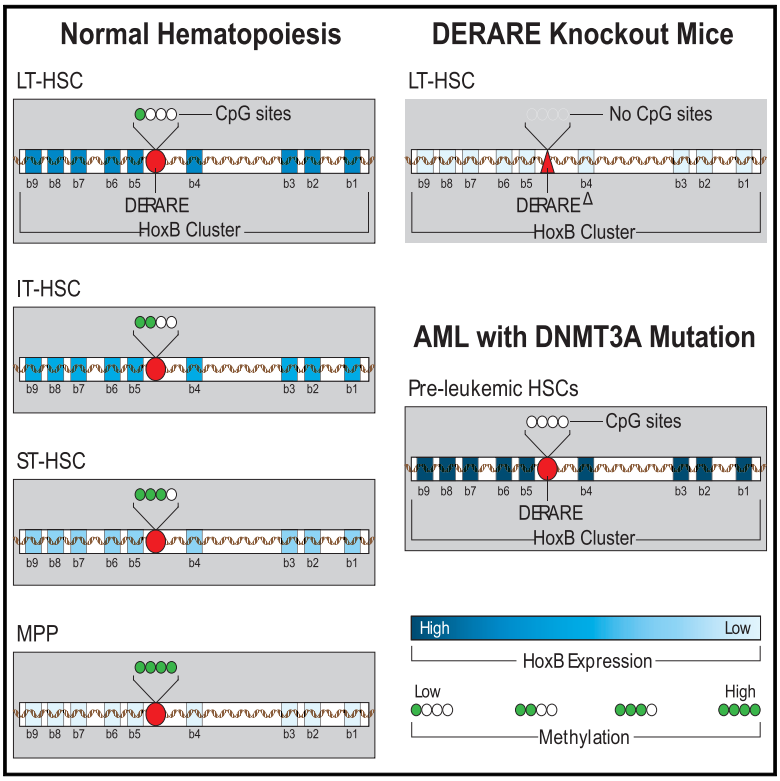
Article

Cell Stem Cell

Retinoid-Sensitive Epigenetic Regulation of the Hoxb Cluster Maintains Normal Hematopoiesis and Inhibits Leukemogenesis

Graphical Abstract



Highlights

- A *cis*-regulatory element *DERARE* works over long range to control *Hoxb* genes in HSCs
- Loss of the *DERARE* abrogates *Hoxb* expression and leads to functional HSC defects
- DNMT enzymes mediate *DERARE* methylation to repress *Hoxb* cluster expression
- dCas9-DNMT3A targeted DNA methylation on the *DERARE* alleviates AML leukemogenesis

Authors

Pengxu Qian, Bony De Kumar, Xi C. He, ..., Hua Li, Robb Krumlauf, Linheng Li

Correspondence

rek@stowers.org (R.K.), lil@stowers.org (L.L.)

In Brief

Hox genes are regulators of HSC maintenance and contributors in leukemogenesis. Li and colleagues elucidate a mechanism for how the retinoid-dependent *cis*-regulatory element *DERARE* maintains normal hematopoiesis but prevents leukemogenesis by coordinate regulation of *Hoxb* cluster genes in a methylation-dependent manner.

Qian et al., 2018, Cell Stem Cell *22*, 740–754
May 3, 2018
https://doi.org/10.1016/j.stem.2018.04.012

CellPress

CellPress

Cell Stem Cell
Article

Retinoid-Sensitive Epigenetic Regulation of the Hoxb Cluster Maintains Normal Hematopoiesis and Inhibits Leukemogenesis

Pengxu Qian,[1,7] Bony De Kumar,[1,8] Xi C. He,[1] Christof Nolte,[1] Madelaine Gogol,[1] Youngwook Ahn,[1] Shiyuan Chen,[1] Zhenrui Li,[1,2] Hanzhang Xu,[3] John M. Perry,[1] Deqing Hu,[1,4] Fang Tao,[1,2] Meng Zhao,[5] Yingli Han,[1] Kate Hall,[1] Allison Peak,[1] Ariel Paulson,[1] Chongbei Zhao,[1] Aparna Venkatraman,[1] Andrew Box,[1] Anoja Perera,[1] Jeffrey S. Haug,[1] Tari Parmely,[1] Hua Li,[1] Robb Krumlauf,[1,6,*] and Linheng Li[1,2,9,*]

[1]Stowers Institute for Medical Research, Kansas City, MO 64110, USA
[2]Department of Pathology & Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66160, USA
[3]Key Laboratory of Cell Differentiation and Apoptosis of the Chinese Ministry of Education, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
[4]Tianjin Medical University School of Basic Medicine, Tian Jin 300070, China
[5]Key Laboratory of Stem Cells and Tissue Engineering, Sun Yat-Sen University, Ministry of Education, Guangzhou, Guangdong 510080, China
[6]Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS 66160, USA
[7]Present address: Center for Stem Cell and Regenerative Medicine, Department of Basic Medical Sciences and Institute of Hematology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China
[8]Present address: The Jackson Laboratory, Bar Harbor, ME 04609, USA
[9]Lead Contact
*Correspondence: rek@stowers.org (R.K.), lil@stowers.org (L.L.)
https://doi.org/10.1016/j.stem.2018.04.012

SUMMARY

Hox genes modulate the properties of hematopoietic stem cells (HSCs) and reacquired *Hox* expression in progenitors contributes to leukemogenesis. Here, our transcriptome and DNA methylome analyses revealed that *Hoxb* cluster and retinoid signaling genes are predominantly enriched in LT-HSCs, and this coordinate regulation of *Hoxb* expression is mediated by a retinoid-dependent *cis*-regulatory element, *distal element RARE* (*DERARE*). Deletion of the *DERARE* reduced *Hoxb* expression, resulting in changes to many downstream signaling pathways (e.g., non-canonical Wnt signaling) and loss of HSC self-renewal and reconstitution capacity. DNA methyltransferases mediate DNA methylation on the *DERARE*, leading to reduced *Hoxb* cluster expression. Acute myeloid leukemia patients with *DNMT3A* mutations exhibit *DERARE* hypomethylation, elevated *HOXB* expression, and adverse outcomes. CRISPR-Cas9-mediated specific DNA methylation at *DERARE* attenuated *HOXB* expression and alleviated leukemogenesis. Collectively, these findings demonstrate pivotal roles for retinoid signaling and the *DERARE* in maintaining HSCs and preventing leukemogenesis by coordinate regulation of *Hoxb* genes.

INTRODUCTION

Hematopoietic stem cells (HSCs) are pivotal for homeostasis of the blood system by maintaining stem cells in quiescent and low metabolic state, and producing lineage cells to meet demands for normal circulation and stress response (Orkin and Zon, 2008). The balance between self-renewal and multilineage differentiation in HSCs is finely tuned by interactions between gene regulatory networks, which integrate the inputs from HSCs and their environment, and transcription factors (TFs) through the organization of their *cis*-regulatory elements (CREs) (Göttgens, 2015). The binding of TFs, such as *Gata2*, *Runx1*, and *C/EBPa*, to appropriate target DNA sequences recruits chromatin-modifying factors and components of the transcription machinery, which modulate gene expression and direct the functionality and lineage specification of HSCs. Epigenetic status, such as DNA methylation and chromosome accessibility of the CREs, plays important roles in these regulatory networks by influencing TF binding and activity (Hon et al., 2013; Thurman et al., 2012). Recent studies have depicted genome-wide maps of transcriptome, DNA methylome, histone modifications, and chromosome accessibility during hematopoietic differentiation and HSC aging, which are important to understand the *cis*-regulatory network in hematopoiesis (Cabezas-Wallscheid et al., 2014; Cimmino et al., 2015; Corces et al., 2016; Jeong et al., 2014; Ji et al., 2010; Lara-Astiaso et al., 2014). However, the specific CREs critical for regulating physiological functions and underlying mechanisms in HSCs are largely unknown.

In mammals, 39 *Hox* TFs are separated into four chromosomal complexes (*Hoxa*, *b*, *c*, and *d*), and their coordinate expression in embryogenesis generates unique positional codes to specify cellular identity (Alexander et al., 2009). Initial activation of *Hox* clusters involves signaling by retinoic acid (RA), Wnt and fibroblastic growth factors (FGFs) (Deschamps and Duboule, 2017), and multiple retinoic acid response element (RAREs) within the *Hox* clusters are important in directly modulating their expression (Nolte et al., 2003; Studer et al., 1994). The subsequent establishment and refinement of their expression patterns

involves extensive auto- and cross-regulatory interactions between *Hox* genes themselves. *Hox* genes are regulators of HSC maintenance and contributors in leukemogenesis (Alharbi et al., 2013; Parker and Krumlauf, 2017). Ectopic expression of *Hoxb4* endows mouse embryonic stem cells (ESCs) with definitive HSC features (Kyba et al., 2002), while *Hoxb5* was recently reported as a unique marker of long-term (LT)-HSC (Chen et al., 2016). During the early phases of progression of acute myeloid leukemia (AML), acquisition of founder mutations in epigenetic modifiers such as DNA methyltransferase (DNMT) 3A in HSCs leads to a clonally expanded pool of pre-leukemic HSCs (Shlush et al., 2014). *HOXB* cluster genes showed markedly diminished DNA methylation and significantly increased expression in *DNMT3A* mutant mouse HSCs and AML patients (Jeong et al., 2014; Yan et al., 2011). Thus, the underlying mechanisms by which *Hoxb* genes are coordinately regulated in HSCs are not only important for understanding their physiologic roles, but they may represent a new target for treating leukemia. RA signaling is known to play important roles in the dormancy, maintenance, and developmental programs of HSCs (Cabezas-Wallscheid et al., 2017; Chanda et al., 2013; Purton et al., 2000), but the underlying regulatory mechanisms are unclear. Here, our analyses reveal that a specific CRE, *distal element RARE* (*DERARE*), which is located between the *Hoxb4* and *Hoxb5* genes, functions in the integration of RA signaling to maintain HSCs and prevent leukemogenesis by coordinate regulation of *Hoxb* cluster genes in a methylation-dependent manner.

RESULTS

Hoxb Cluster and RA Signaling Genes Are Predominantly Enriched in LT-HSCs

We employed transcriptional profiling of purified cell types to identify genes highly enriched in LT-HSCs, as a means to uncover a candidate group of TFs associated with their self-renewal potential. We evaluated the reconstitution capacity of three different adult HSC populations by transplanting 50 of each into lethally irradiated recipient mice: LT-HSC (Lineage$^-$Sca-1$^+$c-Kit$^+$ (LSK)CD34$^-$Flk2$^-$CD49b^{lo}); intermediate-term (IT)-HSC (LSKCD34$^-$Flk2$^-$CD49b^{hi}) and short-term (ST)-HSC (LSKCD34$^+$ Flk2$^-$). LT-HSCs exhibited permanent repopulating ability, whereas IT-HSCs and ST-HSCs failed to engraft after 32 weeks and 4 weeks, respectively (Figure S1A). These three types of HSCs, together with multipotent progenitors (MPPs, LSKCD34$^+$Flk2$^+$), four committed progenitors, and nine mature lineage cells, were sorted from the bone marrow (BM) of C57BL/6J mice, and RNA sequencing (RNA-seq) was conducted on their poly(A)+ RNA (Figure 1A). This yielded expression of 13,783 protein-coding genes. Both unsupervised hierarchical family tree analysis and principal components analysis (PCA) showed that samples from HSCs, progenitors, and lineages clustered together, which validated the high quality of our isolation method and RNA-seq data (Figures 1B and S1B). Based on specificity scores in these 17 cell types, we identified 44 genes that formed a unique fingerprint for expression in LT-HSCs, including multiple *Hoxb* genes (*Hoxb2*, *Hoxb5*, and *Hoxb6*) (Figure 1C; Table S1). We then analyzed the expression patterns of genes from all four *Hox* clusters and found that *Hoxc* and *Hoxd* genes (except for *Hoxc6*) were rarely transcribed in HSCs. *Hoxa* genes were widely expressed in hematopoietic stem and progenitor cells (HSPCs), whereas *Hoxb* genes were predominantly enriched in LT-HSCs (Figure 1D). Concordantly, the enriched expression of *Hoxb* genes was detected in LT-HSCs by checking other hematopoiesis datasets, such as Gene Expression Commons (Seita et al., 2012), Bloodspot (Bagger et al., 2016), and Haemosphere (de Graaf et al., 2016) (Figures S1C–S1E). Together, these data suggest that *Hoxb* cluster genes might be important for modulating the properties of HSCs.

In addition to *Hoxb* genes, the RNA-seq analyses revealed that several genes associated with RA signaling were highly expressed in LT-HSCs (Figures 1E and S1F), consistent with a recent study showing that RA signaling regulates dormant HSCs (Cabezas-Wallscheid et al., 2017). This correlative link between components of RA signaling and the *Hoxb* cluster in HSCs is intriguing, as retinoids are known to modulate *Hox* expression in embryogenesis, and multiple RAREs have been shown to control *Hox* expression in hindbrain segmentation and limb development (de Laat and Duboule, 2013; Parker and Krumlauf, 2017). This suggests potential regulatory roles of RA signaling in directly modulating *Hoxb* expression in HSCs.

Identification of the *DERARE* as a *CRE* within *Hoxb* Cluster in HSCs

Differentially methylated regions (DMRs) have been shown to largely overlap with CREs, such as DNaseI hypersensitivity sites, promoters, and enhancers, and correlate with HSC maintenance and leukemogenesis (Cabezas-Wallscheid et al., 2014; Cimmino et al., 2015; Guillamot et al., 2016; Jeong et al., 2014; Ji et al., 2010; Trowbridge et al., 2009). We previously reported that two germline DMRs, *H19-DMR* and *IG-DMR*, function as a transcription insulator in the *H19-Igf2* and an enhancer in the *Dlk1-Gtl2* imprinting loci respectively, to maintain LT-HSCs in a quiescent and low metabolic state (Qian et al., 2016; Venkatraman et al., 2013). To screen for potential regulatory CREs in the *Hoxb* cluster, we performed whole-genome bisulfite sequencing in HSPCs and identified multiple DMRs within *Hoxb* cluster (Figure 1F), consistent with patterns reported in a previous study using a different strategy to isolate HSCs (Cabezas-Wallscheid et al., 2014). Intriguingly, one DMR located between *Hoxb4* and *Hoxb5* overlapped with the region containing the *DERARE* (Figures 1F and 1G), which has been shown to regulate RA-dependent expression of *Hoxb* genes in the CNS (Ahn et al., 2014; Valarché et al., 1997). The *DERARE* is highly conserved in mouse, human, and other eutherian mammals (Figure S2). When HSCs differentiate, DNA methylation on the *DERARE* was gradually elevated (Figures 1G and S1G), which inversely correlated with *Hoxb* cluster expression (Figure 1H). To explore other properties associated with CREs, we analyzed the recently published transposase-mediated chromatin accessibility using sequencing (ATAC-seq) profiles performed in 14 types of human hematopoietic cells (Corces et al., 2016) and observed gradually decreased chromosome accessibility adjacent to the human *DERARE* during HSC differentiation (Figure S1H). In line with the human ATAC-seq data indicating accessible chromatin near (800 bp) the *DERARE* region, we found a similar profile of accessibility near (710 bp) the mouse *DERARE* and occupancy of retinoic acid receptors (RARs) and retinoid X receptors

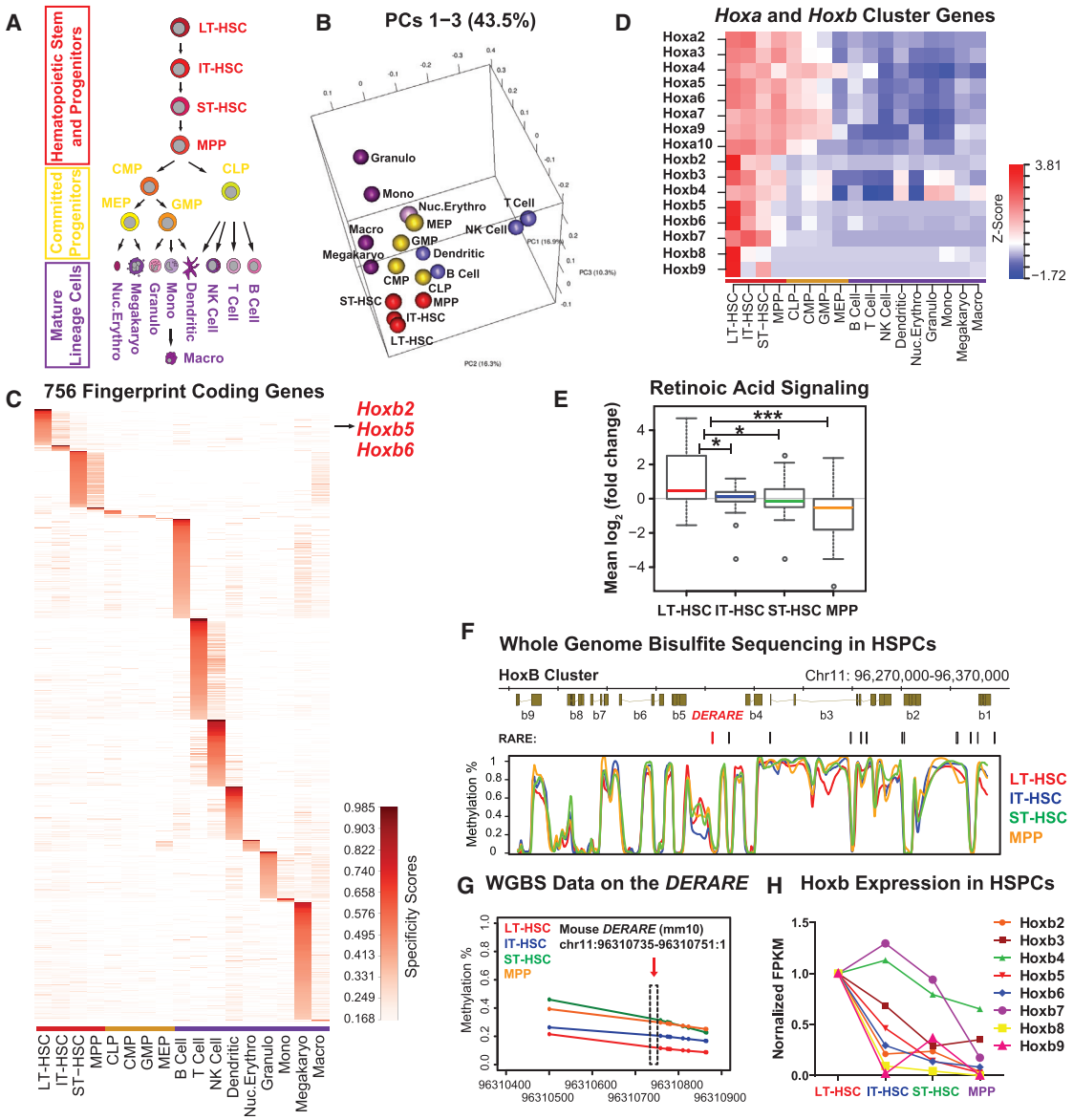


Figure 1. The *DERARE* Is a CRE of *Hoxb* Cluster Genes, which Are Highly Enriched in LT-HSCs

(A) Schematic representation of the hematopoietic hierarchy showing cell types in this study.

(B) Principal-component analysis of 13,783 expressed protein-coding genes shown with the first three principal components (PC1–PC3). HSPCs, red; committed progenitor cells, yellow; myeloid cells, violet; lymphoid cells, blue; nucleated erythrocytes, pink.

(C) Heatmap of 756 cell-type-specific coding genes.

(D) Heatmap of *Hoxa* and *Hoxb* cluster genes from RNA-seq of 17 hematopoietic cell types.

(E) GO enrichment analysis of RA metabolism genes.

(F) Methylation distributions from whole-genome bisulfite sequencing of the *Hoxb* cluster in 4 adult HSPC samples.

(G) Methylation distributions around the *DERARE* region in 4 adult HSPC samples. The dashed line box highlights the location of *DERARE*. Round dots indicate the CpG sites.

(H) RNA-seq analysis of *Hoxb* genes in 4 adult HSPC samples. *p < 0.05; ***p < 0.001.

See also Figures S1 and S2 and Table S1.

(RXRs) on this region in ESC models (De Kumar et al., 2015). Together, these data implied that the *DERARE* might be a conserved retinoid-dependent and methylation-sensitive CRE that participates in long-range regulation of *Hoxb* genes in HSCs.

To test this idea, we used a mutant mouse line carrying a 52-bp deletion containing the *DERARE* (*DERARE*$^{\Delta}$) (Ahn et al., 2014) (Figures 2A and S2C). Both RNA-seq and real-time qPCR analyses showed that expression of *Hoxb* cluster genes was markedly lower in homozygous *DERARE*$^{\Delta}$ mice (abbreviated as *Homo* thereafter) compared to their wild-type (WT) littermate controls (Figures 2B and S3A). These data demonstrate that the *DERARE* is a key

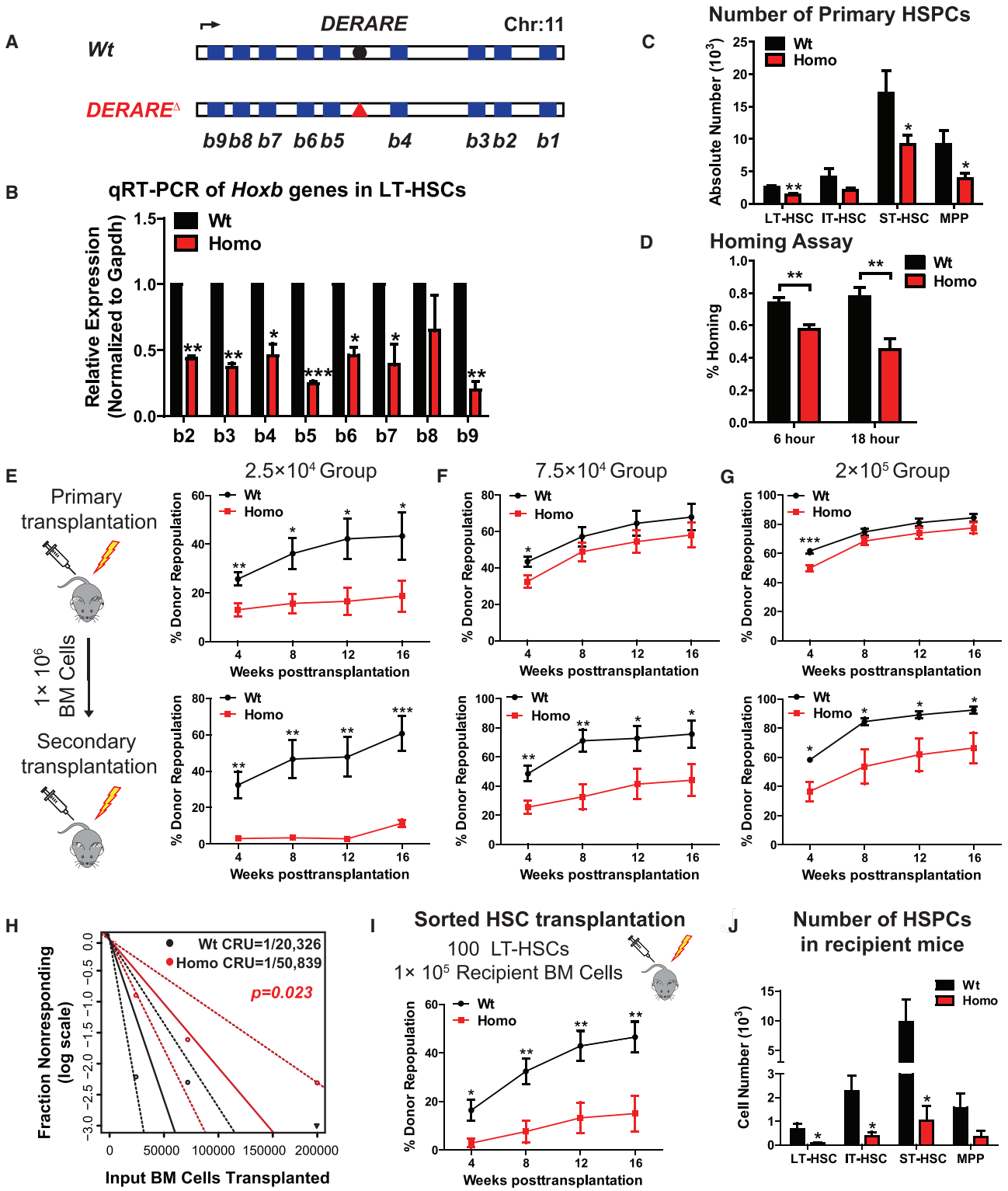


Figure 2. Deletion of the *DERARE* Leads to Phenotypic and Functional Defects in HSCs

(A) Schematic representation of *Hoxb* cluster in *WT* and *DERARE*$^{\Delta}$ mice. Blue rectangles show *Hoxb* genes. Black circle represents the *DERARE* in *WT* mice. Red triangle shows deletion of *DERARE* in *DERARE*$^{\Delta}$ mice. Black arrow indicates the transcription directions of *Hoxb* genes.

(B) Real-time qPCR showing reduced *Hoxb* gene expression in *DERARE*$^{\Delta}$ LT-HSC compared to *WT* LT-HSC. Mouse *Gapdh* was used as internal control.

(C) Absolute number of HSPCs in *WT* (n = 4) and *DERARE*$^{\Delta}$ (n = 5) mice.

(D) 1 × 10^6 CFDA SE-labeled BM cells were transplanted into recipient mice. 6 and 18 hr later, BM was analyzed for homed events (n = 10 mice per group).

(legend continued on next page)

CRE involved in coordinate modulation of *Hoxb* cluster expression in HSCs.

Deletion of the *DERARE* Leads to HSC Defects

Given that *Hoxb* genes are known to contribute to hematopoiesis, we asked whether the *DERARE* plays a functional role in HSC maintenance. Although deletion of the *DERARE* did not alter BM cellularity, we observed a significant decrease in frequency and absolute number of LT-HSCs, ST-HSCs, MPPs, common myeloid progenitors (CMPs), and megakaryocyte-erythroid progenitors (MEPs) in homozygous *DERARE*$^{\Delta}$ mice (Figures 2C and S3B–S3E). Flow cytometry analysis showed an increased number of erythrocytes, myeloid and T cells in BM, and Hemavet analysis revealed an elevated number of white blood cells, platelets, and neutrophils in peripheral blood of *DERARE*$^{\Delta}$ mice (Figures S3F–S3I). This indicates that HSPCs alter their potential for self-renewal and differentiate into mature lineage cells after deletion of the *DERARE*. We next examined LT-HSC cell-cycle status and found a significant decrease in the G0 phase fraction and a concomitant increase in the G1 and S/G2/M phase fractions in *DERARE*$^{\Delta}$ LT-HSCs compared to controls (Figure S3J).

To further characterize functional roles of the *DERARE* on HSCs, we performed homing assays to check short-term engraftment ability by transplantation of 1 × 10^6 carboxyfluorescein diacetate succinimidyl ester (CFDA SE)-labeled BM cells from homozygous mutants or their control littermates into recipient mice. The homing of *DERARE*$^{\Delta}$ BM cells was impaired at 6 (23.7% reduction) and 18 hr (42.4% reduction) after transplantation (Figure 2D). We next executed competitive repopulation assays by transplantation of equal numbers of donor BM cells (CD45.2) together with recipient BM cells derived from the *Ptprc* mutant strain (CD45.1), into lethally irradiated recipient mice. In the 2.5 × 10^4 group, compared to controls, we observed 2.3- and 5.3-fold reduction in overall repopulation rate from *DERARE*$^{\Delta}$ donor cells in primary (1st) and secondary (2nd) recipients, respectively, whereas in the 7.5 × 10^4 and 2 × 10^5 groups, we detected a decrease only in the 2nd recipients (Figures 2E–2G). In the *DERARE*$^{\Delta}$-transplant recipients, we found a marked decrease in frequency, absolute number, and G0 phase percentage of donor HSCs and observed a myeloid lineage bias (Figures S3K–S3S). Limiting dilution, competitive repopulating unit (CRU) analysis revealed a 2.5-fold decrease in *DERARE*$^{\Delta}$ HSCs compared to controls (Figure 2H). To confirm that the observed defect was specific to HSCs, we transplanted 100 sorted LT-HSCs from *WT* and *DERARE*$^{\Delta}$ mice together with 1 × 10^5 rescue BM cells into recipient mice. We observed a 3-fold reduction in overall repopulation rate and a 9-fold decrease in numbers of donor HSCs from mutant mice compared to controls (Figures 2I and 2J). These results show that homozygous deletion of the *DERARE* compromises HSC homing and subsequent reconstitution capacity.

RA Signaling Is Required for *Hoxb* Expression in HSCs

RA signaling is essential for the emergence of definitive HSCs during embryogenesis (Chanda et al., 2013) and facilitates adult HSC maintenance (Cabezas-Wallscheid et al., 2017; Purton et al., 2000). This signaling is mediated by RARs, which form heterodimers with RXRs and bind to RAREs. Our RNA-seq data in 17 hematopoietic cells showed *Rarb* was enriched in LT-HSCs, whereas *Rara* and *Rarg* were enriched in myeloid cells (Figure S1F). To test whether this component of RA signaling affects *Hoxb* cluster expression, we crossed *Rarb* floxed mice with *Mx1-Cre* mice and deleted *Rarb* in the hematopoietic system by administration of pIpC (Figure S4A). We found increased DNA methylation on the *DERARE* and reduced *Hoxb* gene expression in *Rarb*$^{-/-}$ LT-HSCs (Figures S4B and S4C). Similarly, we detected an overall decrease in numbers of primary HSCs, CMPs, MEPs, and an increase in numbers of erythrocytes and myeloid and T cells in BM from *Rarb*$^{-/-}$ mice compared to controls (Figures S4D–S4J). We further observed a reduced overall repopulation rate and number of donor HSCs in recipient mice transplanted with *Rarb*$^{-/-}$ BM cells (Figures S4K and S4L). These data further support the essential roles of RA signaling in maintaining LT-HSCs and suggest that *Rarb* may have a direct regulatory input on *Hoxb* genes through the *DERARE*.

Reduced Expression of *Hoxa* Cluster and Non-canonical Wnt Pathway Genes Contribute to HSC Defects in *DERARE*$^{\Delta}$ Mice

To dissect downstream mechanisms underlying the *DERARE*$^{\Delta}$-mediated HSC deficiency, we performed RNA-seq and identified 3,633 differentially expressed genes in LT-HSCs sorted from *WT* and *DERARE*$^{\Delta}$ mice (Figures 3A and 3B; Table S2). A 3D PCA map showed that, unlike *WT* LT-HSCs, *DERARE*$^{\Delta}$ LT-HSCs were closer to more differentiated MPPs (Figure S5A). We found markedly lower expression of HSC surface markers and TFs essential for HSC self-renewal and specification in *DERARE*$^{\Delta}$ LT-HSCs compared to controls. This held true for the eight TFs recently reported to reprogram committed progenitors into HSCs with multilineage potential (Riddell et al., 2014; Figure 3C). In addition, most of the *Hoxa* cluster genes and several imprinting genes related to HSC maintenance displayed reduced expression in *DERARE*$^{\Delta}$ LT-HSCs (Figures S5B and S5C). The lower levels of expression of *Hoxa* genes in LT-HSCs are likely to be a consequence of the reduction in the levels of *Hoxb* gene expression, as there is evidence for direct cross-regulatory input of *Hoxb* genes into maintaining *Hoxa* gene expression in neural development and HSCs (Bijl et al., 2006; Parker and

(E–G) Indicated numbers of whole bone marrow (WBM) cells, (E) 2.5x10^4, (F) 7.5x10^4, and (G) 2x10^5, were transplanted with 2 × 10^5 rescue cells into recipient mice. At 16 weeks post-transplant, 1 × 10^6 BM cells isolated from 1st recipients (top) were transplanted into 2nd recipients (bottom). Peripheral blood was analyzed for percentage of donor repopulation at the indicated number of weeks after transplantation (2 experiments with a total of 10 recipient mice per group).
(H) CRU frequency determined by extreme limiting dilution assay (ELDA) analysis (n = 10).
(I) 100 LT-HSCs sorted from *WT* or *DERARE*$^{\Delta}$ mice were transplanted with 1 × 10^5 rescue cells into recipient mice. Peripheral blood (PB) analysis for percentage of engrafted donor cells at the indicated number of weeks post-transplantation (n = 10).
(J) Absolute numbers of HSPCs in recipient mice injected with sorted LT-HSCs from *WT* and *DERARE*$^{\Delta}$ mice (n = 6). All panels reflect mean ± SD from 2–3 independent experiments. *p < 0.05; **p < 0.01; ***p < 0.001.
See also Figures S3 and S4.

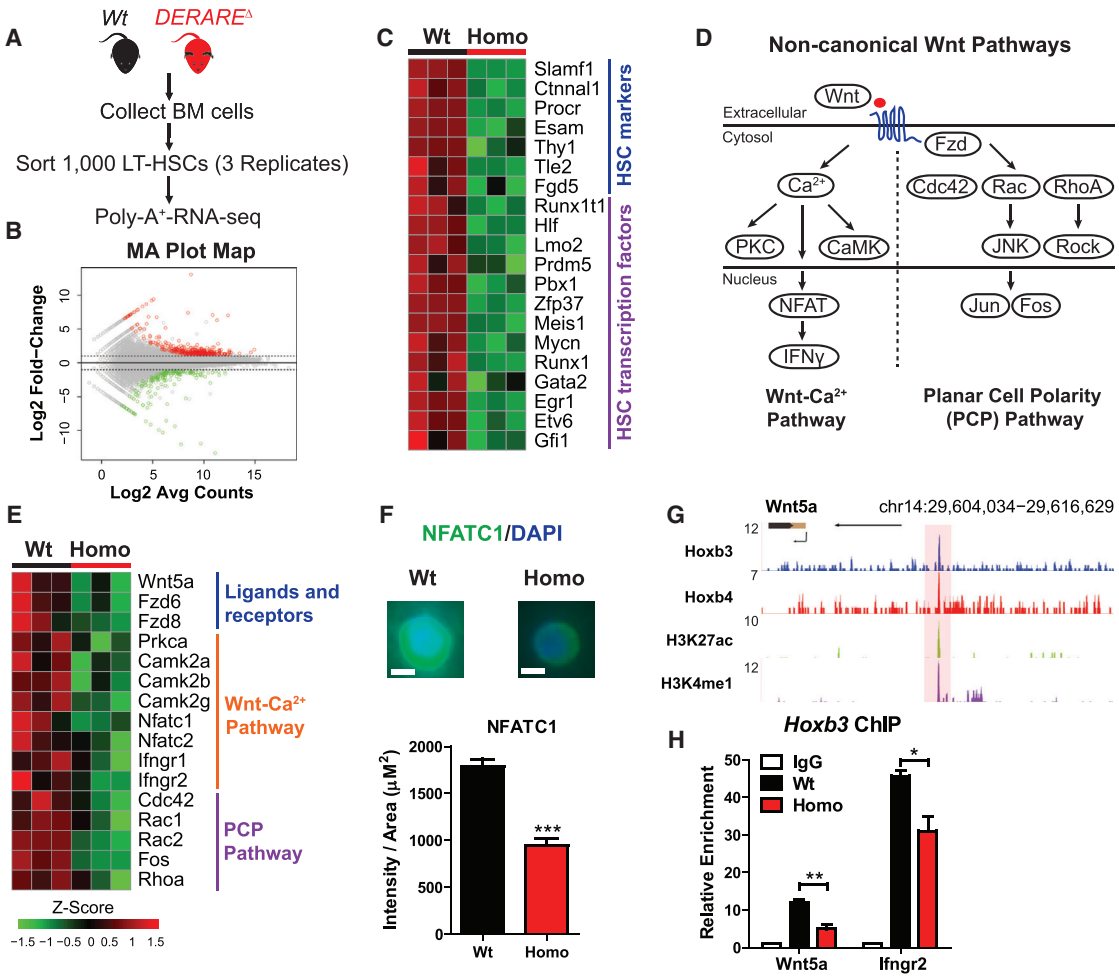


Figure 3. Non-canonical Wnt Pathway Is Directly Targeted by *Hoxb* Genes

(A) Poly(A)+ RNA-seq workflow of 1,000 *WT* or *DERARE*$^{\Delta}$ LT-HSCs.

(B) Log ratio/mean average (MA) scales plot for differential gene expression analysis ($\log_2|FC| > 1$; false discovery rate (FDR) <0.05; multiple t tests with $p < 0.05$) in *WT* and *DERARE*$^{\Delta}$ LT-HSCs.

(C) Heatmap from RNA-seq data showing fragments per kilobase of transcript per million mapped reads (FPKMs) of HSC markers and TF genes lower in *DERARE*$^{\Delta}$ than in *WT* LT-HSCs.

(D) Schematic representation of non-canonical Wnt pathways.

(E) Heatmap from RNA-seq data showing FPKMs of non-canonical Wnt pathway genes were reduced in *DERARE*$^{\Delta}$ compared to *WT* LT-HSCs.

(F) Representative images and quantification of staining intensity of *WT* (n = 71) and *DERARE*$^{\Delta}$ (n = 19) LT-HSCs for *NFATC1*. Scale bars, 5 μm.

(G) ChIP-seq analysis of H3K27ac, H3K4me1, and *Hoxb* genes in mouse KH2 ESCs. Tracks were configured by using windowing function as mean and smoothing windows as 0 pixels in the UCSC Genome Browser.

(H) ChIP-qPCR of *Hoxb3* at the enhancer regions of *Wnt5a* and *Ifngr2* in LT-HSCs sorted from *WT* and *DERARE*$^{\Delta}$ mice. Rabbit immunoglobulin G (IgG) was used as control. All data reflect mean ± SD from 2–3 independent experiments. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

See also Figure S5 and Tables S2 and S3.

Krumlauf, 2017). Hence, our data suggest that the *Hoxa* cluster is a downstream target of *Hoxb* genes in HSC maintenance and along with alterations in *Hoxb* expression may be an important component in contributing to HSCs defects in *DERARE*$^{\Delta}$ mice.

Gene ontology (GO) analysis showed that downregulated terms enriched in the *DERARE*$^{\Delta}$ HSCs were related to hematopoietic or lymphoid organ development and hemopoiesis, while the upregulated terms were enriched for cell death, apoptosis, and oxidative phosphorylation (Figure S5D and Table S3). Notably, the non-canonical Wnt signaling pathway was enriched as a downregulated term in *DERARE*$^{\Delta}$ LT-HSCs. This is consistent with previous observations that non-canonical Wnt signaling can maintain murine and human LT-HSCs and augment their repopulation efficiency (Nemeth et al., 2007; Sugimura et al., 2012). Indeed, we detected lower expression of ligands and receptors, as well as essential components of the non-canonical Wnt-Ca^{2+} and planar cell polarity pathways, in *DERARE*$^{\Delta}$ LT-HSCs than in controls (Figures 3D, 3E, and S5E). Single-cell immunostaining further revealed that *DERARE*$^{\Delta}$ LT-HSCs exhibited 1.9-fold decreased intensity of *NFATC1* nuclear translocation, indicative of non-canonical *Wnt* repression (Figure 3F). We next assessed whether *Hoxb* genes regulate

non-canonical Wnt pathway in HSCs. Due to the technical limitations of rare HSC numbers, we initially performed chromatin immunoprecipitation sequencing (ChIP-seq) of *Hoxb3*, *Hoxb4*, as well as H3K27ac and H3K4me1, in differentiating mouse ESCs and found that *Hoxb3* and *Hoxb4* bound to enhancers or promoters of *Wnt5a*, *Ifngr2*, and *Fos* (Figures 3G, S5F, and S5G). We then performed *Hoxb3* ChIP-qPCR in sorted LT-HSCs and confirmed that *Hoxb3* also bound to enhancer regions of *Wnt5a* and *Ifngr2* in HSCs and that binding was largely abrogated by deletion of the *DERARE* (Figure 3H). This suggests the *Hoxb* proteins may directly target components of the non-canonical Wnt pathway in LT-HSCs.

In light of the key role for the non-canonical Wnt pathway in HSC homeostasis, we tested whether pharmacological restoration of non-canonical Wnt pathway could rescue the HSC defects in *DERARE*$^{\Delta}$ mice. We sorted *WT* and *DERARE*$^{\Delta}$ LT-HSCs, cultured them *ex vivo* with or without Wnt5a for 60 hr, and then transplanted them into lethally irradiated recipient mice (Figure 4A). Wnt5a treatment partially rescued the diminished short-term homing ability, long-term repopulation rate, the decreased number of donor LT-HSCs, and the increased myeloid lineage bias caused by deletion of the *DERARE* (Figures 4B–4E). RNA-seq analysis in sorted LT-HSCs, and the 3D PCA map revealed that Wnt5a treatment shifted the transcriptome of *DERARE*$^{\Delta}$ LT-HSCs closer to that of *WT* HSCs (Figure 4F). Finally, we identified 405 differentially expressed genes, which were rescued at transcription level (Figure 4G; Table S4). GO analysis revealed that most of the GO terms enriched in the *DERARE*$^{\Delta}$ HSCs were partially restored by Wnt5a treatment (Figure 4H; Table S4). Thus, these data suggest that repression of the non-canonical Wnt pathway due to downregulation of *Hoxb* genes is an important component contributing to the functional and molecular defects observed in the *DERARE*$^{\Delta}$ HSCs.

DNMT-Mediated DNA Methylation on the *DERARE* Represses *Hoxb* Cluster Expression

The correlation between increased DNA methylation and diminished chromatin accessibility on the *DERARE* as well as attenuated expression of *Hoxb* genes during HSC differentiation (Figures 1G, 1H, S1G, and S1H) raised the possibility that DNA methylation on the *DERARE* might affect *Hoxb* cluster expression. To investigate upstream mechanisms governing how the *DERARE per se* is potentiated to modulate *Hoxb* cluster expression, we measured the expression levels of DNMT genes in HSCs. *Dnmt1*, *Dnmt3a*, and *Dnmt3b* showed increasing expression during HSC differentiation and LT-HSCs exhibited the lowest global DNA methylation (Figures 5A and 5B). To test the impact of methylation, we cloned a cassette containing the *DERARE* into a CpG-free luciferase vector, performed *in vitro* methylation assay, and then cotransfected them with the pGL3 control vector into 293T cells. DNA methylation on the *DERARE* dramatically inhibited the luciferase activity, which was restored when methylation was low due to lack of methylase SssI or methyl donor S-adenosyl methionine (SAM) (Figure 5C). In addition, mutation of certain CpG sites in the *DERARE* partially abrogated DNA-methylation mediated transcription repression (Figures 5D and 5E).

To ask whether *DNMT* enzymes are required for DNA methylation on the *DERARE*, we analyzed data from a study that reported 1,104 undermethylation regions (called canyons) via whole-genome DNA methylation analysis in HSCs from *Mx1-cre;Dnmt3a*$^{fl/fl}$ (*Dnmt3a* knockout [KO]) and control mice (Jeong et al., 2014). We observed that the *DERARE* is located within one of the canyons. Loss of *Dnmt3a* further reduced DNA methylation on the *DERARE* and resulted in the merger of a group of canyons in the *Hoxb* cluster, including the *DERARE*, into a larger canyon (termed "Grand Canyons"). Consistent with this change, most of the *Hoxb* cluster genes exhibited higher expression in *Dnmt3a* knockout HSCs compared to control HSCs, and several *Hoxb* genes (*Hoxb1*, *b6*, *b8*, *b9*) showed even higher expression in *Dnmt3a* and *Dnmt3b* double-knockout HSCs (Figure 5F). These results show that *DNMT* enzymes mediated DNA methylation on the *DERARE* represses *Hoxb* cluster expression.

Methylation Status on the *DERARE* Correlates with Prognosis in AML Patients

Deletion of *Dnmt3a* upregulates HSC multipotency genes, strengthens self-renewal, and blocks differentiation of HSCs, which expands the pre-leukemic HSC pool and causes hematopoietic malignancies including AML (Challen et al., 2011; Shlush et al., 2014). Indeed, a high frequency of *DNMT3A* mutation has been reported in AML patients (Ley et al., 2010; Yan et al., 2011). We checked expression of *HOXB* and *DNMT* genes in AML by reanalyzing the RNA-seq data of the Leucégène project, consisting of a panel of 17 human CD34^{+} HSPC samples, 43 AML, and 12 acute lymphoblastic leukemia (ALL) patient samples (Macrae et al., 2013). The 2D PCA map showed that different types of samples clustered together (Figure S6A). We observed higher expression of *HOXB* genes and lower expression of *DNMT* genes in AML (79%), but not in the ALL samples, compared to normal HSPCs (Figure 6A). Consistent with our analyses in mice, many essential components of the non-canonical Wnt pathways, such as *WNT5B*, *FZD8*, and *IFNGR2*, were significantly enriched in AML samples (Figure S6B).

Next, we investigated the correlation between DNA methylation on the *DERARE* and leukemogenesis. We analyzed the DNA methylation data obtained by Illumina Infinium HumanMethylation450 BeadChip profiling performed on a cohort of 192 AML patient specimens in The Cancer Genome Atlas (TCGA) project (Ley et al., 2013). Based on DNA methylation status on the *DERARE*, we precluded the samples without methylation information and divided remaining samples into two groups with low (n = 91) or high (n = 88) *DERARE* methylation (Figure 6B; Table S5). Using French-American-British (FAB) classification systems in AML, we found that *DERARE* low methylation patients showed higher percentages of M1 and M5 morphologic subtypes and lower percentages of M0, M2, and M3 subtypes compared to high methylation patients, suggesting the immature and underdeveloped feature of cells from *DERARE* low-methylation patients (Figure 6C). Further, we found that cytogenetic profiles of the *DERARE* low-methylation patients were more likely associated with no fusion, normal karyotype, and intermediate risk (Figures 6D and S6C–S6E), concordant with profiles found in patients with *DNMT3A* mutation (Ley et al., 2010; Yan et al., 2011). We observed strong correlations between *DERARE* low methylation and mutations in *DNMT3A*, *FLT3*, and *NPM1* genes, as well as enhanced expression of

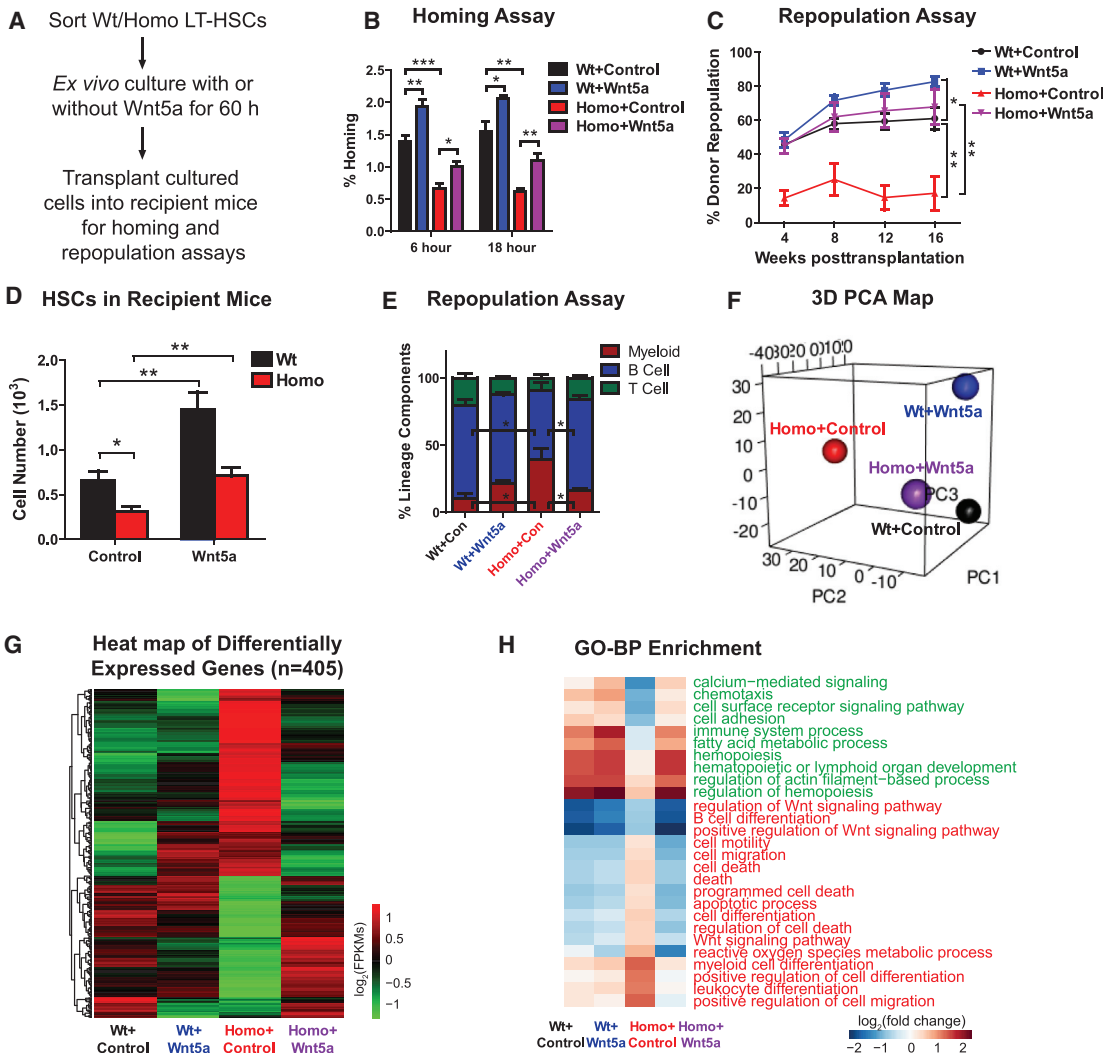


Figure 4. Wnt5a Rescues Functional and Molecular Defects in *DERARE*$^{\Delta}$ LT-HSCs

(A) Workflow of experiments with treatment of Wnt5a.

(B) BM cells from *WT* or *DERARE*$^{\Delta}$ mice were cultured with 500 ng/mL Wnt5a for 60 hr, and 2 × 10^5 CFDA SE-labeled cells were transplanted into irradiated mice. 6 and 18 hr later, BM was analyzed by fluorescence-activated cell sorting (FACS) for homed events (5 mice per group).

(C) Sorted *WT* or *DERARE*$^{\Delta}$ LT-HSCs were cultured with 500 ng/mL Wnt5a for 60 hr, and 500 cultured HSCs along with 1 × 10^5 rescue cells were transplanted into recipients. PB analysis for percentage of donor cells at the indicated number of weeks post-transplantation (n = 6 per group).

(D) Absolute numbers of HSPCs in recipient mice transplanted with Wnt5a-treated HSCs at 16 weeks post-transplantation (n = 6).

(E) At 16 weeks post-transplant, PB of recipient mice transplanted with HSCs was treated with control, or Wnt5a was analyzed for percentage of donor-derived B, T, and myeloid cells (n = 6).

(F) Unsupervised PCA analysis of the transcriptome of sorted cultured HSCs.

(G) Heatmap showing 405 differentially expressed genes using parameters: (1) FPKM > 1; (2) $\log_2$|fold change of FPKMs (*WT+Control* / *Homo+Control*)| > 1; (3) $\log_2$|fold change of FPKMs (*Homo+Wnt5a* / *Homo+Control*)| > 1; (4) FDR <0.05; (5) $p < 0.05$.

(H) GO-term enrichment analysis for Wnt5a rescued LT-HSCs. All data reflect mean ± SD from 2–3 independent experiments. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. See also Table S4.

HOXB genes (Figures 6E, 6F, S6F, and S6G). Finally, prognosis analysis revealed that both overall and event-free survival among patients with *DERARE* low methylation were worse than that among patients with *DERARE* high methylation, which is comparable to the difference between *DNMT3A* Mut and *DNMT3A* WT groups (Figures 6G, 6H, S6H, and S6I). Hence, hypomethylation on the *DERARE* correlates with *HOXB* cluster overexpression and worse prognosis in AML patients.

Targeted DNA Methylation on the *DERARE* Alleviates AML Leukemogenesis

To further validate whether *DERARE* hypomethylation contributes to leukemogenesis, we employed a CRISPR-Cas9-based epigenetic editing tool, which fuses either the active (WT) or inactive (ANV [AlaAsnVal]) *DNMT3A* catalytic domain to the C terminus of deactivated Cas9 (dCas9) nuclease via a short linker to enable DNA methylation adjacent to the single guide

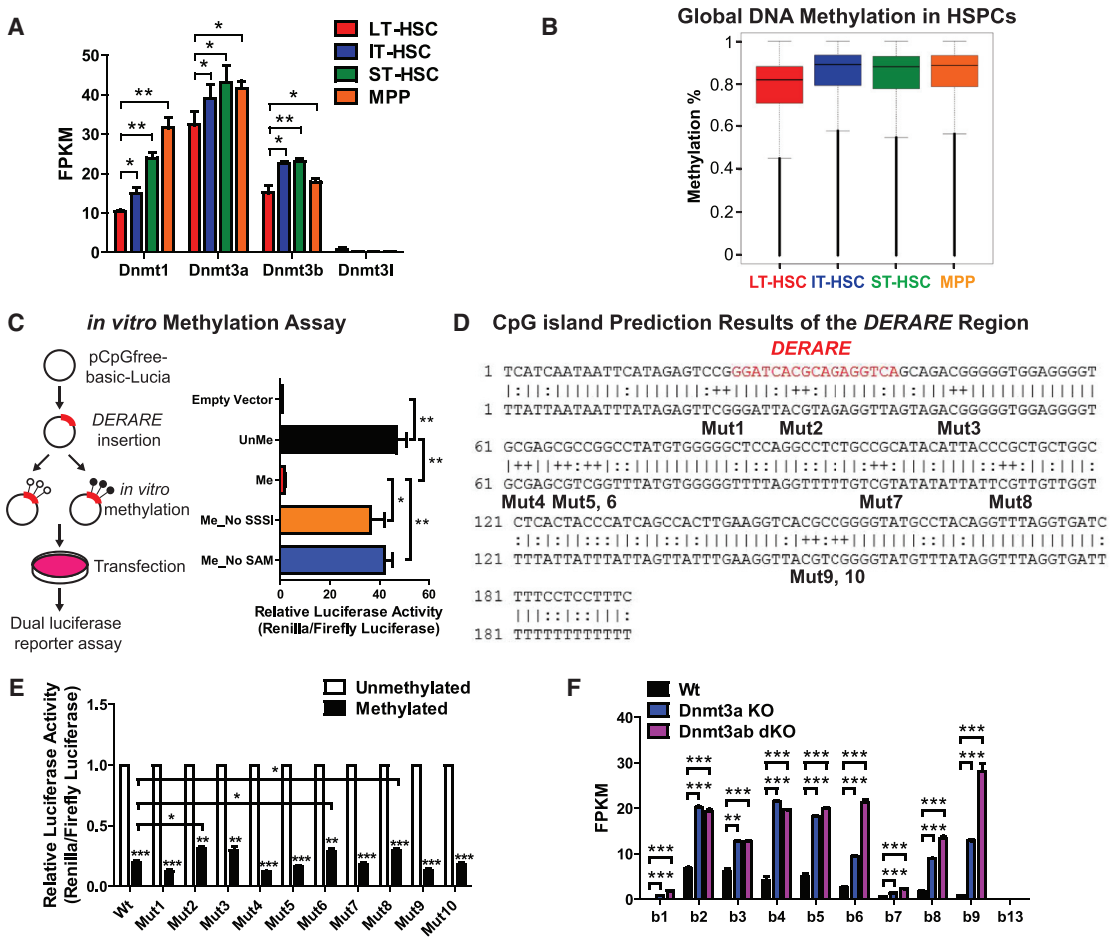


Figure 5. DNMT Enzyme-Mediated DNA Methylation on the *DERARE* Represses *Hoxb* Cluster Expression

(A) RNA-seq analysis of DNMTgenes in 4 adult HSPCs.

(B) Boxplots of the average DNA methylation across the mouse genome in 4 adult HSPCs.

(C) Luciferase activity of pCpGfree-basic-Lucia plasmids with methylated or unmethylated *DERARE* region was determined after co-transfection with pGL3 control vector in 293T cells. The luciferase activity was normalized to firefly luciferase activity. Unme, unmethylated; Me, methylated; SssI, CpG methyltransferase; SAM, S-adenosyl methionine.

(D) Location of CpG sites and their mutation in the mouse *DERARE* region. Upper row, original sequence. Lower row, bisulfite modified sequence. ++, CpG sites: C converted to T in non-CpG sites. Mutation, CpG to TpG. The *DERARE* is shown with red color.

(E) Luciferase activity of pCpGfree-basic-Lucia plasmids with WT or mutated CpG sites in the *DERARE* region was determined after co-transfection with pGL3 control vector in 293T cells. The luciferase activity was normalized to firefly luciferase activity.

(F) RNA-seq analysis of *Hoxb* cluster genes in HSCs sorted from *Dnmt3a* knockout mice and *Dnmt3a*, *Dnmt3b* double-knockout mice. For details, see the STAR Methods. All data reflect mean ± SD from 2 independent experiments. *p < 0.05; **p < 0.01; ***p < 0.001.

RNA (sgRNA) binding site (Vojta et al., 2016) (Figure 7A). We then constructed specific sgRNA targeting human *DERARE* into lentiviral vector and infected two human AML cell lines carrying *DNMT3A* mutation, OCI-AML2, and OCI-AML3. The methylation level at the targeted CpG sites in the *DERARE* region elevated 12- and 10-fold in OCI-AML3 and OCI-AML2 cells, respectively (Figures 7B and S7A). Although the algorithm predicted six potential off-target sites of the *DERARE* sgRNA, either no CpG sites existed in their neighboring region, or the DNA methylation at these off-target sites remained at the same level (Figures S7B–S7E; Table S6). We observed decreased expression of *Hoxb* cluster genes in *DNMT3A* WT-treated AML cells compared to *DNMT3A* ANV-treated cells, indicating that targeted DNA methylation on the *DERARE* efficiently suppresses *Hoxb* cluster expression (Figures 7C and S7F).

We next evaluated the effects of targeted *DERARE* methylation on leukemogenesis. The *in vitro* colony-forming unit (CFU) assay showed significant reduction in colony size and number in *DNMT3A* WT-treated AML cells (Figures 7D and S7G). We then transplanted equal number of *DNMT3A* WT or ANV-treated AML cells into lethally irradiated NOD/SCID/IL-2Rγ-null (NSG) mice and found that *DNMT3A* WT-treated cells had reduced spleen weight and number of leukemia stem cells (Figures 7E, 7F, S7H, and S7I). Moreover, *DNMT3A* WT-treated leukemic cells were more differentiated shown by CD11b and CD14 cell-surface expression and morphology (Figures 7G–7I and S7J–S7L). Finally, *DNMT3A* mediated *DERARE*

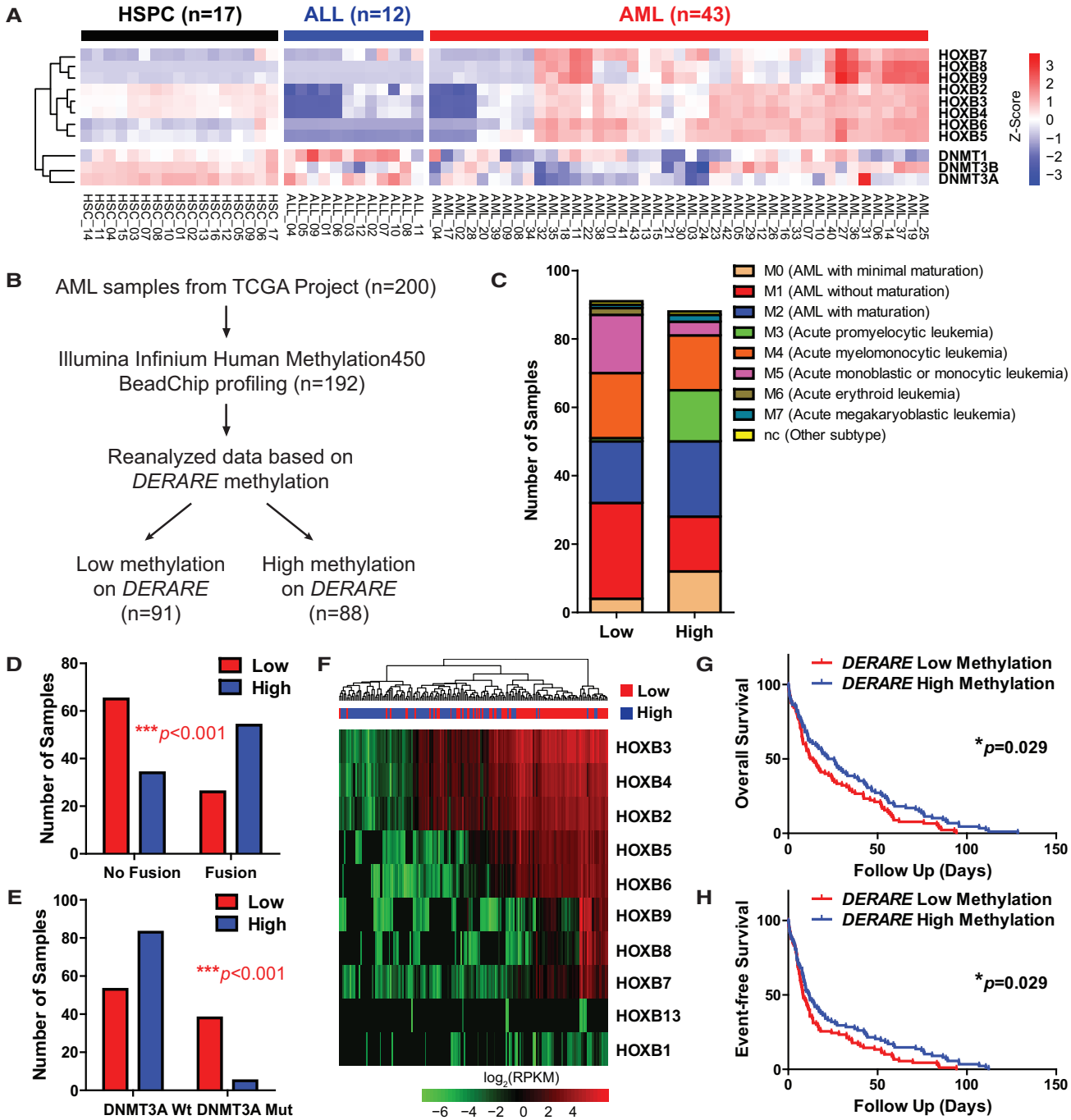


Figure 6. *DERARE* Hypomethylation Correlates with *DNMT3A* Mutation, *HOXB* Cluster Overexpression, and Worse Prognosis in AML Patients

(A) Heatmap of *HOXB* cluster and *DNMT* genes from RNA-seq data of Leucégène project consisting of 17 normal human HSPCs, 12 ALL, and 43 AML patient samples.

(B) Workflow for DNA methylation data analysis of 200 AML patients from TCGA project.

(C) Distribution of French-American-British (FAB) subtypes associated with *DERARE* low or high methylation in AML patients.

(D and E) Correlation between DNA methylation on the *DERARE* and gene fusion (D) or *DNMT3A* mutation (E). Fisher's exact test was used for p values. ***p < 0.001.

(F) Heatmap of unsupervised hierarchical cluster analysis of DNA methylation and transcriptome profile in AML patients.

(G and H) Correlation between DNA methylation on the *DERARE* and 5-year patient overall survival (G) or event-free survival (H) by Kaplan-Meier analysis. Log-rank (Mantel-Cox) test was used to calculate p values. *p < 0.05.

See also Figure S6 and Table S5.

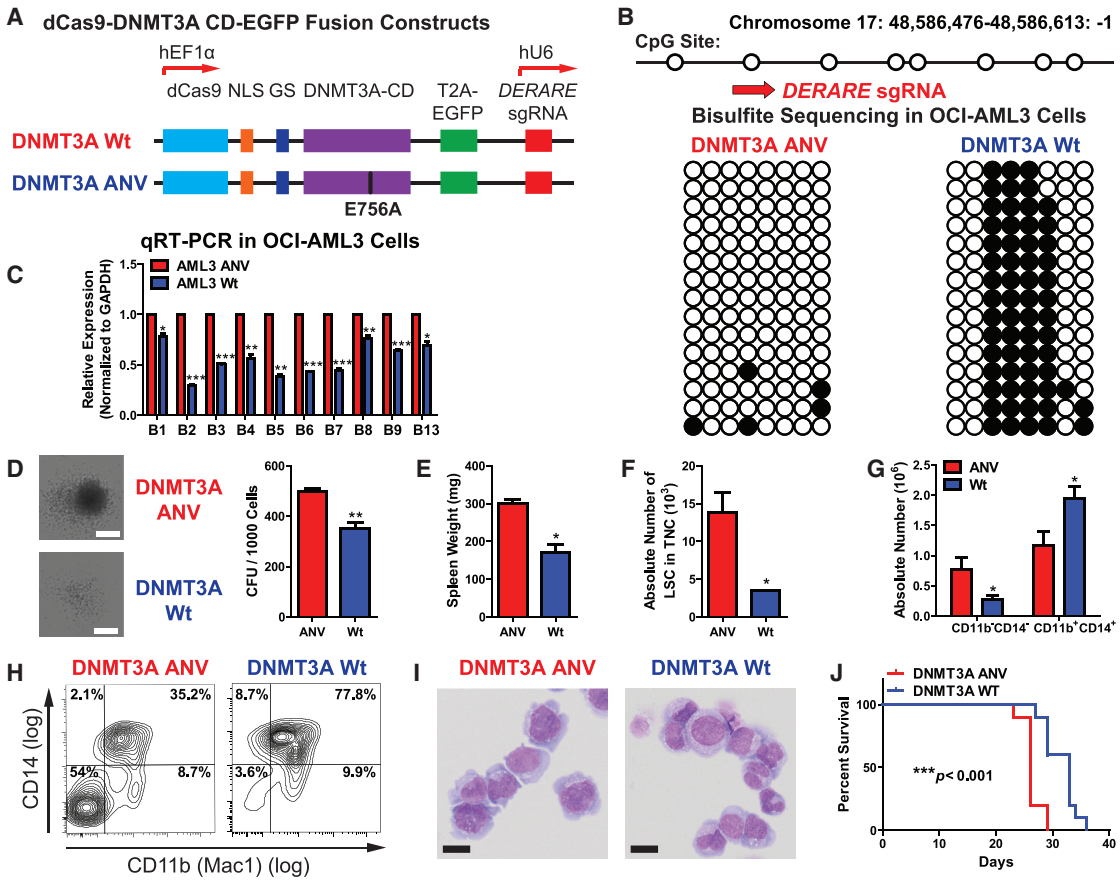


Figure 7. Targeted DNA Methylation on the *DERARE* Alleviates AML Leukemogenesis

(A) Schematic representation of the dCas9-DNMT3A-T2A-EGFP fusion protein and sgRNA targeting *DERARE*. NLS, nuclear localization signal; GS, Gly_4Ser peptide linker; CD, catalytic domain (amino acids P602-V912) of *DNMT3A*; E756A, mutation of ENV to ANV in *DNMT3A*.

(B) Bisulfite sequencing analysis of the CpG island neighboring the human *DERARE* region in OCI-AML3 cells. Arrow indicating the location of the *DERARE* sgRNA. Each clone is shown by a row, and each CpG site is arranged in columns. White and black circles represent unmethylated and methylated cytosine, respectively.

(C) Real-time qPCR showing reduced *HOXB* cluster expression in *DNMT3A* ANV-treated OCI-AML3 cells compared to control. Human *GAPDH* was used as internal control.

(D) (left) Representative images of CFU colonies after 2 weeks culture in methylcellulose medium. Scale bar, 1 mm. (right) Colony number of OCI-AML3 cells.

(E) Spleen weights of recipient mice injected with *DNMT3A* WT or ANV-treated OCI-AML3 cells at sacrifice.

(F) Absolute number of leukemia stem cells ($Lin^-CD34^+CD38^-TIM3^+CD99^+$) in recipient mice injected with *DNMT3A* WT or ANV-treated OCI-AML3 cells.

(G and H) Flow cytometry analysis of BM leukemic cells from recipient mice injected with *DNMT3A* WT or ANV-treated OCI-AML3 cells. Quantification is shown in (G), and representative images are shown in (H).

(I) Wright-Giemsa staining showing signs of maturation in *DNMT3A* ANV-treated OCI-AML3 cells. Scale bar, 10 μm.

(J) Kaplan-Meier survival curves of NSG mice injected with *DNMT3A* WT or ANV-treated OCI-AML3 cells. The p values were calculated by the log-rank (Mantel-Cox) test. All panels reflect mean ± SD from 2–3 independent experiments. *p < 0.05; **p < 0.01; ***p < 0.001.

See also Figure S7 and Table S6.

methylation significantly prolonged the survival of engrafted mice (Figures 7J and S7M). Together, these data suggest that targeted DNA methylation on the *DERARE* could delay the progression and improve survival of mice injected with the *DNMT3A* mutated AML cells.

DISCUSSION

This study has provided important insights into *cis*-regulatory mechanisms underlying the coordinate expression of the *Hoxb* cluster in response to RA and their critical roles in hematopoiesis and leukemogenesis. Transcriptome profiles in 17 hematopoietic types revealed that *Hoxb* cluster and RA signaling genes are highly enriched in LT-HSCs. Our analyses have demonstrated that the *DERARE* is part of a retinoid-dependent *cis*-regulatory module that globally controls expression of multiple *Hoxb* genes to sustain LT-HSC self-renewing potential. HSCs and their derivatives from $DERARE^{\Delta}$ mice become exhausted in serial transplantation assays. In the *Hoxb* complex, we observed a dynamic methylation pattern over the region containing the *DERARE* element as the HSCs exit self-renewal. Hence there is a direct correlation between HSC differentiation, decreasing

Hoxb gene expression and increasing methylation over the *DERARE* region. In AML patients, upregulated *HOXB* expression correlates with adverse clinical outcomes due to *DNMT3A* loss-of-function mutation, which causes low methylation on the *DERARE*. Thus, our findings elucidate a mechanism for how the retinoid-dependent and methylation-sensitive *DERARE* element maintains normal hematopoiesis but prevents leukemogenesis by coordinate regulation of *Hoxb* genes.

In hematopoiesis, it has been shown that rate of RA degradation by the hematopoietic niche is important in modulating the balance between maintenance and differentiation of HSCs (Ghiaur et al., 2013). In this study, we observed that RA signaling genes were enriched in LT-HSCs, and knockout of *Rarb* in HSCs led to reduced expression of *Hoxb* cluster genes and the absolute number as well as reconstitution ability of HSCs. This implies that *Rarb* might be the primary RAR involved in potentiating the *DERARE* element that mediates long-range regulation of *Hoxb* genes in HSCs. *Rarb* is not a DNA methylation-sensitive TF, since no CpG sites are found in RAR-RXR motif within the *DERARE*, and neither CXXC domain nor methyl-CpG-binding domain (MBD) exists in *Rarb*. Thus, hypomethylation over the *DERARE* region might be not a consequence of *Rarb* binding but due to other mechanisms. Upon ligand binding, RARs can change their conformation and recruit coactivators including members of the p160 family and thymine DNA glycosylase (TDG). The coactivators might release Dnmt3a and further recruit additional complexes, such as histone acetyltransferase, TET/TDG, and nucleosome remodeling complexes, which cooperate to activate transcription of target genes (Hassan et al., 2017). It will be worthwhile to explore how RA signaling regulates *Hoxb* gene expression through input into *DNMTs*, perhaps by releasing *Dnmt3a* through binding of *Rarb* and altering DNA methylation on the *DERARE*.

Hox genes have been reported as regulators of HSC maintenance and contributors in leukemogenesis, but the roles of *Hoxb* genes in hematopoiesis remains controversial (Alharbi et al., 2013). Some groups reported that *Hoxb4*-null mice exhibit a decrease in number of HSPCs, and *Hoxb3*/*Hoxb4* double-knockout mice show a greater reduction in number of HSPCs (Björnsson et al., 2003; Brun et al., 2004). Other groups deleted the entire *Hoxb* cluster (b1–b9) and did not observe defects in reconstitution and differentiation potential in fetal liver HSCs (Bijl et al., 2006). In our study, we found that knockout of the *DERARE* reduced expression of *Hoxb* cluster and led to defects in adult HSCs. One possible explanation for this discrepancy is that *Hoxb* gene expression is much higher in adult HSCs than fetal liver HSCs. Furthermore, although we have similar observations that *Hoxa* cluster expression was decreased in $DERARE^{\Delta}$ HSCs, we didn't find upregulation of *Hoxc* or *Hoxd* cluster genes, which is different from previous study in fetal liver HSCs that expression of *Hoxc4*, *Hoxc9*, and *Hoxc11* was upregulated in *Hoxb1–b9* knockout mice. Besides, our RNA-seq data showed reduced expression of HSC surface markers and TFs essential for HSC self-renewal and specification in $DERARE^{\Delta}$ LT-HSCs, which is in line with their defective reconstitution capacity.

DNMT3A heterozygous mutation has been found in 20%–30% of AML patients, with the majority (~60%) carrying a hotspot mutation at arginine 882 (R882) in the catalytic domain (Ley et al., 2010; Yan et al., 2011). The $DNMT3A^{R882H}$ mutation disrupts formation of the *DNMT3A*-associated tetramer and reduces methyltransferase activity in a dominant-negative manner (Russler-Germain et al., 2014), or alternatively it may interact with polycomb repressive complex 1 to block differentiation of HSCs and leukemic cells (Koya et al., 2016). Recent studies reported that $DNMT3A^{R882H}$ induced focal hypomethylation at specific CpGs, particularly at CREs such as enhancers, promotes leukemic transformation (Lu et al., 2016; Yang et al., 2016). In line with this model, we identified the *DERARE* as a retinoid-dependent CRE and elucidated the underlying mechanisms regarding how the *DERARE* controls *Hoxb* cluster expression and downstream non-canonical Wnt signaling in a methylation-dependent manner. Using the CRISPR-Cas9-based epigenetic editing system (Vojta et al., 2016), we provided the direct evidence that targeted DNA methylation on a specific CRE region could alleviate AML leukemogenesis. Since *HOXB* cluster genes showed dramatic increase in *DNMT3A* mutant AML patients, our work provides mechanistic insights into utilization of the DNA methylation on the *DERARE* as potential screening tools for therapeutic drugs against *DNMT3A*-mutated AML.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENTS AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
  - Mice
- METHOD DETAILS
  - HSC Isolation and Flow Cytometry
  - Hemavet Analysis
  - Transplantation Studies
  - PCR Genotyping
  - Cell Cycle Analysis
  - Single Cell Immunostaining
  - *In vitro* Methylation Assay
  - Luciferase Reporter Assay
  - qRT-PCR
  - Lentivirus Production and Transduction
  - Bisulfite Sequencing
  - Cytospin and Wright-Giemsa Staining
  - *In vitro* Colony-Forming Unit (CFU) Assay
  - *In vivo* Leukemia Analysis
  - RNA-Seq
  - ChIP-Seq and PCR Assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Fingerprint Coding Gene Analysis
  - Whole Genome Bisulfite Sequencing and Data Processing
  - Analysis of Published RNA-seq Data
  - Analysis of Published ATAC-seq Data
  - Analysis of TCGA Data from AML Patients
  - Statistical Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.stem.2018.04.012.

ACKNOWLEDGMENTS

We thank Prof. Pierre Chambon at Institut de Génétique et de Biologie Moléculaire et Cellulaire for gifting *Rarb*$^{fl/fl}$ mice, M. Miller for model image modification, K. Tannen for proofreading and editing, and M. Hembree, A. Moran, D. Dukes, H. Marshall, K. Zapien, M. Durnin, T. Corbin, B. Lewis, J. Park, L. Blunk, C. Semerad, M. Katt, K. Delventhal, and R. Egidy for technical support. We are grateful to the Li and Krumlauf lab members for scientific discussions and critical reading of the manuscript. This work was funded by the Stowers Institute for Medical Research and the American Society of Hematology (ASH) Scholar Award.

AUTHOR CONTRIBUTIONS

Conceptualization, P.Q., R.K., and L.L.; Methodology, P.Q. and B.D.K.; Formal Analysis, M.G., S.C., A. Paulson, A.B., H.L., and J.S.H.; Investigation, P.Q., B.D.K., X.C.H., C.N., Y.H., C.Z., T.P., K.H., A. Peak, and A. Perera; Validation, Z.L., H.X., J.M.P., D.H., F.T., M.Z., and A.V.; Resources, Y.A.; Data Curation, M.G., S.C., A. Paulson, and H.L.; Writing, P.Q., R.K., and L.L.; Supervision, L.L. and R.K.; Funding Acquisition, P.Q., R.K., and L.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 10, 2017
Revised: September 15, 2017
Accepted: April 13, 2018
Published: May 3, 2018

REFERENCES

Ahn, Y., Mullan, H.E., and Krumlauf, R. (2014). Long-range regulation by shared retinoic acid response elements modulates dynamic expression of posterior Hoxb genes in CNS development. Dev. Biol. *388*, 134–144.

Alexander, T., Nolte, C., and Krumlauf, R. (2009). Hox genes and segmentation of the hindbrain and axial skeleton. Annu. Rev. Cell Dev. Biol. *25*, 431–456.

Alharbi, R.A., Pettengell, R., Pandha, H.S., and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia *27*, 1000–1008.

Bagger, F.O., Sasivarevic, D., Sohi, S.H., Laursen, L.G., Pundhir, S., Sønderby, C.K., Winther, O., Rapin, N., and Porse, B.T. (2016). BloodSpot: A database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. Nucleic Acids Res. *44* (D1), D917–D924.

Bijl, J., Thompson, A., Ramirez-Solis, R., Krosl, J., Grier, D.G., Lawrence, H.J., and Sauvageau, G. (2006). Analysis of HSC activity and compensatory Hox gene expression profile in Hoxb cluster mutant fetal liver cells. Blood *108*, 116–122.

Björnsson, J.M., Larsson, N., Brun, A.C., Magnusson, M., Andersson, E., Lundström, P., Larsson, J., Repetowska, E., Ehinger, M., Humphries, R.K., and Karlsson, S. (2003). Reduced proliferative capacity of hematopoietic stem cells deficient in Hoxb3 and Hoxb4. Mol. Cell. Biol. *23*, 3872–3883.

Brun, A.C., Björnsson, J.M., Magnusson, M., Larsson, N., Leveén, P., Ehinger, M., Nilsson, E., and Karlsson, S. (2004). Hoxb4-deficient mice undergo normal hematopoietic development but exhibit a mild proliferation defect in hematopoietic stem cells. Blood *103*, 4126–4133.

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D.B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., et al. (2014). Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell *15*, 507–522.

Cabezas-Wallscheid, N., Buettner, F., Sommerkamp, P., Klimmeck, D., Ladel, L., Thalheimer, F.B., Pastor-Flores, D., Roma, L.P., Renders, S., Zeisberger, P., et al. (2017). Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. Cell *169*, 807–823.e19.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat. Genet. *44*, 23–31.

Challen, G.A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., et al. (2014). Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. Cell Stem Cell *15*, 350–364.

Chanda, B., Ditadi, A., Iscove, N.N., and Keller, G. (2013). Retinoic acid signaling is essential for embryonic hematopoietic stem cell development. Cell *155*, 215–227.

Chapellier, B., Mark, M., Bastien, J., Dierich, A., LeMeur, M., Chambon, P., and Ghyselinck, N.B. (2002). A conditional floxed (loxP-flanked) allele for the retinoic acid receptor beta (RARbeta) gene. Genesis *32*, 91–94.

Chen, J.Y., Miyanishi, M., Wang, S.K., Yamazaki, S., Sinha, R., Kao, K.S., Seita, J., Sahoo, D., Nakauchi, H., and Weissman, I.L. (2016). Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature *530*, 223–227.

Cimmino, L., Dawlaty, M.M., Ndiaye-Lobry, D., Yap, Y.S., Bakogianni, S., Yu, Y., Bhattacharyya, S., Shaknovich, R., Geng, H., Lobry, C., et al. (2015). TET1 is a tumor suppressor of hematopoietic malignancy. Nat. Immunol. *16*, 653–662.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203.

de Graaf, C.A., Choi, J., Baldwin, T.M., Bolden, J.E., Fairfax, K.A., Robinson, A.J., Biben, C., Morgan, C., Ramsay, K., Ng, A.P., et al. (2016). Haemopedia: An Expression Atlas of Murine Hematopoietic Cells. Stem Cell Reports *7*, 571–582.

De Kumar, B., Parrish, M.E., Slaughter, B.D., Unruh, J.R., Gogol, M., Seidel, C., Paulson, A., Li, H., Gaudenz, K., Peak, A., et al. (2015). Analysis of dynamic changes in retinoid-induced transcription and epigenetic profiles of murine Hox clusters in ES cells. Genome Res. *25*, 1229–1243.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.

Deschamps, J., and Duboule, D. (2017). Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. Genes Dev. *31*, 1406–1416.

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

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258.

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

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Ghiaur, G., Yegnasubramanian, S., Perkins, B., Gucwa, J.L., Gerber, J.M., and Jones, R.J. (2013). Regulation of human hematopoietic stem cell self-renewal by the microenvironment's control of retinoic acid signaling. Proc. Natl. Acad. Sci. USA *110*, 16121–16126.

Göttgens, B. (2015). Regulatory network control of blood stem cells. Blood *125*, 2614–2620.

Guillamot, M., Cimmino, L., and Aifantis, I. (2016). The impact of DNA methylation in hematopoietic malignancies. Trends Cancer *2*, 70–83.

Hassan, H.M., Kolendowski, B., Isovic, M., Bose, K., Dranse, H.J., Sampaio, A.V., Underhill, T.M., and Torchia, J. (2017). Regulation of active DNA

demethylation through RAR-mediated recruitment of a TET/TDG complex. Cell Rep. *19*, 1685–1697.

Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. *13*, R83.

Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R.V., McConkey, M.E., Thielke, A., Aster, J.C., Regev, A., and Ebert, B.L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. Nat. Biotechnol. *32*, 941–946.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat. Genet. *45*, 1198–1206.

Hu, Y., and Smyth, G.K. (2009). ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods *347*, 70–78.

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

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338–342.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Koya, J., Kataoka, K., Sato, T., Bando, M., Kato, Y., Tsuruta-Kishino, T., Kobayashi, H., Narukawa, K., Miyoshi, H., Shirahige, K., and Kurokawa, M. (2016). DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation. Nat. Commun. *7*, 10924.

Kyba, M., Perlingeiro, R.C., and Daley, G.Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. Cell *109*, 29–37.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

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

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. N. Engl. J. Med. *363*, 2424–2433.

Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A., Hoadley, K., Triche, T.J., Jr., Laird, P.W., Baty, J.D., et al.; Cancer Genome Atlas Research Network (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. *368*, 2059–2074.

Li, L.C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. Bioinformatics *18*, 1427–1431.

Lu, R., Wang, P., Parton, T., Zhou, Y., Chrysovergis, K., Rockowitz, S., Chen, W.Y., Abdel-Wahab, O., Wade, P.A., Zheng, D., and Wang, G.G. (2016). Epigenetic perturbations by Arg882-mutated DNMT3A potentiate aberrant stem cell gene-expression program and acute leukemia development. Cancer Cell *30*, 92–107.

Macrae, T., Sargeant, T., Lemieux, S., Hébert, J., Deneault, E., and Sauvageau, G. (2013). RNA-Seq reveals spliceosome and proteasome genes as most consistent transcripts in human cancer cells. PLoS ONE *8*, e72884.

Nemeth, M.J., Topol, L., Anderson, S.M., Yang, Y., and Bodine, D.M. (2007). Wnt5a inhibits canonical Wnt signaling in hematopoietic stem cells and enhances repopulation. Proc. Natl. Acad. Sci. USA *104*, 15436–15441.

Nolte, C., Amores, A., Nagy Kovács, E., Postlethwait, J., and Featherstone, M. (2003). The role of a retinoic acid response element in establishing the anterior neural expression border of Hoxd4 transgenes. Mech. Dev. *120*, 325–335.

O'Brien, A., and Bailey, T.L. (2014). GT-Scan: Identifying unique genomic targets. Bioinformatics *30*, 2673–2675.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An evolving paradigm for stem cell biology. Cell *132*, 631–644.

Parker, H.J., and Krumlauf, R. (2017). Segmental arithmetic: Summing up the Hox gene regulatory network for hindbrain development in chordates. Wiley Interdiscip. Rev. Dev. Biol. Published online November 6, 2017. https://doi.org/10.1002/wdev.286.

Purton, L.E., Bernstein, I.D., and Collins, S.J. (2000). All-trans retinoic acid enhances the long-term repopulating activity of cultured hematopoietic stem cells. Blood *95*, 470–477.

Qian, P., He, X.C., Paulson, A., Li, Z., Tao, F., Perry, J.M., Guo, F., Zhao, M., Zhi, L., Venkatraman, A., et al. (2016). The Dlk1-Gtl2 locus preserves LT-HSC function by inhibiting the PI3K-mTOR pathway to restrict mitochondrial metabolism. Cell Stem Cell *18*, 214–228.

Riddell, J., Gazit, R., Garrison, B.S., Guo, G., Saadatpour, A., Mandal, P.K., Ebina, W., Volchkov, P., Yuan, G.C., Orkin, S.H., and Rossi, D.J. (2014). Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. Cell *157*, 549–564.

Russler-Germain, D.A., Spencer, D.H., Young, M.A., Lamprecht, T.L., Miller, C.A., Fulton, R., Meyer, M.R., Erdmann-Gilmore, P., Townsend, R.R., Wilson, R.K., and Ley, T.J. (2014). The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. Cancer Cell *25*, 442–454.

Seita, J., Sahoo, D., Rossi, D.J., Bhattacharya, D., Serwold, T., Inlay, M.A., Ehrlich, L.I., Fathman, J.W., Dill, D.L., and Weissman, I.L. (2012). Gene Expression Commons: An open platform for absolute gene expression profiling. PLoS ONE *7*, e40321.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W., et al.; HALT Pan-Leukemia Gene Panel Consortium (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature *506*, 328–333.

Studer, M., Popperl, H., Marshall, H., Kuroiwa, A., and Krumlauf, R. (1994). Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1. Science *265*, 1728–1732.

Sugimura, R., He, X.C., Venkatraman, A., Arai, F., Box, A., Semerad, C., Haug, J.S., Peng, L., Zhong, X.B., Suda, T., and Li, L. (2012). Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. Cell *150*, 351–365.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442–449.

Valarché, I., de Graaff, W., and Deschamps, J. (1997). A 3′ remote control region is a candidate to modulate Hoxb-8 expression boundaries. Int. J. Dev. Biol. *41*, 705–714.

Venkatraman, A., He, X.C., Thorvaldsen, J.L., Sugimura, R., Perry, J.M., Tao, F., Zhao, M., Christenson, M.K., Sanchez, R., Yu, J.Y., et al. (2013). Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. Nature *500*, 345–349.

Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., Klasić, M., and Zoldoš, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. *44*, 5615–5628.

Yan, X.J., Xu, J., Gu, Z.H., Pan, C.M., Lu, G., Shen, Y., Shi, J.Y., Zhu, Y.M., Tang, L., Zhang, X.W., et al. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat. Genet. *43*, 309–315.

Yang, L., Rodriguez, B., Mayle, A., Park, H.J., Lin, X., Luo, M., Jeong, M., Curry, C.V., Kim, S.B., Ruau, D., et al. (2016). DNMT3A loss drives enhancer hypomethylation in FLT3-ITD-associated leukemias. Cancer Cell *29*, 922–934.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-mouse CD3-PE-Cy5 (clone: 145-2C11)	eBioscience	Cat# 15-0031-83, RRID: 4277752
anti-mouse CD4-PE-Cy5 (clone: RM4-5)	eBioscience	Cat# 15-0042-83, RRID: E06075-1632
anti-mouse CD8-PE-Cy5 (clone: 53-6.7)	eBioscience	Cat# 15-0081-83, RRID: E06085-1632
anti-mouse Mac1-PE-Cy5 (clone: M1/70)	eBioscience	Cat# 15-0112-83, RRID: E06098-1632
anti-mouse Gr1-PE-Cy5 (clone: RB6-8C5)	eBioscience	Cat# 15-5931-82, RRID: E06253-1630
anti-mouse B220-PE-Cy5 (clone: RA3-6B2)	eBioscience	Cat# 15-0452-83, RRID: E06148-1634
anti-mouse IgM-PE-Cy5 (clone: II-41)	eBioscience	Cat# 15-5790-82, RRID: 4272444
anti-mouse Ly-6A/E (Sca1)-PE-Cy7 (clone: D7)	eBioscience	Cat# 25-5981-82, RRID: 4274408
anti-mouse CD117 (cKit)-APC (clone: 2B8)	eBioscience	Cat# 17-1171-83, RRID: E07204-1635
anti-mouse CD34-FITC (clone: RAM34)	eBioscience	Cat# 11-0341-85, RRID: E00265-1634
anti-mouse CD135 (Flk2)-Biotin (clone: A2F10)	eBioscience	Cat# 13-1351-85, RRID: 4278456
anti-mouse CD49b-PE (clone: HMα2)	BioLegend	Cat# 103506, RRID: B178546
anti-mouse CD45.2-FITC (clone: 104)	eBioscience	Cat# 11-0454-85, RRID: E00316-1630
anti-mouse CD45.2-V500 (clone: 104)	BD Biosciences	Cat# 562129, RRID: 5107958
anti-mouse CD45.1-PE-Cy5 (clone: A20)	eBioscience	Cat# 15-0453-82, RRID: E06151-1634
anti-mouse Mac1-PE-Cy7 (clone: M1/70)	eBioscience	Cat# 25-0112-82, RRID: 4289817
anti-mouse Gr1-AF780 (clone: RB6-8C5)	eBioscience	Cat# 47-5931-82, RRID: 4291984
anti-mouse CD3-APC (clone: 145-2C11)	eBioscience	Cat# 17-0031-83, RRID: 4277756
anti-mouse B220-PE (clone: RA3-6B2)	eBioscience	Cat# 12-0452-83, RRID: E01248-1633
anti-human Lineage Cocktail-FITC	BioLegend	Cat# 348701, RRID: B199910
anti-human CD34-APC (clone: 581)	BD Biosciences	Cat# 555824, RRID: 4338604
anti-human CD38-PE-Cy7 (clone: HIT2)	BD Biosciences	Cat# 560677, RRID: 7079938
anti-human CD366 (TIM3)-BV421 (clone: 7D3)	BD Biosciences	Cat# 565562, RRID: 6244943
anti-human CD99-PE (clone: TÜ12)	BD Biosciences	Cat# 555689, RRID: 7121967
anti-human CD14-FITC (clone: M5E2)	BD Biosciences	Cat# 557153, RRID: 6173924
anti-human CD11b-APC-Cy7 (clone: ICRF44)	BD Biosciences	Cat# 557754, RRID: 7082517
FITC Mouse Anti-Ki-67 Set	BD Biosciences	Cat# 556026, RRID: 4156624
Streptavidin APC-eFluor 780	eBioscience	Cat# 47-4317-82, RRID: E08466-1640
Rabbit IgG Isotype Control	Invitrogen	Cat# 02-6102, RRID: 630554A
HoxB3 (H-50) Antibody	Santa Cruz Biotechnology	Cat# sc-28606X, RRID: B1612
NFATc1 Antibody (7A6)	Santa Cruz Biotechnology	Cat# sc-7294, RRID: I2214
Donkey anti-Mouse IgG, Alexa Fluor 488	Invitrogen	Cat# A-11001, RRID: 1182671
Bacterial and Virus Strains		
5-alpha Competent *E. coli*	New England Biolabs	Cat# C2987H
JM109 Competent E.coli Cells	Promega	Cat# L2001
GT115 Competent E.coli cells	InvivoGen	Cat# gt115-11
Chemicals, Peptides, and Recombinant Proteins		
7-AAD	Thermo Fisher Scientific	Cat# A1310
DAPI	Sigma-Aldrich	Cat# D9542
Poly (I:C)	Sigma-Aldrich	Cat# P9582
Polybrene	Sigma-Aldrich	Cat# AL-118
CpG Methyltransferase (M.SssI)	New England Biolabs	Cat# M0226L
CFDA-SE	Molecular Probes	Cat# V12883

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TRIzol	Thermo Fisher Scientific	Cat# 15596018
Lipofectamine 2000	Thermo Fisher Scientific	Cat# 11668019
DMEM, high glucose	Thermo Fisher Scientific	Cat# 10566016
a-MEM	Thermo Fisher Scientific	Cat# 32561-037
Superscript III Reverse Transcriptase	Thermo Fisher Scientific	Cat# 8080044
Ambion RNase Inhibitor	Thermo Fisher Scientific	Cat# AM2682
Fetal Bovine Serum	Thermo Fisher Scientific	Cat# 10082139
MethoCult H4034 Optimum	StemCell Technologies	Cat# 04034
Recombinant Murine SCF	PeproTech	Cat# 250-03
Recombinant Murine TPO	PeproTech	Cat# 315-14
Recombinant Human/Mouse Wnt-5a Protein	R&D Systems	Cat# 645-WN-010
Critical Commercial Assays		
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69504
QIAGEN Plasmid Mini Kit	QIAGEN	Cat# 12125
QIAGEN Plasmid Midi Kit	QIAGEN	Cat# 12145
QIAquick PCR Purification Kit	QIAGEN	Cat# 28106
EpiTect Bisulfite Kit	QIAGEN	Cat# 59104
REDExtract-N-Amp Tissue PCR Kit	Sigma-Aldrich	Cat# XNAT-100RXN
Dual-Glo Luciferase Assay System	Promega	Cat# E2920
pGEM-T Easy Vector Systems	Promega	Cat# A1380
ZymoTaq PreMix	Zymo Research	Cat# E2004
Pico Methyl-Seq Library Prep Kit	Zymo Research	Cat# D5456
True MicroChIP kit	Diagenode	Cat# C01010130
SMART-Seq Ultra Low Input RNA Kit	Clontech Laboratories	Cat# 634936
Wright-Giemsa Stain Kit	American MasterTech	Cat# KTWGIPT
ABI High Capacity cDNA Reverse Transcription Kit	Applied Biosystems	Cat# 4368814
Power SYBR Green PCR Master Mix	Thermo Fisher Scientific	Cat# 4367659
Deposited Data		
DERARE knockout, Wnt5a Rescue RNA-seq	This paper	GEO: GSE87556
Methyl-seq	This paper	GEO: GSE89490
HSPCs, Progenitor, and Lineage Cell RNA-seq	(Qian et al., 2016)	ArrayExpress: E-MTAB-2923, E-MTAB-3079
Leucégène Project RNA-seq	(Macrae et al., 2013)	GEO: GSE48173
Dnmt3a knockout and Dnmt3a, Dnmt3b Double-knockout RNA-seq	(Challen et al., 2014)	GEO: GSE50793
Human Hematopoietic Cell ATAC-seq	(Corces et al., 2016)	UCSC Genome Browser Track Hub (https://s3-us-west-1.amazonaws.com/chang-public-data/2016_NatGen_ATAC-AML/hub.txt)
TCGA Data from AML Patients	(Ley et al., 2013)	https://gdc.cancer.gov/
Experimental Models: Cell Lines		
293T	ATCC	CRL-3216
OCI-AML2	DSMZ	ACC-99
OCI-AML3	DSMZ	ACC-582
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	The Jackson Laboratory	Stock No: 000664
Mouse: B6.SJL-*Ptprc*^{a} *Pepc*^{b}/BoyJ (*ptprc*)	The Jackson Laboratory	Stock No: 002014
Mouse: NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG)	The Jackson Laboratory	Stock No: 005557
Mouse: B6.Cg-Tg(Mx1-cre)1Cgn/J (Mx1-cre)	The Jackson Laboratory	Stock No: 003556

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: DERARE-KO	(Ahn et al., 2014)	N/A
Mouse: $Rarb^{fl/fl}$	(Chapellier et al., 2002)	N/A
Oligonucleotides		
See Table S7 for list of primers.	This paper	N/A
Recombinant DNA		
pCpG-free basic-Lucia	InvivoGen	Cat# pcpgf-baslc
pGL3-Control	Promega	Cat# E1741
pL-CRISPR-EFS-GFP	(Heckl et al., 2014)	Addgene #12259
dCas9-DNMT3A-EGFP	(Vojta et al., 2016)	Addgene #71666
dCas9-DNMT3A (E756A)-EGFP	(Vojta et al., 2016)	Addgene #71685
pL-CRISPR-EF1α-Cas9-DNMT3A-EGFP-*DERARE*-sgRNA	This paper	*DNMT3A* Wt
pL-CRISPR-EF1α-Cas9-DNMT3A (E756A)-EGFP-*DERARE*-sgRNA	This paper	*DNMT3A* ANV
Software and Algorithms		
Data analysis (FACS)	FlowJo 10, TreeStar	http://docs.flowjo.com/vx/faq/general-faq/tree-star-flowjo/
Data analysis (Imaging)	ImageJ	https://imagej.nih.gov/ij/index.html
Data analysis	Prism 5, Graph Pad	https://www.graphpad.com/scientific-software/prism/
Tophat 2.0.13	(Trapnell et al., 2009)	https://ccb.jhu.edu/software/tophat/index.shtml
Cufflinks (v2.1.1)	(Trapnell et al., 2010)	http://cole-trapnell-lab.github.io/cufflinks/releases/v2.1.1/
GOstats tool	(Falcon and Gentleman, 2007)	https://omictools.com/gostats-tool
bowtie2 (2.2.0)	(Langmead and Salzberg, 2012)	https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.2.0/
MACS2	(Feng et al., 2012)	https://github.com/taoliu/MACS/tree/master/MACS2
Bismark Methylation Extractor	(Krueger and Andrews, 2011)	https://www.bioinformatics.babraham.ac.uk/projects/bismark/
R language (3.2.2) with bsseq (1.6.0)	(Hansen et al., 2012)	https://www.r-project.org/
GenomeGraphs (1.30.0)	(Durinck et al., 2009)	https://bioconductor.riken.jp/packages/3.2/bioc/html/GenomeGraphs.html
STAR	(Dobin et al., 2013)	https://github.com/alexdobin/STAR/releases/tag/STAR_2.4.0h
R package NMF	(Gaujoux and Seoighe, 2010)	https://cran.r-project.org/web/packages/NMF/index.html
Other		
Gene Expression Commons	(Seita et al., 2012)	https://gexc.riken.jp/
Bloodspot	(Bagger et al., 2016)	http://servers.binf.ku.dk/bloodspot/
Haemosphere	(de Graaf et al., 2016)	http://haemosphere.org/

CONTACT FOR REAGENTS AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Linheng Li (lil@stowers.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

293T cell line was obtained from the American Type Culture Collection (ATCC) and cultured as recommended. The human *DNMT3A*-mutated AML cell lines OCI-AML2 and OCI-AML3 were purchased from the Leibniz-Institute DSMZ (https://www.dsmz.de) and cultured with 80% αMEM+ 20% h.i. FBS. All procedures involving human cell lines were approved by the Institutional Biosafety Committee (IBC) of Stowers Institute for Medical Research.

Mice

DERARE knockout (*DERARE*$^{\Delta}$) mice were established as previously described (Ahn et al., 2014), in which a 52 bp fragment containing the conserved RAR-RXR consensus binding site and a spacer of 5 bases (DR5-RARE) was deleted (Sequence of deleted region is shown in Figure S2C). All the mice were backcrossed onto C57BL/6J background. *Rarb*$^{fl/fl}$ mice were kindly gifted from Prof. Pierre Chambon at Institut de Génétique et de Biologie Moléculaire et Cellulaire (Chapellier et al., 2002) and crossed to *Mx1-Cre* mice. Deletion of floxed alleles was carried out by 7 intraperitoneal injections (250 μg per mouse) of pIpC (Sigma) in PBS every other day. *Ptprc* and NOD/SCID/IL-2Rγnull (NSG) mice were purchased from Jackson Laboratory. Male and female mice from 8 to 12 weeks old were used for all studies. All mice were housed in the AAALAC-accredited animal facility at the Stowers Institute for Medical Research (SIMR) and handled according to SIMR and National Institutes of Health guidelines. All procedures involving experimental animals were approved by the Institutional Animal Care and Use Committee (IACUC) of SIMR under protocols 2016-0164 and 2015-0148.

METHOD DETAILS

HSC Isolation and Flow Cytometry

HSPCs, progenitors, and lineage cells were harvested from BM (femur and tibia) and peripheral blood of C57BL/6J mice in PBS with 2% heat inactivated fetal bovine serum (FBS). Red blood cells were lysed using a 0.16 M ammonium chloride solution, and the cells were filtered with 70 μm strainers to generate single cell suspensions. For HSC identification, cells were stained with antibodies against Sca-1 (D7), c-Kit (2B8), CD34 (RAM34), Flk2 (A2F10), CD49b (HMα2), together with lineage cocktail including CD3 (145-2C11), CD4 (RM4-5), CD8 (53-6.7), Mac-1 (M1/70), Gr1 (RB6-8C5), CD45R (B220, RA3-6B2), IgM (II-41) and Ter119 (TER-119). For progenitors and lineage cells, cells were stained with antibodies as previously described (Qian et al., 2016). For leukemic cell analysis, cells were stained with antibodies against CD34 (581), CD38 (HIT2), TIM3 (7D3), CD99 (TÜ12), CD14 (M5E2), CD11b (ICRF44), together with human lineage cocktail (CD3, CD14, CD19, CD20, CD56). All the antibodies were purchased from BD Biosciences, Biolegend and eBioscience. Antibody staining was performed at 4°C for 30 min, and then stained with the viability dyes 7-Aminoactinomycin D (7-AAD, 0.1 mg/ml) to exclude dead cells. Cell sorting and analyses were performed on MoFlo (Dako), InFlux Cell Sorter (BD Biosciences), CyAn ADP (Dako), and/or MACSQuant (Miltenyi Biotec). Data analysis was performed using FlowJo software V7.0.

Hemavet Analysis

30 ul anticoagulated peripheral blood samples were collected from the submandibular vein of mice and then analyzed on the Hemavet HV 1700FS hematology analyzer (Drew Scientific) programmed with mouse hematology settings.

Transplantation Studies

In vivo *Homing Assay*

Whole bone marrow (WBM) cells or cultured cells from CD45.2 mice was labeled with 5 μM 5-(and −6)-carboxyfluorescein diacetate succinimidyl ester (CFDA SE) (Molecular Probes) at 37°C for 10 min, washed for 3 times, and 1 × 10^6 WBM or 2 × 10^5 cultured cells were transplanted into lethally irradiated *ptprc* mice. After 6 and 18 hours, femurs and tibias were flushed, and CFDA SE$^+$ cells were determined.

Competitive Reconstitution Assay

Competitive reconstitution assays were performed by intravenously transplantation of a series (25,000, 75,000 and 200,000) of donor-derived WBM cells or 100 sorted LT-HSCs (LSKCD34$^-$Flk2$^-$CD49b^{lo}) from wt or *DERARE*$^{\Delta}$ mice (CD45.2), together with 2 × 10^5 rescue cells (CD45.1) into groups of ten lethally irradiated (10 Gy) *Ptprc* recipient mice. Baytril water was given to recipient mice 3 days before irradiation and continued for another 2 weeks after irradiation. Competitive repopulation unit (CRU) frequency was measured using ELDA (Extreme limiting dilution assay) software (Hu and Smyth, 2009), in which successful engraftment was defined as the presence of a distinct CD45.2$^+$ CD45.1$^-$ population (≥1%) above a background set by parallel analysis of animals transplanted with only rescue BM., primary transplant recipients were sacrificed. BM cells were dissected from femur and tibia, and then transplanted mouse-to-mouse at a dosage of 1 × 10^6 cells into irradiated secondary recipient mice.

Wnt5a Rescue Transplantation

For Wnt5a rescue transplantation, 5,000 LT-HSCs were sorted from wt or *DERARE*$^{\Delta}$ mice and cultured with or without mouse Wnt5a (500 ng/ml, R&D system) in HSC culture medium (StemSpan media supplemented with 10 ng/ml recombinant mouse stem cell factor and 20 ng/ml recombinant mouse thrombopoietin). After 60 hours, cells were collected and counted using EC800 analyzer (Sony

Biotechnology Inc), and 500 cultured HSCs along with 1 × 10^5 rescue BM cells were injected into recipient ptprc mice. Donor-derived engraftment was determined at 4-week intervals after transplantation, followed by collection of peripheral blood, red blood cell lysis, and staining with CD45.1 (recipient) and CD45.2 (donor) antibodies. Multilineage reconstitution was calculated on CyAn ADP (Dako) by CD3 (for T lymphoid), B220 (for B lymphoid) and Gr1, Mac-1 (for myeloid) gating on donor cells.

PCR Genotyping

DNA was extracted from tail samples and PCR was performed using Extract-N-Amp Tissue PCR Kit (Sigma-Aldrich) according to the manufacturer's instructions.

Cell Cycle Analysis

Cell cycle analysis was performed with FITC mouse anti-human Ki67 set (BD PharMingen) according to the manufacturer's instructions. Briefly, 5 × 10^6 BM cells were isolated and stained with HSC antibodies as described above. Cells were fixed by 4% paraformaldehyde at 4°C overnight or room temperature (RT) for 1 hour, and then permeabilized with 0.2% Triton X-100 on ice for 15 mins. After washing with PBS containing 2% FBS, cells were incubated with Ki-67 antibody at RT for 1 hour in the dark, and 0.1 mg/ml 4',6-diamidino-2-phenylindole (DAPI) at RT for another 5 min, followed by flow cytometric analysis with InFlux Cell Sorter.

Single Cell Immunostaining

HSCs were sorted onto Poly-L-lysine coating slides, which were placed in a moisture chamber on ice for 30 min to permit cells to settle onto the slides. Cells were fixed with chilled methanol at RT for 10 min, blocked with universal blocking reagent (BioGenex) at RT for 30 min, and stained with mouse *NFATC1* antibody (7A6) or mouse IgG control (Abcam) at 4°C overnight. Cells were then stained with Alexa Fluor 488 donkey anti-mouse IgG (Thermo Fisher Scientific) at 4°C for 30 mins. Images were taken on Axiomager Z1 (Zeiss) with AxioVision 4.7.2.0, and staining intensity per image was quantified by ImageJ program.

In vitro Methylation Assay

A 402-bp fragment containing *DERARE* was cloned from mouse genomic DNA using BamHI and HindIII-linked primers, and subcloned into pCpG-free basic-Lucia plasmid (InvivoGen). Cloned vectors were isolated by QIAGEN plasmid midi kit (QIAGEN), and then methylated *in vitro* using M.SssI CpG methyltransferase (NEB) according to manufacturer's instructions. Briefly, 2.5 ug of plasmid DNA was incubated at 37°C for 4 hours with M. SssI methyltransferase (4 units/ul) and 160 uM S-adenosylmethionine (methylated), or without (mock-methylated). Methylated DNA was then purified using the QIAquick PCR purification kit (QIAGEN) and quantified by NanoDrop (Thermo Scientific). Methylation was confirmed by digestion with the methylation-sensitive restriction enzymes HhaI and HpaII.

Luciferase Reporter Assay

293T cells (5 × 10^4/well) were seeded into 24-well plates one day before transfection. 300 ng methylated or unmethylated, wt or mutated pCpG-free plasmid was co-transfected with 100 ng firefly luciferase reporter vector, pGL3-control (Promega) using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. Cells were harvested 48 hours post-transfection, and then Firefly and Renilla luciferase activities were measured using the dual-luciferase reporter kit (Promega) on Victor3 Multilabel Plate Counter (Perkin Elmer).

qRT-PCR

RNA was extracted from 10,000 wt and *DERARE*$^{-4}$ LT-HSCs or 2 × 10^6 AML cells and treated at 37°C for 30 min with 2 μL RNase-free DNase-1 (2 U/μl; Ambion), followed by reverse transcription using ABI High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instruction. qRT-PCR was performed using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific) in an StepOne & StepOnePlus Real-Time PCR Systems (Thermo Fisher Scientific). Data were normalized relative to mouse or human *GAPDH*. Sequences of primers are listed in Table S7.

Lentivirus Production and Transduction

pL-CRISPR-EFS-GFP was used as the lentiviral backbone (Addgene plasmid #12259). The promoter of minimal size hEF1α (EFS) was initially replaced by a promoter of full size hEF1α (EF1α). To target methylation on specific CpG sites in the genome, we further replaced the S. pyogenes Cas9 with either the dCas9-DNMT3A-EGFP (Addgene #71666) or dCas9-DNMT3A (E756A)-EGFP (Addgene #71685) established by a recent study (Vojta et al., 2016). Protospacer sequence targeting human *DERARE* sgRNA and the cognate off-target sites were analyzed by GT-Scan (O'Brien and Bailey, 2014). The lentiviruses were generated by transfection of pL-CRISPR-EF1α-Cas9-DNMT3A-EGFP constructs together with the pspax2 and pMD2.G plasmids at a ratio of 10:7.5:2.5 into 293T cells using calcium phosphate transfection. The virus particles were harvested 48, 72, and 96 hours after transfection, filtered by 0.45 μm filter unit (Millipore), and centrifuged at 18,000 RPM, 4°C for 2 hours. OCI-AML2 and OCI-AML3 cells were infected with lentivirus-transducing units in the presence of 4 μg/mL polybrene (Sigma). 48 hours after infection, GFP$^+$ cells were sorted and cultured for experiments.

Bisulfite Sequencing

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (QIAGEN) and bisulfite-conversion was carried out by EpiTect Bisulfite Kit (QIAGEN) according to manufacturer's instructions. Bisulfite specific PCR was performed with purified DNA (50 ng), ZymoTaq PreMix (Zymo Research) and region-specific primers designed by Methprimer software (Li and Dahiya, 2002). PCR product was gel purified and cloned into pGEM-T Easy vector (Promega), followed by transformation and mini-preps. Plasmids isolated from at least 10 colonies for each sample were sequenced. The primer sequences for bisulfite sequencing are listed in Table S7.

Cytospin and Wright-Giemsa Staining

Cells were collected, washed, and resuspended in PBS at a concentration of 2×10^5 /ml. Cytospin (Thermo Scientific Shandon) preparations were performed at 1,000 rpm for 1 min, and the cells were allowed to air dry. Cells were stained in 100% Wright-Giemsa (American MasterTech) for 5 min and washed in PBS buffer (PH = 6.8) for 1 min. Stained cells were rinsed in deionized water, and coverslips were affixed with Permount prior to microscopy.

In vitro Colony-Forming Unit (CFU) Assay

DNMT3A Wt or ANV treated AML cells were counted and plated in triplicate (1,000 viable cells/well) in methylcellulose media (H4034, StemCell Technologies, Inc.). After 12 to 14 days, plates were scored for colony number and morphology on Celigo Imaging Cytometer.

In vivo Leukemia Analysis

2×10^6 of *DNMT3A* Wt or ANV treated AML cells were transplanted into lethally irradiated (3.25 Gy) NSG mice via tail vein injection. Mice were monitored for leukemia development and leukemia was verified after the mice were sacrificed upon symptoms of illness (weight loss below 16.75 g, partial paralysis, malaise, palpable spleen, hunched back, and weakness in hind limbs and walking). The spleen was removed and weighed. The femur and tibia were removed and analyzed by flow cytometry.

RNA-Seq

20,000 each of HSPCs, progenitors, and lineage cells were sorted from BM of 8-12-week-old C57BL/6J mice, RNA-seq (ArrayExpress: E-MTAB-2923, E-MTAB-3079) was performed, and fingerprint coding genes were identified by setting the cutoff of tissue specificity scores at $\geq$ 0.4 as previously described (Qian et al., 2016). For low cell number RNA-seq, 1,000 LT-HSCs ($LSKCD34^{-}Flk2^{-}CD49b^{lo}$) were sorted into 96-well plates from BM of $DERARE^{\Delta}$, $Rarb^{fl/fl}$ mice and their littermate controls. First-strand cDNA synthesis and cDNA libraries were constructed using the SMARTer ultra-low input RNA kit for sequencing - v3 (Clontech) following the manufacturer's instructions. cDNA quality was determined by Agilent high sensitivity DNA kit on Agilent 2100 BioAnalyzer (Agilent Technologies), and then sequenced on Illumina HiSeq 2500 to obtain single-end 50bp reads. Raw transcriptome sequence data were aligned to UCSC mm10 using Tophat 2.0.13 to produce a reference-guided transcript assembly, and expression levels for each sample were quantified to fragments per kilobase of transcript per million mapped reads (FPKM) using Cufflinks (v2.1.1) . Three biological replicates were sequenced for each population. Z-scores were calculated based on the normalized RNA expression values of each population. Functional enrichment analysis was performed using GOstats tool, and $-\log_{10}(p \text{ value})$ were plotted to show term significance.

ChIP-Seq and PCR Assays

ChIP-seq assays of *Hoxb3*, *Hoxb4*, H3K27ac and H3K4me1 were performed in mouse KH2 embryonic stem (ES) cells as previously described (De Kumar et al., 2015). ChIP-qPCR in HSCs were performed using the True MicroChIP kit (Diagenode) according to the manufacturer's instructions. Briefly, 10,000 HSPCs were sorted into PBS buffer and cross-linked with 1% formaldehyde at RT for 10 min, and 0.125 M glycine was added to stop the reaction at RT for 5 min, followed by washing once with ice-cold Hank's buffered salt solution (HBSS) containing protease inhibitor cocktail (PIC). The cell pellet was lysed in 25 μL of lysis buffer tL1 (10 mM Tris, pH 7.5, 1 mM EDTA, 1% SDS) on ice for 5 mins. After dilution with 75 μL of HBSS with PIC, DNA was sonicated by Bioruptor (Diagenode) for 8 cycles (30 s "ON" and 30 s "OFF") to generate fragments of 200-500 bp in length. Sheared chromatin was centrifuged at 14,000 × g, 4°C for 10 min, transferred to fresh low retention 1.5 mL EP tubes, and added with 100 ul of complete ChIP buffer tC1. 1/10 volume (20 μl) was set aside as input control. Samples were incubated with ChIP-qualified Hoxb3 antibody (Santa Cruz Biotechnology, sc-28606X) or rabbit IgG (Invitrogen, 02-6102) at 4°C for 16 hours. Then, samples were incubated with 10 μL of pre-washed protein A magnetic beads (Dynal, Invitrogen) at 4°C for 2 hours, and washed with 4 different types of wash buffers. Beads-DNA complex and input DNA was eluted at RT for 30 mins in 200 μL of buffer tE1 and transferred to a fresh tube. Samples were added with 8 ul tE2, and incubated in a thermomixer at 1300 rpm, 65°C for 4 hours. Precipitated DNA was purified by phenol chloroform extraction and eluted in 12 μL of elution buffer. ChIP-qPCR reactions were performed using SYBR Green PCR master mix kit with primers targeting enhancer regions of non-canonical Wnt and retinoic acid signaling genes, as listed in Table S7.

QUANTIFICATION AND STATISTICAL ANALYSIS

Fingerprint Coding Gene Analysis

Tissue specificity (TS) scores per gene were calculated as preciously described (Qian et al., 2016), using $\log_2$ adjusted FPKMs. To prevent minor differences in FPKM distributions between samples from dominating tissue-specificity scoring, we first performed between-samples quantile normalization using the limma package in R. All samples were normalized to the ST-HSC sample, which had the clearest bimodality in FPKMs. We then found that a TS cutoff of 0.4 provided a good tradeoff between the number of protein-coding genes which were specific to one sample, i.e., having a score > = 0.4 in only one sample, and the evenness of fingerprint gene number across samples, as measured by coefficient of variation. We also set up the threshold of FPKM > = 1 in at least one sample to discard the genes with extremely low expression.

Whole Genome Bisulfite Sequencing and Data Processing

20-30 ng of genomic DNA was isolated from sorted HSPCs using the QIAamp DNA blood mini kit (QIAGEN). Libraries were constructed using the Pico Methyl-Seq library prep kit (Zymo Research), and each library was subjected to 50 bp single-end sequencing on a single lane of a HiSeq 2500 device (Illumina). Reads were trimmed by trim galore (0.4.0) with default parameters to remove the reads containing adapters and showing low quality. Trimmed reads were aligned using Bismark Methylation Extractor against the mouse reference genome mm10 with option –non-directional and parameters: –report –ignore 10 –ignore_3prime 1 –bedGraph –counts –buffer_size 10G –cytosine_report. R language (3.2.2) with bsseq (1.6.0) was used to smooth methylation data and generate methylation coverage data, while GenomeGraphs (1.30.0) was used to make genome-wide methylation maps.

Analysis of Published RNA-seq Data

RNA-seq data from the Leucégène project were downloaded from GEO: GSE48173, which was performed by Illumina HiSeq 2000 and consisted of a panel of 17 $CD34^+$ human cord blood samples (n = 5 of $CD34^+$; n = 12 of $CD34^+CD45RA^-$), 43 acute myeloid leukemia (AML), and 12 acute lymphoblastic leukemia (ALL) patient samples (Macrae et al., 2013). The downloaded fastq files were aligned to UCSC hg38 reference genome using STAR. Read counts per gene were converted to RPKMs (reads per kilobase of transcript per million reads mapped) for downstream analysis. Heatmaps and clusters were drawn based on $\log_2$(z-score values of RPKM). RNA-seq data of HSCs (Side population$^+$Lin$^-$Sca-1$^+$c-Kit$^+$CD150$^+$) from the *Dnmt3a* knockout and *Dnmt3a*, *Dnmt3b* double-knockout mice were downloaded from previous study (Challen et al., 2014) with GEO: GSE50793.

Analysis of Published ATAC-seq Data

ATAC-seq data were obtained from a recently published study (Corces et al., 2016) in human $CD34^+$ hematopoietic stem and progenitor cells (HSPCs), committed progenitors, and lineage cells using the UCSC Genome Browser Track Hub (https://s3-us-west-1.amazonaws.com/chang-public-data/2016_NatGen_ATAC-AML/hub.txt). Genomic coordinate of human *DERARE* region is chr. 17: 46,664,800-46,665,700:-1. All the coordinates referenced here are for the hg19 human reference genome.

Analysis of TCGA Data from AML Patients

The RNA-seq data (Affymetrix Human Genome U133 Plus 2.0 Array) and DNA methylation data (Illumina Infinium HumanMethylation450 BeadChip profiling) from a cohort of 200 AML patient specimens were obtained from The Cancer Genome Atlas (TCGA) project (https://gdc.cancer.gov/) (Cancer Genome Atlas Research, 2013). For gene expression, read counts were obtained from RPKM matrix (laml.rnaseq.179_v1.0_gaf2.0_rpkm_matrix.txt. tcgaID.txt) and processed by the *R* package Limma (3.22.7). For DNA methylation on *DERARE* region, we calculated the mean methylation beta values of ± 5 kb around the *DERARE* from the file LAML.HumanMethylation27.Level_3.1.5.0. After removing 21 AML samples without methylation information, we divided remaining samples into two groups with either low (n = 91) or high (n = 88) *DERARE* methylation based on mean methylation beta values. Heatmap between expression of *HOXB* genes and *DERARE* methylation was made using the R package NMF with parameter -annCol to add *DERARE* methylation status bar onto the heatmap.

Statistical Analysis

All values are shown as the mean ± s.d. as indicated. All statistical analyses were generated using GraphPad Prism 5 (GraphPad Software). Student's t test was used for comparisons between two groups, and one-way ANOVA followed by Tukey's post hoc tests was used for comparisons among multiple groups. Fisher's exact test was used to evaluate the *p* values of relationship between *DERARE* methylation and cytogenetic profiles and other clinical characteristics in AML patients. Log-rank (Mantel-Cox) test was used for Kaplan-Meier survival curves. Statistical significance was defined as $p < 0.05$.

DATA AND SOFTWARE AVAILABILITY

The accession number for sequencing data reported in this paper is SuperSeries GEO: GSE89492, which is composed of the SubSeries of GEO: GSE87556 (RNA-seq) and GEO: GSE89490 (Methyl-seq). Original data underlying this manuscript can be accessed from the in Stowers Original Data Repository at http://www.stowers.org/research/publications/LIBPB-1119.