NPM1 mutation mediated HOXBLINC lncRNA activation promotes AML leukemogenesis

Ganqian Zhu
University of Texas Health San Antonio
Huacheng Luo
Pennsylvania State University College of Medicine
Yang Feng
University of Florida College of Medicine
Olga Guryanova
University of Florida
Jianfeng Xu
Baylor College of Medicine
Shi Chen
University of Texas Health San Antonio
Qian Lai
Pennsylvania State University College of Medicine
Bing Xu
The First Affiliated Hospital of Xiamen University
Zhigang Zhao
Department of Hematology and Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin
Hongyu Ni
University of Illinois at Chicago https://orcid.org/0000-0002-7597-2037
David Claxton
Penn State University College of Medicine
Ying Guo
University of Texas Health San Antonio
Yi Qiu
Penn State University https://orcid.org/0000-0002-3282-6072
Feng-Chun Yang
Sylvester Comprehensive Cancer Center, Department of Biochemistry and Molecular Biology, University of Miami Miller School of Medicine
Wei Li
University of California, Irvine https://orcid.org/0000-0001-9931-5990
Stephen Nimer

University of Miami

Suming Huang

Pennsylvania State University College of Medicine

Mingjiang Xu (xum1@uthscsa.edu)

University of Texas Health San Antonio

Article

Keywords: Leukemogenesis, cancer, HOXBLINC lncRNA activation, NPM1c+, acute myeloid leukemia

Posted Date: August 6th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-49109/v1

Version of Record: A version of this preprint was published at Nature Communications on March 29th, 2021. See the published version at https://doi.org/10.1038/s41467-021-22095-2.

NPM1 mutation mediated *HOXBLINC* lncRNA activation promotes AML leukemogenesis

Ganqian Zhu[1,3#], Huacheng Luo[6#], Yang Feng[10], Olga A. Guryanova[10], Jianfeng Xu[11], Shi Chen[1,3], Qian Lai[6], Bing Xu[13], Zhigang Zhao[14], Hongyu Ni[15], David Claxton[7,9], Ying Guo[2,3], Yi Qiu[8,9], Feng-Chun Yang[1,3,5,16], Wei Li[12], Stephen D. Nimer[4,5], Suming Huang[6,9*] and Mingjiang Xu[1,3,5,16*]

[1]Department of Molecular Medicine, [2]Department of Cell System & Anatomy, University of Texas Health San Antonio, San Antonio, TX 78229, USA

[3]Department of Biochemistry and Molecular Biology, [4]Department of Medicine, [5]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL 33136, USA

[6]Department of Pediatrics, [7]Division of Hematology/Oncology, Department of Medicine, [8]Department of Cellular & Molecular Physiology, [9]Penn State Cancer Institute, Pennsylvania State University College of Medicine, Hershey, PA 17033, USA

[10]Department of Pharmacology and Therapeutics, University of Florida College of Medicine, Gainesville, FL 32610, USA

[11]Department of Molecular and Cellular Biology, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

[12] Division of Computational Biomedicine, Department of Biological Chemistry, School of Medicine, University of California, Irvine, CA 92697, USA

[13]Department of Hematology, The First Affiliated Hospital of Xiamen University, Xiamen 361003, China

[14]Department of Hematology and Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin 300060, China

[15]Department of Pathology, University of Illinois at Chicago, Chicago, IL 60612, USA

[16]Mays Cancer Center, University of Texas Health San Antonio, San Antonio, TX 78229, USA

[#]Co-first authors

*Correspondence: xum1@uthscsa.edu (M.X.) and shuang4@pennstatehealth.psu.edu (S.H.)

Lead Contact:
Mingjiang Xu, MD, Ph.D.
Department of Molecular Medicine, Mays Cancer Center, University of Texas Health San Antonio, San Antonio, TX 78229, USA
Email: xum1@uthscsa.edu

Running title: NPM1c$^+$ activates *HOXBLINC* to drive leukemogenesis

Abstract

Nucleophosmin (*NPM1*) is the most commonly mutated gene in acute myeloid leukemia (AML) resulting in aberrant cytoplasmic translocation of the encoded nucleolar protein (NPM1c$^+$). NPM1c$^+$ maintains a unique leukemic gene expression program, characterized by activation of *HOXA*/*B* clusters and *MEIS1* oncogene to facilitate leukemogenesis. However, the mechanisms by which NPM1c$^+$ controls such gene expression patterns to promote leukemogenesis remain largely unknown. Here, we show that the activation of *HOXBLINC,* a *HOXB* locus-associated long non-coding RNA (lncRNA), is a critical downstream mediator of NPM1c$^+$-associated leukemic transcription program and leukemogenesis. *HOXBLINC* loss attenuates NPM1c$^+$-driven leukemogenesis by rectifying the signature of NPM1c$^+$ leukemic transcription programs. Furthermore, overexpression of *HoxBlinc* (*HoxBlinc*Tg) in mice enhances HSC self-renewal and expands myelopoiesis, leading to the development of AML-like disease, reminiscent of the phenotypes seen in the *Npm1* mutant knock-in (*Npm1$^{c/+}$)* mice. *HoxBlinc*Tg and *Npm1$^{c/+}$* HSPCs share significantly overlapped transcriptome and chromatin structure. Mechanistically, *HoxBlinc* binds to the promoter regions of NPM1c$^+$ signature genes to control their activation in *HoxBlinc*Tg HSPCs, via MLL1 recruitment and promoter H3K4me3 modification. Our study reveals that *HOXBLINC* lncRNA activation plays an essential oncogenic role in *NPM1c$^+$* leukemia. *HOXBLINC* and its partner MLL1 are potential therapeutic targets for *NPM1c$^+$* AML.

Introduction

Nucleophosmin (*NPM1*) mutations are the most frequently recurring genetic abnormalities in patients with acute myeloid leukemia (AML), occurring in approximately 50% of adult and 20% of childhood AML with normal karyotypes [1,2]. *NPM1*-mutated AML has been included as a distinct AML entity in the World Health Organization (WHO) classification [3]. *NPM1* encodes a protein that is normally located in the nucleolus and has multiple functions such as biogenesis of ribosomes and maintenance of genomic stability [4]. *NPM1* mutations result in cytoplasmic mislocalization of the mutant protein (NPM1c$^+$), which is critical for its role in leukemogenesis [5,6]. As an AML-initiating lesion, NPM1c$^+$ maintains a distinctive transcriptional signature in AML cells, characterized by upregulation of *HOXA* and *HOXB* cluster genes and their oncogenic cofactor *MEIS1* [7,8]. However, the precise mechanisms by which NPM1c$^+$ drives the leukemic gene expression programs remain unclear.

Dysregulation of *HOXA/B* genes is a dominant mechanism of leukemic transformation and hematopoietic stem/progenitor cell (HSPC) deregulation [9]. A wide variety of molecular determinants including transcriptional factors, epigenetic regulators (e.g. polycomb and trithorax proteins), microRNAs, chromatin structure, and long non-coding RNAs (lncRNAs) are known to control *HOX* gene expression. However, their relationship with each other to fine-tune the *HOXA/B* gene expression pattern in AML remains to be elucidated. Recently two *HOXA/B* loci associated lncRNAs, *HOTTIP* and *HOXBLINC*, were shown to regulate transcription of *HOXA/B* genes through influencing epigenetic landscape [10-12]. *HOXBLINC* has been reported to play a critical role in hematopoietic specification during development through its *cis*-acting function to coordinate anterior *HOXB* gene expression via recruitment of the SETD1A/MLL1 histone H3K4 methyltransferase complexes [10].

Located in the anterior *HOXB* locus and serving as a regulator of *HOXB* gene transcription, the role of *HOXBLINC* in HSPC biology and leukemogenesis remains unknown. In this study, we demonstrate that *HOXBLINC* is upregulated restrictively in *NPM1c*$^+$ AML. *Npm1* mutant knock-in (*Npm1*$^{c/+}$) and *HoxBlinc*Tg HSPCs share significantly overlapped chromatin signatures and gene expression profiles in their upregulated genes as compared to WT HSPCs, including the NPM1c$^+$ signature *HoxA*/*B* cluster genes and homeobox oncogene *Meis1*. Transgenic overexpression of *HoxBlinc* lncRNA in hematopoietic cells led to the development of an AML-like disease by triggering HSC self-renewal and expanding myelopoiesis, similar to the

phenotypes displayed by *Npm1*$^{c/+}$ mice, while inhibition of *HOXBLINC* in NPM1c$^+$ AML cells mitigates leukemogenesis. Importantly, *HoxBlinc* overexpression in HSPCs increases its binding to NPM1c$^+$ signature genes and drives the leukemic specific transcription program in HSPCs by recruiting the MLL1 complex to reorganize local chromatin signatures. Together, our studies provide compelling evidence for the potent oncogenic role of *HOXBLINC* in NPM1c$^+$-mediated leukemogenesis. *HOXBLINC* lncRNA and MLL1 could serve as potential therapeutic targets for the treatment of *NPM1c*$^+$ AML.

Results

***HOXBLINC* is specifically upregulated in *NPM1c*$^+$ AML.** Up-regulation of *HOX* genes, especially *HOXA* and *HOXB* cluster genes are not only a characteristic but also a dominant mechanism for the pathogenesis of AML [9]. *HoxBlinc* lncRNA has been shown to be required to activate anterior *HoxB* gene transcription during development [10]. To determine whether *HOXBLINC* is aberrantly expressed along with the *HOXB* genes in AML, we performed RT-qPCR on bone marrow mononuclear cells (BMMNCs) from a cohort of AML patients (*NPM1c*$^+$**,** n=25**;** and *NPM1*-wt, n=40; see patient information in Table S1) as compared to both BMMNCs (n=16) and CD34$^+$ cells (n=11) from normal individuals. Interestingly, a dramatic upregulation of *HOXBLINC* was observed specifically in *NPM1c*$^+$ AML patients (Figure 1A) as compared to *NPM1*-wt patients and normal CD34$^+$ cells. When the RNA-seq data from TCGA-LAML datasets consisting of a cohort of 181 AML patients was analyzed for *HOXBLINC* expression, significantly higher *HOXBLINC* expression was observed in *NPM1c*$^+$, but not *MLL*-rearranged (*MLLr*$^+$) AML patients as compared to *NPM1c*$^-$*MLLr*$^-$ patients (Figure 1B). The expression of *HOXBLINC* was positively correlated with the expression levels of NPM1c$^+$ signature genes including anterior *HOXB* genes, *HOXA9* and *MEIS1*, but not *HOXB13* in this AML cohort (Figure S1A). Interestingly, AML patients with high *HOXBLINC* expression (the top thirty percentile of patients) had a significantly shortened survival as compared to patients with low *HOXBLINC* expression (the bottom thirty percentile, Figure S1B). Consistently, *HOXBLINC* was highly expressed in *NPM1c*$^+$ OCI-AML3 and IMS-M2 AML cells, but not *NPM1*-wt AML cells such as *MLLr*$^+$ MOLM-13, MV4-11, THP-1, NOMO-1 and OCI-AML2 cells, as well as *BCR-ABL1*$^+$ K562 and *JAK2V617F*$^+$ SET-2 cells (Figure S1C). These data collectively indicate that *HOXBLINC* is upregulated specifically in *NPM1c*$^+$ AML patients.

Loss of *HOXBLINC* perturbs NPM1c$^+$-mediated transcription programs and leukemogenesis. To confirm *HOXBLINC* activation is a downstream event of NPM1c$^+$ and determine the role of *HOXBLINC* in NPM1c$^+$-mediated transcription regulation and leukemogenesis, we performed RNA-seq analysis on LSK cells isolated from the *Npm1* mutant knock-in (*Npm1$^{c/+}$*) mice [13]. As compared to WT LSK cells, *Npm1$^{c/+}$* LSK cells had 871 down-regulated genes and 980 upregulated genes, including *HoxBlinc* and the NPM1c$^+$ signature genes *HoxB2-5*, *HoxA7,9-11*, *Meis1* and *Runx1* (Figure 1C, S1D), some of which are confirmed by qPCR (Figure S1E). Gene Ontology (GO) and gene set enrichment (GSEA) analyses revealed that the upregulated genes in *Npm1$^{c/+}$* vs. WT LSK cells are enriched with cell fate commitment, cell cycle, myeloid cell proliferation, stem cell maintenance, Wnt and Jak-STAT signaling pathways, pathways in cancer, as well as AML *NPM1*-mutated and HOXA9 oncogenic pathway (Figure S1F,G).

We next examined the effect of *HOXBLINC* loss on NPM1c$^+$-mediated transcription regulation and leukemogenesis using *NPM1*c$^+$ OCI-AML3 cells. We created CRISPR-dCas9-KRAB mediated *HOXBLINC* epigenetic silencing clones (*HOXBLINC*i) by targeting the KRAB repressive domain to the *HOXBLINC* promoter (Figure S2A). We then compared genome-wide transcriptome changes between control and *HOXBLINC*i OCI-AML3 cells by performing RNA-seq analysis. Consistently, *NPM1*c$^+$ OCI-AML3 cells exhibited high expression of *HOXBLINC* lncRNA and the common NPM1c$^+$ AML signature genes (Figure 1D). Interestingly, inhibition of *HOXBLINC* in OCI-AML3 cells significantly impaired the transcription of many NPM1c$^+$ signature genes such as *HOXB2-5, HOXA9-11, RUNX1* and *MEIS1* (Figure 1D). GSEA and GO analyses revealed that loss of *HOXBLINC* affects the pathways and genes involved in AML with *NPM1*-mutated, HOXA9 pathway, pathways in cancer, cell cycle, cell fate commitment, myeloid cell differentiation, and Wnt and JAK-STAT signaling pathways (Figures 1E&F, S2B), similar to those observed in the upregulated genes of *Npm1$^{c/+}$* vs. WT LSK cells (Figure S1F,G).

We further explored whether *HoxBlinc* is required for NPM1c$^+$-mediated leukemogenesis by generating two *Tet-ON* inducible *HOXBLINC* shRNA knockdown (KD) OCI-AML3 clones (Figure S2C). *HOXBLINC* KD significantly impaired OCI-AML3 cell proliferation compared to the scramble control (Figure 1G). Cell cycle analysis revealed that *HOXBLINC* KD increased the sub-G0 cell population, suggesting *HOXBLINC* perturbation induced apoptosis (Figure S2D).

When we transplanted a Dox-inducible *HOXBLINC* KD OCI-AML3 clone into *NOD-scid IL2Rγ^{null}* (NSG) mice (2x10^5 cells/mouse) followed by Dox induction or vehicle treatment, the Dox-treated recipient mice had a significantly prolonged survival as compared to the recipients without Dox treatment (Figure 1H). At 30 days after transplantation, Dox-treated mice had significantly lower chimerism of the hCD45$^+$ cell population in the BM, spleen and peripheral blood (PB) of the recipients compared to the untreated animals (Figure S2E). These results indicate that *HOXBLINC* KD suppresses OCI-AML3 leukemic cell proliferation both *in vitro* and *in vivo,* likely through the normalization of *NPM1*-mutation induced abnormal gene expression patterns (Figure 1D). In addition, we silenced *HOXBLINC* expression in primary AML cells with or without NPM1c$^+$ mutation (#1315: *NPM1c*$^+$;*FLT3wt*, #921: *NPM1c*$^+$;*FLT3mu*, #LPP4: *NPM1*wt*;MLLr*$^+$) by the *CRISPR-dCas9-KRAB* (*HOXBLINC*i), and then xenografted them into NSG mice. Both #1315 and #921 exhibited high *HOXBLINC* expression, while #LPP4 had low *HOXBLINC* expression (data not shown). In line with cell line xenograft results, mice receiving *HOXBLINC*i primary AML cells with *NPM1c*$^+$;*FLT3wt* (#1315) or *NPM1c*$^+$;*FLT3mu* (#921) had significantly prolonged survival as compared to mice transplanted with control cells (Figure 1I). FACS analysis revealed that *HOXBLINC*i dramatically decreased the hCD45$^+$ cell chimerism in BM, spleen and PB of recipients (Figure S2F). In contrast, *HOXBLINC*i neither prolonged the survival nor decreased hCD45$^+$ cell chimerism in mice transplanted with *NPM1*wt*;MLLr*$^+$ (#LPP4) AML cells (Figures 1I, S2F). Thus, *HOXBLINC* perturbation decreased tumor burden and attenuated leukemic progression *in vivo* most likely specific for *NPM1*c$^+$ AML patients.

Transgenic expression of *HoxBlinc* in hematopoiesis leads to lethal AML-like disease in mice. It has been shown that activation of a humanized *NPM1*c$^+$ knock-in allele in mouse HSCs (*NPM1*$^{c/+}$) causes *Hox* gene overexpression, enhanced self-renewal and expanded myelopoiesis, as well as development of delayed-onset AML [14]. Since NPM1c$^+$ activates *HOXBLINC* which is critical for NPM1c$^+$-mediated transcription program and leukemogenesis, it is important to determine whether *HOXBLINC* activation is sufficient to cause abnormal hematopoieiss and myeloid malignancies similar to the *NPM1*$^{c/+}$ mice. We first examined the *HoxBlinc* expression pattern along the HSC differentiation hierarchy. *HoxBlinc* expression was high in long-term (LT) and short-term (ST) HSCs, decreased in progenitor cells (MPP, CMP and GMP) and was further

decreased in the mature lineage cell populations except the $B220^+$ B cells (Figure 2A). The expression pattern of *HoxBlinc* in hematopoiesis suggests that this lncRNA might play an important role in regulating HSPC function. To investigate the impact of *HoxBlinc* activation on normal hematopoiesis and leukemogenesis *in vivo*, we generated a *HoxBlinc* transgenic (Tg) mouse model in which full-length mouse *HoxBlinc* cDNA was inserted into mouse genome under the control of *Vav1* promoter and enhancer (HS321/45-vav vector) to ensure the expression of transgene specifically in hematopoiesis (Figure 2B). Two founder *HoxBlinc*Tg mice were obtained. The transgene was inserted into the intron of the *Bin1* gene on chromosome 18q for Tg Line #1 (Figure 2C). The expression levels of *HoxBlinc* RNA in BM cells were ~18- and 3-folds greater than the endogenous *HoxBlinc* expression in Tg Line #1 and #2, respectively (Figure 2D).

Monitoring of a cohort of *HoxBlinc*Tg (Line #1) mice showed that within 1 year of age, 67% of *HoxBlinc*Tg mice (10 of 15) died or were sacrificed because of a moribund condition, whereas none of the WT mice (n=12) died (Figure 2E). Moribund *HoxBlinc*Tg mice exhibited weight loss, hepatosplenomegaly, enlarged lymph nodes as well as pale footpads and femurs as compared to WT (Figures 2F, S3A). Peripheral blood examination revealed marked leukocytosis due to elevated immature myeloid cells and neutrophils, thrombocytopenia and severe anemia in these moribund *HoxBlinc*Tg mice (Figure S3B). Morphologically, May-Grünwald-Giemsa stained PB smears showed significantly increased blasts (Figure 2G, *left*). BM cell cytospin preparations also demonstrated a predominance of myeloid cells with increased immature myeloid precursors (Figure 2G, *right*). Flow cytometric analyses of the BM cells revealed increased c-Kit^+ (consistently >20%) and Gr-1^{low} immature myeloid cell populations, as well as decreased lymphoid and erythroid cell populations (Figure S3C,D). Morphologic evaluation of BM histologic sections revealed myeloid hyperplasia with increased immature myeloid precursors, which were myeloperoxidase (MPO) positive indicating myeloid origin (Figures 2H, S3E). In addition, histologic evaluation of the *HoxBlinc*Tg spleen, liver and lymph node sections showed distortion of organ normal architecture with infiltration of MPO positive myeloid cells (Figures 2H,I & S3F). The sections of liver demonstrated a sinusoidal infiltration pattern with clusters of immature myeloid cells (Figures 2I & S3F). Consistently, Line #2 *HoxBlinc*Tg mice also developed AML-like disease similar to Line #1, characterized by shortened survival and >20% c-Kit^+ myeloid cells in the BM, PB and spleen (Figure S3G-I). The longer survival

exhibited in Line #2 than Line #1 *HoxBlinc*Tg mice suggests a dosage effect of the *HoxBlinc* expression on hematopoiesis and transformation. These data demonstrate that similar to *NPM1*$^{c/+}$ mice, *HoxBlinc* overexpression in mice is sufficient to cause abnormal hematological characteristics resembling AML.

Transgenic expression of *HoxBlinc* enhances HSC self-renewal and expands myelopoiesis. The capacity of NPM1c$^+$ such as enhancing HSC self-renewal and expanding myelopoiesis had been well clarified [14]. To further understand the role of *HoxBlinc* in AML pathogenesis, we then determined the effect of *HoxBlinc* overexpression on HSPC function. Flow cytometric analyses on BM cells of young *HoxBlinc*Tg mice (8-10 weeks, Line #1) showed dramatically increased proportions of Gr1$^+$/Mac1$^+$ granulocytic/monocytic and B220$^+$ B cells as well as decreased proportions of CD4$^+$/CD8$^+$ T and Ter119$^+$ erythroid cells as compared to age-matched WT mice (Figure S4A). Importantly, *HoxBlinc*Tg BM cells contained a significantly greater proportion of Lin$^-$Scal-1$^+$c-Kit$^+$ (LSK) cells, while comparable Lin$^-$Scal-1$^-$c-Kit$^+$ (LK) cells as compared to WT mice (Figure 3A). When the total number of LT-HSCs, ST-HSCs and multipotent progenitor cells (MPPs) per femur were calculated based on their proportions within the LSK cell population and BM cellularity, the pools of both LT- and ST-HSCs, but not MPP were significantly expanded, although only the proportion of ST-HSCs within LSK cells were increased (Figures 3B, S4B). When each myeloid progenitor population was analyzed within the LK cells, a significantly higher percentage of GMP, but lower percentages of MEP/CMP cell populations were observed in *HoxBlinc*Tg mice than WT mice (Figures 3C, S4C). Consistent with the increased frequency of LSK and GMP, colony-forming unit cell (CFU-C) assays revealed significant higher frequencies of CFU-Cs, especially CFU-GM in the BM of *HoxBlinc*Tg mice than WT mice (Figures S4D). Similar to Line #1, significantly increased proportions of Gr1$^+$/Mac1$^+$ and decreased proportions of erythroid cells were observed in the BM of Line #2 *HoxBlinc*Tg mice compared to WT mice (Figure S4E). Line #2 *HoxBlinc*Tg BM cells also contained higher percentages of LSK and GMP cells than WT (Figure S4F,G). The similar hematological and disease phenotypes in both lines of *HoxBlinc*Tg mice indicate that these observed phenotypes are induced by the *HoxBlinc* transgenic expression but not the positional effect. Therefore, overexpression of *HoxBlinc* dysregulates HSPC pools with skewed hematopoiesis towards myeloid lineage *in vivo*.

We next examined the effect of *HoxBlinc* overexpression on the self-renewal and repopulation capacity of HSCs using *in vitro* replating and paired-daughter cell assays and *in vivo* competitive transplantation. A significantly higher replating potential was observed in each of the four successive rounds of replating in *HoxBlinc*Tg LSK cells than WT cells (Figure 3D). Both symmetric and asymmetric cell divisions are required for the preservation of normal HSC pool and continuous production of sufficient blood cells. Paired-daughter cell assays using WT and *HoxBlinc*Tg primitive $CD34^-$LSK cells showed a higher proportion of *HoxBlinc*Tg $CD34^-$ LSK cells with symmetric self-renewal capacity, while the cells that underwent symmetric differentiation or asymmetric self-renewal were reduced as compared to WT (Figure 3E). Competitive transplantation assays showed that the donor cell ($CD45.2^+$) chimerism in the PB of recipients transplanted with *HoxBlinc*Tg BM cells steadily increased, reaching ~80% 7 months after transplantation, while the $CD45.2^+$ cell population in the PB of mice receiving WT BM cells remained ~50% (Figure 3F). Furthermore, *HoxBlinc*Tg BM cells contributed to greater proportions of $Gr\text{-}1^+/Mac1^+$ cells in the PB than WT cells in recipient mice (data not shown). Interestingly, mice receiving *HoxBlinc*Tg BM cells became moribund or died 2.5-7 months after transplantation (Figure 3G). These *HoxBlinc*Tg BM recipients displayed similar hematological phenotypes as the primary *HoxBlinc*Tg mice, including high WBC counts, dramatically elevated immature myeloid cells and neutrophils, severe anemia, and decreased platelet counts (Figure S5A). Flow cytometric analyses of the donor ($CD45.2^+$) and competitor ($CD45.1^+$) derived BM cells revealed significantly greater percentages of $c\text{-}Kit^+$ and $Gr\text{-}1^+/Mac1^+$ cells and lower proportions of $CD3^+$, $B220^+$ and $Ter119^+$ cells in *HoxBlinc*Tg vs. WT recipients (Figure S5B,C). Strikingly, the BM $CD45.2^+$ Lin^- cells of *HoxBlinc*Tg recipients are comprised of significantly higher LK and LSK cells than that of $CD45.1^+Lin^-$ cells in *HoxBlinc*Tg recipients and Lin^- cells in WT recipients (Figure S5B). Consistently, a higher proportion of immature myeloid cells were observed in the BM of *HoxBlinc*Tg recipients (Figure S5D,E). The similar phenotypes displayed in recipients transplanted with *HoxBlinc*Tg BM cells and primary *HoxBlinc*Tg mice indicate that the aberrant HSPC function and AML-like disease induced by *HoxBlinc* overexpression in mice are transferable and HSPC cell-autonomous. Collectively, transgenic expression of *HoxBlinc* in mice enhances HSC self-renewal and expands myelopoiesis, leading to AML-like disease, reminiscent of the phenotypes seen in the $NPM1^{c/+}$ mice.

Overexpression of *HoxBlinc* activates NPM1c$^+$ signature genes via increased enhancer/promoter chromatin accessibility in HSPCs. To further delineate whether *HOXBLINC* is a downstream mediator of NPM1c$^+$ to maintain NPM1c$^+$-signature gene activation in HSPCs**,** RNA-seq was performed with WT and *HoxBlinc*Tg LSK cells from 8-week old mice. Comparison of gene expression profiles of *HoxBlinc*Tg LSK cells to those of WT LSK cells identified 1,281 differentially expressed genes (P<0.05 and fold-change ≥ 2,0, Figure 4A). Among the DEGs, 718 were up-regulated and 563 were down-regulated. Strikingly, 27.3% of the up-regulated genes and 21% of the down-regulated genes were overlapped with the up- and down-regulated genes in *NPM1*$^{c/+}$ v.s. WT LSK cells, respectively (Figure 4B). These common up-regulated genes of *HoxBlinc*Tg and *NPM1*$^{c/+}$ LSK cells (Table S2) included the NPM1c$^+$ signature genes *HoxB2-5, HoxA9-10*, *Meis1*, and *Runx1*, which were confirmed by RT-qPCR (Figure S6A). Consistently, GO and GSEA analyses of the differentially expressed genes in *HoxBlinc*Tg vs. WT LSK cells revealed association/enrichment of similar transcription signatures and pathways with those in *NPM1*$^{c/+}$ vs. WT LSK cells, including *NPM1*-mutated signature, HOXA9 oncogenic pathway, HSC proliferation, cell fate commitment, myeloid differentiation, Wnt and JAK-STAT signaling pathways (Figures 4C,D & S6B).

Since *HoxBlinc* promotes the expression of anterior *HoxB* genes by recruiting SETD1A and MLL1 complexes and organizing active chromatin domain in the anterior *HoxB* locus[15], *HoxBlinc* upregulation caused by either NPM1c$^+$ or *HoxBlinc*Tg could activate its target genes in HSPCs by enhancing enhancer/promoter chromatin accessibility. To test this, we carried out an assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq) using LSK cells from WT, *NPM1*$^{c/+}$ and *HoxBlinc*Tg mice (Figure S6C,D). Coincidently, *NPM1*$^{c/+}$ and *HoxBlinc*Tg LSK cells shared significant portions of genes exhibiting gain (28.9%) or loss (24.3%) of promoter accessibility (Figure 4E, Table S3). In addition, significant portions of the genes with promoter accessibility gain by either NPM1c$^+$ or *HoxBlinc*Tg were also upregulated (Figure S6E). These genes included the NPM1c-signature genes *HoxB2-5, HoxA9-10*, *Meis1*, and *Runx1,* as well as other target genes such as *Stat1* and *Cdr2* (Figures S1D; 4F,G & S6F,G,) that also play critical roles in HSC regulation and leukemogenesis. As a control, no significant changes in chromatin accessibility were observed in the *HoxD* and *Lypla1* loci, and the expression of *HoxD* or *Lypla1* genes was not altered by *HoxBlinc* upregulation caused by either NPM1c$^+$ or *HoxBlinc*Tg (Figure S6G & data not shown). These results suggest that

overexpression of *HoxBlinc* lncRNA specifically activates NPM1c$^+$ signature genes via enhancing enhancer/promoter chromatin accessibility in HSPCs.

***HoxBlinc* directly binds to target genes and mediates chromatin interactions to drive gene regulatory networks in HSPCs.** CTCF boundaries facilitate enhancer/promoter interactions within confined topologically associated domains (TADs). We recently reported that a CTCF boundary in the posterior HOXA locus establishes and maintains an active TAD to drive posterior *HOXA* gene expression [16]. To examine whether *HoxBlinc* overexpression affects CTCF defined anterior TAD domain and enhancer/promoter regulatory networks in the anterior *HoxB* locus, circular chromosome conformation capture using high throughput sequencing (4C-seq) was performed using the several *HoxB* locus CTCF binding sites (CBSs) as viewpoints in *HoxBlinc*Tg vs. WT Lin$^-$c-Kit$^+$ cells (Figure 5A). When the CBS located between *HoxB4* and *B5* (CBS4/5, which overlaps *HoxBlinc* gene) was used as a viewpoint, CBS4/5 interacted with the +43Kb CBS (+43CBS, Figure 5B). CBS4/5 also contacted each of the anterior *HoxB* genes (Figure 5B), suggesting that either CBS4/5 or more likely *HoxBlinc* communicates with anterior *HoxB* gene promoters. When +43CBS was used as a viewpoint, the interaction of CBS4/5 and +43CBS was confirmed and +43CBS communicated with anterior *HoxB* genes too (Figure 5B). Interestingly, *HoxBlinc* overexpression intensified each of these long-range interactions within the anterior *HoxB* locus mediated by CBS4/5 and/or +43CBS (Figure 5B). In contrast, +73Kb CBS (+73CBS) and *HoxB13* CBS (CBS13) did not interact with the anterior *HoxB* genes, although +73CBS interacted with CBS5/6 and CBS8/9, which however was not affected by *HoxBlinc* overexpression (Figure 5B). Furthermore, *HoxBlinc* overexpression also induced the long-range interactions of CBS4/5 and/or +43CBS with the promoter regions of *HoxBlinc* target genes such as *Stat1*, *Crd2* and posterior *HoxA* genes, but not non-*HoxBlinc* target *HoxD* genes (Figures 5C, S7A). These data indicate that *HoxBlinc* coordinated with the CBS4/5 and +43CBS in HSPCs to facilitate and maintain long range chromatin interactions with NPM1c$^+$ signature gene loci for their activation.

To completely understand the mechanism by which *HoxBlinc* overexpression regulates hematopoietic transcription program in HSPCs, we carried out ChIRP-seq (Chromatin Isolation by RNA Purification combined with deep sequencing) to map the genomic *HoxBlinc* binding sites in WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells. Overexpression of *HoxBlinc* significantly increased

its binding to the promoter regions of anterior *HoxB* genes and other *trans* targets, such as *Stat1*, *Cdr2*, *Wnt5a, Runx1, Meis1* and posterior *HoxA* genes (Figures 5D, S7B & data not shown). The *HoxBlinc* binding to anterior *HoxB* and *Runx1* promoters were confirmed by ChIRP-qPCR (Figure S7C,D). The global *HoxBlinc*-binding site distribution in $Lin^{-}c\text{-}Kit^{+}$ cell genome revealed that *HoxBlinc* mainly interacted with noncoding regions, including intergenic regions, introns, and promoters. Emphatically, *HoxBlinc* overexpression markedly increased its occupancy with promoters and UTRs (Figure 5E). Furthermore, GO analysis of *HoxBlinc*-bound genes are enriched with pathways important for HSPC regulation and leukemogenesis such as *Hox* genes, AML, HSC proliferation, Wnt signaling, and cell cycle (Figure S7E). Integration of ChIRP-seq, RNA-seq, and ATAC-seq datasets from WT and *HoxBlinc*Tg HSPCs revealed that around 74% of the genes with increased *HoxBlinc* binding exhibited a ≥2 folds increase in gene expression levels (Figure 5F) and 44.7% of them showed increased promoter chromatin accessibility (Figure 5G). These data revealed that *HoxBlinc* acts as an epigenetic regulator to control target gene expression through remodeling promoter chromatin accessibility. Transcription motif analysis showed that the top *HoxBlinc* bound motifs in *HoxBlinc*Tg HSPCs are transcription factors important for hematopoiesis, such as CTCF and PU.1 (Figure S7F, Table S4). These data demonstrated that *HoxBlinc* directly binds to hematopoietic specific target genes, mainly the $NPM1c^{+}$ signature genes and mediates long range chromatin interactions to drive gene regulatory networks in HSPCs.

Recruitment of MLL1 is critical for *HoxBlinc* overexpression mediated target gene expression and leukemogenesis. As *HoxBlinc* recruits SETD1A and MLL1 complexes to organize active chromatin domain in the *HoxB* loci in mouse ECSs derived primitive erythroid progenitor cells[15], we performed RIP-qPCR analysis which showed much greater *HOXBLINC* enrichment in the immunoprecipitates of anti-MLL1 and anti-SETD1A, but not control IgG and anti-LSD1 antibodies in high *HoxBlinc* expressing OCI-AML3 cells as compared to the low *HoxBlinc* expressing OCI-AML2 cells (Figure S8A), confirming the interaction of human *HOXBLINC* with MLL1 and SETD1A. We next examined whether SETD1A and MLL1 are important for *HoxBlinc* overexpression mediated abnormal HSPC function and leukemogenesis both *in vitro* and *in vivo*. However, knockdown *Mll1*, but not *Setd1a* in *HoxBlinc*Tg LSK cells was capable of mitigating the abnormal replating potential mediated by *HoxBlinc* overexpression

(Figures 6A, S8B). Furthermore, genetic deletion of one *Setd1a* allele in *HoxBlinc*Tg mice (*Setd1a*$^{+/-}$;*HoxBlinc*Tg) did not rescue the abnormal hematologic phenotypes induced by *HoxBlinc* overexpression *in vivo* (Figure S8C-F), indicating that SETD1A is not likely a key player in *HoxBlinc* overexpression mediated abnormal hematopoiesis. When transplantation was performed using *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells transduced with lentivirus expressing scramble control or sh*Mll1*, *Mll1* KD significantly prolonged the survival of recipients receiving sh*Mll1*-expressing *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells as compared to sh*Scramble* expressing *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells (Figure 6B). And the aberrant expansion of CD117$^+$/CD11b$^+$ immature myeloid cells and GMPs as well as anemia in *HoxBlinc*Tg mice were also largely restored by *Mll1* KD (Figure 6C,D). These data indicate that *Mll1* KD is capable of mitigating the AML development induced by *HoxBlinc* overexpression.

Since MLL1 is critical for *HoxBlinc* overexpression mediated abnormal hematopoiesis, *HoxBlinc* overexpression in HSPCs might activate its target genes by increasing MLL1 recruitment and thereby enhancing H3K4me3 occupancy to facilitate enhancer/promoter chromatin accessibility. To confirm this, we carried out MLL1 and H3K4me3 ChIP-seq using WT and *HoxBlinc*Tg LSK cells. Combined ChIP-seq and CHIRP-seq analyses revealed extremely high overlap for the genomic binding sites of *HoxBlinc* and MLL1 (Figures 6E-G, S9A). Impressively, 51.2% of the genes with increased *HoxBlinc* binding exhibited an elevated MLL1 recruitment and most of which (31.7%) also showed increased H3K4me3 occupancy, including the NPM1c$^+$-signature genes such as anterior *HoxB,* posterior *HoxA*, *Meis1* and *Runx1,* as well as other hematopoietic genes such as *Stat1* and *Cdr2* (Figure 6E-G, S9A, Table S5). In contrast, no significant changes in MLL1 recruitment and H3K4me3 occupancy were observed in the control *Lypla1* (no change in *HoxBlinc* binding)*, Eno1* (no *HoxBlinc* binding) and *Car6* (no *HoxBlinc*/MLL1 binding) loci (Figure S9B); and the expression of such genes was not altered by *HoxBlinc* overexpression. These data demonstrate that *HoxBlinc* overexpression mediates target gene expression via increased recruitment of MLL1 and subsequent enhancement of H3K4me3 occupancy.

Discussion

In a humanized *NPM1c*$^+$ knock-in mouse model, NPM1c$^+$ enhanced HSC self-renewal and expanded myelopoiesis leading to 1/3rd of the animals developed late-onset AML [14]. Brunetti

et al recently reported that specific reduction of NPM1c$^+$ lessens key features of the leukemic program [17]. Thus, *NPM1* mutation is an AML-driving lesion and maintains leukemia through a gain-of-function by the NPM1c$^+$. NPM1c$^+$-mediated leukemogenesis has been shown to depend on the unique gene expression signatures such as *HOXA/B* and *MEIS1* activation [18]. However, how this aberrant gene expression program is driven and maintained is largely unknown. In this study, we show that NPM1c$^+$ deregulates it signature genes, perturbs hematopoiesis and promotes leukemogenesis via the activation of a critical lncRNA, *HOXBLINC*. *HOXBLINC* overexpression is strongly association with *NPM1* mutations in AML and NPM1c$^+$ expression leads to *HoxBlinc* activation in HSPCs. Although *HoxBlinc* is involved in normal hematopoietic development,[13] *HoxBlinc* overexpression plays an essential and sufficient oncogenic role in NPM1c$^+$-mediated signature gene expression, HSPC deregulation and leukemogenesis. Therefore, our studies identify *HoxBlinc* activation as a novel and critical downstream mediator for NPM1c$^+$.

In addition to mutations and/or aberrant expression in protein-coding genes, misregulation of lncRNAs perturbs cellular physiology in multiple ways and plays important roles in the development and progression of various cancers [19]. However, how lncRNAs affect the initiation and progression of malignant myelopoiesis remains to be determined. Accordingly, lncRNAs have been profiled in various myeloid leukemias in order to identify potential oncogenic lncRNAs [20-22]. Recent studies have shown that *LncHSC-2* and *Hottip* lncRNAs contribute to the control of critical signaling pathways in HSC regulation [12,23]. However, a direct link between lncRNAs and oncogenesis remains elusive in malignant myelopoiesis. Furthermore, the detailed molecular mechanisms underlying lncRNA dysregulation-mediated myeloid malignancy development remain largely unknown. Our RNA-seq analyses on *HOXBLINC* ko vs. control *NPM1c*$^+$ OCI-AML3 cells and *HoxBlinc*Tg vs. *NPM1*$^{c/+}$ or WT LSK cells demonstrate that *HOXBLINC* regulates NPM1c$^+$ signature genes including anterior *HOXB* genes where *HOXBLINC* resides and genes located on other chromosomes such as *HOXA9-10*, *MEIS1* and *RUNX1*. ChIRP-seq analyses reveal that *HoxBlinc* binds to the promoters of both resident anterior *HoxB* cluster genes and distant target genes such as posterior *HoxA, Meis1* and *Runx1*. These results indicate that the regulation of the NPM1c$^+$ signature genes by *HoxBlinc* is achieved through the direct *HoxBlinc* binding via *cis* and *trans* actions. Indeed, GSEA and GO analyses signify enrichment of genes for *NPM1*-mutated signature, HOXA9 oncogenic pathway, Wnt and

JAK-STAT signaling pathways in *HOXBLINC* ko vs. control OCI-AML3 cells and *HoxBlinc*Tg or *NPM1*$^{c/+}$ vs. WT LSK cells. Additional functional analyses demonstrate that normal hematopoiesis requires a tightly controlled *HOXBLINC* expression, and its misregulation caused by NPM1c$^+$ is an oncogenic event in leukemogenesis. Our findings on the direct oncogenic role of *HOXBLINC* in *NPM1*-mutated AML could serve as a blueprint for implicating lncRNAs in AML leukemogenesis.

LncRNAs show a high versatility in their mechanism-of-action, influencing many cellular processes such as spatial conformation of chromosomes, chromatin modifications and RNA transcription [24]. *HoxBlinc* recruits MLL1 and SETD1a to anterior *HoxB* loci and controls their gene expression by regulating chromatin states during development [10]. Recently, SETD1A was demonstrated to act as a positive epigenetic regulator of HSC function during hematopoiesis [25,26]. However, genetic deletion of one *Setd1a* allele could not rescue *HoxBlinc* overexpression mediated abnormal HSPC function and leukemogenesis. On the other hand, our *Mll1* KD experiments revealed that loss of MLL1 mitigates *HoxBlinc* overexpression induced abnormal HSPC function and leukemogenesis *in vivo*, demonstrating MLL1 as a dependency in *NPM1c*$^+$/*HoxBlinc*-overexpressing AML. We further showed that *HoxBlinc* overexpression increases *HoxBlinc* occupancy (in *cis* and *trans* actions) and MLL1 recruitment at the promoters of its target genes including the NPM1c$^+$ signature genes to induce aberrant long-range chromatin interaction networks and promote their expression. These results are in line with a recent study by Kuhn et al showing that the chromatin binding of MLL1 is critical for *NPM1*-mutated leukemias and Menin-Mll1 interaction controls the expression of *HOX*, *MEIS1* and *FLT3* genes in *NPM1*-mutated AML [18]. Interestingly, *NPM1* mutations are mutually exclusive with several genetic abnormalities such as the *MLL*-rearrangement and *MLL1*-partial tandem duplication in AML [27,28]. Thus, in *NPM1*-mutated AML, NPM1c$^+$ achieves and maintains its signature gene expression program via *HoxBlinc* overexpression, which increases MLL1 recruitment and induces aberrant chromatin structure at NPM1c$^+$ signature genes.

LncRNAs are promising candidates for targeted cancer therapy, especially when they act as tissue-specific drivers of cancer. However, to the best of our knowledge, there are no therapeutic examples targeting lncRNAs in AML to date. *HOXBLINC* is highly expressed in HSCs. Given the profound impacts of *HOXBLINC* overexpression on promoting HSC self-renewal and AML initiation, targeted therapeutics could be developed to suppress the

overexpressed *HOXBLINC* lncRNA in *NPM1*-mutated AML. To support this, we show that knockdown of *HOXBLINC*, likely through the normalization of the aberrant NPM1c$^+$ signature gene expressions, prolongs the survival of mice transplanted with *NPM1c*$^+$ AML cells. In line with our findings, Qian et al recently reported that CRISPR-Cas9-mediated specific DNA methylation at *DERARE* attenuates *HOXB* gene expression and alleviates leukemogenesis [29].

In summary, we show that *HOXBLINC* overexpression is a critical event to drive leukemogenesis by establishing aberrant NPM1c$^+$ signature gene expression program via controlling the MLL1 recruitment, chromatin domains and promoter accessibility in *cis* and *trans* actions. Our studies, therefore not only provide novel molecular insights into the biology of HSC and *NPM1*-mutated AML, but also create a unique opportunity for the identification of novel drug targets for *NPM1c*$^+$ AML.

Supplemental Information

Supplemental information includes 9 figures and 8 tables.

Competing Interest

The authors have declared that no competing interests exist.

Author Contribution

G.Z., H.L., Y.Q. F-C.Y., S.H. and M.X. conceived and designed experiments. G.Z., H.L, S.C., Q.L. and Y.G. performed experiments. H.L. performed bioinformatics and statistical analysis for mouse NGS data. J.X., Z.Z., D.C., B.X., and W.L. analyzed primary human patient samples and performed TCGA AML data analysis. H.N analyzed the histological slides from mouse tissues. Y.P. and O.A.G. provided *Npm1c*$^+$ knock in mouse and related samples. M.X. and S.H. wrote the original draft, O.A.G., S.D.N. and F-C.Y. revised and edited the manuscript.

Acknowledgments

This work was supported by grants from the National Institute of Health (R01HL141950 to M.X. and S.H., R01CA172408 to M.X. and F-C.Y., NIH R00CA178191 and R01DK121831 to O.A.G., and R01DK110108 to S.H.), as well as the Florida Academic Cancer Center Alliance (FACCA) Award (S.H. and M.X.). We also thank Rachael Mills at the Penn State Hershey College of Medicine for editing the manuscript.

Figure legends

Figure 1. *HOXBLINC* is activated in *NPM1c*$^+$ AML patients and loss of *HOXBLINC* perturbs NPM1c$^+$-mediated transcription program and leukemogenesis. (A) RT-qPCR analysis of *HOXBLINC* RNA expression in BMMNCs (n=24) and/or CD34$^+$ cells (n=14) from healthy individuals and in BMMNCs from *NPM1*wt (n=30) or *NPM1*c$^+$ (n=40) AML patients. Data is presented as dot plot with Median, *P*-value was calculated by Mann Whitney nonparametric test. **(B)** TCGA database (GSE62944) was used to retrieve RNA expression levels of *HOXBLINC* in BM cells from AML patients without *NPM1* mutations and *MLL* rearrangements (*NPM1c*$^-$*/MLLr*$^-$, n=128) or AML patients with *MLL* rearrangements (*MLL*r$^+$, n=11), or AML patients with *NPM1* mutations (*NPM1*c$^+$, n=42). Violin plots show mean, interquartile, and 1.5x interquartile. The width shows the probability density. **(C)** Volcano plot of RNA-seq analysis of differentially expressed genes in *Npm1*c$^+$ knock-in (*Npm1*$^{c/+}$) vs. WT LSK cells. **(D)** Heat map of RNA-seq analysis shows the up- and down-regulated genes in *HOXBLINC*-KRAB vs. WT OCI-AML3 cells. Arrows: dysregulated genes implicated in leukemogenesis. **(E)** The *HOXBLINC*-KRAB affected genes in OCI-AML3 cells were analyzed and annotated by the Gene Ontology (GO) analysis. **(F)** Enrichment of *HOXBLINC*-KRAB dysregulated genes in *NPM1*-mutated (*top*) and HOXA9 (*bottom*) oncogenic pathways by Gene Set Enrichment Analysis (GSEA). **(G)** Proliferation curves of WT, *shScramble*-expressing, and two different *shHOXBLINC*-expressing OCI-AML3 cells upon doxycycline (Dox) treatment for 5 days. Data are presented as mean ± SD of total viable cells. $^{**}P<0.01$. **(H)** Kaplan-Meier survival curve of inducible *shHOXBLINC* OCI-AML3 cell transplanted mice treated with or without Dox ($P<0.01$, 5 mice/group). **(I)** Kaplan-Meier survival curves of mice transplanted with control or *HOXBLINC*-KRAB BM cells from three AML patients with different gene mutations, 1315# (*NPM1*c$^+$;*FLT3*wt, *left*), 921# (*NPM1*c$^+$; *FLT3*mu, *middle*), LPP4 (*NPM1*wt;*MLL*r+, *right*), (4 mice/group). The Log-rank test was used to analyze differences between the survival curves.

Figure 2. Transgenic overexpression of *HoxBlinc* in mice lead to lethal AML. (A) RT-qPCR analyses of *HoxBlinc* expression levels on FACS-sorted BM cell populations along the hematopoiesis hierarchical differentiation tree of WT mice. **(B)** Diagram of the *Vav1* promoter/enhancer driven *HoxBlinc* transgene strategy. **(C)** PCR based transgenic integrating

location identification (TAIL) assay maps the *Vav1-HoxBlinc* transgene integration site to the mouse chromosome 18 in Line #1 *HoxBlinc*Tg mice. **(D)** RT-qPCR analysis of the *HoxBlinc* RNA expression in BM cells of WT and 2 lines of *HoxBlinc*Tg mice. **(E)** Kaplan-Meier survival curve of WT (n=15) and *HoxBlinc*Tg (Line #1, n=15) mice up to 1 year of age. **(F)** Gross appearance of spleens, femur, livers, feet, and lymph nodes of representative WT and moribund *HoxBlinc*Tg mice. **(G)** May-Giemsa stained PB smears and BM cytospins prepared from representative WT and moribund *HoxBlinc*Tg mice. **(H)** H&E stained femur and spleen sections of representative WT and moribund *HoxBlinc*Tg mice. **(I)** MPO stained liver sections from representative WT and moribund *HoxBlinc*Tg mice. Scale bar, 20μm.

Figure 3. Transgenic expression of *HoxBlinc* enhances HSC self-renewal and expands myelopoiesis. (A) FACS analysis of LSK ($Lin^-Sca\text{-}1^+c\text{-}Kit^+$) and LK ($Lin^-Sca\text{-}1^-c\text{-}Kit^+$) cell populations in the BM Lin^- cells of representative young WT and *HoxBlinc*Tg mice (*left*). Quantitation of the percent LSK and LK cells in the total BM cells of each genotype of mice is shown (*right*). **(B)** FACS analysis of LT-HSC, ST-HSC and MPP cell populations in the BM LSK cells of representative young WT and *HoxBlinc*Tg mice (*left*). The total number of LT- and ST-HSCs per femur in WT and *HoxBlinc*Tg mice are shown (*right*). **(C)** FACS analysis of GMP, MEP and CMP populations within BM LK cells of representative young WT and *HoxBlinc*Tg mice (*left*). Quantitation of the percent GMP and MEP cell populations in the BM LK cells of each genotype of mice is shown (*right*). (WT mice, n=7; *HoxBlinc*Tg mice, n=11; 8-10 weeks). **(D)** The number of colonies per 100 WT or *HoxBlinc*Tg BM LSK cells are shown (1st). Colonies were replated every 7 days for 4 times (2nd-5th). **(E)** Paired-daughter cell assays were performed on $CD34^-$LSK cells clone-sorted from BM cells of WT and *HoxBlinc*Tg mice, and each cell was analyzed for symmetric self-renewal (blue), asymmetric cell division (orange) or symmetric differentiation (gray). **(F)** Kinetic flow cytometric analyses of CD45.2 chimerism in the PB of recipients transplanted with WT or *HoxBlinc*Tg BM cells. **(G)** Kaplan-Meier survival curve of recipient mice (8 mice/genotype) receiving WT or *HoxBlinc*Tg BM cells. Data in (A-F) are presented as mean ± SEM; $^{**}P<0.01$ and $^{***}P<0.001$ by two-tailed unpaired Student's t-test.

Figure 4. Overexpression of *HoxBlinc* activates $NPM1c^+$ signature genes via enhancing promoter chromatin accessibility in LSK cells. (A) Heatmap of RNA-seq analysis shows the up- and down-regulated genes in *HoxBlinc*Tg vs. WT LSK cells. Red arrows: up-regulated genes

implicated in HSPC regulation/leukemogenesis. **(B)** Overlap of up- (*top*) or down- (*bottom*) regulated genes between *HoxBlinc*Tg vs. WT and *NPM1*$^{c/+}$ vs. WT LSK cells. **(C)** Enrichment of upregulated genes involved in *NPM1*-mutated signature (*top*) and HOXA9 (*Bottom*) oncogenic pathway upon overexpression of *HoxBlinc* in LSK cells by GSEA. **(D)** The *HoxBlinc* overexpression dysregulated genes in LSK cells were analyzed and annotated by GO analysis. **(E)** Overlap of global gain (*top*) or loss (*bottom*) of promoter chromatin accessibilities between ATAC-seq data of *HoxBlinc*Tg vs. WT and *NPM1*$^{c/+}$ vs. WT LSK cells. **(F)** RNA-seq (chromatin accessibility, *top 4 tracks*) and ATAC-seq (gene expression, *bottom 4 tracks*) analysis of WT and *Npm1*$^{c/+}$ LSK cells in the *HoxB* gene locus. **(G)** RNA-seq analysis (*top 4 tracks*) and ATAC-seq analysis (*bottom 2 tracks*) of WT and *HoxBlinc*Tg LSK cells in the *HoxB* gene locus.

Figure 5. *HoxBlinc* directly binds to target genes and mediates chromatin interactions to drive gene regulatory networks in HSPCs. (A) Schematic diagram showing the CTCF binding sites (CBS), location of the 4C baits in *HoxB* locus of the mouse genome. **(B, C)** Long-range chromatin interactions with *HoxB* locus (4 baits) as determined by 4C-seq analysis in WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells. *HoxBlinc* overexpression increased the interactions of CBS4/5 and +43CBS with the *HoxB* genes (B), *HoxBlinc* overexpression also enhanced the interaction of CBS4/5 or +43CBS with *Stat1* promoter region (C), ChIP-seq analysis of CTCF binding sites was obtained from the NCBI GEO public database (GSM918748). **(D)** ChIRP-seq analysis of *HoxBlinc* bindings to the *HoxB* and *Stat1* gene loci in WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells. **(E)** The pie chart shows the distribution of promoter, exon, intron, UTR and intergenic region within the total *HoxBlinc* binding sites in WT (*left*) and *HoxBlinc*Tg (*right*) Lin$^-$c-Kit$^+$ cells as identified by ChIRP-seq. **(F)** Overlap of genes with *HoxBlinc* binding gain to their promoter regions as identified by ChIRP-seq and the upregulated genes as identified by RNA-seq in *HoxBlinc*Tg vs. WT HSPCs. **(G)** Overlap of genes with *HoxBlinc* binding gain to their promoter regions as identified by ChIRP-seq and genes with promoter accessibility gain as identified by ATAC-seq in *HoxBlinc*Tg vs. WT HSPCs.

Figure 6. Recruitment of MLL1 is critical for *HoxBlinc* overexpression mediated target gene expression and abnormal HSPC function. (A) Number of colonies per 100 sorted GFP$^+$ *HoxBlinc*Tg BM LSK cells transduced with Sh*Scramble*, sh*Setd1a* or sh*Mll1* lentivirus are shown (1st). Colonies were replated every 7 days for 4 times (2nd-5th). **(B)** Kaplan-Meier

survival curve of recipient mice receiving Sh*Scramble* or sh*Mll1* LSK cells (P=0.0067, 4 mice/group). **(C)** Gross appearance of femur dissected from mice transplanted with sh*Scramble* or sh*Mll1* LSK cells. **(D)** FACS analysis of BM CD11b$^+$/CD117$^+$ (*top*) and GMP (*bottom*) cell populations from recipient mice (4 mice/condition) transplanted with sh*Scramble* or sh*Mll1* LSK cells. Data in (A, D) are presented as mean ± SEM; *P<0.05 and **P<0.01 by two-tailed unpaired Student's t-test. **(E)** ChIP-seq analysis of H3K4me3 (*top 2 tracks*) and MLL1 (*middle 2 tracks*), as well as ChIRP-seq analysis of *HoxBlinc* (*bottom 2 tracks*) in the *HoxB* loci of WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells. **(F)** The overlap among genes with gain of *HoxBlinc,* H3K4me3 and MLL1 bindings to their promoter regions based on ChIRP-seq and ChIP-seq analyses. **(G)** Heatmap (*right*) and peaks density (*left*) to show the overlap of *HoxBlinc* ChIRP-seq, MLL1 and H3K4me3 ChIP-seq.

Supplementary Figure legends:

Supplementary Figure 1, **related to Figure 1.** (**A**) Correlation between the expression of *HOXBLINC* and anterior *HOXB* genes, *HOXA9*, *MEIS1* and *HOXB13* in the TCGA AML datasets (GSE62944). **(B)** Survival analysis in AML patients (n=179) of the AML TCGA dataset stratified by *HOXBLINC* expression. *HOXBLINC*^{Hi}, the top thirty percentile of AML patients based on *HOXBLINC* expression; *HOXBLINC*^{Lo}, the bottom thirty percentile (n=49/group). **(C)** RT-qPCR analyses of *HOXBLINC* expression levels in AML cell lines with various gene mutations (See Table S8). **(D)** Heatmap of RNA-seq analysis shows the up- and down-regulated genes in *Npm1*c$^+$ knock in (KI, *Npm1*$^{c/+}$) vs. WT LSK cells. Red arrows: up-regulated genes implicated in HSPC regulation and/or leukemogenesis. **(E)** RT-qPCR analysis of *HoxBlinc*, *HoxB2* and *HoxB5* expression levels in WT and *Npm1*$^{c/+}$ LSK cells. **(F)** The *Npm1*c$^+$ KI affected genes in LSK cells were analyzed and annotated by the Gene Ontology (GO) analysis. **(G)** Enrichment of *Npm1*c$^+$ KI dysregulated genes in AML with *NPM1* mutated, HOXA9, JAK-STAT, and Cancer pathways by Gene Set Enrichment Analysis (GSEA).

Supplementary Figure 2, **related to Figure 1. *HOXBLINC* is critical for maintaining the leukemic state of *NPM1c*$^+$ AML. (A)** RT-qPCR analysis of *HOXBLINC* levels in WT and *HOXBLINC*-dCas9-KRAB OCI-AML3 cells to determine the *HOXBLINC* inhibition efficiency.

(B) Enrichment of *HOXBLINC*-KRAB dysregulated genes in WNT (*left*) and JAK-STAT (*right*) signaling pathways by GSEA. **(C)** RT-qPCR analysis of *HOXBLINC* levels in WT, sh*Scramble*, sh*HOXBLINC*_1#, and sh*HOXBLINC*_2# OCI-AML3 cells after Dox treatment to determine the *HOXBLINC* knockdown (KD) efficiency. **(D)** FACS analysis of cell cycle using propidium iodide shows that *HOXBLINC* KD increases the sub-G0 cell population in OCI-AML3 cells as compared to *shScramble* cells. **(E)** FACS analysis of human $CD45^+$ cell chimerism in bone marrow (BM), spleen (SP), and peripheral blood (PB) cells from NSG mice 30 days after receiving *shHOXBLINC* OCI-AML3 cells. Animals were treated with or without Doxycycline (Dox, 5 mice/group). Representative FACS plots from the BM cells of recipient mice treated with or without Dox are shown (*left*). **(F)** FACS analysis of human $CD45^+$ cell chimerism in BM (representative scatter plots were shown at *top*), SP, and PB of NSG mice transplanted with control or *HOXBLINC* inhibited primary BM cells from AML patients with different gene mutations. 1315# (*NPM1*c^+;*FLT3*wt, *left*), 921# (*NPM1*c^+;*FLT3*mu, *middle*), LPP4 (*NPM1*wt;*MLLr*$^+$, *right*), (4 mice/group). Data in (A), (B), (E) and (F) are presented as mean ± SEM; **P*<0.05, ***P*<0.01 and ****P*<0.001 by two-tailed unpaired Student's *t*-test.

Supplementary Figure 3, **related to Figure 2. Development of AML in *HoxBlinc*Tg mice. (A)** Spleen/body weight ratio for age-matched WT (n=8) and diseased *HoxBlinc*Tg (n=10) mice. **(B)** Parameters of PB counts were summarized from diseased *HoxBlinc*Tg (n=10) and age-matched WT (n=8) mice. WBC: white blood cells, NE: neutrophils, LY: lymphocytes, Hb: hemoglobin, RBC: red blood cells, PLT, platelets. **(C)** Quantitation of c-Kit^+ cells within total BM cells of age-matched WT (n=8) and diseased *HoxBlinc*Tg (n=10) mice. **(D)** Representative FACS analysis of c-Kit/Gr-1 (*top*), Ter119/B220 (*middle*) and CD4/CD8 (*bottom*) cell populations in the BM of diseased *HoxBlinc*Tg and age-matched WT mice. **(E)** MPO stained femur sections from representative WT and moribund *HoxBlinc*Tg mice. **(F)** H&E stained lymph nodes and liver sections of representative diseased *HoxBlinc*Tg and WT mice. **(G)** Representative FACS analysis of c-Kit^+/$CD11b^+$ cell populations in the BM of diseased Line#2 *HoxBlinc*Tg (L2) and age-matched WT mice. **(H)** Kaplan-Meier survival curve of WT (n=10) and Line#2 *HoxBlinc*Tg (n=14) mice up to 1 year of age. **(I)** May-Giemsa stained PB smears, and cytospins of BM and spleen (SP) cells prepared from representative moribund Line#2 *HoxBlinc*Tg mice. Scale bar, 20μm. ****P*<0.001, two-tailed unpaired Student's *t*-test.

Supplementary Figure 4, **related to Figure 3. Transgenic expression of *HoxBlinc* enhances HSC self-renewal and expands myelopoiesis. (A)** Quantitation of the percentage of the lineage cell populations in the BM of young WT (n=7) and *HoxBlinc*Tg (n=11) mice (8-10 weeks old) as determined by flow cytometry analysis. FACS plots from representative WT and *HoxBlinc*Tg mice are shown (*right*). **(B)** Quantitation of the total number of MPP cells per femur from these young WT and *HoxBlinc*Tg mice. **(C)** Quantitation of the percent CMP cells in the BM LK cells from these young WT and *HoxBlinc*Tg mice. **(D)** Frequencies of CFU-Cs in the BM cells from WT and *HoxBlinc*Tg mice (4 mice/genotype). GM: granulocytes/macrophages; E: burst forming unit-erythrocyte; GEMM: mixed colonies of GM and E and/or megakaryocytic cells. **(E-G)** FACS analysis of myeloid (Gr-1/CD11b) (E), LSK/LK (Lin^- cell gated) (F), and GMP/CMP/MEP (LK cell gated) (G) cell populations in BM of WT (n=4) and Line#2 *HoxBlinc*Tg (L2, n=7) mice (8-12 weeks old). **P*<0.05, ***P*<0.01, ****P*<0.001; two-tailed unpaired Student's *t*-test.

Supplementary Figure 5, **related to Figure 3. The aberrant HSPC function and AML-like disease induced by *HoxBlinc* overexpression in mice are transferable. (A)** Parameters of PB counts were summarized from recipient mice transplanted with WT (n=5) or *HoxBlinc*Tg (n=8) BM cells. **(B)** Flow cytometric analyses showing CD45.2 vs. CD45.1 chimerism as well as their respective lineage distribution (Gr-1/CD11b, CD3/B220, Gr-1/c-Kit) and percent LSK/LK cell populations (within Lin^- cells) in the BM of representative mice receiving WT or *HoxBlinc*Tg BM cells. **(C)** Quantitation of the percent lineage cell populations in the BM $CD45.2^+$ cell populations from recipient mice transplanted with WT (n=5) or *HoxBlinc*Tg (n=8) BM cells. **(D)** Images of May–Grunwald–Giemsa stained cytospin preparations of BM cells from representative mice receiving WT or *HoxBlinc*Tg BM cells. **(E)** H&E stained femur sections of representative recipients receiving WT or *HoxBlinc*Tg BM cells. Scale bar, 20μm. ***P*<0.01, ****P*<0.001; two-tailed unpaired Student's *t*-test.

Supplementary Figure 6, **related to Figure 4. *Npm1*c$^+$ knock in or transgenic expression of *HoxBlinc* remodels chromatin structure and alters hematopoietic transcription programs. (A)** RT-qPCR analysis shows the relative mRNA levels of *HoxB2-5*,9, *Meis1* and *Runx1* genes in *HoxBlinc*Tg LSK cells to those in WT. ***P*<0.01, ****P*<0.001; two-tailed unpaired Student's *t*-test. **(B)** Enrichment of upregulated genes involved in WNT (*Top*) and JAK-STAT (*Bottom*)

signaling pathways upon overexpression of *HoxBlinc* in LSK cells by GSEA. **(C)** ATAC-seq promoter density map of WT and *HoxBlinc*Tg LSK cells. **(D)** ATAC-seq promoter density map of WT and *Npm1*$^{c/+}$ LSK cells. **(E)** Overlap of upregulated genes and genes with promoter accessibility gain based on RNA-seq and ATAC-seq datasets of *NPM1*$^{c/+}$ vs. WT (*right*) and *HoxBlinc*Tg vs. WT (*left*) LSK cells. **(F)** RNA-seq (chromatin accessibility, *top 4 tracks*) and ATAC-seq (gene expression, *bottom 4 tracks*) analysis of WT and *Npm1*$^{c/+}$ LSK cells in the *HoxA* gene locus. **(G)** RNA-seq analysis (*top 4 tracks*) and ATAC-seq analysis (*bottom 2 tracks*) of WT and *HoxBlinc*Tg LSK cells in the *HoxA*, *Stat1*, *Cdr2* and *Lypla1* gene loci.

Supplementary Figure 7, **related to Figure 5. Transgenic expression of *HoxBlinc* alters long-range chromatin interactions and enhancer/promoter accessibility in Lin$^-$c-Kit$^+$ cells.** **(A)** Long-range chromatin interactions (viewed from 4 different baits of *HoxB* loci) with *Cdr2* and *HoxA* loci as determined by 4C-seq analysis in WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells. *HoxBlinc* overexpression increased the interactions of CBS4/5 and +43CBS with the *Cdr2* (*top*) and posterior *HoxA* (*bottom*) genes. ChIP-seq analysis of CTCF binding sites was obtained from the NCBI GEO public database (GSM918748). **(B)** ChIRP-seq analysis of *HoxBlinc* bindings to the *Cdr2*, *Wnt5a* and *HoxA* gene loci in WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells. **(C)** ChIRP-qPCR analysis of the bindings of *HoxBlinc* to the promoter regions of selected genes in WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells. The bindings of *HoxBlinc* to the *HoxB1-6* and *Runx1* promoter regions are enhanced by *HoxBlinc* overexpression in Lin$^-$c-Kit$^+$ cells. LacZ probes were used as control. **(D)** RT-qPCR analysis of RNA retrieved by the complementary *HoxBlinc* tiling probes in *HoxBlinc*Tg and WT Lin$^-$c-Kit$^+$ cells. **P*<0.05, ***P*<0.01, ****P*<0.001; two-tailed unpaired Student's *t*-test. **(E)** The genes which *HoxBlinc* binds to their promoter regions in Lin$^-$c-Kit$^+$ cells were analyzed and annotated by the Gene Ontology analysis. **(F)** List of top 10 significant transcription factor binding motifs enriched in the *HoxBlinc* binding sites in *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells.

Supplementary Figure 8, **related to Figure 6. *Setd1a* is not a critical player in *HoxBlinc* overexpression mediated hematopoietic phenotypes.** **(A)** RIP-qPCR analysis of *HoxBlinc* RNA retrieved by antibodies against MLL1, SETD1A and LSD1 from OCI-AML3 and OCI-AML2 cells. **(B)** RT-qPCR analysis of *Setd1a* and *Mll1* mRNA levels in sorted GFP$^+$LSK cells transduced with Sh*Scramble*, sh*Setd1a* or sh*Mll1* lentivirus. **(C)** RT-qPCR analysis of *Setd1a*

mRNA levels in the BM cells from *Setd1a*$^{+/-}$ or *Setd1a*$^{+/-}$;*HoxBlinc*Tg mice to determine the efficiency of pIpC for inducing *Setd1a* gene inactivation. **(D)** FACS analysis of LSK and LK cell populations in the BM Lin$^-$ cells of WT, *HoxBlinc*Tg, *Setd1a*$^{+/-}$, and *Setd1a*$^{+/-}$;*HoxBlinc*Tg mice. **(E)** FACS analysis of GMP, MEP and CMP populations within BM LK cells of WT, *HoxBlinc*Tg, *Setd1a*$^{+/-}$, and *Setd1a*$^{+/-}$;*HoxBlinc*Tg mice. **(F)** FACS analysis of CD117$^+$/CD11b$^+$ cells in BM cells of WT, *HoxBlinc*Tg, *Setd1a*$^{+/-}$, and *Setd1a*$^{+/-}$;*HoxBlinc*Tg mice. Data in D-F were from 4-5 mice/genotype. n.s. means $P>0.05$ by two-tailed unpaired Student's *t*-test.

Supplementary Figure 9, **related to Figure 6. *HoxBlinc* overexpression increases MLL1 recruitment and H3K4me3 occupancy to its target genes. (A-B)** ChIP-seq analysis of H3K4me3 (*top 2 tracks*), MLL1 (*middle 2 tracks*) and ChIRP-seq analysis of *HoxBlinc* (*bottom 2 tracks*) at the *Stat1, Cdr2, Lypla1, Eno1 and Car6* gene loci of WT and *HoxBlinc*Tg Lin$^-$c-Kit$^+$ cells.

Supplementary Table Information:

Supplementary Table 1: Patient information list

Supplementary Table 2: Gene list of common regulated genes by transgenic *HoxBlinc* and *Npm1*c$^+$ KI based on RNA-seq

Supplementary Table 3: List of overlapping genes with gain or loss of promoter accessibilities caused by transgenic *HoxBlinc* and *Npm1*c$^+$ KI based on ATAC-seq data

Supplementary Table 4: List of *HoxBlinc* ChIRP transcription motifs

Supplementary Table 5: List of overlapping genes of *HoxBlinc* ChIRP-seq, H3K4me3 ChIP-seq, and MLL1 ChIP-seq

Supplementary Table 6: Sequences of primers and sgRNAs

Supplementary Table 7: Key resources

Supplementary Table 8: Detailed information of AML cell lines

References

1 Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med* **352**, 254-266, doi:10.1056/NEJMoa041974 (2005).

2 Schlenk, R. F. *et al.* Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *N Engl J Med* **358**, 1909-1918, doi:10.1056/NEJMoa074306 (2008).

3 Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391-2405, doi:10.1182/blood-2016-03-643544 (2016).

4 Grisendi, S. *et al.* Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* **437**, 147-153, doi:10.1038/nature03915 (2005).

5 Falini, B. *et al.* Immunohistochemistry predicts nucleophosmin (NPM) mutations in acute myeloid leukemia. *Blood* **108**, 1999-2005, doi:10.1182/blood-2006-03-007013 (2006).

6 Bolli, N. *et al.* Born to be exported: COOH-terminal nuclear export signals of different strength ensure cytoplasmic accumulation of nucleophosmin leukemic mutants. *Cancer Res* **67**, 6230-6237, doi:10.1158/0008-5472.CAN-07-0273 (2007).

7 Alcalay, M. *et al.* Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* **106**, 899-902, doi:10.1182/blood-2005-02-0560 (2005).

8 Spencer, D. H. *et al.* Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. *Leukemia* **29**, 1279-1289, doi:10.1038/leu.2015.6 (2015).

9 Alharbi, R. A., Pettengell, R., Pandha, H. S. & Morgan, R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia* **27**, 1000-1008, doi:10.1038/leu.2012.356 (2013).

10 Deng, C. *et al.* HoxBlinc RNA Recruits Set1/MLL Complexes to Activate Hox Gene Expression Patterns and Mesoderm Lineage Development. *Cell reports* **14**, 103-114, doi:10.1016/j.celrep.2015.12.007 (2016).

11 Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120-124, doi:10.1038/nature09819 (2011).

12 Luo, H. *et al.* HOTTIP lncRNA Promotes Hematopoietic Stem Cell Self-Renewal Leading to AML-like Disease in Mice. *Cancer Cell* **36**, 645-659 e648, doi:10.1016/j.ccell.2019.10.011 (2019).

13 Mallardo, M. *et al.* NPMc+ and FLT3_ITD mutations cooperate in inducing acute leukaemia in a novel mouse model. *Leukemia* **27**, 2248-2251, doi:10.1038/leu.2013.114 (2013).

14 Vassiliou, G. S. *et al.* Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. *Nat Genet* **43**, 470-475, doi:10.1038/ng.796 (2011).

15 Deng, C. *et al.* HoxBlinc RNA Recruits Set1/MLL Complexes to Activate Hox Gene Expression Patterns and Mesoderm Lineage Development. *Cell Rep* **14**, 103-114, doi:10.1016/j.celrep.2015.12.007 (2016).

16 Luo, H. *et al.* CTCF boundary remodels chromatin domain and drives aberrant HOX gene transcription in acute myeloid leukemia. *Blood*, doi:10.1182/blood-2017-11-814319 (2018).

17 Brunetti, L. *et al.* Mutant NPM1 Maintains the Leukemic State through HOX Expression. *Cancer Cell* **34**, 499-512 e499, doi:10.1016/j.ccell.2018.08.005 (2018).

18 Kuhn, M. W. *et al.* Targeting Chromatin Regulators Inhibits Leukemogenic Gene Expression in NPM1 Mutant Leukemia. *Cancer Discov* **6**, 1166-1181, doi:10.1158/2159-8290.CD-16-0237 (2016).

19 Huarte, M. The emerging role of lncRNAs in cancer. *Nat Med* **21**, 1253-1261, doi:10.1038/nm.3981 (2015).

20 Trimarchi, T. *et al.* Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* **158**, 593-606, doi:10.1016/j.cell.2014.05.049 (2014).

21 Garzon, R. *et al.* Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci U S A* **111**, 18679-18684, doi:10.1073/pnas.1422050112 (2014).

22 Schwarzer, A. *et al.* The non-coding RNA landscape of human hematopoiesis and leukemia. *Nat Commun* **8**, 218, doi:10.1038/s41467-017-00212-4 (2017).

23 Luo, M. *et al.* Long non-coding RNAs control hematopoietic stem cell function. *Cell Stem Cell* **16**, 426-438, doi:10.1016/j.stem.2015.02.002 (2015).

24 Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298-1307, doi:10.1016/j.cell.2013.02.012 (2013).

25 Li, Y. *et al.* Setd1a and NURF mediate chromatin dynamics and gene regulation during erythroid lineage commitment and differentiation. *Nucleic Acids Res* **44**, 7173-7188, doi:10.1093/nar/gkw327 (2016).

26 Arndt, K. *et al.* SETD1A protects HSCs from activation-induced functional decline in vivo. *Blood* **131**, 1311-1324, doi:10.1182/blood-2017-09-806844 (2018).

27 Falini, B. *et al.* NPM1 mutations and cytoplasmic nucleophosmin are mutually exclusive of recurrent genetic abnormalities: a comparative analysis of 2562 patients with acute myeloid leukemia. *Haematologica* **93**, 439-442, doi:10.3324/haematol.12153 (2008).

28 Cancer Genome Atlas Research, N. *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).

29 Qian, P. *et al.* Retinoid-Sensitive Epigenetic Regulation of the Hoxb Cluster Maintains Normal Hematopoiesis and Inhibits Leukemogenesis. *Cell Stem Cell* **22**, 740-754 e747, doi:10.1016/j.stem.2018.04.012 (2018).

30 Yang, H. *et al.* Gain of function of ASXL1 truncating protein in the pathogenesis of myeloid malignancies. *Blood* **131**, 328-341, doi:10.1182/blood-2017-06-789669 (2018).

31 Li, J. *et al.* Loss of Asxl2 leads to myeloid malignancies in mice. *Nat Commun* **8**, 15456, doi:10.1038/ncomms15456 (2017).

32 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).

33 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

34 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

35 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515, doi:10.1038/nbt.1621 (2010).

36 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

37 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

38 Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693, doi:10.1126/science.1192002 (2010).

39 Deng, C. *et al.* USF1 and hSET1A mediated epigenetic modifications regulate lineage differentiation and HoxB4 transcription. *PLoS Genet* **9**, e1003524, doi:10.1371/journal.pgen.1003524 (2013).

40 Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* **44**, 667-678, doi:10.1016/j.molcel.2011.08.027 (2011).

41 Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet Journal* **17**, 10-12 (2011).

42 Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* **7**, 1338, doi:10.12688/f1000research.15931.2 (2018).

43 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

44 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

45 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

46 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

47 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

48 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

49 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).

50 Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* **8**, 509-524, doi:10.1038/nprot.2013.018 (2013).

51 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

52 van de Werken, H. J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* **9**, 969-972, doi:10.1038/nmeth.2173 (2012).

53 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

54 Huang, Y. *et al.* cis-Regulatory Circuits Regulating NEK6 Kinase Overexpression in Transformed B Cells Are Super-Enhancer Independent. *Cell Rep* **18**, 2918-2931, doi:10.1016/j.celrep.2017.02.067 (2017).

55 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-29, doi:10.1002/0471142727.mb2129s109 (2015).

56 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).

57 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

58 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389, doi:10.1038/nature10730 (2012).

59 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978, doi:10.1038/nmeth.4401 (2017).

60 Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176**, 361-376 e317, doi:10.1016/j.cell.2018.11.022 (2019).

Methods

Generation of the *HoxBlinc* transgenic (Tg) mouse model. All studies were conducted in accordance with the regulatory guidelines by the Institutional Animal Care and Use Committee (IACUC) at the UT Health San Antonio and University of Miami Miller School of Medicine. Full-length mouse *HoxBlinc* cDNA was cloned into downstream of *Vav1* promoter (HS321/45-vav vector) followed by *Vav1* enhancer to ensure transgene expression solely in hematopoiesis [30]. The plasmid DNA was digested with SacII to remove the PbsIISK backbone and was used for injection into the pronuclei of fertilized eggs from C57BL/6 mice. Two *HoxBlinc*Tg founder mice were obtained by PCR screening of the tail genomic DNAs with P1 (to detect both endogenous and transgenic *HoxBlinc* gene) and P2 (to specifically recognize the transgenic *HoxBlinc* gene) primer sets (Table S6), the positive band is 589bp. Transgenic founder mice were crossed with WT C57BL/6 mice. *HoxBlinc* negative siblings of the *HoxBlinc*Tg mice were used as controls throughout the study. Two *HoxBlinc*Tg lines were used for this study. The mHoxBlinc-set1 and mHoxBlinc-set2 were used as real-time PCR primers to recognize both endogenous and exogenous *HoxBlinc* lncRNA (Figure 2D, Table S6). The levels of transgenic expression of *HoxBlinc* were also confirmed by RNA-seq analysis (Figure 6A).

Morphological and histological analyses of the hematopoietic organs. PB was collected by tail vein bleeding and was subjected to an automated blood count (Hemavet System 950FS). PB smears were subjected to May-Grünwald-Giemsa staining for morphological and lineage differential analysis. Morphological evaluation of BM and spleen cells were performed on cytospins followed by May-Grünwald-Giemsa staining. For histopathological analyses, femurs were fixed in 10% Neutral Buffered Formalin (10% NBF) and demineralized in a solution of 10% EDTA for 1-2 weeks. The specimens and other soft tissues (spleens, lymph nodes and livers) were fixed in 10% Neutral Buffered Formalin and then dehydrated using ethanol and cleared in xylenes. The specimens were then embedded in melted paraffin and allowed to harden. Thin sections (5μm) were cut and floated onto microscope slides. For routine assessment, slides were stained with hematoxylin and eosin (H&E) staining. For MPO and hCD45 immunohistochemical staining, the tissue was rehydrated followed by heat-induced epitope retrieval, peroxidase and serum blocking. Samples were then incubated with MPO (R&D, #MAB3174) or hCD45 antibody (BD, #555485) overnight at 4°C followed by staining with the biotinylated second antibody. Slides were visualized under a Nikon TE2000-S microscope. Images were taken by a QImaging camera and QCapture-Pro software (Fryer Company Inc.). Chemicals were obtained from Sigma (St. Louis, MO) unless otherwise indicated.

Flow cytometry analysis, cell sorting, and colony assay. Total white blood cells were obtained after lysis of red blood cells with red blood cell lysis buffer (QIAGEN 1045722). Single-cell suspensions from BM, spleen and PB were stained with panels of fluorochrome-conjugated antibodies (listed in Table S7). The analyses were performed using BD LSRII or LSR Fortessa flow cytometer. All data were analyzed by FlowJo.V10 software. Purified LSK cells were used for the colony and replating assays. Briefly, BM cells from 6-8 weeks old mice were pre-enriched with lineage depletion beads (MiltenyiBiotec, Bergisch Gladbach, Germany) and then stained with c-Kit, lineage, and Sca-1 antibodies (listed in Table S7) and then sorted by BD FACSAriaII. The purity of selected LSK cells were routinely over 98%. The purified LSK were incubated in RPMI1640 containing 30% FBS, 2% BSA, and a combination of cytokines (mG-CSF, 10ng/mL; mIL-3, 5ng/mL; mEPO, 4U/mL; hTPO, 100ng/mL; and mSCF, 100ng/mL). Colonies were scored at 8-10 days. For replating assays, CFU assays were performed with LSK cells in methylcellulose medium supplemented with the same cytokine cocktails. Colonies were passaged every 7 days for 4 sequential replatings. Antibodies used are listed in Table S7.

Competitive repopulation assay. The competitive repopulation assay was performed by transplanting a total of 1×10^6 BM cells at 1:1 ratio of CD45.2 (WT or *HoxBlin*Tg): CD45.1 (B6.SJL) into lethally irradiated (950cGy) B6.SJL recipients (CD45.1) by tail vein injection. The contribution of $CD45.1^+$ vs. $CD45.2^+$ cells in the PB was monitored every month for 6 months after transplantation.

Suspension Culture. For $Lin^-c\text{-}Kit^+$ (LK) cell selection, BM cells were pre-enriched with lineage depletion beads (MiltenyiBiotec, Bergisch Gladbach, Germany) and then stained with c-Kit and lineage antibodies (listed in Table S7) and sorted by BD FACSAriaII. The purity of selected LK cells was routinely over 98%. The LK cells were incubated in RPMI1640 containing 30% FBS, 2% BSA, and a combination of cytokines (mG-CSF, 10ng/mL; mIL-3, 5ng/mL; mEPO, 4U/mL; hTPO, 100ng/mL; and mSCF, 100ng/mL). At weekly intervals, cultures were mixed by pipetting and half of the culture media were removed, which was then replaced by the newly prepared medium with the same combinations of cytokines. Cells in the collected media were counted and used for flow cytometric analysis. Total CFUs generated at each time point in the suspension culture were evaluated by culturing a fraction of the expanded cells in the colony assay as described above.

Paired-daughter cell assay. To examine the frequency of HSCs to undergo self-renewal and differentiation, we performed paired-daughter cell assays [31]. Single $CD34^-$LSK cells from BM of WT and *HoxBlinc*Tg mice were clone-sorted into 96-well plates. The cells were maintained in RPMI1640 media supplemented with mSCF (100ng/mL) and hTPO (50ng/mL). After the first cell division, the two daughter cells were separated, one per well for an additional 12 days in the media supplemented

with mSCF, hTPO, mEPO, mIL-3 and mG-CSF. The self-renewal and differentiation capabilities of cultured $CD34^-$LSK cells were determined by morphological analyses of the progenies of the two daughter cells microscopically following May-Grünwald-Giemsa staining. A total of 192 single cells (two 96-well plates) were analyzed to calculate the percentage of symmetric/asymmetric cell divisions.

Human AML samples and patient data analysis. All human samples from healthy donors and patients with primary AML were obtained after informed consent following the guidelines of the Institutional Review Board of the Institute of Hematology and Blood Disease Hospital, Tianjin, China. BM low-density mononuclear cells (MNCs) were purified using Ficoll-Hypaque. MNCs of AML patients or healthy controls were treated with RNase-free DNase to remove contaminating genomic DNA. First-strand cDNA was synthesized. Real-time PCR was performed using Fast SYBR Green master mix. PCR amplifications were performed in triplicate for each gene of interest along with parallel measurements of human *GAPDH* (internal controls). The expression of *HOXBLINC* lncRNA was also assessed using the TCGA AML dataset, containing a list of 179 AML patients.

RNA isolation, quantitative RT-PCR, as well as RNA-sequencing and data analysis. Total RNAs from WT or *HoxBlinc*Tg LSK cells were purified with the RNeasy mini-isolation kit according to the manufacturer's instructions (Qiagen, MD, USA). A total of 2μg RNA was subjected to reverse-transcription with Superscript II Reverse Transcriptase (Invitrogen) and analyzed by a real-time polymerase chain reaction (PCR) Detection System (Bio-Rad). Primer sequences for qPCR are listed in Table S6 and key reagents are listed in Table S7. RNA library was prepared with the IlluminaTruSeq strand-specific mRNA sample preparation system. Paired-end RNA-seq was performed by the UF ICBR core facility according to standard protocols. All of the sequencing reads were processed and aligned to the mouse genome assembly (mm9) or human genome assembly (hg19) using TopHat (version 2.0) and Bowtie2 [32-34]. To prevent false positives, a stringent approach was taken to identify differentially expressed genes. First, FPKM (paired-end fragments per kilobase of exon model per million mapped reads) was calculated for each gene and further normalized (RMS-FPKM). To prevent false positives due to the fluctuation of detection among genes with low expression levels, only genes with 50 or more reads in one of the conditions (WT control or *HoxBlinc*Tg) were included in the analysis. Differential expression was determined according to abundance estimations (FPKMs) processed with Cufflinks v2.2.1 and Cuffdiff [35]. Differentially expressed genes were identified if the ratio of RMS-FPKM in the two conditions was greater than 2.0-fold, or undetectable in one condition but detectable by more than 50 reads in the other. The scatter plot was based on the $\log_2$ transformation of the RMS-FPKM values. Expression level increased or decreased genes were marked with red or blue, respectively. Gene Ontology analysis with the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool

(https://david.ncifcrf.gov/, Version 6.8) [36] . Gene set enrichment analysis (GSEA) [37] was conducted according to recommended parameters (http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html) using gene sets obtained from the Molecular Signatures Database. The normalized expression data was loaded to Integrated Genomic Viewer (IGV) for comparison of different *HOX* loci. The sequence reads have been deposited in the NCBI GEO dataset (GSE115096). Key software and algorithms used are listed in Table S7.

RNA immunoprecipitation (RIP) assay. The RNA-IP protocol was performed according to the previous reported [15,38]. The OCI-MAL2 and OCI-AML3 cells were harvested and washed with PBS (e.g. 10^7 cells in 2 mL PBS), and then resuspended in freshly nuclear isolation buffer (1.28 M sucrose, 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 4% Triton X-100), and then kept on ice for 20 min (with frequent mixing). After that, nuclei were precipitated by centrifugation at 2,500 g for 15 min, and then resuspended in freshly lysis buffer (10 mM HEPES-KOH pH7, 150 mM KCl, 5 mM $MgCl_2$, 5 mM EDTA, 0.5% IGEPAL-CA-630, 0.5 mM dithiothreitol, 0.2 mg/mL Heparin, 100 U/mL RNase OUT, 100 U/mL Superase IN, protease inhibitor tablet adding before use). Chromatin shearing was performed with sonication. The suspension was spun down for three times at 14,000 g at 4 °C for 10 min, and supernatant was harvested and precipitated with antibody (2-10 μg) overnight at 4 °C with rotation. The precipitant was captured by the equilibrated Protein G magnetic beads followed by washing four times in ice-cold NT2 buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% IGEPAL-CA-630) supplemented with 0.02 mg/mL heparin. The RNA-protein complexes were eluted twice with 500 mL SDS-EDTA (50 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA, 1% SDS) for 10 min at 65°C. Coprecipitated *HoxBlinc* RNA was purified by resuspending beads in TRIzol RNA extraction reagent, eluted with nuclease-free water, treated with DNaseI, and then examined with RT-qPCR.

Chromatin immunoprecipitation (ChIP). ChIP was performed as described previously[39]. Briefly, Nuclei were sonicated with the Bioruptor™ UCD200. Chromatin samples prepared from 5×10^6 cells were immunoprecipitated with antibodies against MLL and H3K4me3, separately. The immunoprecipitates were subjected to a series of washing steps to remove non-specific binding materials. After reverse-crosslinking, the DNA samples were purified and then analyzed by RT-qPCR. Final results represent the percentage of input chromatin and error bars (SEM) through triplicate experiments. The MLL1 and H3K4me3 ChIP-DNA libraries were prepared using Illumina's TruSeq ChIP Sample Preparation Kit according to the manufacturer's instructions (Cat #IP-202-1012). The quality of the library was checked with Qubit and Agilent Bioanalyzer. Final libraries were submitted to paired-end sequencing of 100 bp length on an Illumina HiSeq 3000.

Chromatin Isolation by RNA Immunoprecipitation (ChIRP). ChIRP assay was carried out based on our previously described with some modifications [40]. Briefly, 20 million cells were collected and cross-linked in 20 ml of PBS buffer containing 1% glutaraldehyde (Sigma, Cat# G5882) at room temperature for 10 mins. Cross-linked cells were washed by chilled PBS twice, then per 100 mg of the pellet of cells were lysed in 1ml lysis buffer (50mM Tris-Cl pH 7.0, 10mM EDTA, 1% SDS, PMSF, DTT, P.I. and SUPERase were added before use), and sonicated using a Bioruptor (Diagenode) to prepare chromatin. Chromatin was diluted twice using hybridization buffer (750mM NaCl, 1%SDS, 50mM Tris-Cl 7.0, 1.0mM EDTA, 15% Formamide, add DTT, PMSF, P.I, and SUPERase-in fresh), hybridized with 100 pmol of biotinylated DNA probes targeting *HoxBlinc* or *LacZ* (sequences are listed in Table S6), and incubated with 100ul of Streptavidin-magnetic C1 beads (Invitrogen). RNA and DNA were purified after beads were washed 5 times with washing buffer (2x SSC, 0.5% SDS) and subjected to analysis by RT-qPCR. Then DNA was made to a library for CHIRP-seq. Libraries were prepared using Illumina's TruSeq ChIP Sample Preparation Kit according to the manufacturer's instructions (Catalog: #IP-202-1012). The quality of the library was checked with Qubit and Agilent Bioanalyzer. Final CHIRP libraries were submitted to paired-end sequencing of 100bp length on an Illumina HiSeq 2500.

ChIP-seq and ChIRP-seq data analysis. The ChIP-seq or ChIRP-seq raw data were trimmed through cutadapt (http://cutadapt.readthedocs.io, version 1.2.0) to remove adaptors and low quality reads[41]. Cutadapt-filtered reads aligned to mouse reference genome (mm9) using Bowtie2 with default parameters [33], and the quality of these trimmed data was evaluated by FastQC program [42]. After alignment, SAM files were converted to BAM files and sorted using Samtools [43]. Peak calling was performed using peak calling algorithm MACS2 [44]. Peaks were transformed to visualized files with deepTools[45], including control and experimental datasets. All sequencing tracks were visualized using the Integrated Genomic Viewer [46]. Peaks annotation was performed with the command "annotatePeaks.pl" with HOMER package [47]. Differential Peaks calling was performed with getDifferentialPeaks in Homer software. For ChIRP-seq binding motif analysis, the *de novo* motif analysis was performed by the "findmotifsgenome.pl" from the HOMER motif discovery algorithm [47]. The genes and pathways regulated by the *HoxBlinc* bound regions were analyzed and annotated by the Gene Ontology analysis with the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool (https://david.ncifcrf.gov/, Version 6.8) [48,49]. Each GO term with a p-value more than 1 × 10^-3 is used for cutoff (threshold: 10^-3). All genomics datasets were deposited in the NCBI GEO under accession number (GSE115096).

Circular Chromosome conformation capture (4C) assays. The 4C-seq assay was performed as previously described [50] with minor modifications. In brief, the 3C library prepared with 400U of *BglII* enzyme (NEB) digestion at 37°C overnight and 6000 U of T4 DNA ligase (NEB) ligation at 16°C

overnight, was then digested with 300U of *NlaIII* (NEB) at 37°C for 4 hours with shaking. The reaction was stopped by adding SDS to a final concentration of 1.6% at 65°C for 20 minutes. The digested chromatin/DNA was diluted in ligation buffer containing 6000 U of T4 DNA ligase at 16°C. The ligated circularized chromatin was then reverse crosslinked by adding Proteinase K (Invitrogen) overnight and the 4C DNA was purified by phenol-chloroform followed by Qiagen PCR kit and amplified by inverse PCR using bait-specific primers. The invert PCR products were cloned into pGEM®-T Easy Vector Systems (Promega) for Sanger sequencing. Libraries for 4C-seq were constructed by adding barcoded Illumina adapters to the 5' end of each primer (Table S6). PCR reactions were performed using the Expand Long Template PCR System (Roche), and DNA was purified and quantified before sequencing. The bar-coded DNA libraries were sequenced as 150bp pair-end reads using the Illumina Nextseq500 platform. Reads were aligned to the reference human genome (build hg19) with Bowtie2 2.2.9 [51]. 4C-seq data was analyzed either using the 4cseq_pipeline [52] and normalized using DESeq2 [53]. Statistical analysis for differential interactions between genotypes was performed using DESeq2. Spearman correlation of each genotype was performed using R [54]. The 4C-sequencing sequence reads have been deposited in the NCBI GEO database (GSE115096).

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq). ATAC-seq was performed as described previously [12,55]. In Brief, 5×10^4 cells were used for library preparation. Washed cells were re-suspended in lysis buffer containing 10 mM Tris-HCL (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1 % NP-40. After washing with cold 1x phosphate buffered saline (PBS) buffer, these cells were fragmented with Tn5 Transposes for transposition reaction at 37°C for 30 min. Then these DNA fragments were purified using the MinElute Kit (QIAGEN). Library fragments were amplified using 1x NEB next PCR master mix and 1.25 μM indexed Nextra PCR primers (Ad1_noMX and Ad2.1-2.4 barcoded primers) with following PCR conditions: 72 °C for 5 min, 98 °C for 30 s, followed by thermocycling at 98 °C for 10 s, 63 °C for 30 s and 72 °C for 1 min. The eluted DNA was used in a quantitative PCR (qPCR) reaction to estimate the optimum number of amplification cycles. Libraries were quantified using qPCR (Kapa Library Quantification Kit for Illumina, Roche), and libraries were purified with AMPure XP beads (Beckman Coulter), and then the quality of the DNA library was tested by Agilent Bioanalyzer 2100 prior to sequencing with 2x100 bp paired-end reads on an Illumina HiSeq 2500. Each sample includes two replicates for statistical analysis.

ATAC-seq analysis. Two biological replicates of ATAC-seq experiments were carried out and analyzed according to our previous reports[12,16]. For quality control, first, each replicate should have 50 million reads for paired-end sequencing. Second, the alignment rate of each replicate is more than 95%. Third, we also removed the mitochondrial related reads from total reads after alignment and PCR duplicates were

also removed. Finally, non-uniquely aligned reads were filtered based on MAPQ scores with samtools (MAPQ > 30), and plotPCA from BiocGenerics package in R package (R/3.6.1) was carried out to identify the variance between control and treatment groups. Moreover, fragSizeDist from ATACseqQC package in R package was carried out to show the fragment size distribution for control and treatment groups. Briefly, all of the raw fastq files were filtered through cutadapt (http://cutadapt.readthedocs.io, version 1.2.0) to remove adaptors and low quality reads[41]. Cutadapt-filtered reads aligned to mouse genome (mm9) using Bowtie2 with default parameters (version Bowtie 2/2.2.6)[33], and the quality of these trimmed data was evaluated by FastQC program (version 0.11.8)[42]. After alignment, SAM files were converted to BAM files and sorted using Samtools (version 1.8.0) [43]. PCR duplicates were removed using Picard MarkDuplicates (version 2.0.1), and mitochondrial reads were removed with Samtools [56]. ENCODE blacklist regions were filtered (https://sites.google.com/site/anshulkundaje/projects/blacklists). Peak calling was performed using peak calling algorithm MACS2 with parameters ("-g mm -p 1e-9 –nolambda -f BAMPE –nomodel –shiftsize=100 --extsize 200")[44]. bedGraphToBigWig program was employed to generate the bigWig file of fragment or read coverages, including control and experimental datasets (https://www.encodeproject.org/software/bedgraphtobigwig/). All sequencing tracks were viewed using the Integrated Genomic Viewer (IGV/2.4.19)[46]. Peaks annotation was carried out with the command "annotatePeaks.pl" from HOMER package (version 4.8) [47] and GREAT[57]. DEseq2 (Benjamini-Hochberg adjusted $p < 0.05$; FoldChange≥ 2) were also performed to find the differential binding sites between two peak files, including control and treatment groups with C+G normalized and "reads in peaks" normalized data [58]. The *de novo* motif analysis was performed by the "findmotifsgenome.pl" from the HOMER package [47]. For each genomic feature (peaks or chromVAR annotation), we calculated the chromatin accessibility median deviation z-score (for chromVAR features) or fragment counts (for peaks) in control and treatment groups with chromVAR package in R language [59,60]. Pearson's correlation coefficient and Pearson's $\chi 2$-test were carried out to calculate overall similarity between the replicates of ATAC-seq global open chromatin signatures. All genomics datasets were deposited in the NCBI GEO under accession number (GSE115096).

dCas9-mediated inactivation of *HoxBlinc* in AML cells. Guide RNA plasmid targeting the promoters of *HOXBLINC* were designed using the Zhang laboratory web tool (http://crispr.mit.edu), and sgRNA was subcloned into the pLKO5.sgRNA.EFS.GFP vector (Addgene#57822). The gRNA plasmids encoding GFP and puromycin resistance were co-transfected with a repressive plasmid encoding dCas9-KRAB (pHR-SFFV-dCas9-BFP-KRAB, Addgene plasmid #46911). 24 hrs after transfection, OCI-AML3 cells were selected with 2μg/mL of puromycin for another 48hrs, and then GFP^+ cells were sorted by FACS. RNA was extracted from GFP positive cells and was analyzed by RT-qPCR.

Xenotransplantation of human OCI-AML3 or primary leukemic cells. Adult NOD.*Cg-Prkdc*^{scid}*Il2rg*^{tm1Wjl}/SzJ (NSG) mice (6–8 weeks old) were pretreated with 280cGy total body irradiation, then 5x10^5 viable AML cells (in 300µl of PBS) were transplanted into each NSG mice by tail-vein injection. After transplantation, the recipients were administered with doxycycline (for Dox inducible *HoxBlinc* KD) in the drinking water (Sigma D-9891, 1 mg/ml, 1% sucrose, newly prepared every other day) until being sacrificed. Daily monitoring of the mice for symptoms (ruffled coat, hunched back, weakness and reduced motility) and survival time. Human CD45 chimerism in BM, spleen and PB cells were analyzed by flow cytometry as described above.

Quantification and statistical analysis. Differences between experimental groups were determined by the Student's *t*-test or analysis of variance (ANOVA) followed by Newman-Keuls multiple comparison tests as appropriate. $P<0.05$ is considered significant. For *in vivo* experiments, the sample size chosen was based on the generalized linear model with Bonferroni multiple comparison adjustments; with the proposed sample size of at least five mice/group/genotype. Animals were randomly assigned to each study. For all *in vitro* experiments, at least three independent experiments with more than three biological replicates for each condition/genotype were performed to ensure adequate statistical power.

Figures

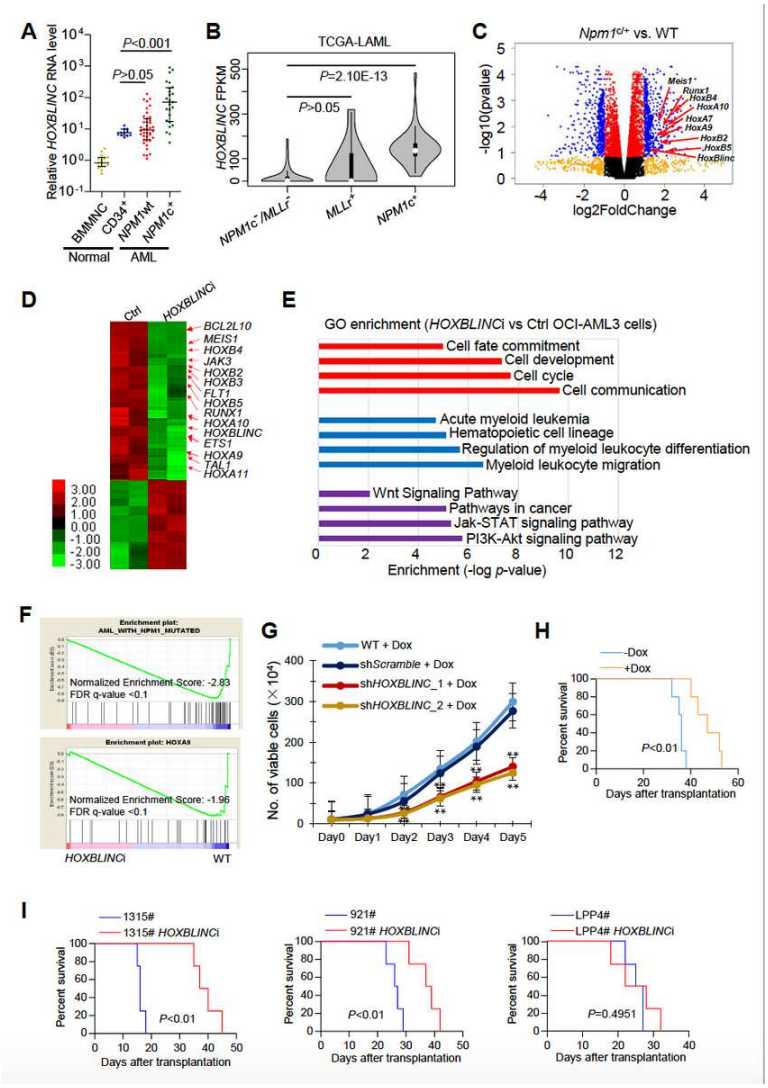

Figure 1

HOXBLINC is activated in NPM1c+ AML patients and loss of HOXBLINC perturbs NPM1c+-mediated transcription program and leukemogenesis. (A) RT-qPCR analysis of HOXBLINC RNA expression in BMMNCs (n=24) and/or CD34+ cells (n=14) from healthy individuals and in BMMNCs from NPM1wt (n=30) or NPM1c+ (n=40) AML patients. Data is presented as dot plot with Median, P-value was calculated by Mann Whitney nonparametric test. (B) TCGA database (GSE62944) was used to retrieve RNA expression levels of HOXBLINC in BM cells from AML patients without NPM1 mutations and MLL rearrangements (NPM1c-/MLLr-, n=128) or AML patients with MLL rearrangements (MLLr+, n=11), or AML patients with NPM1 mutations (NPM1c+, n=42). Violin plots show mean, interquartile, and 1.5x interquartile. The width shows the probability density. (C) Volcano plot of RNA-seq analysis of differentially expressed genes in Npm1c+ knock-in (Npm1c/+) vs. WT LSK cells. (D) Heat map of RNA-seq analysis shows the up- and down-regulated genes in HOXBLINC-KRAB vs. WT OCI-AML3 cells. Arrows: dysregulated genes implicated in leukemogenesis. (E) The HOXBLINC-KRAB affected genes in OCI-AML3 cells were analyzed and annotated by the Gene Ontology (GO) analysis. (F) Enrichment of HOXBLINC-KRAB dysregulated genes in NPM1-mutated (top) and HOXA9 (bottom) oncogenic pathways by Gene Set Enrichment Analysis (GSEA). (G) Proliferation curves of WT, shScramble-expressing, and two different shHOXBLINC-expressing OCI-AML3 cells upon doxycycline (Dox) treatment for 5 days. Data are presented as mean ± SD of total viable cells. **P<0.01. (H) Kaplan-Meier survival curve of inducible shHOXBLINC OCI-AML3 cell transplanted mice treated with or without Dox (P<0.01, 5 mice/group). (I) Kaplan-Meier survival curves of mice transplanted with control or HOXBLINC-KRAB BM cells from three AML patients with different gene mutations, 1315# (NPM1c+;FLT3wt, left), 921# (NPM1c+; FLT3mu, middle), LPP4 (NPM1wt;MLLr+, right), (4 mice/group). The Log-rank test was used to analyze differences between the survival curves.

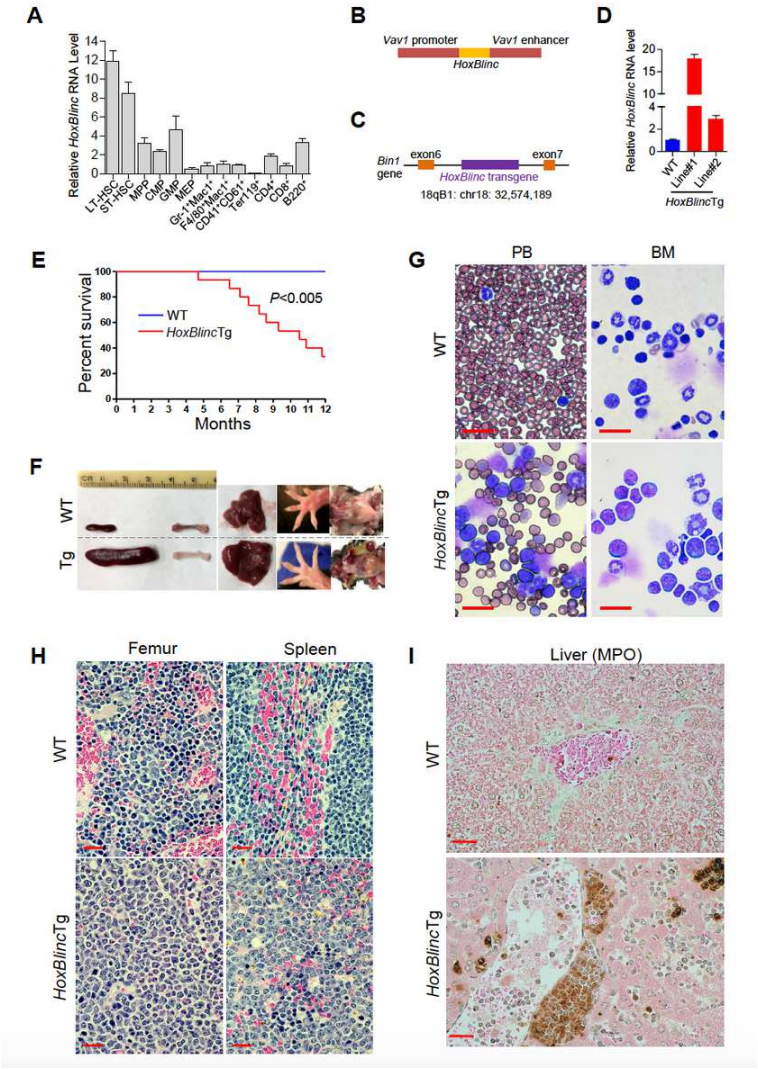

Figure 2

Transgenic overexpression of HoxBlinc in mice lead to lethal AML. (A) RT-qPCR analyses of HoxBlinc expression levels on FACS-sorted BM cell populations along the hematopoiesis hierarchical differentiation tree of WT mice. (B) Diagram of the Vav1 promoter/enhancer driven HoxBlinc transgene strategy. (C) PCR based transgenic integrating location identification (TAIL) assay maps the Vav1-HoxBlinc transgene integration site to the mouse chromosome 18 in Line #1 HoxBlincTg mice. (D) RT-qPCR analysis of the HoxBlinc RNA expression in BM cells of WT and 2 lines of HoxBlincTg mice. (E) Kaplan-Meier survival curve of WT (n=15) and HoxBlincTg (Line #1, n=15) mice up to 1 year of age. (F) Gross appearance of spleens, femur, livers, feet, and lymph nodes of representative WT and moribund HoxBlincTg mice. (G) May-Giemsa stained PB smears and BM cytospins prepared from representative WT and moribund HoxBlincTg mice. (H) H&E stained femur and spleen sections of representative WT and moribund HoxBlincTg mice. (I) MPO stained liver sections from representative WT and moribund HoxBlincTg mice. Scale bar, 20μm.

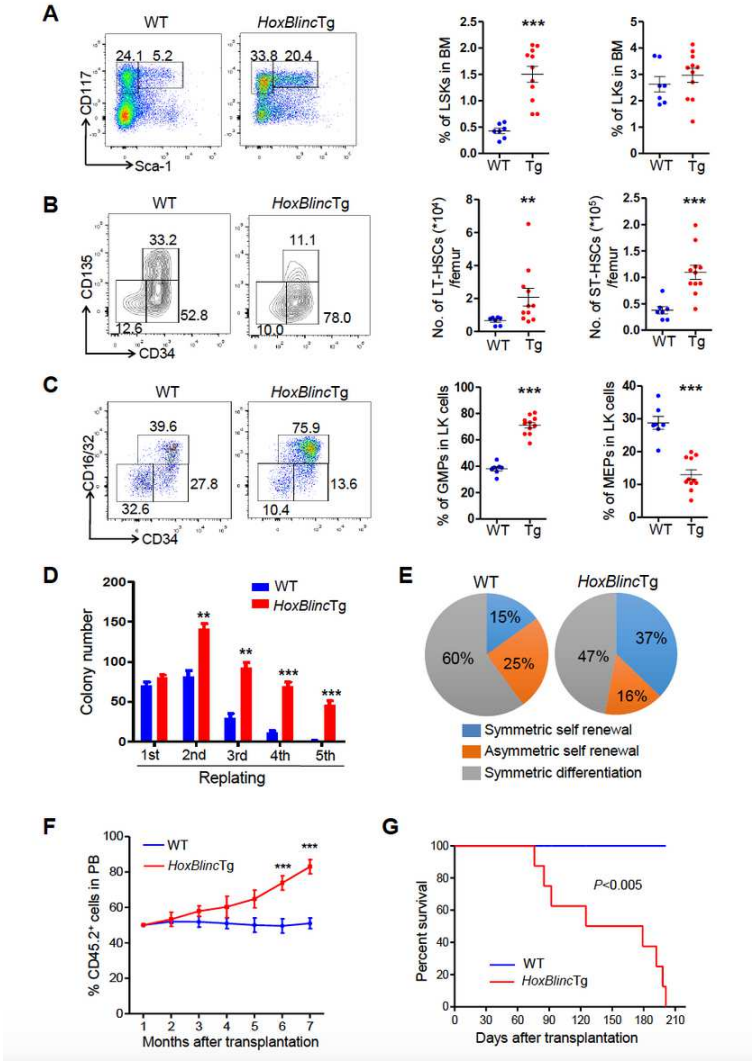

Figure 3

Transgenic expression of HoxBlinc enhances HSC self-renewal and expands myelopoiesis. (A) FACS analysis of LSK (Lin-Sca-1+c-Kit+) and LK (Lin-Sca-1-c-Kit+) cell populations in the BM Lin- cells of representative young WT and HoxBlincTg mice (left). Quantitation of the percent LSK and LK cells in the total BM cells of each genotype of mice is shown (right). (B) FACS analysis of LT-HSC, ST-HSC and MPP cell populations in the BM LSK cells of representative young WT and HoxBlincTg mice (left). The total number of LT- and ST-HSCs per femur in WT and HoxBlincTg mice are shown (right). (C) FACS analysis of GMP, MEP and CMP populations within BM LK cells of representative young WT and HoxBlincTg mice (left). Quantitation of the percent GMP and MEP cell populations in the BM LK cells of each genotype of mice is shown (right). (WT mice, n=7; HoxBlincTg mice, n=11; 8-10 weeks). (D) The number of colonies per 100 WT or HoxBlincTg BM LSK cells are shown (1st). Colonies were replated every 7 days for 4 times (2nd-5th). (E) Paired-daughter cell assays were performed on CD34-LSK cells clone-sorted from BM cells of WT and HoxBlincTg mice, and each cell was analyzed for symmetric self-renewal (blue), asymmetric cell division (orange) or symmetric differentiation (gray). (F) Kinetic flow cytometric analyses of CD45.2 chimerism in the PB of recipients transplanted with WT or HoxBlincTg BM cells. (G) Kaplan-Meier survival curve of recipient mice (8 mice/genotype) receiving WT or HoxBlincTg BM cells. Data in (A-F) are presented as mean ± SEM; **P<0.01 and ***P<0.001 by two-tailed unpaired Student's t-test.

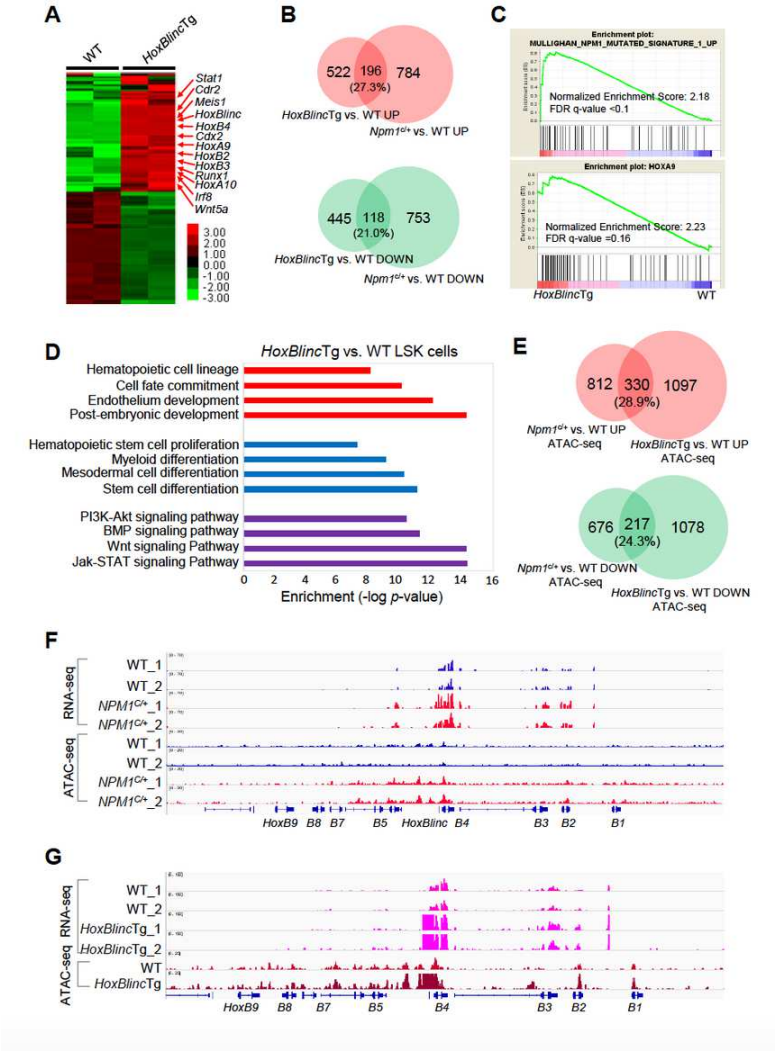

Figure 4

Overexpression of HoxBlinc activates NPM1c+ signature genes via enhancing promoter chromatin accessibility in LSK cells. (A) Heatmap of RNA-seq analysis shows the up- and down-regulated genes in HoxBlincTg vs. WT LSK cells. Red arrows: up-regulated genes implicated in HSPC regulation/leukemogenesis. (B) Overlap of up- (top) or down- (bottom) regulated genes between HoxBlincTg vs. WT and NPM1c/+ vs. WT LSK cells. (C) Enrichment of upregulated genes involved in NPM1-mutated signature (top) and HOXA9 (Bottom) oncogenic pathway upon overexpression of HoxBlinc in LSK cells by GSEA. (D) The HoxBlinc overexpression dysregulated genes in LSK cells were analyzed and annotated by GO analysis. (E) Overlap of global gain (top) or loss (bottom) of promoter chromatin accessibilities between ATAC-seq data of HoxBlincTg vs. WT and NPM1c/+ vs. WT LSK cells. (F) RNA-seq (chromatin accessibility, top 4 tracks) and ATAC-seq (gene expression, bottom 4 tracks) analysis of WT and Npm1c/+ LSK cells in the HoxB gene locus. (G) RNA-seq analysis (top 4 tracks) and ATAC-seq analysis (bottom 2 tracks) of WT and HoxBlincTg LSK cells in the HoxB gene locus.

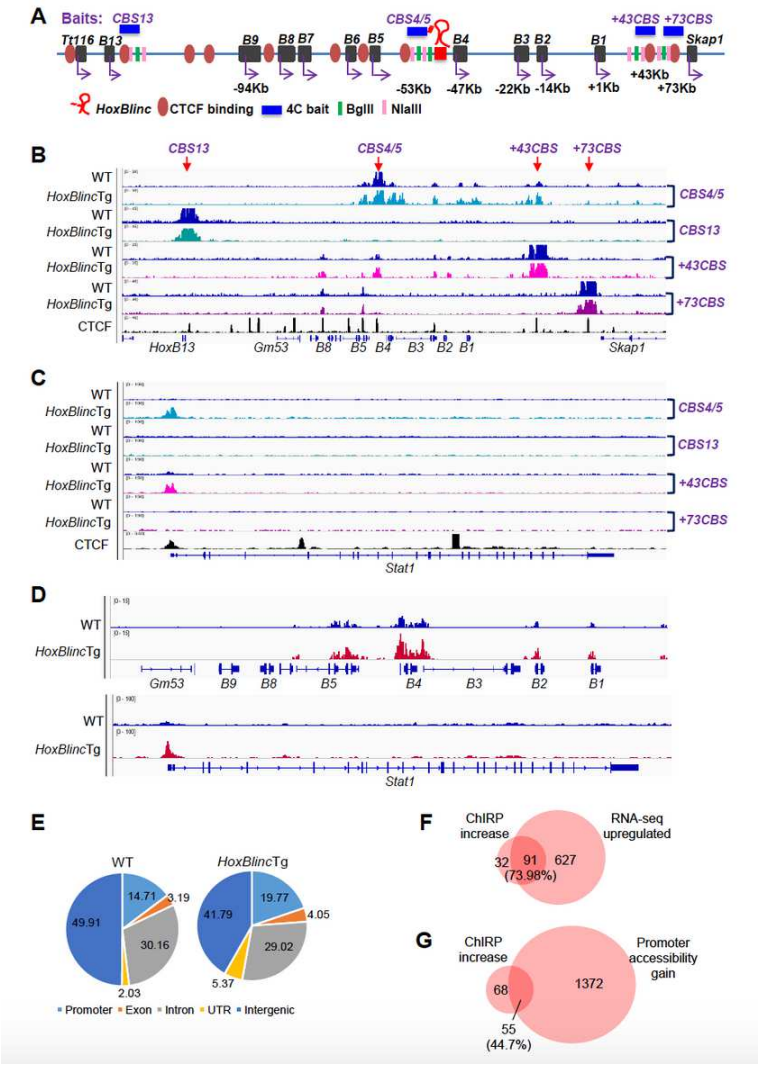

Figure 5

HoxBlinc directly binds to target genes and mediates chromatin interactions to drive gene regulatory networks in HSPCs. (A) Schematic diagram showing the CTCF binding sites (CBS), location of the 4C baits in HoxB locus of the mouse genome. (B, C) Long-range chromatin interactions with HoxB locus (4 baits) as determined by 4C-seq analysis in WT and HoxBlincTg Lin-c-Kit+ cells. HoxBlinc overexpression increased the interactions of CBS4/5 and +43CBS with the HoxB genes (B), HoxBlinc overexpression also enhanced the interaction of CBS4/5 or +43CBS with Stat1 promoter region (C), ChIP-seq analysis of CTCF binding sites was obtained from the NCBI GEO public database (GSM918748). (D) ChIRP-seq analysis of HoxBlinc bindings to the HoxB and Stat1 gene loci in WT and HoxBlincTg Lin-c-Kit+ cells. (E) The pie chart shows the distribution of promoter, exon, intron, UTR and intergenic region within the total HoxBlinc binding sites in WT (left) and HoxBlincTg (right) Lin-c-Kit+ cells as identified by ChIRP-seq. (F) Overlap of genes with HoxBlinc binding gain to their promoter regions as identified by ChIRP-seq and the upregulated genes as identified by RNA-seq in HoxBlincTg vs. WT HSPCs. (G) Overlap of genes with HoxBlinc binding gain to their promoter regions as identified by ChIRP-seq and genes with promoter accessibility gain as identified by ATAC-seq in HoxBlincTg vs. WT HSPCs.

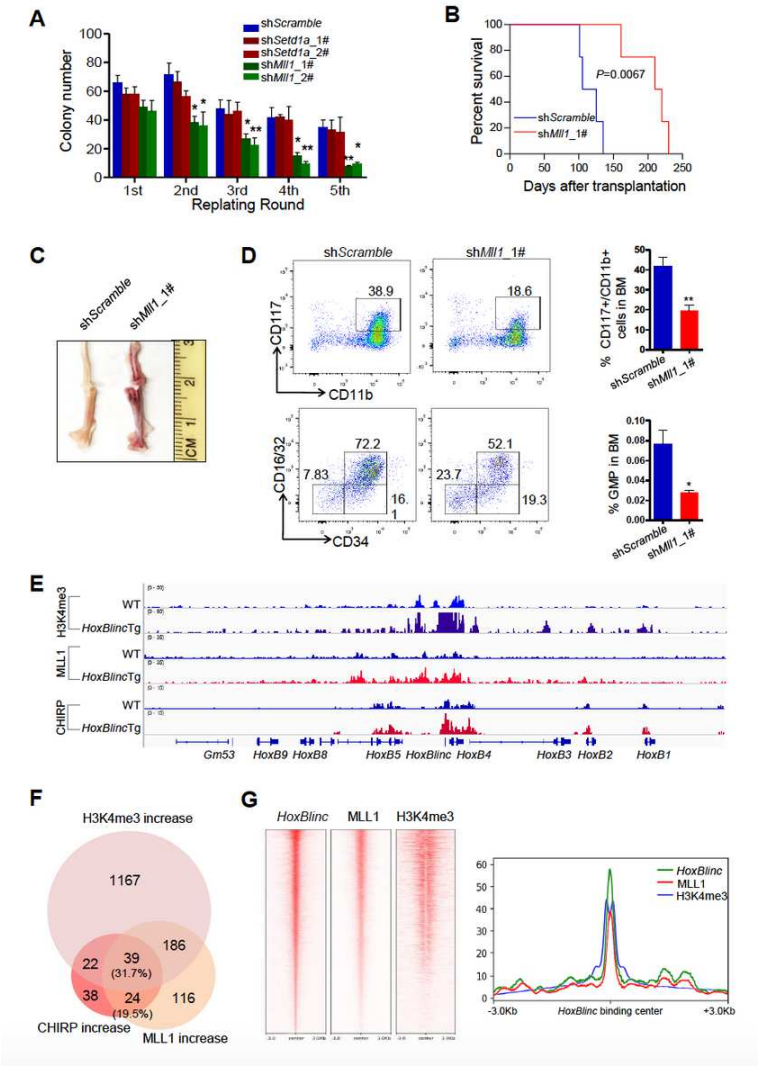

Figure 6

Recruitment of MLL1 is critical for HoxBlinc overexpression mediated target gene expression and abnormal HSPC function. (A) Number of colonies per 100 sorted GFP+ HoxBlincTg BM LSK cells transduced with ShScramble, shSetd1a or shMll1 lentivirus are shown (1st). Colonies were replated every 7 days for 4 times (2nd-5th). (B) Kaplan-Meier survival curve of recipient mice receiving ShScramble or shMll1 LSK cells (P=0.0067, 4 mice/group). (C) Gross appearance of femur dissected from mice transplanted with shScramble or shMll1 LSK cells. (D) FACS analysis of BM CD11b+/CD117+ (top) and GMP (bottom) cell populations from recipient mice (4 mice/condition) transplanted with shScramble or shMll1 LSK cells. Data in (A, D) are presented as mean ± SEM; *P<0.05 and **P<0.01 by two-tailed unpaired Student's t-test. (E) ChIP-seq analysis of H3K4me3 (top 2 tracks) and MLL1 (middle 2 tracks), as well as ChIRP-seq analysis of HoxBlinc (bottom 2 tracks) in the HoxB loci of WT and HoxBlincTg Lin-c-Kit+ cells. (F) The overlap among genes with gain of HoxBlinc, H3K4me3 and MLL1 bindings to their promoter regions based on ChIRP-seq and ChIP-seq analyses. (G) Heatmap (right) and peaks density (left) to show the overlap of HoxBlinc ChIRP-seq, MLL1 and H3K4me3 ChIP-seq.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tableS120200522Patientinformationlist.pdf
- tableS220200522GenelistofcommonregulatedgenesbytransgenicHoxBlincandNpm1cKIbasedonRNAseq.pdf
- tableS320200522ListofcommongeneswithgainorlossofpromoteraccessibilitiescausedbytransgenicHoxBlincandNpm1cKIbasedonATACseqdata.pdf
- tableS420200522ListofHoxBlincChIRPtranscriptionmotifanalysis.pdf
- tableS520200522ListofcommongenesofHoxBlincChIRPseqH3K4me3ChIPseqandMLL1ChIPseq.pdf
- tableS620200522SequencesofprimersandsgRNAs.pdf
- tableS720200522experimenalresources.pdf
- tableS820200522DetailedinformationofAMLcelllines.pdf

- HoxblincSupplfigures.pdf