Long Non-Coding RNAs Control Hematopoietic Stem Cell Function

Authors
Min Luo, Mira Jeong, ..., Wei Li, Margaret A. Goodell

Correspondence
wl1@bcm.edu (W.L.),
goodell@bcm.edu (M.A.G.)

In Brief
Luo et al. perform deep RNA sequencing of hematopoietic stem cells (HSCs), identifying previously unannotated long non-coding RNAs (lncRNAs) enriched in this cell population. Functional characterization of several HSC-enriched lncRNAs demonstrated roles in regulating HSC differentiation and self-renewal and revealed genomic binding to hematopoietic transcription factor binding sites.

Highlights
- We identified 159 unannotated lncRNAs enriched in mouse HSCs (LncHSCs)
- LncHSC expression is epigenetically regulated and altered with HSC functional decline
- LncHSC-1 and LncHSC-2 knockdown impacts HSC self-renewal and differentiation
- Genomic LncHSC-2 binding sites are enriched for TFs and repeats

Accession Numbers
GSE63276
GSE53928
GSE47817
GSE50775

CrossMark
April 2, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.stem.2015.02.002
Long Non-Coding RNAs Control Hematopoietic Stem Cell Function

Min Luo,1,4 Mira Jeong,1,4 Deqiang Sun,5,4 Hyun Jung Park,2,4 Benjamin A.T. Rodriguez,2,4 Zheng Xia,2 Liubin Yang,1 Xiaotian Zhang,1 Kuanwei Sheng,1 Gretchen J. Darlington,3 Wei Li,2,* and Margaret A. Goodell1,2,*

1Stem Cells and Regenerative Medicine Center, Department of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
2Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA
3Huffington Center on Aging, Baylor College of Medicine, Houston, TX 77030, USA
4Co-first author
*Correspondence: wl1@bcm.edu (W.L.), goodell@bcm.edu (M.A.G.)
http://dx.doi.org/10.1016/j.stem.2015.02.002

SUMMARY

Hematopoietic stem cells (HSCs) possess unique gene expression programs that enforce their identity and regulate lineage commitment. Long non-coding RNAs (lncRNAs) have emerged as important regulators of gene expression and cell fate decisions, although their functions in HSCs are unclear. Here we profiled the transcriptome of purified HSCs by deep sequencing and identified 323 unannotated lncRNAs. Comparing their expression in differentiated lineages revealed 159 lncRNAs enriched in HSCs, some of which are likely HSC specific (LncHSCs). These lncRNA genes share epigenetic features with protein-coding genes, including regulated expression via DNA methylation, and knocking down two LncHSCs revealed distinct effects on HSC self-renewal and lineage commitment. We mapped the genomic binding sites of one of these candidates and found enrichment for key hematopoietic transcription factor binding sites, especially E2A. Together, these results demonstrate that lncRNAs play important roles in regulating HSCs, providing an additional layer to the genetic circuitry controlling HSC function.

INTRODUCTION

Hematopoietic stem cells (HSCs) continuously regenerate all blood and immune cell types throughout life and are also capable of self-renewal. Protein-coding genes specifically expressed in HSCs (HSC “fingerprint” genes (Chambers et al., 2007) have been identified by microarray studies, and many have been shown to be functionally critical for HSC function (reviewed in Rossi et al., 2012). Similarly, microRNAs can regulate HSC function (Lechman et al., 2012; O’Connell et al., 2008, 2010).

Recent whole transcriptome sequencing has revealed a large number of putative long non-coding RNAs (lncRNAs). The function of some lncRNAs has been established in a limited scope of biological processes, such as cell cycle regulation, embryonic stem cell (ESC) pluripotency, lineage differentiation, and cancer progression (Guttman et al., 2011; Hung et al., 2011; Klattenhoff et al., 2013; Prensner et al., 2011). In the hematopoietic system, only a few IncRNAs have been identified to be involved in differentiation or quiescence. Xist-deficient HSCs exhibit aberrant maturation and age-dependent loss (Yildirim et al., 2013), and maternal deletion of the H19 regulatory elements reduced HSC quiescence (Venkatraman et al., 2013). In addition, LncRNA erythroid prosurvival (lincRNA-EPS) has been found to promote terminal differentiation of mature erythrocytes by inhibiting apoptosis (Hu et al., 2011), whereas HOTAIRM1 and eosinophil granule ontogeny (EGO) are involved in granulocyte differentiation (Wagner et al., 2007; Zhang et al., 2009). Furthermore, recent genomic profiling identified thousands of lncRNAs expressed in erythroid cells. Some of them have been shown to play a role in erythroid maturation and erythro-megakaryocyte development (Alvarez-Dominguez et al., 2014; Paralkar et al., 2014).

Nevertheless, lncRNAs function in HSCs still remains largely unknown. Considering that lncRNAs usually exhibit cell type- or stage-specific expression and HSCs are rare (~0.01% of bone marrow), we reasoned that many HSC-specific lncRNAs may not have been identified and annotated yet. Notably, Cabezaz-Wallscheid et al. (2014) recently identified hundreds of IncRNAs expressed in HSCs and compared their expression with that in lineage-primed progenitors. However, without expression validation, comparison of expression in differentiated lineages, and functional studies, their specificity and regulatory role remains unclear. Therefore we aimed here to identify the full complement of lncRNAs expressed in HSC with extremely deep RNA sequencing to determine IncRNAs specific to HSCs relative to representative differentiated lineages and also to perform an initial analysis of their relevance to HSC function.

RESULTS

Identification of HSC-Specific IncRNAs

To identify unannotated putative lncRNAs, we purified the most primitive long-term HSCs (SP-KSL-CD150+, hereafter termed HSCs) from mouse bone marrow by fluorescence-activated cell sorting (FACS). To uncover IncRNAs expressed in HSCs across different ages, we performed RNA sequencing (RNA-seq) on HSCs from 4-month-old (m04), 12-month-old (m12), and 24-month-old (m24) mice (Sun et al., 2014), generating
368, 311, and 293 million mapped reads for m04, m12, and m24 HSCs, respectively. To achieve the greatest power to detect unannotated transcripts, we also included RNA-seq data from Dnmt3a knockout (KO) HSCs (Jeong et al., 2014) to reach a total of 1,389 million mapped reads for the HSC transcriptome. Although Dnmt3a KO HSCs differentiate inefficiently, they retain many features of normal self-renewing HSCs, adding power to novel gene discovery. In addition, we performed RNA-seq on sorted bone marrow B cells (B220+) and granulocytes (Gr1+) for comparison. We then performed a stringent series of filtering steps to identify lncRNAs in different ages of wild-type (WT) HSCs, including a minimum length of 200 bases and multiple exons (Figure 1A).

We first verified the high quality of our data by confirming the lineage-specific expression of known protein-coding “fingerprint” genes (Chambers et al., 2007), such as Myct1, Ebf1, and Cldn1, with HSC-, B cell-, and granulocyte-specific expression, respectively (Figure S1A). Next, we identified 2,614 transcripts annotated previously as non-coding RNAs by the University of California, Santa Cruz (UCSC) Known Gene, RefSeq, or Ensemble databases. Comparing their expressions in these three cell types revealed that 154, 57, and 81 lncRNAs were enriched to HSCs, B cells, and granulocytes (Table S1), such as AK018427, AK156636, and AK089406 (Figure S1B).

With known genes filtered out, we focused on the remaining unannotated and multiply spliced transcripts, which resulted in 503 unannotated genes in HSCs. Comparison of the expression of these transcripts in HSCs, B cells, and granulocytes revealed that almost one-third were HSC-specific (Figure 1B). Comparing their expression in 20 additional tissues (RNA-seq data from A CB HSC 4mo) revealed that 1,056 novel lncRNAs (LncHSCs) were enriched in other blood cell types (Table S1).

See also Figure S1 and Table S1.
Figure 2. Transcriptional Regulation of LncHSCs

(A) Heatmap depicting the expression of 159 LncHSCs: 4-month-old HSCs versus 12-month-old HSCs and 24-month-old HSCs (left) and WT versus Dnmt3a KO HSCs (right).

(legend continued on next page)
One example is located in a canyon with a methylation canyon (UMR >3.5 kb) undermethylated regions (UMRs) in their promoter regions, and ure S1G) contribute to about 35% of their genomic sequences. For DNA methylation, whole-genome DNA methylation studies (Sun et al., 2014; Jeong et al., 2014). For LncHSCs, (H3K36me3), in purified WT HSCs were obtained from previous methylation (H3K4me3) and histone H3 Lys36 trimethylation for two activation-associated histone marks, histone H3 Lys4 tri-

We next examined the features of the set of HSC-enriched IncRNAs (named LncHSCs). The LncHSCs generally have fewer exons and lower expression but similar transcript lengths and conservation (PhastCon) scores compared with protein-coding genes (Figures S1C–S1F), in line with previous reports (Derrien et al., 2012). Because retrovirus-related transposon elements (TEs) have been shown to be enriched in ESC IncRNAs (Kelley and Rinn, 2012), we considered whether this was also applicable to LncHSCs. Consistent with a previous report (Kapusta et al., 2013), we found that TEs cover about 40% of the genomic sequence, 15% of known IncRNAs, but 5% of protein coding genes. For LncHSCs, TEs (particularly LTR-associated; Figure S1G) contribute to about 35% of their genomic sequences. These results suggest that LncHSCs here are distinct from protein-coding genes and likely to act as IncRNAs.

We further examined the chromatin state of the LncHSCs. Chromatin immunoprecipitation sequencing (ChIP-seq) data for two activation-associated histone marks, histone H3 Lys4 trimethylation (H3K4me3) and histone H3 Lys36 trimethylation (H3K36me3), in purified WT HSCs were obtained from previous studies (Sun et al., 2014; Jeong et al., 2014). For LncHSCs, H3K4me3 was typically located at their predicted transcriptional start site (TSS) and H3K36me3 along their gene bodies (Figure 1E). For DNA methylation, whole-genome DNA methylation analysis in HSCs showed that 62% of LncHSCs are marked by undermethylated regions (UMRs) in their promoter regions, and some are even located in a methylation canyon (UMR >3.5 kb) (Jeong et al., 2014). One example is located in a canyon with a broad H3K4me3 peak (Figure 1F). Another example is a LncHSC whose transcription originates from the promoter of an active protein coding gene, but in the opposite direction (Figure S1H), consistent with behavior previously noted as common among IncRNAs (Sigova et al., 2013).

**LncHSCs Showed Altered Expression with HSC Functional Decline**

To gain insights into the function of LncHSCs in HSC self-renewal and differentiation, we compared their expression between different HSC ages and between WT and Dnmt3a KO HSCs because our previous studies showed that aged HSCs exhibited a repopulation defect and myeloid-biased differentiation, and Dnmt3a KO HSCs exhibit defective differentiation and enhanced self-renewal (Challen et al., 2012). There is a small subset of LncHSCs (29 of 159) whose expression changes between 4- and 24-month-old HSCs (false discovery rate [FDR] < 1E-04). However, almost 58% of LncHSCs (92 of 159) showed significant expression changes (FDR < 1E-04) between age-matched m12 WT HSCs and Dnmt3a KO HSCs (Figure 2A). To examine the basis for these differences at the epigenetic level, whole-genome bisulfate sequencing of WT and Dnmt3a KO HSCs (Jeong et al., 2014) was analyzed. Interestingly, a majority of LncHSCs showed strong loss of DNA methylation at their TSS regions after Dnmt3a KO, suggesting a role for Dnmt3a in their regulation. For example, LncHSC-1 to LncHSC-6 all show decreased methylation at their TSS regions. However, decreased methylation was not always correlated with expression changes (Figure S2A). LncHSC-1 and LncHSC-2 exhibit decreased expression, LncHSC-3 and LncHSC-4 show upregulation and are accompanied by increased H3K4me3 (Figure 2B), and LncHSC-5 and LncHSC-6 do not show significant expression changes (Figure S2B). This relatively poor correlation between DNA methylation changes and gene expression alteration resembles the observations from protein-coding genes (Challen et al., 2012).

We also examined the promoters (TSS ± 5 kb) of LncHSCs for critical HSC-associated transcription factor (TF) binding. Using published ChIP-seq data for ten key HSC TFs, including Erg, Flil1, Lmo2, Meis1, Gata2, Runx1, Pu.1, Scl, Lyl1, and Gata2, across a variety of blood lineages (more than ten) (Hannah et al., 2011), we found that 51% of LncHSCs contain at least one or more TF binding sites on their promoters (e.g., LncHSC-1 and LncHSC-2; Figure S2C). Among these ten TFs, Erg, Flil1, and Pu.1 are the top three factors, exhibiting binding sites near 38%, 29%, and 25% of LncHSCs, respectively. This percentage is comparable with protein-coding genes but much higher than the genome random control, suggesting that the expression of LncHSCs may be precisely regulated by hematopoietic TFs (Figure S2D).
Next, we selected LncHSC-1 through LncHSC-6 for expression validation by quantitative RT-PCR. Among them, two overlapped with unannotated expressed sequence tags (ESTs): LncHSC-1 (AK039852) and LncHSC-2 (D032622). Notably, LncHSC-2 is represented as an EST on Affymetrix microarrays (MOE430 V2.0) and has been identified previously as an HSC fingerprint gene (Chambers et al., 2007; Figure S2E). qRT-PCR confirmed that LncHSC-1 through LncHSC-6 are highly expressed in stem and progenitor cell populations (c-Kit+Sca-1+Lin [KSL]) but not, or at very low levels, in six other terminally differentiated blood lineages (CD4, CD8, B220, Gr1, Mac1, and Ter119; Figure 2C). For LncHSC-4, we detected a higher expression in erythroid cells (Ter119) than other blood lineages. Indeed, LncHSC-4 (aka Lincred1; Tallack et al., 2012) has been shown to be regulated by Klf1, Gata1, and Tal1 and, possibly, to be involved in erythroid differentiation.

To gain insights into how LncHSCs control HSC function, we focused on LncHSC-1 and LncHSC-2, which exhibit several features of interest. These two transcripts are highly expressed in WT HSC (FPKM > 3) but suppressed in Dnmt3a KO HSCs (Figure S2B), and their promoter regions are bound by multiple TFs in hematopoietic progenitor cells (Figure S2C). Further expression analysis in HSC and different progenitors revealed that LncHSC-1 is more HSC-specific, and LncHSC-2 is expressed in HSCs and also different progenitors but not terminally differentiated cells (Figure S2F). Moreover, we found that LncHSC-1 and LncHSC-2 are transcribed from enhancer regions, marked by histone H3 Lys4 acetylation (H3K27ac) or histone H3 Lys4 monomethylation (H3K4me1) but not H3K4me3 and H3K27me3 (Figure 2D). LncHSC-1 is located close to two functionally important coding genes, Zfp36l2 and Thada. Genomic translocation has been reported within Thada and distal to the Zfp36l2 locus in various myeloid malignancies (Trubia et al., 2009). In addition, heterozygous mutations of Zfp36l2 have been detected in leukemias (Iwanaga et al., 2011). Zfp36l2 homozygous knockout mice die from HSC failure within 2 weeks of birth (Stumbo et al., 2009) and, recently, it has been reported that Zfp36l2 is required for self-renewal of erythroid progenitors (Zhang et al., 2013). The human synteny block, including LncHSC-1, is also located close to THADA and ZFP36L2 on chromosome 2. RNA-seq data from bone marrow of The Cancer Genome Atlas (TCGA) patients (Ley et al., 2010) indicated that there are several unannotated transcripts expressed in this region. However, based on sequence homology, we could not identify specific orthologs to LncHSC-1, consistent with generally poor conservation for lncRNAs across species (Ultisky et al., 2011). In mouse HSCs, there are four LncHSCs (LncHSC-1, LncHSC-3, LncHSC-82, and LncHSC-13) close to the Thada and Zfp36l2 genes, all of which showed expression changes after Dnmt3a KO, with LncHSC-1 and LncHSC-13 downregulated and LncHSC-3 and LncHSC-82 upregulated (Figure S3A). To examine how the transcripts in this region were impacted by human DNMT3A mutations, we reconstructed and selected, from TCGA acute myeloid leukemia (AML) patient data, one abundantly expressed transcript at this region (corresponding to the EST tag AF150238; Figure S3B), and found that patients with a DNMT3A mutation showed increased expression (p value = 0.01; Figure S3C). These data suggest a similar regulation of putative lncRNAs by DNA methylation in this syntenic region despite the lack of clear sequence homology. For LncHSC-2, it is close to a protein-coding gene, Pkn2, and sequence comparison by nucleotide-nucleotide BLAST (BLASTN) revealed that it is highly homologous (87.6%) to a 3-kb region in its human syntenic block, which is also close to the PKN2 gene. However, we did not detect expressed transcripts at this region in TCGA patients.

LncRNA function is also dependent on subcellular localization. Enhancer-associated lncRNAs are more enriched in the nucleus, whereas lncRNAs involved in other functions such as posttranscriptional and translational processes tend to be more cytoplasmic. We therefore performed RNA fluorescence in situ hybridization (FISH) to determine the localization of LncHSCs. LncHSC-1 is mainly located in the HSC nucleus compared with the control 18S rRNA (Figure 2E). In parallel, LncHSC-2 is also located in the HSC nucleus, suggesting that LncHSC-1 and LncHSC-2 are likely functional noncoding RNAs. To confirm their specificity, we also examined one granulocyte-enriched LncRNA (LncGr-1), which was found to be exclusively expressed in granulocytes but not in HSCs (Figure 2F).

LncHSCs Control HSC In Vitro and In Vivo Differentiation

To characterize the functions of LncHSC-1/2, we generated retrovirally expressed constructs to knock down their expression (Figure 3A). In stem and progenitor cells (Sca-1+), the knockdown (KD) constructs led to 50%–70% reduction of their expression by RT-PCR (Figure 3B). To examine their effects on HSC self-renewal and differentiation, retrovirally transduced KSL-GFP+ cells were sorted after 2 days of in vitro culture and plated in methylcellulose for colony-forming unit (CFU) assays. Knockdown of those transcripts had no effect on colony number or lineage specificity after the first plating (Figure S3D). However, after the second plating, KD of LncHSC-1 significantly increased the colony numbers compared with the control, suggesting that progenitors with reduced LncHSC-1 in the first plating had not undergone terminal differentiation (Figure 3C). Indeed, KD of LncHSC-1 led to an increase in cells expressing the HSC/progenitor marker c-Kit (Figure S3D), and cells from second-plating colonies had a more homogeneous morphology (Figure 3D).

Next we performed transplantation to examine the function of LncHSC1/2 in vivo. Because we observed that KD of LncHSC-1 increased the myeloid colony number in vitro, we also generated retroviral constructs to overexpress LncHSC-1 in stem/progenitor cells. However, after transplantation for 16 weeks, even though the LncHSC-1 transcript level increased almost 500-fold in the GFP+ (Lin- c-Kit+ Sca-1+) KSL cell population, there was no difference in lineage differentiation (Figures S3E and S3F). Meanwhile, we transplanted stem/progenitor cells transduced with the LncHSC-1/2 KD constructs. 16 weeks after transplantation, the percentages of donor-derived cells (CD45.2+) in the peripheral blood (PB) were similar between the groups. However, although the initial transduction efficiency (Figure S4A) and donor engraftment efficiency (Figure S4B) are similar, the percentage of the GFP+ population varied significantly between different groups (Figure 4A), possibly because of the effects of LncHSC on HSC self-renewal. To determine their impact on lineage differentiation, we compared the percentage of different lineages within the GFP+ population. We found that...
KD of LncHSC-1 significantly increased myeloid differentiation at the expense of B cells compared with control KD, in alignment with the in vitro findings. In contrast, KD of LncHSC-2 significantly increased T cell lineage and decreased B cell output. As a control, the CD45.2+GFP– population showed similar lineage distributions between different groups (Figure 4B). To confirm the KD efficiency in vivo, we isolated bone marrow (BM) GFP+KSL cells 20 weeks after transplantation for RT-PCR and confirmed that LncHSCs were knocked down (Figure S4C). We further performed lineage, progenitor, and HSC analyses in the bone marrow after 20 weeks. Notably, the bone marrow GFP+ population in the LncHSC-1 KD-1 and LncHSC-2 KD-2 groups was too low for detailed HSC and progenitor analyses, so we focused on the LncHSC-1 KD-2 (short hairpin RNA [shRNA] #2) and LncHSC-2 KD-1 (shRNA#1) groups. We found that there were no significant differences for the granulocyte-macrophage progenitor (GMP), common myeloid progenitor (CMP), and megakaryocyte-erythroid progenitor (MEP) population or the

Figure 3. LncHSCs Regulate HSC Differentiation In Vitro
(A) Flow chart depicting knockdown of LncHSC for in vitro and in vivo functional studies. 5-FU, 5-fluorouracil.
(B) Quantitative RT-PCR showing LncHSCs knockdown. Sca-1+ cells were transduced with knockdown constructs and cultured in vitro for 2 days, and then 20,000 GFP+ cells were sorted for RT-PCR (n = 3, mean ± SD).
(C) Methylcellulose CFU assay using 200 KSL-GFP+ cells (transduced by LncHSC-1 KD-1 and LncHSC-2 KD-1). Sorted cells were put into one well of 6-well plate containing MethoCult3434, and the average colony numbers were counted after 14 days. For the second plating, 2,000 live cells from the colonies obtained in the first plating were plated as before and cultured for 14 days. **p < 0.01. (n = 3, mean ± SD). Data are representative of three experiments.
(D) Morphology of cells from the colonies at the second plating by cytospin.
See also Figure S3.
Figure 4. LncHSCs Control HSC Function In Vivo

(A) Contribution of retrovirally transduced donor HSCs (CD45.2+GFP+) to recipient mouse PB after primary transplantation. *p < 0.05, **p < 0.01. Error bars represent mean ± SEM (n = 10 for control KD and 5–8 for LncHSCs KD).

(B) Analysis of HSC differentiation in peripheral blood at 16 weeks post-primary transplant. The percentage of the indicated lineages within CD45.2+GFP+/C0 or CD45.2+GFP+ cell compartments are shown. Myeloid cells (Mye) were defined as Gr1+ and Mac1+, B cells (B) are B220+, and T cells (T) are CD4+ and CD8+. **p < 0.01, ***p < 0.001. Error bars represent mean ± SEM (n = 10 for control KD and 5–8 for LncHSCs KD).

(C) Contribution of donor HSCs (CD45.2+GFP+ or CD45.2+GFP+) to recipient mouse PB after secondary transplantation. Data are mean ± SEM (n = 5–6). For secondary transplantation, 500 CD45.2+GFP+ KSL cells from primary recipients were re-sorted 20 weeks after transplantation, mixed with 250,000 CD45.1 WBM cells, and injected into new lethally irradiated CD45.1 recipients.

(legend continued on next page)
long-term (LT)-HSC, short-term HSC, and multipotent progenitor populations after KD (Figure S4D). However, we observed that KD of LncHSC-1 led to increased myeloid cells and lineage-negative c-kit-positive (LK) cells (myeloid progenitors) and that KD of LncHSC-2 led to more T cells, consistent with the peripheral analysis (Figure S4E).

To examine HSC self-renewal activity, 500 BM GFP+ KSL cells from primary recipients of LncHSC-1 (KD-2) and LncHSC-2 (KD-1) were sorted and transplanted into secondary recipients. Peripheral blood (PB) analysis showed that GFP+ levels were comparable between different groups (Figure 4C). KD of LncHSC-1 increased myeloid differentiation, and KD of LncHSC-2 increased T cell differentiation, consistent with primary transplantation results (Figure 4D). For the bone marrow analysis 16 weeks after secondary transplantation, interestingly, the percentage of the side population (SP) and KSL cells was decreased for LncHSC-2 KD, suggesting that LncHSC-2 is involved in HSC long-term self-renewal (Figure 4E).

Although the bone marrow of primary recipients of LncHSC-2 KD-2-transduced cells had too few GFP+ cells for a detailed analysis, we were able to isolate enough KSL GFP+ cells from a pool of these mice to perform secondary transplantation to verify the effect on self-renewal. Again, we observed that the percentage of GFP+ cells in the peripheral blood was very low at 4 weeks and almost undetectable at 16 weeks (Figure S4F). This precluded us from performing a lineage analysis with this KD construct. In the bone marrow at 16 weeks, we observed that percentage of the side population (SP) and KSL cells was decreased for LncHSC-2 KD, suggesting that LncHSC-2 is involved in HSC long-term self-renewal (Figure S4E).

To better understand the functions of LncHSCs, we sought to determine their binding sites by chromatin isolation by RNA purification sequencing (ChIRP-seq) (Chu et al., 2011; Engreitz et al., 2013; Simon et al., 2011). Given the technical challenges because of the limited number of primary HSCs, we utilized HPC5 cells, a mouse bone marrow-derived multipotent progenitor line (Pinto do O et al., 2002) that expresses LncHSC-2 at levels comparable with primary HSCs. We therefore performed ChIRP-seq to identify LncHSC-2 binding sites using HPC5 cells. After pull-down, RT-PCR showed that more than 90% of LncHSC-2 RNA was pulled down. For the negative control, less than 1% of GAPDH RNA was pulled down (Figure S5A). From ChIRP-seq, we identified 264 LncHSC-2 binding sites concordant in three of four biological replicates and absent in a LacZ negative control (for peak coordinates, see Table S3). Similar to transcription factors, LncHSC-2 binding sites were focal (median size, 284 base pairs [bp]), and most did not spread beyond 600 bp. The distribution of binding sites showed that ~11% were localized to promoter/5’ UTR elements (Figure 5A, left), representing a 3- to 7-fold enrichment over the genome background (Figure 5A, right). The remaining peaks occurred primarily in intronic and intergenic regions.

Next we asked whether LncHSC-2 accesses the genome through specific DNA sequences. A motif analysis of LncHSC-2 binding sites identified four core motifs (Table S4), suggesting that specific DNA motifs may be involved in LncHSC-2 occupancy. To further characterize the motifs, we quantified their similarity to known DNA sequence motifs. This revealed a significantly enriched bHLH motif corresponding to a transcription factor E2A isoform encoded by Tcf3 (Figure 5B, Table S4). E2A proteins act to promote the developmental progression of the entire spectrum of early hematopoietic progenitors, including LT-HSC, MPP, and common lymphoid progenitors (Semerad et al., 2009). To gain insights into potential LncHSC-2-mediated chromatin states, we tested the overrepresentation of its occupied sites (relative to the LacZ control) among the ChiP-seq profiles of hematopoietic transcriptional regulators and epigenetic marks in LT-HSC, multipotent progenitors (HPC-7 cells), as well as tissues (bone marrow, thymus, and spleen). LncHSC-2 sites were characterized by significant enrichment of undermethylated CpG regions (UMRs), the active histone marks H3K4me3/H3K27ac, and the TFs Erg/Fli1/Meis1/Pu.1 (Figure 5C; Table S5). Remarkably, a Genomic Regions Enrichment of Annotations Tool (GREAT) analysis of mouse genotype-phenotype associations showed that gene and promoter proximal binding sites were significantly enriched almost exclusively for hematopoietic and immune system phenotypes (14 of 16 terms with binomial test, q < 0.05), including abnormal lymphopoiesis (Table S5).

Having identified potential associations between LncHSC-2 and individual transcriptional regulators and epigenetic marks, we next analyzed occupancy patterns of regions bound by LncHSC-2 and enriched factors by hierarchical clustering. This analysis separated LncHSC-2-bound sites into two major clusters: undermethylated promoter proximal regions associated with activating chromatin marks (H3K4me3 and H3K27ac) and...
hematopoietic TFs and promoter-distal intergenic/intragenic regions associated with insulator CTCF, enhancers, or E2A binding motifs (Figure S5D). One LncHSC-2-occupied site containing an E2A motif mapped to the intronic region of the Pml (promyelocytic leukemia protein) gene locus (Figure S5E). As a tumor suppressor, Pml is essential for HSC maintenance, and its deficiency affects all hematopoietic lineages in recipient mice after BM transplantation (Tio et al., 2008). The core promoter of ltpkb is a site of potential co-occupancy by LncHSC-2 and the TFs Erg, Pu.1, Fli1, and Meis1 (Figure S5E). Mice lacking ltpkb, the B isoform of the Lns(1,4,5)P3 3-kinase, have a complete and specific T cell deficiency because of a developmental block at the double-positive thymocyte stage (Pouillon et al., 2003). Other LncHSC-2 co-occupied promoter regions include Cox5b, Itgb2, Tnf, and Sicc35c2 (Figure S5B).

Because a motif analysis showed that LncHSC-2 binding sites are highly enriched for E2A binding, we wondered whether LncHSC-2 is involved in recruiting E2A to its target sites. To address this question, we analyzed previous ChIP-seq data for E2A binding in HPC7 cells and found that there are almost 20 binding peaks overlapped between E2A and LncHSC-2 (Table S5). From them, we selected three sites with the highest scores of enrichment, which are close to the genes Nln, Sicc35c2, and Itgb2, respectively. Interestingly, ChIP qPCR showed that E2A binding on these sites was abrogated after LncHSC-2 KD (Figure S5F), suggesting that LncHSC-2 is directly involved or responsible for E2A binding on some target sites.

Recent studies have implicated transposable elements such as endogenous retroviruses (ERVs) and long terminal repeats (LTRs) in the evolution, regulation, and function of lncRNAs (Kapusta et al., 2013; Kelley and Rinn, 2012). To measure whether LncHSC-2 was enriched at any classes of repetitive elements, we performed peak calling again with both unique and multiple-mapped paired-end reads (including up to two alignments). The results show that LncHSC-2 binding sites are specifically enriched for the ERVL-mammalian apparent LTR retrotransposon LTR families of repeats and depleted of long interspersed nuclear elements (L1), short interspersed nuclear elements (B4 and Alu), and Simple repeats (Figure S5C; Table S6).

**DISCUSSION**

In this study, we carried out a comprehensive RNA-seq analysis in purified HSCs, differentiated B cells, and granulocytes. We discovered 2,614 known lncRNAs and almost 500 unannotated transcripts expressed in HSCs. This list contains almost all of the lncRNAs identified from in a previous study (Cabezas-Wallscheid et al., 2014) but is more comprehensive. Furthermore, we performed a series of analyses to characterize those lncRNAs, including examining their conservation, overlap with repeats, and correlation with DNA methylation and histone marks.

Although the known lncRNAs may play important functions for HSCs, in this study we specifically focused on previously unannotated transcripts and identified 159 high-confidence LncHSCs compared with the representative differentiated lineages of B cells and granulocytes. Among them, we demonstrated that LncHSC-1 and LncHSC-2 are located in the nucleus and expressed differentially between WT and Dnmt3a KO HSCs. KD of LncHSC1/2 revealed that LncHSC-1 is involved in myeloid differentiation and that LncHSC-2 is involved in HSC self-renewal and T cell differentiation. Moreover, we determined that LncHSC-2 bind sites are enriched for the hematopoietic-specific TF binding sites, especially E2A, which is a well-recognized regulator of hematopoietic differentiation.

**How Complete Is Our Catalog of Potential HSC-Specific LncRNAs?**

Here we used extremely deep sequencing data (> 1.3 billion HSC reads when combined) to detect lncRNA expression in HSCs. The number of transcripts that are truly unique to HSCs could be reduced when similarly deep sequencing was performed across additional hematopoietic lineages and when lncRNAs shared with progenitors were eliminated. On the other hand, our filtering criteria were highly stringent, including size, splicing, and expression level criteria, and we excluded putative lncRNAs.
that overlapped with protein-coding genes and their extended 3’ UTRs, even when they were predicted by splice motif analysis to be transcribed in the opposite direction of the associated coding gene. In this regard, all IncRNAs identified in our study are intergenic, which may underestimate the number of bona fide HSC-specific IncRNAs. Finally, our use of poly-A+ RNA and filtering criteria likely excludes many enhancer RNAs (eRNAs) (Natoli and Andrau, 2012), another interesting set of non-coding RNAs. Therefore, with this comprehensive but conservative approach, we can expect that these 159 LncHSCs, because of their low expression in most other tissues (Figure 1), are unlikely to have been discovered using any other approach. This relatively small number (a total of 300 HSC-specific IncRNAs including those annotated previously) is aligned with the small number of protein-coding genes (~300) thought to be uniquely expressed in HSCs compared with other blood lineages (Chambers et al., 2007) and is larger than the number of β cell- or granulocyte-specific IncRNAs, perhaps suggesting their particular roles in primitive cells.

Functional Characterization of LncHSCs

More data on the functional relevance of IncRNAs will be needed to understand their importance relative to protein-coding genes. About two-thirds (116 of 152) of reported protein-coding gene KOs result in some degree of hematopoietic defect after HSC or bone marrow transplantation (Rossi et al., 2012). Here both LncHSCs we tested in vivo showed an impact on lineage differentiation, and LncHSC-2 showed an effect on self-renewal after KD and transplantation. Although the IncRNAs we used here provide an efficient strategy for initial screening, further confirmation of these phenotypes using complete ablation and rescue experiments would be of value in the future.

In addition to functional studies, mapping the binding sites of LncHSC-2 using ChiRP-seq revealed that they are enriched for TF binding sites. KD of LncHSC-2 blocked E2A binding on some target sites, suggesting that LncHSC-2 is involved in TF binding. Whether LncHSC-2 binds directly to TFs or through other complexes to recruit them would need to be further examined.

Although IncRNAs are recognized as being less conserved across species than coding genes, we identified a human syntenic region with several putative IncRNAs that changed in DNMT3A mutant AML patients in concordance with changes in similarly localized transcripts in mouse Dnmt3a KO HSCs. Whether these LncHSCs contribute to disease development remains to be determined, but the frequent mutation of DNMT3A in AML (Ley et al., 2010; Yan et al., 2011) and other hematologic malignancies (Goodell and Godley, 2013), and the observation that >50% of LncHSCs change in expression after Dnmt3a KO, suggest that this relationship warrants further investigation.

EXPERIMENTAL PROCEDURES

See Supplemental Information for more extensive methods.

HSC Purification

All procedures were IUCAC-approved and conducted in accordance with institutional guidelines. Whole bone marrow cells were isolated from mouse femurs, tibiae, pelvis, and humeri. LT-HSCs were purified using the SP method, as described previously (Goodell et al., 1996), in conjunction with the following cell surface markers: Lineage- (CD3, CD4, CD8, B220, Gr1, Mac1, and T119) and Sca-1+ c-Kit+ CD150+.

RNA Sequencing

RNA was isolated from FACS-sorted HSCs with the RNeasy micro kit (QiAGEN). Paired-end libraries were generated with the Illumina TruSeq RNA kit. Alignment was performed by RNA-seq unified mapper (Grant et al., 2011). Cufflinks and Cuffdiff (Trapnell et al., 2010) were used for transcript reconstruction, quantification, and differential expression analysis.

shRNA Cloning and Viral Transduction

Oligos targeting each desired transcript were cloned with the BLOCK-iT Polii miRNA expression vector kit (Invitrogen). The oligos were further recombined into the retroviral MSCV-RFB vector. For retroviral transduction of hematopoietic progenitors, the suspension was spin-infected at 250 × g at room temperature for 2 hr in the presence of polybrene (4 μg/ml). For in vivo transplantation, cells were incubated for a further 1 hr at 37°C. For in vitro assays, transduced cells were cultured in fresh transduction medium for a further 2 days.

In Vivo Transplantation

C57Bl/6 CD45.1 mice were transplanted by retro-orbital injection following a split dose of 10.5 Gy of lethal irradiation. 50,000 Sca-1+ (CD45.2) donor cells were injected into the recipient mice. For secondary transplantation, 500 CD45.2+GFP+ KSL cells from primary recipients were re-sorted 20 weeks after transplantation, mixed with 250,000 CD45.1 whole bone marrow (WBM) cells, and injected into new lethally irradiated recipients.

FISH

Single-molecule RNA FISH was performed using the QuantGene ViewRNA ISH cell assay according to the manufacturer’s instructions (Affymetrix). Images were taken on an API Deltavision deconvolution microscope (Applied Precision).

CHIRP

CHIRP was performed as described previously (Chu et al., 2011).

ACCESSION NUMBERS

The GEO accession numbers for the data associated with this paper are GSE63276, GSE53928, GSE47817, and GSE50775.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2015.02.002.

AUTHOR CONTRIBUTIONS

M.L. and M.J. designed and performed the experiments and wrote the manuscript. D.S. and H.J. analyzed the RNA-seq data. B.R. analyzed the RNA-seq and CHIRP-seq data and wrote the manuscript. Z.X. analyzed the TCGA RNA-seq data. L.Y. and X.T. performed the validation experiments. K.S. made the overexpression constructs. G.A.D. supervised the research and data interpretation. W.L. supervised the bioinformatic analyses. M.A.G. supervised the study and wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by NIH grants 5T32AI007495, AG036562, AG28865, CA126752, DK092883, and DK084259; by CPRIT grant RP110028; by the Samuel Waxman Cancer Research Foundation; and by the Edward P. Evans Foundation. W.L. was supported by CPRIT grants RP110471 and RP150292 and NIH grant R01HG007538. We thank the Cytometry and Cell Sorting (grants A003621, CA125123, and RR024574), the Genomic and RNA Profiling (grant CA125123), and Integrated Microscopy (grants HD007495, DK56338, and CA125123) cores at Baylor College of Medicine.
REFERENCES


