally, there was an increase of *in vitro* 5hmC levels following the addition of ascorbate, which was not observed in azacitidine-only treated cells (Fig. 7C). To evaluate 5hmC changes at specific gene loci, we calculated the 5hmC fold change at each gene in *TET2* KO cells compared with *AAVS1* control cells and compared these values to *TET2* KO cells following treatment with azacitidine (A), ascorbate (V), or A + V. We then used these changes to determine whether 5hmC abundance in treated *TET2* KO cells was reverted to levels consistent with *AAVS1* controls (Fig. 7D and E). In the *in vitro* studies, a significant proportion of genes in *TET2* KO cells had 5hmC reverted to levels matching the unmodified *AAVS1* control cells after having been treated with ascorbate or A + V (Fig. 7D and E). A number of key hematopoietic and erythroid-specific regulators including *HOX* and *GATA* family genes became rehydroxymethylated after pharmacologic treatment (Supplementary Fig. S13B). The lack of global 5hmC reversion with azacitidine may reflect the relatively short time course of these experiments.

At erythroid-associated loci, 5hmC levels were significantly and specifically restored following treatment of *TET2* KO cells with ascorbate (Fig. 7F). Myeloid loci in *TET2* KO cells also showed an increase in 5hmC following the addition of ascorbate (Supplementary Fig. S13C). 5hmC at *GATA1* and *GATA2* enhancer motif loci was also restored following the addition of ascorbate (Supplementary Fig. S13D). These data support a mechanism by which *TET2* KO erythroid differentiation is restored by ascorbate through the re-5-hydroxymethylation of erythroid-associated gene pathways.

We investigated 5hmC distribution in the *in vivo* experiments, which identified a more pronounced functional effect of azacitidine in reducing the competitive advantage of *TET2*

Figure 7. Azacitidine and ascorbate revert 5hmC profiles of *TET2* KO cells *in vitro* and *in vivo*. **A**, Box plot displaying the 5hmC gene score fold change between *AAVS1* control and *TET2* KO cells for gene sets representing cell type-specific gene-expression signatures. Gene sets represent genes that have expression enriched in the cell type of interest compared with all other cell types. A reference set of genes was derived by randomly selecting 100 genes from the entire genome without replacement. The dashed red line represents the mean 5hmC gene score fold change across all genes in *TET2* KO cells vs. *AAVS1* control. Two-way *t* test. **B**, Transcription factor motif deviations at enhancer regions as determined by ChromVAR. The 5hmC signal was used to determine differences in motif enrichment. Deviations for each biological cord blood replicate are scaled independently to control for between-cord effects. The top motifs by FDR-corrected *P* are plotted. **C**, *In vitro* median 5hmC abundance in *AAVS1* control DMSO, *AAVS1* control drug, *TET2* KO DMSO, and *TET2* KO drug conditions. From left to right, the drug conditions shown are azacitidine (A), ascorbate (V), and A + V. Three cords were used per condition, and a paired *t* test was performed controlling for between-cord differences. *N* = 3 for all conditions. **D**, Dot plot depicting drug treatment effects on gene-level 5hmC abundance in *TET2* KO cells. The 5hmC log₂ fold change for each gene in *TET2* KO cells compared with *AAVS1* control is shown on the x-axis. The 5hmC log₂ fold change in *TET2* KO cells after being treated with drug vs. DMSO is shown on the y-axis. Genes falling in the upper left quadrant had 5hmC levels decreased after *TET2* knockout and 5hmC levels increased after drug treatment. Genes in the lower right quadrant had 5hmC levels increased after *TET2* knockout and levels decreased after drug treatment. A negative trend on the dot plot indicates an overall reversion of *TET2* KO 5hmC profiles to *AAVS1* control levels upon drug treatment. **E**, *In vitro* 5hmC drug treatment dot plots for azacitidine (A), ascorbate (V), and A + V. *N* = 3 for all conditions. **F**, Normalized 5hmC abundance in *AAVS1* control, *TET2* KO, and *TET2* KO + ascorbate (V) conditions for all erythroid-associated genes. Values represent an average across all biological replicates (*n* = 3 for each condition). *P* values were derived using a paired *t* test. **G**, *In vivo* 5hmC drug treatment dot plots for azacitidine (A) and ascorbate (V). *N* = 2 for all conditions. **H**, Proposed mechanism for *TET2* KO mediated erythroid differentiation block, myeloid skew, and rescue by azacitidine (Aza) and/or ascorbate (VitC). Following *TET2* disruption, erythroid-associated genes are specifically depleted in 5hmC, corresponding to the decreased activity of these programs and aberrant differentiation. Treatment with Aza and/or VitC restores 5hmC and results in balanced cellular differentiation down both erythroid and myeloid trajectories. All the cells were edited by the Hit and Run method and used for the experiments. ns = not significant; *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001; ****, *P* < 0.0001. Error bars, ± SD.



KO cells (Fig. 6C). In these conditions, 5hmC was reverted to basal levels in *TET2* KO cells treated with azacitidine or ascorbate with hematopoietic differentiation-associated genes being among the top genes with restored 5hmC abundance (Fig. 7G; Supplementary Fig. S13E and S13F). GSEA of 5hmC gene scores indicated that hematopoiesis-associated genes that were depleted in 5hmC after *TET2* KO had levels restored after drug treatment (Supplementary Fig. S13G). There were some inconsistencies between 5hmC changes at TF gene bodies and 5hmC at TF binding sites. Notably, however, 5hmC at these features was not strongly correlated across TFs, indicating that more complex biological regulation underlies the 5hmC patterns at distal regions (Supplementary Fig. S13H). These data suggest that *in vivo*, azacitidine and ascorbate rescue the 5hmC effects of *TET2* KO, resulting in the restoration of normal HSPC gene expression, function, differentiation, and proliferation.

We sought to determine how ascorbate restores phenotypes in *TET2* KO cells. Prior studies have suggested a functional redundancy between *TET1* or *TET3* and *TET2* (40). Therefore, we assayed the expression of *TET1*, *TET2*, and *TET3* in *AAVS1* control and *TET2* KO cells following editing, as well as after treatment with pharmacologic agents. *TET1* and *TET3* were expressed in *TET2* KO cells at levels comparable with *AAVS1* control, whereas there was very low, but detectable, *TET2* expression in *TET2* KO cells, presumably due to small numbers of residual unedited cells (Supplementary Fig. S14A). In *TET2* KO cells, neither *TET1*, *TET3*, nor residual *TET2* expression was changed following 10 days of culture with ascorbate (Supplementary Fig. S14B). Based on these results, the mechanism by which ascorbate restores 5hmC in *TET2*-deficient cells may be through *TET1*, *TET3*, and/or residual *TET2*.

The *in vitro* and *in vivo* functional, drug treatment, and 5hmC studies of *TET2* KO HSPCs elucidate mechanisms by which *TET2* disruption causes cellular dysfunction in preleukemia. Loss of *TET2* function in these cells results in specific depletion of 5hmC at erythroid gene programs, which is associated with a decreased ability of these cells to differentiate into the erythroid lineage. Conversely, 5hmC at myeloid gene programs is spared, in line with the myeloid-skewed differentiation observed in the *TET2* KO cells. Treatment with ascorbate and/or azacitidine results in the restoration of 5hmC at these loci and the ultimate reversion of aberrant cellular phenotypes and differentiation (Fig. 7H).

DISCUSSION

Numerous studies have elucidated the importance of *TET2* in gene regulation, tumor suppression, and cell differentiation, particularly within the hematopoietic system. However, the role of 5hmC in cell differentiation and tumorigenesis, its genome-wide distribution and function as an epigenetic mark, and role in mediating *TET2*-mutant phenotypes, especially in human models, is poorly understood. Critically, it is not known if loss of 5hmC in *TET2*-mutant cells can be restored by pharmacologic agents active against these cells. In this study, we addressed these questions by creating a 5hmC genome-wide map of normal hematopoietic development and integrating these data with ATAC-seq

and RNA-seq to uncover associations between 5hmC, chromatin accessibility, and transcript abundance at both gene bodies and distal enhancers. We then engineered *TET2*-KO human HSPCs and demonstrated myeloid-skewed differentiation and competitive advantage associated with loss of 5hmC and increased stem cell gene programs. Finally, we demonstrated that azacitidine and ascorbate restored 5hmC and slowed or reverted the expansion of *TET2*-mutant clones. Our work provides key insights into the epigenetic regulation of hematopoietic development and *TET2*-mutant preleukemia/CH in human models and pinpoints agents that may be used to prevent disease progression in an epigenetic-dependent manner.

5hmC has been identified as an intermediate step in the demethylation of 5mC, initially suggesting that it passively affects epigenetic regulation through its effects on DNA methylation. However, recent studies suggest that 5hmC has an independent epigenetic and regulatory role (3, 24–26, 41). Eventually, it will be crucial to use 5hmC site-specific editing to assess the role of 5hmC in epigenetic regulation, as has been done with dCas9 and other epigenome editing techniques (42). Finally, further integration of genome-wide 5hmC and 5mC profiles will also be needed to assess their potential independent regulatory roles.

In our study, we observed that 5hmC features are highly cell type-specific and enriched at key hematopoietic regulators, TFs, and distal enhancers. Furthermore, the 5hmC signal was positively correlated with transcript abundance and chromatin accessibility within and between cell types. These patterns are consistent with a direct epigenetic regulatory role for 5hmC, in addition to its role as an intermediate in active demethylation. Based on our data, we speculate that 5hmC may facilitate active transcription through interactions with transcription machinery and other regulatory proteins. Recent studies have indicated that *TET2* facilitates TF recruitment in hematopoietic cells and our results suggest that 5hmC may be an element through which that recruitment occurs (43).

Here, we utilized two methods for CRISPR/Cas9-mediated genome editing of the endogenous *TET2* locus in primary human HSPCs and demonstrated that these cells exhibit the key features of preleukemia/CH. Namely, they exhibited myeloid-skewed differentiation and a competitive advantage in *in vivo* xenograft assays. Notably, *in vivo*, these cells adopted a CMML-like phenotype upon long-term engraftment and serial transplantation. These studies demonstrate the potential of engineering primary human HSPCs for disease modeling as a parallel approach to genetically engineered mouse models. Notably, the use of the HDR method with fluorescent reporters allows the tracking and isolation of populations of defined genotypes and has the potential to enable multiplexed editing of multiple loci (44).

Our models demonstrate two important distinctions between this work and prior studies (7, 45). First, human models of *TET2*-mutant CH develop at a much slower rate and have milder phenotypes than previously studied mouse models. This observation is consistent with data from human CH; individuals with *TET2* mutations only develop measurable CH over the course of years whereas mouse models develop CH phenotypes in a much shorter time span. Second,

our results highlight the importance of considering CH as a distinct entity compared with leukemia. *TET2*-mutant CH phenotypes are mild and take time to observe relative to leukemogenesis, and human models display distinct phenotypes relative to their murine counterparts.

Disruption of *TET2* in specific HSPC subpopulations revealed that the effects of *TET2* KO on *in vitro* differentiation are dependent on the cellular context and are restricted to the immature HSC/MPP compartment (Supplementary Fig. S9A–S9C). This is most striking in the case of erythroid differentiation where *TET2* KO in HSC/MPP, but not megakaryocyte-erythroid progenitors (MEP), results in a differentiation block. The underlying mechanisms for this specificity remain to be elucidated. Similar cell-context effects have been identified for cohesin mutations in HSPC differentiation (46), suggesting that the epigenetic context within which mutations occur can be key to resulting phenotypic effects. Regardless, these findings reveal the importance of the specific cell type subject to initiating mutations in preleukemia and CH.

TET2 KO HSPCs displayed reduction and redistribution of 5hmC levels, reduction in transcript levels, and disruption of hematopoietic pathways associated with the aberrant differentiation and growth phenotypes observed in those cells. These data suggest that *TET2* loss of function results in 5hmC loss at gene loci, in turn disrupting the transcription of genes driving normal cellular function. Notably, there were differences in 5hmC at key differentiation-associated motifs at enhancers in *TET2* KO cells compared with controls. These results exemplify the complex role of 5hmC in mediating *TET2*-associated phenotypes and raise the need for further studies into 5hmC's role in malignant hematopoiesis.

Our results showing that azacitidine and/or high-dose ascorbate suppress the growth and revert the phenotypes of human *TET2* KO cells build on previous work showing that ascorbate can restore *TET2* enzymatic function and prevent the expansion of *TET2*-mutant mouse cells (29). Recent work has shown that *TET2* inactivation impairs erythroid gene expression and spontaneous differentiation of the human erythroleukemia cell line TF-1. This phenotype was associated with 5hmC and 5mC dynamics at erythroid gene enhancers, and the phenotype of *TET2* KO was partially and transiently corrected by azacitidine exposure (47). In our work, we note that not only did ascorbate restore the phenotypes of *TET2* KO cells, but it reverted the 5hmC landscape of the cells to basal levels, linking the restoration of 5hmC to the functional effects of these pharmacologic agents. Notably, treatment of *TET2* KO cells with azacitidine alone or in combination with ascorbate significantly reduced the number of live cells compared with untreated *TET2* KO cells. In addition, these agents restored erythroid differentiation *in vitro*, reduced the competitive advantage of *TET2* KO cells *in vivo*, and restored locus-specific 5hmC *in vivo* (Fig. 6A–C and Fig. 7C–H). We attribute the observed effects of these agents to the ability to restore 5hmC landscapes to states similar to controls. We speculate that the 5hmC-restoring effects of azacitidine and ascorbate depend on *TET1* and/or *TET3* function and show that these proteins are expressed at levels comparable with unedited controls, indicating that they may be acted upon by azacitidine and/or ascorbate. Additionally, the lack of

in vitro efficacy of azacitidine may be due to the short-term nature of the *in vitro* assay and/or the ability of azacitidine to preferentially target more immature cells, sparing *in vitro* differentiated cultured cells.

Overall, this study charts the 5hmC landscape of normal human hematopoietic differentiation, characterizes human *TET2*-mutant phenotypes and the underlying 5hmC states, and reveals the possibility of treating *TET2*-mutated clones with epigenetically active drugs. Our observations exemplify that dysregulated HSPC-specific lineage programs contribute to disease development. Targeting these pathways with azacitidine and/or ascorbate may provide therapeutic avenues for patients with preleukemia/CH carrying *TET2* mutations.

METHODS

Healthy Human Samples

Normal donor human BM cells were obtained fresh from AllCells, and peripheral blood cells were obtained fresh from the Stanford Blood Center. All normal blood cell populations were sorted fresh. Human cord blood samples were collected with written informed consent from the mothers at delivery at the Lucile Packard Children's Hospital (the Binns Program for Cord Blood Research, Stanford IRB #33818) or purchased from the New York Blood Center and Stem-Express. Human BM samples (CD34⁺ cell enriched) were purchased from AllCells. Mononuclear cells from peripheral blood and cord blood were isolated by Ficoll (GE Healthcare) separation. Cord blood cells were resuspended in 90% FBS + 10% dimethyl sulfoxide (Sigma-Aldrich) and cryopreserved in liquid nitrogen for future use. All analyses conducted here on cord blood cells used freshly thawed cells.

Animal Care

All mouse experiments were conducted in accordance with a protocol approved by the Institutional Animal Care and Use Committee (Stanford Administrative Panel on Laboratory Animal Care #22264) and in adherence with the U.S. NIH's Guide for the Care and Use of Laboratory Animals.

Flow Cytometry/FACS

Cells were washed with FACS buffer (PBS, 2% FBS, 2 mmol/L EDTA) and stained with antibodies for 30 minutes on ice in a 50 μ L total volume. Cells were then washed and stained with propidium iodide (PI) at a final concentration of 1 μ g/mL or DAPI at a final concentration of 0.1 μ g/mL right before the analysis or sorting. All cell-sorting steps were validated using post-sort analyses to verify the purity of the sorted cell populations. All antibodies used for flow cytometry are detailed in Supplementary Tables S4–S7. Cell analysis was performed on a FACSCanto II (Becton Dickinson and Company (BD)), and cell sorting was performed on a FACSria II (BD).

Genomic DNA Extraction

All the samples were collected into 1.5 mL microcentrifuge tubes and snap-frozen until further processing. Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturer's instructions. Genomic DNA eluates were quantified using the Qubit dsDNA High-Sensitivity assay (Thermo Fisher Scientific) and stored at -20°C until further processing. Prior to sequencing library construction, genomic DNA was fragmented to a modal 150 base-pair size using an ME220 focused ultrasonicator (Covaris, Inc.). Modal fragmented DNA sizes were verified using the TapeStation 2200 dsDNA high-sensitivity assay (Agilent Technologies) and quantified as described above prior to commencing library construction.

5hmC Sequencing Library Preparation

Preparation of 5hmC-enriched sequencing libraries was performed as described previously (24). For each sample, two libraries were generated: a library derived from 5hmC-enriched reads and a library derived from input-unenriched DNA. Fragmented DNA was normalized to 20 ng total input for each assay and ligated to sequencing adapters. Adapter ligated products were either prepared for whole-genome sequencing using the KAPA Hyper Prep kit (Roche, Switzerland) according to the manufacturer's instructions or 5hmC bases were biotinylated via a two-step chemistry and subsequently enriched by binding to Dynabeads M270 Streptavidin (Thermo Fisher Scientific). All libraries were quantified using a Qubit dsDNA High-Sensitivity Assay (Thermo Fisher Scientific) and normalized in preparation for sequencing. Fragment size profiles of sequencing libraries were verified with the Bioanalyzer dsDNA High-Sensitivity assay (Agilent Technologies) prior to sequencing.

Bulk 5mC/5hmC Abundance Assays

5mC and 5hmC analyses in Supplementary Fig. S11A were quantified using the MethylFlash Global DNA Hydroxymethylation (5mC) ELISA Easy Kit (Epigentek) and the MethylFlash Global DNA Hydroxymethylation (5hmC) ELISA Easy Kit (Epigentek), respectively, according to the manufacturer's instructions.

DNA Sequencing and Preprocessing

Seventy-five base-pair, paired-end DNA sequencing was completed using a NextSeq550 instrument with NovaSeq Hi Output version 2.5 reagent kits (Illumina). Raw data processing and demultiplexing to sample-specific FASTQ output were performed using the Illumina BaseSpace Sequence Hub. Sequencing reads were aligned to the hg19 reference genome using default parameters of BWA-MEM. Duplicate reads were marked and removed using Picard MarkDuplicates. An overall schematic of bioinformatic pipeline steps is depicted in Supplementary Fig. S1C.

Peak Calling and Merging

Peaks were called using MACS2 using standard parameters and a *P* value cutoff of $1e-05$ (48). A consensus peak set of healthy blood cell type samples was generated by concatenating peaks of all samples into a single bed file and was merged using "bedtools merge." Peaks more than 500 bp apart were not merged, peaks that contained signals from less than three samples were removed, and peaks falling within black list regions (<https://sites.google.com/site/anshulkundaje/projects/blacklists>) were removed. This procedure resulted in a consensus peak set of ~570,000 peaks.

5hmC Gene Scores

We observed that across all cell types, the 5hmC signal was enriched within 10 kb of gene bodies. We, therefore, generated 5hmC gene scores for each gene by quantifying the amount of 5hmC signal falling within a region spanning 10 kb upstream of the TSS to 10 kb downstream of the TTS. In other words, a genomic region was defined for each gene by extending the gene body by 10 kb on either side. If the 10 kb extension ran into an adjacent gene, the region was truncated to exclude the adjacent gene. 5hmC signal was quantified within these regions by counting the number of reads falling into each region for each sample (see below).

Count Matrix Generation and Normalization

Count matrices were generated using either the healthy hematopoiesis consensus peakset or the extended gene bodies regions. In order to count reads falling into each genomic feature (either merged peak or extended gene body), bam files for each sample were converted to bed files using bedtools bamtobed. The number of reads in

each feature was then calculated using the CountOverlaps function from the Iranges package in R. Count matrices were then converted to CPM space using edgeR.

5hmC Sample Similarity

5hmC profile similarity between all healthy hematopoiesis samples was quantified by calculating the Pearson product-moment correlation between reads in consensus peaks. A technical and sample QC was performed by calculating between-sample similarities for all biological replicates. All cell types displayed high levels of concordance (Supplementary Fig. S1D). The effectiveness of 5hmC in discriminating between cell types was evaluated using a "cluster purity" metric that was calculated as described previously (12).

5hmC Abundance Plot

To quantify the 5hmC signal in regions within and flanking gene bodies, all genes across all samples were evaluated. 100 kb flanking regions for each gene were binned into 500 bp regions, and one hundred 500 bp regions were tiled along the gene body. This procedure resulted in a set of 500 bins per gene that captured the 5hmC abundance in gene body regions. The number of reads falling into each of these bins was quantified and normalized. This analysis was performed for each sample, and all samples were averaged together. The normalized abundance was averaged across all genes, smoothed using a spline function, and plotted.

5hmC Genome Tracks

BigWig files for each blood cell type were generated using the 5hmC data by sampling the bam files for all biological replicates to equal depth and merging them into a single bam file. Next, deepTools bamCoverage was used to convert the merged bam files to bigWig format. These bigWig files were then visualized in IGV.

Healthy Hematopoiesis ATAC-seq and RNA-seq Data

ATAC-seq and RNA-seq from healthy hematopoietic cell types were obtained from Corces and colleagues [ref. 12; Gene-Expression Omnibus (GEO) accession GSE75384].

5hmC and ATAC-seq TSS Plots

TSS plots for 5hmC and ATAC-seq data were generated by quantifying the insertions per base (ATAC-seq) or reads mapped per base (5hmC-seq) for all samples in 20 kb windows centered at all TSSs. Signal was averaged across all TSS and all samples, smoothed using a spline function, and plotted.

Enhancer Regions

To generate a set of genomic regions associated with active enhancers, we utilized ATAC-seq data from healthy hematopoietic cell types (ref. 12; GEO accession GSE75384) and H3K27ac ChIP-seq data from CD34⁺ HSPCs (ENCODE experiment ENCSR891KSP; ref. 32). The following steps were used to generate the enhancer region set. (i) We identified distal ATAC peaks (>2 kb distance to gene) and isolated peaks that overlapped with H3K27ac peaks. (ii) Regions from (i) that were separated by a distance of <5 kb were merged together. This step was performed because enhancer regions often spanned multiple ATAC-seq peaks. (iii) Regions from (ii) were extended by 1 kb in both directions in order to ensure the adequate capture of both 5hmC and ATAC signals at each locus. This procedure resulted in a set of 5,677 regions associated with active enhancers in HSPCs. The number of reads (5hmC) and the number of insertions (ATAC) within each region were counted for all samples and utilized for downstream analyses.

Two 5hmC Gene Body Signatures

Bins were defined at the center and flanks of each gene and quantified the normalized read depth in each bin across all samples. Genes

were categorized into “centrally enriched” and “centrally depleted” groups based on whether the ratio of central 5hmC to flanking 5hmC was above or below the median ratio across all genes (Supplementary Fig. S4A).

Culture Assays on HSPCs

CD34⁺ MACS (Miltenyi Biotech)-enriched HSPCs or CRISPR/Cas9-edited cells were plated into either HSPC-expansion medium [StemSpan SFEM II (STEMCELL Technologies) with 20 ng/mL SCF, TPO, FLT3L, and IL6, 35 nmol/L UM-171, and 500 nmol/L SR-1], HSPC retention medium (StemSpan SFEM II with 20 ng/mL SCF, TPO, and FLT3L), myeloid differentiation medium (Myelocult H5100 with 20 ng/mL SCF, TPO, FLT3L, IL3, IL6, G-CSF, GM-CSF, and 0.5 µg/mL hydrocortisone), erythroid differentiation medium (StemSpan SFEM II with Erythroid Expansion Supplement; STEMCELL Technologies), or megakaryocyte differentiation medium [StemSpan SFEM II with human low-density lipoproteins (LDL) (STEMCELL Technologies) and Megakaryocyte Expansion Supplement; STEMCELL Technologies]. Cells were cultured (37°C in 5% CO₂) for 10 days with medium changes as necessitated by cellular proliferation. See Supplementary Fig. S6 for the details of culture assays.

CRISPR/Cas9 Editing

To generate *TET2* KO in human HSPCs, we used CRISPR/Cas9 and recombinant adeno-associated virus 6 (rAAV6), following the methodology previously published (44). The HDR approach utilized an sgRNA targeting *TET2* exon 7, which encodes the first portion of the catalytic core domain (33), followed by transduction with rAAV6 encoding a GFP or mCherry expression cassette flanked by arms of homology, so that HDR of introduced dsDNA breaks led to the simultaneous disruption of the *TET2* gene and insertion of a trackable marker (Supplementary Fig. S5A and S5B). When both viruses were used, 10% to 30% of CD34⁺ cells underwent dual HDR, resulting in GFP and mCherry double-positive cells with the predominantly biallelic disruption of *TET2* (Supplementary Fig. S5C; ref. 44). “In-Out” PCR confirmed that the reporters were correctly integrated into the targeted gene locus (Supplementary Fig. S5E), and Western blot confirmed loss of TET2 protein in FACS-purified double-positive cells (Supplementary Fig. S5D). The second approach, termed “Hit and Run,” utilized two sgRNAs to introduce a 65 bp deletion in exon 3, the first coding exon of the *TET2* gene (Supplementary Fig. S5F). In this case, indel frequencies on both alleles were measured by TIDE analysis with an average of 95.8% ($R^2 \geq 0.94$; *in vivo* data; Supplementary Fig. S8H), and Western blot confirmed loss of TET2 protein (Supplementary Fig. S5G). See also Supplementary Table S3 for the details of the HDR and Hit and Run methods.

Single-guide RNAs (sgRNA) were designed to target *TET2*. sgRNAs were chemically modified with three terminal nucleotides at both the 5' and 3' ends containing 2' O-Methyl 3' phosphorothioate (Synthego) and precomplexed with purified Cas9 protein (Integrated DNA Technologies IDT) at a molar ratio of 1:2.5 (Cas9: sgRNA) at 25°C for 10 minutes immediately prior to use. Cord blood-derived CD34⁺ cells were expanded in the HSPC-expansion medium for 48 hours and then electroporated using the Lonza Nucleofector 4D (program #DZ-100) in P3 buffer and Cas9 ribonucleoprotein. AAV vector plasmids were cloned in the pAAV-MCS plasmid (Agilent Technologies) containing ITRs from AAV serotype 2 (AAV2). Vectors contained 400 bp homology arms flanking the CRISPR/Cas9 cut site and a reporter gene (either turboGFP or mCherry) driven by a SFFV (AAVS1 control) or a UBC (*TET2* KO) promoter followed by BGH polyA. AAV6 particles were produced in 293FT cells transfected using standard PEI transfection with ITR-containing plasmid and pDGM6 containing the AAV6 cap genes, AAV2 rep genes and adenovirus five helper genes, harvested after 72 hours, purified using the AAVpro

Purification Kit (Takara Bio Inc.), according to the manufacturer's instructions and then stored at -80°C until further use. Cells were cultured for an additional 3 days (including 6-hour incubation with AAV6 after electroporation), after which the GFP and mCherry double-positive population was sorted for both the AAVS1 control and *TET2* KO cells using a FACS Aria II (BD) with a post-sort analysis to verify the purity of >98%.

Overexpression of TFs in HSPCs

We used the same HDR method to overexpress *GATA1* (GFP), *GATA2* (mCherry), *MAX* (NGFR), and *PBX1* (KusabiraOrange) in HSPCs (Supplementary Fig. S12B and S12C). cDNA of each gene was inserted between polyA and RHA in the same construct we used in Supplementary Fig. S5B targeting the *TET2* locus.

Western Blot

Cells were lysed with standard RIPA buffer with protease inhibitors. Cell extracts were fractionated with SDS-polyacrylamide gel electrophoresis (NuPAGE 4%–12% Bis-Tris Gel, Invitrogen) and transferred onto 0.45-µm nitrocellulose membranes (Amersham Protan, GE Healthcare Life Sciences). Primary antibodies utilized were TET2 mouse monoclonal antibody C.15200179 (Novus Biologicals) and β-Actin mouse monoclonal antibody clone 8H10D10 (Cell Signaling Technology, Inc.). Primary antibodies were detected with horseradish peroxidase-conjugated secondary antibodies (anti-rabbit and anti-mouse, Cell Signaling Technology). Antibodies were detected using Clarity Western ECL Substrate (Bio-Rad).

Apoptosis Assay

FACS analysis for apoptosis was performed on cultured cord blood-derived CD34⁺ HSPCs. The cells were washed twice with annexin-binding buffer (10 mmol/L HEPES, 140 mmol/L NaCl, and 2.5 mmol/L CaCl₂, pH 7.4) and stained with Annexin V Alexa Fluor 647 (1:50; cat. no. A23204; Thermo Fisher Scientific) and DAPI at room temperature for 15 minutes.

Colony Assays

2,400–3,200 CRISPR/Cas9-edited CD34⁺ HSPCs (we sorted GFP and mCherry double-positive cells for HDR method) were added to 4 mL of MethoCult H4435 (STEMCELL Technologies) and plated in triplicate (600–800 sorted cells/plate). Plates were incubated at 37°C and 5% CO₂. The number of colonies in each sample was scored at 12 to 16 days to determine the numbers of CFU/BFU-Es and CFU-GEMMs/GMs. After counting the colonies, we harvested and replated the cells again with 100,000 cells/plate.

Mice

Six- to 8-week-old male or female NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice were used (Jackson Laboratory). All mice were housed in a pathogen-free animal facility in microisolator cages, and the experimental protocol was approved by Stanford University's Administrative Panel on Lab Animal Care (APLAC #22264).

In Vivo Experiments

One day after electroporation/transduction, Hit and Run-edited cells were transplanted into one femur of sublethally irradiated mice (200 rad, 2–24 hours before transplant). Mice were randomly assigned to each experimental group and analyzed in a blinded fashion. We used mouse BM aspirates from the transplanted femur for FACS analysis and sorting. For secondary transplantation, the human cell-engrafted mice were sacrificed, and all BM was harvested by crushing the bones. Nonspecific antibody binding was blocked, and cells were stained (30 minutes, 4°C, dark) with the engraftment antibody panel in Supplementary Table S5.

Microscopy

Aspirated mouse BM cells were resuspended at 20,000 cells/mL in 0.1% bovine serum albumin in PBS and cytospun onto slides. For Wright-Giemsa staining, slides were fixed in 100% methanol and subsequently stained with Hamacolor solution 2 [cat #1.11956 (red), Sigma-Aldrich], washed with DI water, and stained with Hamacolor solution 3 [cat. #1.11957 (blue), Sigma-Aldrich] before washing and mounting with a glass cover slip.

Drug Experiments

In Vitro. Azacitidine and ascorbate were purchased from Sigma Aldrich, dissolved in PBS (azacitidine) or distilled water (ascorbate), respectively, and serially diluted in a culture medium. The final concentrations for azacitidine and ascorbate were 250 nmol/L and 250 μ mol/L, respectively. Cells were seeded at 500 to 1,000 cells per well for the erythroid differentiation assay in 96-well culture plates with a final volume of 200 μ L culture media 24 hours prior to the addition of the first drug dose. Freshly prepared azacitidine or DMSO vehicle was added to cultures every 3 days for a total of three doses starting on day 1. Media were replaced on days 1, 4, and 7 before the analysis on day 10.

In Vivo. Azacitidine (#A2385) and ascorbate (#A7631) were purchased from Sigma Aldrich, freshly dissolved in PBS (azacitidine) or distilled water (ascorbate) at a concentration of 2.5 mg/kg and 4 g/kg, respectively, and administered once a day using i.p. injection. Drug treatments were initiated 3 months after the transplantation of engineered CD34⁺ HSPCs and confirmation of engraftment by flow cytometry of hCD45⁺/HLA-ABC⁺ staining of BM cells. Azacitidine was administered on days 1 to 5 of a 14-day cycle (total 3 cycles). Ascorbate was administered on days 1 to 5 of all cycles. Mice were treated for a total of 6 weeks.

Before and after the drug administration, sorted populations of cells (CD45⁺/HLA-ABC⁺) were analyzed for the *TET2* indel fraction using TIDE (Supplementary Table S3). The high dose of ascorbate should have sufficient serum levels in the blood with i.p. or oral gavage administration per prior reports (49).

TET2 5hmC Gene Scores Matrix

CRISPR/Cas9-edited samples were subjected to the same 5hmC sequencing and bioinformatic procedures as described above. A gene scores matrix was then generated containing the 5hmC abundance at each gene in each edited sample.

TET2 Global 5hmC Normalization

We utilized a methodology commonly implemented in ChIP-seq experiments to estimate the relative global 5hmC abundance in all CRISPR/Cas9-edited samples (50). For each sample, we derived a direct estimate of global 5hmC abundance or “normalization factor” using the sequencing data from both the 5hmC-enriched library and the control WGS library. Background regions of sequencing pulldown libraries (regions of no signal) have been shown to contain an approximately linear relationship with the same regions in the control sample (51). Therefore, the normalized signal in these regions can be used to estimate the relative enrichment of 5hmC signal compared with the background control in each sample. Because the background control library is unenriched and therefore contains the same signal across samples (proportional to total genomic DNA), an approach that estimates 5hmC enrichment relative to control libraries can be used to compare global 5hmC abundance between samples. To estimate the normalization factor or r , we divided the reference genome into nonoverlapping 10 kb bins. Let $n1i$ and $n2i$ be the CPM value in the i th bin for the 5hmC and control libraries, respectively. Let $ni = n1i + n2i$. We identified a set of background regions with no

5hmC enrichment for each sample by ranking bins by ni and selecting the lowest 1% of bins. These bins represent a sample-specific set of regions that contain no enrichment of 5hmC. In order to estimate r , we summed the reads in background regions for the 5hmC library and background library and calculated the ratio of 5hmC background signal to control background signal. This procedure resulted in a single r value for each sample, which was then used for downstream analysis. To correct CPM values for global 5hmC abundance in the edited samples, we multiplied the 5hmC library CPM values by the normalization factor for each sample. This normalization process is schematized in Supplementary Fig. S15.

TET2 5hmC Differential Analysis

We used DESeq2 to perform differential analysis of 5hmC read counts at each gene between *AAVS1*- and *TET2*-edited samples. We performed a paired differential test to control for differences across cords. In order to control for global 5hmC abundance, we used custom scaling factors in the DESeq2 normalization model. We replaced the scaling factors by multiplying the DESeq-derived library size scaling factors by our custom-derived global 5hmC normalization factors (see TET2 global 5hmC normalization). We utilized adjusted P values from the DESeq model to determine significance. Other than the aforementioned changes, we used standard DESeq2 parameters for all differential analyses.

RNA Extraction, Library Preparation, and Sequencing

Total RNA was extracted from CRISPR/Cas9-edited samples using the QIAGEN RNeasy micro kit following the manufacturer's protocol. Purified RNA was submitted to the Novogene commercial sequencing facility for Bioanalyzer quality control analysis, Illumina Next Generation Sequencing, and data quality control analysis. All submitted samples had an RNA integrity number (RIN) >8.0. Between 3.8 and 5.5 million paired-end sequence reads per sample were delivered from Novogene. Novogene also provided an evaluation of the sequencing quality (Phred score), sequencing error rate, A/T/G/C base distribution, and raw data filter; all samples were ascertained to be of high quality and low error before being released and used for analysis.

qPCR

RNA was isolated using the MiniPrep Plus Kit (Qiagen) and reverse transcribed using the SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific). TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific) was used to prepare the reaction mixture, and RNA levels were recorded on the QuantStudio 7 Flex Real-Time PCR Instrument (Applied Biosystems). TaqMan Primer-Probes were purchased from Thermo Fisher Scientific: GAPDH (Hs.PT.39a.222114836), TET1 (Hs.PT.58.27802060), TET2 (Hs.PT.58.26892089), TET3 (Hs.PT.58.4763348), GATA1 (Hs.PT.58.21050378), GATA2 (Hs00231119_m1), MAX (Hs.PT.58.50438546), and PBX1 (Hs.PT.58.18829984).

RNA-seq Analysis

FASTQ sequencing reads were quantified using kallisto pseudoalignment to the reference human genome hg38. The kallisto output was imported to DESeq2 for differential analysis where a paired differential test was performed to control for differences between cords. The R package tximport was used to derive transcripts per million (TPM) values from the kallisto output for subsequent analyses.

5hmC vs. DMR Analysis

DMRs were identified between *TET2*-mutant and wild-type samples using Illumina 450k array data from the TCGA LAML database. Next, the number of 5hmC regions that overlapped with DMRs was calculated for the global 5hmC peak set and for

5hmC peaks that were deemed to be differential between *TET2* KO and *AAVS1* control cells. To determine 5hmC changes associated with either hypermethylated or hypomethylated regions, 5hmC fold change was calculated for each peak that overlapped significantly with hypermethylated regions or hypomethylated regions in *TET2*-mutant samples.

Cell Type-Specific Gene Signatures

Cell type-specific gene-expression signatures were derived using differential gene-expression analysis with the Limma R package. RNA-seq data for sorted healthy and leukemia subpopulations were obtained from GSE74246¹². Raw counts were imported into the Limma pipeline and the standard workflow was performed. A contrast matrix comparing each cell type to all other cell types was constructed in order to identify the most differentially expressed genes. The final gene signature was determined by identifying the top 100 genes per cell type based on significance, and a log fold change of greater than three.

Enhancer Motif Enrichment

To perform the TF motif enrichment of the 5hmC signal at enhancers in *AAVS1* control and *TET2* KO cells, the 5hmC read counts enhancer matrix for *AAVS1* control and *TET2* KO samples was used as the input to ChromVAR (52). Motif deviations were calculated using standard parameters. Notably, each biological cord blood replicate was scaled independently in order to avoid between-cord effects. A *t* test was performed to identify significantly differential motifs between conditions. An adjusted *P* value threshold of 0.05 was utilized to filter nonsignificant motifs. Motif deviations for the top 20 significant, positively enriched and depleted motifs in *TET2* KO cells were plotted.

Statistical Analysis and Figure Generation

All bioinformatic and statistical analyses were performed in R version 4.0.1. For all *in vitro* and *in vivo* experiments, Prism8 (GraphPad Software) was used to perform statistical analysis. A paired/unpaired *t* test was used to define statistical significance (*, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001; ****, *P* < 0.0001). One-way ANOVA tests were performed for experiments with more than two conditions. Experiments were performed with at least 3 biological replicates, with technical duplicates or triplicates per biological sample unless otherwise noted.

Availability of Sequencing Data

All sequencing data are available through the GEO via accession GSE198908 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198908>). Processed counts data used in the study are additionally available at this accession.

Authors' Disclosures

Y. Nakauchi reports other support from Bluestar Genomics during the conduct of the study. R. Sharma reports grants from the American Cancer Society during the conduct of the study; and is a current employee and equity holder in Graphite Bio. D. Karigane reports other support from JSPS during the conduct of the study. R. Dutta reports personal fees from Acuta Capital outside the submitted work. Y. Ning reports other support from Bluestar Genomics during the conduct of the study; other support from Bluestar Genomics outside the submitted work. G.D. Guler reports personal fees from Bluestar Genomics during the conduct of the study; personal fees from Bluestar Genomics outside the submitted work. A. Bergamaschi reports other support from Bluestar Genomics during the conduct of the study. S. Levy reports being an employee and shareholder of the company Bluestar Genomics that partially funded this study.

R. Majeti reports grants from Bluestar Genomics, NIH, Ludwig Institute, Leukemia and Lymphoma Society, Mark Foundation, and Paul G. Allen Foundation during the conduct of the study; other support from CircBio Inc., Kodikaz Therapeutic Solutions, Gilead Sciences, Pheast Therapeutics, Myelogene, and RNAC Therapeutics and personal fees from Syros Pharmaceuticals outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

Y. Nakauchi: Conceptualization, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing—original draft, project administration, writing—review and editing. **A. Azizi:** Data curation, software, formal analysis, validation, investigation, methodology, writing—original draft, writing—review and editing. **D. Thomas:** Supervision, validation, investigation, writing—review and editing. **M. Corces:** Supervision, validation, writing—review and editing. **A. Reinisch:** Supervision, validation, investigation, methodology. **R. Sharma:** Data curation, software, validation, methodology. **D. Cruz Hernandez:** Supervision, validation, investigation, visualization. **T. Kohnke:** Data curation, software, supervision, validation. **D. Karigane:** Supervision, validation, investigation, visualization. **A. Fan:** Software, supervision, validation, investigation. **D. Martinez-Krams:** Software, supervision, writing—review and editing. **M. Stafford:** Resources, supervision, investigation, methodology. **S. Kaur:** Data curation, software. **R. Dutta:** Data curation, software, formal analysis. **P. Phan:** Data curation, software, formal analysis, supervision. **A. Ediriwickrema:** Data curation, software, formal analysis, supervision. **E. McCarthy:** Resources, data curation, software, formal analysis, supervision. **Y. Ning:** Data curation, software, formal analysis, investigation, methodology. **T. Phillips:** Data curation, software, formal analysis. **C.K. Ellison:** Data curation, software, formal analysis. **G.D. Guler:** Data curation, software, formal analysis. **A. Bergamaschi:** Data curation, software, formal analysis. **C. Ku:** Data curation, software, formal analysis. **S. Levy:** Resources, data curation, validation, investigation, methodology, writing—review and editing. **R. Majeti:** Conceptualization, resources, supervision, funding acquisition, investigation, methodology, writing—original draft, project administration, writing—review and editing.

Acknowledgments

We are grateful to Feifei Zhao for lab management. We thank all of the members of the Majeti lab for help, support, encouragement, and inspiration over the years. We also thank Shogo Minamikawa for preparing part of the illustrations. We acknowledge the Flow Cytometry Core of the Stanford Stem Cell Institute and the Binns Program for Cord Blood Research, especially for Sruthi Mantri, Adam Sheikali, and Sherah Talia Menezes. Parts of the computing for this project were performed on the Sherlock cluster at SLAC. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that have contributed to these research results. This work was supported by NIH grants 1R01HL142637 and 1R01CA251331, the Stanford Ludwig Center for Cancer Stem Cell Research and Medicine, and the Blood Cancer Discoveries Grant program through The Leukemia and Lymphoma Society, The Mark Foundation for Cancer Research, and The Paul G. Allen Frontiers Group, all to R. Majeti. R. Majeti is a recipient of a Leukemia and Lymphoma Society Scholar Award. T. Kohnke is a special fellow of The Leukemia and Lymphoma Society. Y. Nakauchi was supported by the Nakayama Foundation for Human Science and a Stanford University School of Medicine Dean's Postdoctoral Fellowship. A. Azizi was supported by the American Society of Hematology HONORS award.

Received August 12, 2021; revised February 7, 2022; accepted May 4, 2022; published first June 2, 2022.

REFERENCES

- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009;324:930–5.
- 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;466:1129–33.
- Hahn MA, Qiu R, Wu X, Li AX, Zhang H, Wang J, et al. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep* 2013;3:291–300.
- Cimmino L, Abdel-Wahab O, Levine RL, Aifantis I. TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell* 2011;9:193–204.
- Huang Y, Rao A. Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet* 2014;30:464–74.
- Ko M, An J, Rao A. DNA methylation and hydroxymethylation in hematologic differentiation and transformation. *Curr Opin Cell Biol* 2015;37:91–101.
- Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 2011;20:11–24.
- Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 2010;468:839–43.
- Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014;371:2488–98.
- Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014;371:2477–87.
- Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR, et al. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* 2012;4:149ra118.
- Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* 2016;48:1193–203.
- Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 2013;500:477–81.
- Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nat Commun* 2015;6:8489.
- Farlik M, Halbritter F, Müller F, Choudry FA, Ebert P, Klughammer J, et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell* 2016;19:808–22.
- Weisenberger DJ. Characterizing DNA methylation alterations from The Cancer Genome Atlas. *J Clin Invest* 2014;124:17–23.
- Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ, et al. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet* 2016;48:1014–23.
- Rasmussen KD, Jia G, Johansen JV, Pedersen MT, Rapin N, Bagger FO, et al. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev* 2015;29:910–22.
- Yamazaki J, Jelinek J, Lu Y, Cesaroni M, Madzo J, Neumann F, et al. TET2 mutations affect non-CpG island DNA methylation at enhancers and transcription factor-binding sites in chronic myelomonocytic leukemia. *Cancer Res* 2015;75:2833–43.
- Solary E, Bernard OA, Tefferi A, Fuks F, Vainchenker W. The ten-eleven translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia* 2014;28:485–96.
- 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:R54–8.
- Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PW, D'Souza Z, et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol* 2016;18:225–33.
- Tsagaratou A, Äijö T, Lio C-WJ, Yue X, Huang Y, Jacobsen SE, et al. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proc Natl Acad Sci U S A* 2014;111:E3306–15.
- Song C-X, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res* 2017;27:1231–42.
- Chiu BC-H, Zhang Z, You Q, Zeng C, Stepniak E, Bracci PM, et al. Prognostic implications of 5-hydroxymethylcytosines from circulating cell-free DNA in diffuse large B-cell lymphoma. *Blood Adv* 2019;3:2790–9.
- Yang Y, Zeng C, Lu X, Song Y, Nie J, Ran R, et al. 5-Hydroxymethylcytosines in circulating cell-free DNA reveal vascular complications of type 2 diabetes. *Clin Chem* 2019;65:1414–25.
- Minor EA, Court BL, Young JI, Wang G. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J Biol Chem* 2013;288:13669–74.
- Agathocleous M, Meacham CE, Burgess RJ, Piskounova E, Zhao Z, Crane GM, et al. Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature* 2017;549:476–81.
- Cimmino L, Dolgalev I, Wang Y, Yoshimi A, Martin GH, Wang J, et al. Restoration of TET2 function blocks aberrant self-renewal and leukemia progression. *Cell* 2017;170:1079–95.
- Bejar R, Lord A, Stevenson K, Bar-Natan M, Pérez-Ladaga A, Zaneveld J, et al. TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients. *Blood* 2014;124:2705–12.
- Cui X-L, Nie J, Ku J, Dougherty U, West-Szymanski DC, Collin F, et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat Commun* 2020;11:1–11.
- Zhang J, Lee D, Dhiman V, Jiang P, Xu J, McGillivray P, et al. An integrative ENCODE resource for cancer genomics. *Nat Commun* 2020;11:3696–11.
- Weissmann S, Alpermann T, Grossmann V, Kowarsch A, Nadarajah N, Eder C, et al. Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia* 2012;26:934–42.
- Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 2014;42:e168–8.
- Oceguera-Yanez F, Kim S-I, Matsumoto T, Tan GW, Xiang L, Hatani T, et al. Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. *Methods* 2016;101:43–55.
- Yan H, Wang Y, Qu X, Li J, Hale J, Huang Y, et al. Distinct roles for TET family proteins in regulating human erythropoiesis. *Blood* 2017;129:2002–12.
- Wang B, Xia M, Chen T, Li M, Shi D, Wang X, et al. Loss of Tet2 affects platelet function but not coagulation in mice. *Blood Science* 2020;2:129–36.
- Smith AE, Mohamedali AM, Kulasekararaj A, Lim Z, Gäken J, Lea NC, et al. Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value. *Blood* 2010;116:3923–32.
- Selimoglu-Buet D, Wagner-Ballon O, Saada V, Bardet V, Itzykson R, Bencheikh L, et al. Characteristic repartition of monocyte subsets as a diagnostic signature of chronic myelomonocytic leukemia. *Blood* 2015;125:3618–26.
- An J, González-Avalos E, Chawla A, Jeong M, López-Moyado IF, Li W, et al. Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat Commun* 2015;6:10071.
- Li W, Zhang X, Lu X, You L, Song Y, Luo Z, et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res* 2017;27:1243–57.
- Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 2015;33:510–7.

43. Rasmussen KD, Berest I, Kessler S, Nishimura K, Simon-Carrasco L, Vassiliou GS, et al. TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. *Genome Res* 2019;29:564–75.
44. Bak RO, Dever DP, Reinisch A, Cruz Hernandez D, Majeti R, Porteus MH. Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *eLIFE* 2017;6:e27873.
45. Quivoron C, Couronné L, Valle Della V, Lopez CK, Plo I, Wagner-Ballon O, 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.
46. Mazumdar C, Shen Y, Xavy S, Zhao F, Reinisch A, Li R, et al. Leukemia-associated cohesin mutants dominantly enforce stem cell programs and impair human hematopoietic progenitor differentiation. *Cell Stem Cell* 2015;17:675–88.
47. Reilly BM, Luger T, Park S, Lio C-WJ, González-Avalos E, Wheeler EC, et al. 5-Azacytidine transiently restores dysregulated erythroid differentiation gene expression in TET2-deficient erythroleukemia cells. *Mol Cancer Res* 2021;19:451–64.
48. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137–9.
49. Chen Q, Espey MG, Sun AY, Pooput C, Kirk KL, Krishna MC, et al. Pharmacologic doses of ascorbate act as a prooxidant and decrease growth of aggressive tumor xenografts in mice. *Proc Natl Acad Sci U S A* 2008;105:11105–9.
50. Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* 2008;26:1293–300.
51. Liang K, Keleş S. Normalization of ChIP-seq data with control. *BMC Bioinform* 2012;13:1–10.
52. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* 2017;14:975–8.