Long noncoding RNAs: emerging regulators of normal and malignant hematopoiesis

Yi Qiu,[1,2] Mingjiang Xu,[3,4] and Suming Huang[2,5]

[1]Department of Cellular and Molecular Physiology and [2]Penn State Cancer Institute, Pennsylvania State University College of Medicine, Hershey, PA; [3]Department of Molecular Medicine and [4]Mays Cancer Center, Joe R. & Teresa Lozano Long School of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX; and [5]Division of Pediatric Hematology/Oncology, Department of Pediatrics, Pennsylvania State University College of Medicine, Hershey, PA

Genome-wide analyses have revealed that long noncoding RNAs (lncRNAs) are not only passive transcription products, but also major regulators of genome structure and transcription. In particular, lncRNAs exert profound effects on various biological processes, such as chromatin structure, transcription, RNA stability and translation, and protein degradation and localization, that depend on their localization and interacting partners. Recent studies have revealed that thousands of lncRNAs are aberrantly expressed in various cancer types, and some are associated with malignant transformation. Despite extensive efforts, the diverse functions of lncRNAs and molecular mechanisms in which they act remain elusive. Many hematological disorders and malignancies primarily result from genetic alterations that lead to the dysregulation of gene regulatory networks required for cellular proliferation and differentiation. Consequently, a growing list of lncRNAs has been reported to be involved in the modulation of hematopoietic gene expression networks and hematopoietic stem and progenitor cell (HSPC) function. Dysregulation of some of these lncRNAs has been attributed to the pathogenesis of hematological malignancies. In this review, we summarize current advances and knowledge of lncRNAs in gene regulation, focusing on recent progress on the role of lncRNAs in CTCF/cohesin-mediated 3-dimensional genome organization and how such genome folding signals, in turn, regulate transcription, HSPC function, and transformation. This knowledge will provide mechanistic and translational insights into HSPC biology and myeloid malignancy pathophysiology.

Introduction

One of the most compelling findings of the genome era of research is that the genome is extensively transcribed from noncoding regions and that these noncoding transcripts possess potential regulatory function.[1,2] Noncoding transcripts include long noncoding RNAs (lncRNAs), which are >200 nucleotides in length with very little or no protein coding ability, and other distinct classes of noncoding RNAs, including microRNAs, small nuclear RNAs, and small nucleolar RNAs. Depending on their locations, lncRNAs can function in the nucleus, nucleolus, and cytoplasm. This review focuses on the nuclear function of several newly emerged lncRNAs with respect to the CCCTC binding factor (CTCF)/cohesin-associated genomic structural and transcriptional regulation in hematopoiesis and leukemogenesis.

Many functions of lncRNAs have been discovered to involve diverse biological processes, such as imprinting, X-chromosome inactivation, apoptosis, cellular proliferation and differentiation, stem cell biology, and tumorigenesis/leukemogenesis.[3–8] The role of lncRNAs in transcriptional regulation involves recruitment or decoy of epigenetic regulators, transcription factors (TFs), and accessory proteins. However,

how and where lncRNAs exert these diverse functions, including with what they interact, have become very important questions and an important research area in order to gain mechanistic insights into this class of RNA molecules. Acute myeloid leukemia (AML) is a heterogeneous disease with genetic or epigenetic alterations that systematically alter gene regulatory networks required for hematopoietic stem and progenitor cell (HSPC) proliferation and differentiation. A growing list of lncRNAs that are involved in controlling hematopoietic gene expression networks has been discovered, and some are linked to the dysregulation of HSPC function in AML.[9,10] In this review, we focus on molecular mechanisms through which lncRNAs systematically regulate hematopoietic transcription pathways via their specific roles in 3-dimensional (3D) hematopoietic/leukemic genome organization.

Overview of mechanisms by which lncRNAs regulate gene transcription

In the nucleus, lncRNAs can regulate gene transcription in *cis* and *trans* actions via physical interactions with target DNA and associated proteins. lncRNAs can regulate epigenetic landscapes, enhancer/promoter activity, and RNA polymerase II machinery by diverse mechanisms.[11,12]

Scaffold *Xist* lncRNA-mediated X-chromosome inactivation is a well-known example of lncRNA-driven transcriptional silencing (Figure 1A), one of the critical mechanisms leading to the gene dosage compensation in female cells. As a key regulator of this process, *Xist* lncRNA, via the orchestration of 3D chromatin conformation, spreads and coats polycomb repressive complex 2 (PRC2) and resulting H3K27me3 modifications on the future inactive X-chromosome for stable silencing of the entire X-chromosome.[5,13] It is interesting to note that deletion of *Xist* in the hematopoietic compartment of female mice impaired hematopoietic stem cell (HSC) function, and mutant females developed aggressive myeloproliferative neoplasm and myelodysplastic syndrome that resulted from a dysregulated myeloerythroid fate decision of HSCs.[14] Notably, the X-linked master erythroid TF Gata-1 was upregulated in *Xist*[−/−] females, perhaps as a result of X-linked gene overdosage in mutant females.[14] Consistent with *Xist* RNA–dependent PRC2 complex repressive function, recent studies using RNA-dependent chromatin immunoprecipitation sequencing analysis revealed that the PRC2 complex generally requires RNA binding to guide chromatin occupancy for spreading of the H3K27me3 mark and its gene-repressive effects.[15] Thus, the role of the PRC2-mediated chromatin repressive complex may require diverse scaffold/architectural RNA species.

Recruitment/guide *Hox* genes are critical for maintaining the balance between the self-renewal and differentiation of HSCs. Their dysregulation is a dominant mechanism underlying AML pathogenesis.[16] Recently, it was reported that hundreds of lncRNAs are transcribed from *HOX* gene clusters.[17] Many *HOX* genes can be regulated by specific lncRNAs via *cis-* or *trans-* action. One such lncRNA, *HOTTIP*, is transcribed from the distal *HOXA* locus and activates genes in *cis* by recruiting and guiding the MLL1/WDR5 complex to the posterior *HOXA* genes[18] (Figure 1B); this suggests that lncRNAs can modulate chromatin dynamics by recruiting histone-modifying and chromatin-remodeling complexes to specific genomic loci. We recently discovered and cloned a unique *HoxB* locus–associated lncRNA, *HoxBlinc*, which is expressed during early hematopoietic differentiation and is strongly concomitant with H3K4me3 patterning and anterior *Hoxb* gene activation. *HoxBlinc* associates with the Setd1a/MLL1 histone methyltransferase complex to coordinate activation of the anterior *Hoxb* genes and specify hematopoietic cell fates during early hematopoietic differentiation.[3] It is clear that lncRNAs can act as epigenetic regulators to modulate gene transcription.

Stabilization Transcription regulation is central in controlling cellular proliferation and differentiation. Many proteins that are known to be involved in eukaryotic transcription can interact with DNA elements, as well as are involved in RNA interaction. A significant number of TFs can be modulated by specific lncRNAs, and their interaction may stabilize TF in chromatin template (Figure 1C). One example of such an lncRNA is *PVT1*, an lncRNA required by MYC in MYC-dependent cancers.[19] Depletion of *PVT1* led to disproportionately decreased MYC protein levels in MLL-AF9–driven AML, whereas enforced expression of *Myc* rescued the effects caused by *Pvt1* knockdown.[20] Threonine 58 phosphorylation enhances MYC protein degradation.[21] The underlying molecular action of *PVT1* in AML involves stabilization of MYC protein by inhibiting its T58 phosphorylation, thereby enhancing MYC-mediated transcription.

Decoy lncRNAs can also inhibit the genomic binding of TFs by acting as a "decoy molecule." The lncRNA growth arrest-specific 5 (*Gas5*) is involved in starvation-associated repression of the glucocorticoid receptor (GR)-mediated transcription.[22] *GAS5* used a decoy mechanism to sequester GR from the glucocorticoid-responsive genes by the formation of an RNA motif in one of its stem-loop secondary structures that mimics the hormone response element–like DNA motif. Therefore, *GAS5* competes for the binding of GR to the glucocorticoid-response elements in the promoters of glucocorticoid-responsive genes[22] (Figure 1D).

Bridge Enhancer RNAs (eRNAs) are a specific class of lncRNAs that are transcribed from enhancer DNA sequences and may have a profound impact on gene transcription. The p53-associated eRNAs were transcribed from the p53-bound enhancer regions (p53BERs) that induce p53-independent intrachromosomal interactions, bringing p53BERs into close proximity of the p53 target genes for efficient transcription enhancement of the associated p53 target genes and leading to p53-dependent cell cycle arrest[23] (Figure 1E).

In T-cell acute lymphoblastic leukemia (T-ALL), mutation in the *TAL1* gene is one of the most frequent gain-of-function mutations found in patients with T-ALL. The TAL1 transcription complex consists of multiple TFs, such as TAL1, GATA-3, RUNX1, MYC, and MYB, and acts as an oncogenic regulator circuit in TAL1 activation–driven T-ALL. One of the major targets of the TAL1 transcription complex is the proto-oncogene *ARID5B*, which supports the survival of human T-ALL cells and potential tumor formation in the thymus. TAL1 transcription complex binds to the −135 kb enhancer of the *ARID5B* gene and induces transcription of the eRNA *ARIEL*, which, in turn, stabilizes the enhancer-occupied TAL1 transcription complex and facilitates enhancer/promoter communications, leading to transcription activation of *ARID5B* and a positive feed-forward regulatory loop that promotes leukemogenesis.[24]

The RNA/DNA hybrid: R-loop or triple helix As epigenetic and transcriptional regulators, lncRNAs need to access specific genomic loci, including enhancers and promoters, to exploit their regulatory effects. lncRNAs can use the R-loop or triple helix to access specific genomic regulatory elements and exert transcription regulation.[4,25,26] R-loops are 3-stranded nucleic acid structures comprising RNA hybridized with the DNA template, leaving the nontemplate DNA single stranded. The presence of nascent RNA-dependent R-loops in the gene promoter region has been linked to transcriptional regulation through the modulation of promoter accessibility, RNA polymerase II recruitment, and RNA splicing.[27–30] Recently, it was shown that *APOLO* lncRNA recognizes and coordinates expression of several target genes by the formation of R-loops in the *Arabidopsis thaliana* genome,[25] suggesting that the R-loop underlies lncRNA-mediated genome accessibility for transcriptional regulation (Figure 2A). On the other hand, oncogenic lncRNA *Khps1* can directly bind to the promoter of the proto-oncogene *SPHK1* by formation of a DNA-RNA triplex through a noncanonical *Hoogsteen* base-pairing mechanism. *Khps1* lncRNA then recruits and targets p300/CBP to the *SPHK1* promoter, leading to local changes in chromatin structure and the activation of *SPHK1* expression in an E2F1-dependent manner (Figure 2B).[4] In the
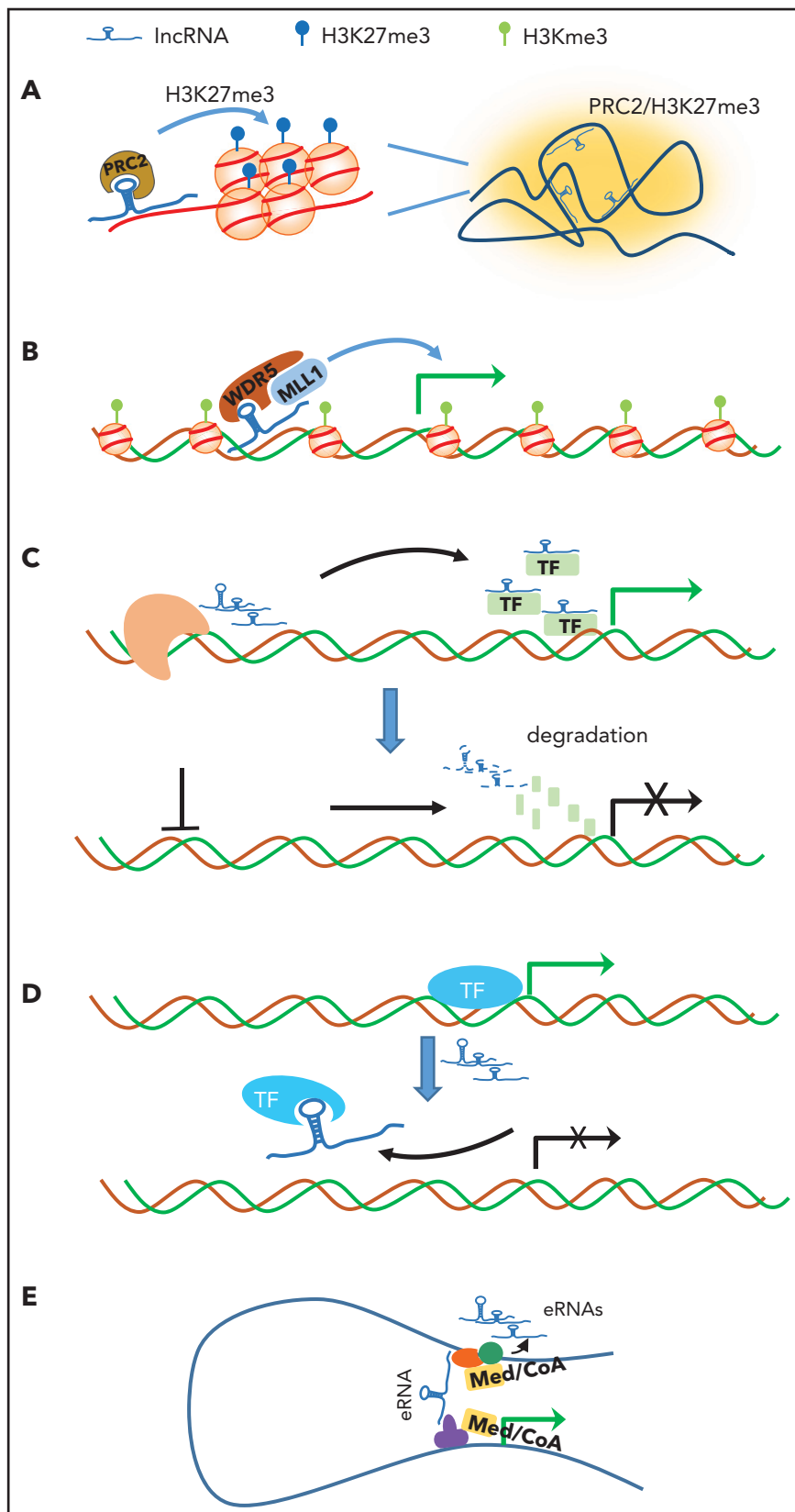
QIU et al

Figure 1. Schematic illustration of the potential mechanisms by which lncRNAs regulate chromatin structure and gene transcription. (A) Scaffold: lncRNA acts as a scaffold to orchestrate active or repressive chromatin (eg, PRC2 complex and H3K27me3 in Xi) over a large distance. (B) Recruitment/guide: lncRNA acts a mediator to recruit and guide the coactivator complex to enhancer/promoter for gene transcription. (C) Stabilization: lncRNAs produced in proximity of gene specifically associate with TFs or cofactors to facilitate gene transcription (upper panel). Depletion of lncRNA (dashed line) leads to degradation of TFs and cofactors, resulting in transcription silencing (lower panel). (D) Decoy: lncRNA acts as molecular decoy that dissociates promoter-bound TFs from the gene, leading to transcriptional inhibition. (E) Bridge: lncRNA or nearby enhancer-produced eRNA binds to and stabilizes the enhancer and promoter loop, leading to transcriptional activation. CoA, coactivator; Med, mediator.
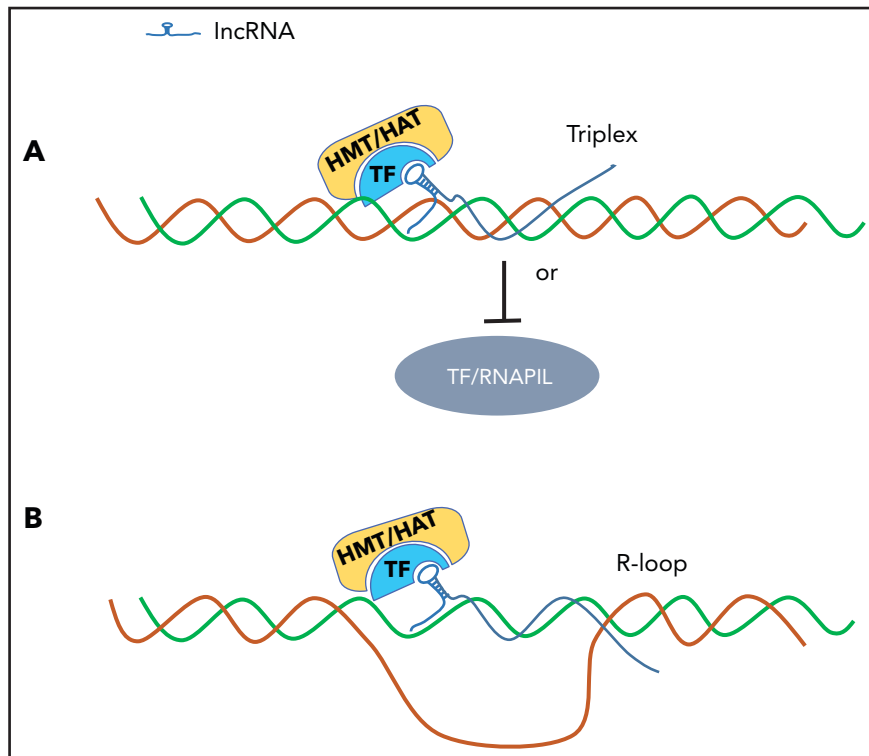
Figure 2. LncRNA binds and accesses the regulator DNA elements by triplex or R-loop mechanism. (A) LncRNA forms a triple helix structure via a Hoogsteen base pair mechanism to access genomic DNA regulatory elements. (B) lncRNA forms an R-loop structure with the genomic DNA regulatory elements.

β-globin locus, an RNA transcript embedded in the second intron of the human *β-globin* gene forms a triplex with *hypersensitive site 2* of the *β-globin* locus control region. The triplex structure displaces TFs and the RNA polymerase II complex, resulting in downregulation of genes controlled by hypersensitive site 2 in the locus.[31] Thus, the direct binding of lncRNAs to enhancers and promoters may serve as a signal to initiate the transcription or silencing process.

Role of lncRNAs in shaping local chromatin structure and enhancer/promoter interactions

Genes and their regulatory elements reside nonrandomly in the nucleus of cells. A special genome organization involved in the positioning and organization of chromatin to facilitate the transcriptional machinery within the nucleus is essential for overall genome function. It seems that these topological associated chromatin domains (TADs) require structural/regulatory protein complexes, various genetic elements, and a nuclear envelope, as well as lncRNAs. Recent genomic studies revealed that many lncRNAs play an important role in local chromatin domain organization and enhancer/promoter interactions, and the list is still growing (Table 1). The *Xist* lncRNA–driven X-chromosome inactivation throughout the female life to equalize X-linked gene dosage with male is an excellent example of an lncRNA serving as a driving force for large-scale chromosome organization. *Xist* initiates and spreads PRC2 and the resulting H3K27me3 accumulation along the X-chromosome by orchestrating a 3D genome topology,[5] as well as targets CTCF to specific genome loci to mediate long-range chromosomal interactions in a locus-specific manner during X-chromosome inactivation.[32] Mechanistically, *Xist* lncRNA actively repels the cohesin complex from the

inactive X-chromosome, because deletion of *Xist* restores cohesin binding. This results in reorganization of TADs that resemble an active X-chromosome,[33] suggesting a central role for lncRNA in the 3D organization of the mammalian genome, perhaps mediated by CTCF and its associated cohesin complex.

The notion that lncRNAs are involved in CTCF-mediated genome organization stems from a seminal finding by Felsenfeld and colleagues that DEAD-box RNA helicase p68 and steroid receptor RNA activator (SRA) interact with CTCF and cohesin. Depletion of p68 or SRA impaired chromatin insulator activity by disrupting CTCF/cohesin interaction,[34] suggesting that p68/SRA stabilizes the CTCF/cohesin complex and that RNA may be required for proper chromatin boundary function. Given that many lncRNAs are *cis* or *trans* acting[35] and associated with CTCF,[32] lncRNAs may contribute to the recruitment and stabilization of the CTCF complex in genomic loci. In support of this notion, 2 back-to-back publications in *Molecular Cell* demonstrated that the zinc finger (ZF) domains of CTCF are required for CTCF clustering and CTCF-mediated long-range chromatin interactions.[36,37] The RNA-binding regions (RBRs) of CTCF are located within the ZF1 and ZF10 domains. RBRs are dispensable for DNA recognition and binding, but deletion of RBR disrupts half of the CTCF loops.[37] Thus, it seems that CTCF contributes to genome topological regulation via an RNA-dependent mechanism. However, the type(s) of RNAs that contribute to CTCF genome action, how RNA species access specific genome regulatory loci, and more importantly, the identification of the underlying molecular mechanisms and biological consequences remain to be determined.

Table 1. lncRNAs involved in genome topology

Name	Role in genome organization	Function	Reference
HOTTIP	Topological organization of posterior HOXA and canonical Wnt/β-catenin target genes	HSC self-renewal AML leukemogenesis	7
Xist	Scaffold and 3D topology for PRC2/H3K27me3	X-inactivation	33
APOLO	R-loop–mediated decoy of PRC1 and chromatin loop modulation	Gene transcription root development	25
Firre	Spatial organization of multiple chromosomal regions	Adipogenesis ESC pluripotency	67
HOXBLINC	3D organization of anterior HOXB genes	Hemagiogenic mesoderm NPM1^{C+}-driven AML	3, 44
PAIR	PAIR elements/Eμ long-range interactions	VDJ recombination	68
MYMLR	Long-range enhancer/promoter interaction	MYC transcription	69
H19X	Long-range enhancer/promoter interaction	Transcription silencing	70
p53BER eRNA	p53BER and p53 target gene long-range contacts	p53-driven transcription	23
ARIEL eRNA	Long-range enhancer and promoter interaction	Proto-oncogene transcription T-ALL survival pathway	24

ESC, embryonic stem cell; T-ALL, T-cell acute lymphoblastic leukemia.

Role of lncRNAs in HSC biology

Compared with messenger RNA, the expression levels of lncRNAs are relatively low, yet the expression of many lncRNAs is highly cell-type/lineage specific, differentiation stage specific, or even disease state specific,[38-40] suggesting that lncRNAs act as determinants of lineage commitment and differentiation, perhaps by acting coordinately with lineage-specific TFs. HSCs require a unique gene-regulatory program that reinforces their cellular identity and controls cell fate decisions. Recent comprehensive genome-wide studies on HSCs and lineage-primed multipotent progenitors revealed that hundreds of lncRNAs are expressed together with lineage-specific TFs in these cells.[41,42] Many of these lncRNAs are closely associated with lineage-specific TFs that are required for hematopoietic differentiation and cell fate decision,[41,42] suggesting that lncRNAs may provide another layer of regulation operating within the HSC hierarchy.

Consistently, Goodell and colleagues used extremely deep RNA sequencing (>1.3 billion combined HSC reads), which overcomes the relatively low expression levels of lncRNAs compared with protein coding transcripts in rare HSC populations, to identify lncRNAs that are specifically associated with mouse long-term HSCs.[43] They identified and annotated 159 candidate HSC-specific lncRNAs with high confidence scores. Some of them modulate HSC self-renewal and differentiation by binding to hematopoietic TF binding sites.[43] Among them, lncHSC-2 controls HSC self-renewal and differentiation by modulating the activity of basic helix-loop-helix TF, TCF3/E2A, through binding to the TCF3-bound E-box elements at genes that are important for HSC homeostasis. Loss of lncHSC-2 in HSPCs resulted in impaired HSC self-renewal and skewed T-cell differentiation by abrogating TCF3 binding, a TF that is required for HSC function and T-cell development.[43] Thus, it is likely that lncRNAs act as an additional layer of genetic circuitry by modulating the accessibility and activity of TFs at their bound DNA motifs.

The expression of HOX genes, particularly HOXA and HOXB, is tightly regulated throughout hematopoiesis, with the highest expression in HSPCs that gradually decreases to minimum levels in mature lineages. Two HOX gene–associated lncRNAs, HOXBLINC and HOTTIP, were recently shown to play critical roles in HSC self-renewal and lineage commitments.[3,7,44] HOXBLINC, embedded in the anterior HOXB locus, is highly expressed in long- and short-term HSCs, and its expression diminishes in mature hematopoietic lineages.[3,44] The expression pattern of HOXBLINC in hematopoiesis suggests a role in HSC maintenance and regulation. During embryonic development, induction of HoxBlinc transcription initiates anterior HoxB genes activation by recruiting the Setd1a/MLL1 complex to the promoters of anterior HoxB genes and promoting 3D enhancer/promoter interactions (Figure 3A). This regulatory process initiates and stimulates hemangiogenic Flk1$^+$ mesoderm differentiation and the hematopoietic fate decisions.[3] HOXBLINC was aberrantly activated in patients with specific subtypes of AML carrying nucleophosmin 1 (NPM1^{C+}) mutations. Interestingly, transgenic overexpression of HoxBlinc in the hematopoietic compartment results in aberrant HSC self-renewal and expanded myelopoiesis, leading to the development of AML-like disease that resembles NPM1^{C+}-driven AML. Mechanistically, HOXBLINC acts as a downstream effector of NPM1^{C+} to recruit the MLL1 complex to the promoters of NPM1^{C+} signature genes and activate signature NPM1^{C+}-associated transcription networks, including anterior HOXB genes.[44]

On the other hand, HOTTIP, which is transcribed from the posterior HOXA locus, activates multiple homeotic TF-encoded genes, including posterior HOXA genes, the canonical Wnt/β-catenin pathway, and various key hematopoietic regulators, by orchestrating CTCF-defined hematopoietic-associated TADs. Aberrant expression of HOTTIP in the mouse bone marrow hematopoietic compartment enhanced HSC self-renewal, leading to leukemic transformation of HSCs via aberrant binding of HOTTIP to hematopoietic/leukemic-specific TFs or CTCF motifs.[7]

In addition to those lncRNAs involved in HSC self-renewal and differentiation, maternal, but not paternal, deletion of *H19* lncRNA and its associated upstream differentially methylated region (*DMR*) reduced adult HSC quiescence, which plays a key role in long-term HSC maintenance. The expression of *H19* is maternally imprinted (expressed exclusively by the maternally inherited allele) and is regulated by binding of CTCF to the unmethylated *DMR*. Loss of maternally expressed *H19* upregulated the paternally associated *Igf2* gene pathway that resulted in FoxO3-mediated cell cycle arrest.[6]

Involvement of HOX-associated lncRNAs in AML leukemic genome organization

AML is a heterogeneous hematopoietic stem or progenitor cell disease that results from genetic alterations or somatic mutations in genes that are required for HSPC biology. Given that lncRNAs are crucial regulators of gene expression in HSCs that determine cell fate and commitment decisions, the expression landscapes of lncRNAs that exhibit an HSPC signature are likely shared with AML blasts.[45] Many AML-associated lncRNAs are essential for the maintenance of leukemic transcription programs that promote leukemic stem cell signatures.[20] Thus, the expression landscape of a specific lncRNA class may be of great prognostic value and clinical relevance for patients with AML.

NPM1 is the most commonly mutated gene and accounts for 30% of AML patients with a normal cytogenetic karyotype. $NPM1^{C+}$ mutation results in aberrant cytoplasmic translocation of the encoded mutant nucleolar protein, which maintains a unique expression pattern of *HOXA/B* genes and the *MEIS1* oncogene to facilitate AML leukemogenesis. An anterior *HOXB*-derived lncRNA, *HOXBLINC*, which is aberrantly activated in patients with AML carrying $NPM1^{C+}$, mediates and maintains $NPM1^{C+}$-derived leukemic gene signatures and leukemogenesis.[44] HoxBlinc activation is also required to promote HSPC transcription signatures and determine HSPC fate and commitment.[3] In HSCs and AML, *HOXBLINC* RNA exploits an active 3D interactome within the anterior *HOXB* domain to recruit the MLL1/Setd1a complex and deposit H3K4me3 marks on each of the anterior *HoxB* genes for their activation.[3,44] Given that NPM1 interacts with CTCF and provides nuclear matrix for anchoring CTCF-mediated chromatin loops,[46] it would be interesting and
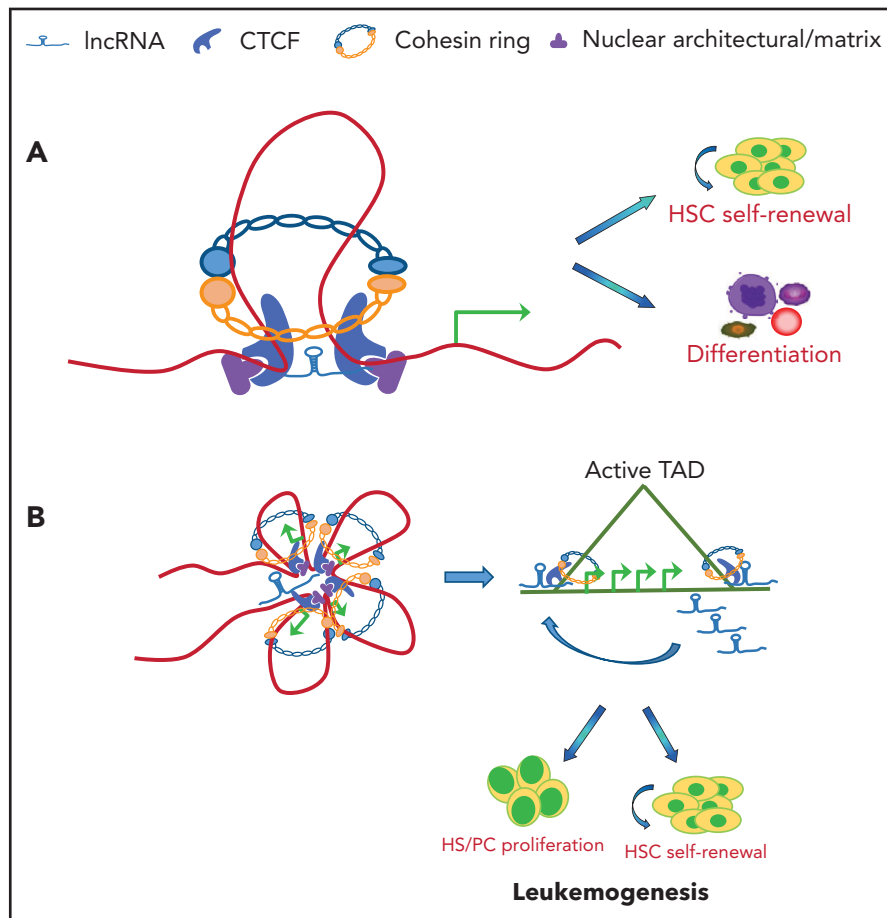

Figure 3. Schematic illustration of lncRNA involved in CTCF/cohesin-mediated chromatin interaction and genome topology. (A) CTCF and cohesin are frequently localized at enhancers and promoters in the genome. lncRNAs may tether and stabilize CTCF/cohesin-mediated enhancer/promoter communication by forming chromosomal loops, which lead to activation of genes required for HSC self-renewal or lineage commitment and differentiation. (B) lncRNA may coordinate with CTCF/cohesin complexes to form TAD domains in hematopoietic/AML genome. In both cases, chromatin organization results in systematic changes in gene regulatory networks and transcription programs that contribute to the promotion of hematopoietic lineage differentiation or blockage of lineage differentiation and leukemogenesis.

QIU et al

important to determine how $NPM1^{C+}$ alters hematopoietic genome topology and whether subsequent activation of HOX-BLINC reorganizes and facilitates HSPC survival pathways upon cytoplasmic translocation of the $NPM1^{C+}$ mutant.

The posterior HOXA locus–associated lncRNA, HOTTIP,[18] exemplifies the importance of lncRNAs in leukemic genome organization. HOTTIP coordinates with CTCF to promote aberrant posterior HOXA TADs and chromatin signature in AML.[7] Posterior HOXA genes are aberrantly activated in specific types of AML, especially in those patients and cells carrying mixed lineage leukemia (MLL) gene rearrangements ($MLLr^+$) or $NPM1^{C+}$ mutations.[47,48] HOXA and HOXB genes are critical for maintaining the balance between self-renewal and differentiation of HSCs.[49-51] Dysregulation of HOXA and/or HOXB genes is a dominant mechanism of leukemic transformation.[52] Epigenetic analysis of the HOX gene loci in $MLLr^+$ or $NPM1^{C+}$ AML cells revealed that distinct chromatin/epigenetic signatures were defined by CTCF boundaries in the posterior HOXA TAD domain, which contribute to aberrant HOXA gene expression in a HOTTIP-dependent manner.[53,54] Attenuation of CBS7/9, a CTCF boundary located between the HOXA7 and HOXA9 genes in the beginning of the posterior HOXA TAD, alters posterior HOXA TADs and inhibits posterior HOXA gene signatures, including HOTTIP lncRNA, to alleviate AML leukemogenesis.[53] Interestingly, depletion of HOTTIP also impairs the posterior HOXA TAD and oncogenic homeotic transcription program by impairing HOTTIP chromatin binding, leading to blockage of AML leukemogenesis.[7] Overexpression of HOTTIP restores CTCF-mediated HOXA TADs and leukemogenesis in attenuated AML cells in the CBS7/9 boundary,[7] supporting a feed-forward mechanism by which collaboration of HOTTIP and CTCF stratifies the CTCF boundary and enforces oncogenic posterior HOXA TADs and gene activation to promote aberrant HSPC self-renewal and transformation (Figure 3B). To support the collaboration between HOTTIP and CTCF in the TAD boundary, analyses of HOTTIP chromatin-binding motifs showed a significant enrichment of core CTCF consensus-binding motifs; loss of HOTTIP resulted in the decrease in a subset of CTCF-defined TADs and chromatin signatures in loci involved in homeotic oncogenic TFs and hematopoiesis.[7] However, it remains unknown through which mechanisms lncRNAs, such as HOTTIP, are targeted to and/or recognize specific CBSs. It is also important to determine how widely required RNAs are for CTCF-mediated genome organization and the molecular basis by which lncRNAs act on CTCF-mediated topological genome regulation. Unbiased isolation of the HOTTIP-associated protein complex in AML cells would provide mechanistic insight into its action in CTCF-mediated genome organization.

lncRNAs and their pathways serve as biomarkers and potential therapeutic targets for leukemia
Many lncRNAs exhibit lineage-, differentiation stage-, or disease-specific expression, making them excellent candidates for biomarkers or therapeutic applications. A study evaluated the lncRNA expression profiles from 148 untreated older AML patients with normal cytogenetics to determine whether lncRNAs are associated with clinical features and certain recurrent mutations in these patients. Interestingly, distinct lncRNA expression profiles were highly associated with specific mutations, such as FLT3-ITD, NPM1, CEBPA, IDH2, ASXL1, and RUNX1 genes, whereas

another subset of lncRNAs is highly correlated with treatment response and survival.[55] Recently, RNA sequencing, followed by univariate and multivariate time-to-event analyses, of 274 Swedish patients with AML after intensive chemotherapy found that 33 lncRNAs were associated with overall survival, and their expression profiles predicted outcomes.[56] These lncRNAs likely play oncogenic roles via their impact on 1 or several hallmarks of leukemia, including self-renewal, antiapoptosis, proliferation, and differentiation. The importance of some of these lncRNAs in reshaping the epigenetic landscape of the leukemic genome is beginning to be recognized. These AML-associated specific characteristics of lncRNAs make them potential markers for clinical prognostic outcomes and potential targets for treatment.

HOTTIP and HOXBLINC are HOX-associated oncogenic lncRNAs that are aberrantly activated in patients with AML carrying $MLLr^+$ or $NPM1^{C+}$ mutations.[7,44] HOTTIP lncRNA activation promotes HSC/leukemic stem cell (LSC) self-renewal and AML progression by activating HOXA9 and the canonical Wnt/β-catenin pathway,[7] which is uniquely required for LSC self-renewal.[57,58] Mechanistically, HOTTIP acts downstream of $MLLr^+$ or $NPM1^{C+}$ by coordinating TAD organization of the AML genome, perhaps by regulating CTCF chromatin boundary activity via direct binding to a subset of CTCF core motifs in the AML genome. Analyses of HOTTIP DNA binding motifs in the AML genome revealed a significant enrichment of CTCF core motifs,[7] supporting that HOTTIP is involved in CTCF boundary function. Indeed, depletion of HOTTIP attenuates AML progression by impairing leukemic TADs and HOXA9- and/or β-catenin–driven transcription programs. In light of understanding the downstream pathways that mediate HOTTIP oncogenic action in $MLLr^+$ AML, pharmacological targeting of the canonical Wnt pathway to suppress the β-catenin–mediated transcription program blocked $MLLr^+$ LSC self-renewal and HOTTIP activation-driven leukemogenesis.[7] Similarly, HOXBLINC activation in $NPM1^{C+}$ AML cells specifically promotes the $NPM1^{C+}$-driven transcription signatures to facilitate $NPM1^{C+}$-derived AML leukemogenesis.[44] Thus, small interfering RNAs/short hairpin RNAs or CRISPR/Cas9-targeted degradation or silencing of HOTTIP or HOXBLINC may enable systematic reversal of the genome TAD and transcription program driven by MLL fusions or $NPM1^{C+}$. In addition, dead Cas9 targeted interference of interactions between lncRNAs and genetic elements or regulatory proteins at specific loci may be able to block specific leukemic transcription programs (eg, HOXA9 or β-catenin), leading to efficient targeted therapy

Summary and perspectives
Novel technologies have been used extensively in genome research to interrogate the genetic targets and associated protein complexes of specific lncRNAs in cells. As such, the roles of lncRNAs in lineage- or disease-specific genome organization have broadened in the past several years. However, the detailed mechanisms involved in lncRNA-dependent genome topology remain to be explored. Although HOTTIP lncRNA is shown to mediate TAD formation in several critical hematopoietic/leukemogenic loci during AML progression,[7] many questions remain unanswered. Mechanistically, it remains unclear how lncRNA regulates TAD and TAD boundaries. It also remains to be determined whether and how lncRNAs recruit, bridge, guide, or facilitate the CTCF/cohesin complex to specific genomic regions to

alter genome topological structure. CTCF is a master regulator of mammalian genome organization.[59] CTCF modulates the chromatin TAD boundary, as well as enhancer/promoter contacts within TADs, in cohesin-dependent or -independent manners.[60,61] Although the CTCF core motif is highly enriched in *HOTTIP*-bound genomic regions,[7] it remains to be determined whether lncRNAs directly regulate the ability of CTCF to access chromatin DNA or its association with the cohesin complex. The cohesin complex frequently colocalizes with CTCF in the genome, and their binding sites are mutational hotspots in the noncoding cancer genome.[62] In addition, the cohesin genes are frequently mutated in AML that exhibits a strong association with the *NPM1* mutation.[63] Therefore, it appears that the CTCF/cohesin-mediated function plays a critical role in HSPC regulation and AML leukemogenesis, perhaps through its actions in hematopoietic specific genome topology. A recent study showed that architectural RNA is essential for PRC2 complex chromatin occupancy,[15] suggesting that RNA provides an additional regulatory layer and may have a wide role in chromatin domain organization beyond the PRC2-repressive complex. Given the general role of RNAs in chromatin and gene regulation, it will be interesting to determine whether the role of CTCF/cohesin requires general action from RNA, such as architectural RNA and regulatory lncRNAs in different AML subtypes, and whether such RNAs control CTCF DNA binding or CTCF/cohesin interaction. With new genomic techniques surging and increasing evidence of lncRNA's role in AML genome organization, it will become clear whether CTCF/cohesin-mediated genome organization generally depends on RNAs to facilitate their activity in inducing aberrant genome topology in lineage- or leukemic-specific loci to promote leukemogenesis. More importantly, it remains to be determined whether lncRNAs play architectural or regulatory roles in CTCF/cohesin chromatin boundary activity and CTCF-mediated genome topology.

The advances in new RNA centric genomic/proteomic technologies, such as chromatin immunoprecipitation by RNA purification in combination with deep sequencing or chromatin immunoprecipitation by RNA purification in combination with mass spectrometry, identification of direct RNA-interacting proteins (iDRiP), R-ChIP for robust mapping of R-loops genome-wide, and DNA-RNA immunoprecipitation combined with deep sequencing,[27,33,64–66] enable researchers to map global binding of a particular lncRNA in the genome and identify the associated protein complexes. These studies aid in the investigation of the molecular mechanisms by which lncRNAs access and regulate the genome in high resolution. Furthermore, refined CRISPR-Cas9–mediated genome editing of lncRNA and targeting of RNA-mediated structures (eg, R-loops or triplex) at specific genomic locations will provide important functional links and validations. Eventually, it is anticipated that increasing mechanistic understanding of lncRNA-driven topological genome organization will solidify its link with HSPC regulation and leukemogenesis in the next few years.

Authorship

Contribution: Y.Q., M.X., and S.H. wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: Y.Q., 0000-0002-3282-6072; M.X., 0000-0002-7063-6253; S.H., 0000-0002-4788-2085.

Correspondence: Suming Huang, Pennsylvania State University College of Medicine, 500 University Dr, H085, Room T3402, Hershey, PA 17033; e-mail: shuang4@pennstatehealth.psu.edu.

REFERENCES

1. Hon CC, Ramilowski JA, Harshbarger J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*. 2017; 543(7644):199-204.

2. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015; 47(3):199-208.

3. Deng C, Li Y, Zhou L, et al. HoxBlinc RNA recruits Set1/MLL complexes to activate Hox gene expression patterns and mesoderm lineage development. *Cell Rep*. 2016;14(1):103-114.

4. Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L, et al. LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol Cell*. 2015;60(4):626-636.

5. Engreitz JM, Pandya-Jones A, McDonel P, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 2013; 341(6147):1237973.

6. Venkatraman A, He XC, Thorvaldsen JL, et al. Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. *Nature*. 2013;500(7462): 345-349.

7. Luo H, Zhu G, Xu J, et al. HOTTIP lncRNA promotes hematopoietic stem cell self-renewal leading to AML-like disease in mice. *Cancer Cell*. 2019;36(6):645-659.e8.

8. Yoon JH, Abdelmohsen K, Srikantan S, et al. LincRNA-p21 suppresses target mRNA translation [published correction appears in *Mol Cell*. 2013;50(2):303]. *Mol Cell*. 2012; 47(4):648-655.

9. Paralkar VR, Weiss MJ. Long noncoding RNAs in biology and hematopoiesis. *Blood*. 2013;121(24):4842-4846.

10. Alvarez-Dominguez JR, Lodish HF. Emerging mechanisms of long noncoding RNA function during normal and malignant hematopoiesis. *Blood*. 2017;130(18): 1965-1975.

11. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell*. 2011; 43(6):904-914.

12. Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription? *Sci Adv*. 2017;3(9):eaao2110.

13. Simon MD, Pinter SF, Fang R, et al. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*. 2013;504(7480):465-469.

14. Yildirim E, Kirby JE, Brown DE, et al. Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell.* 2013;152(4):727-742.

15. Long Y, Hwang T, Gooding AR, Goodrich KJ, Rinn JL, Cech TR. RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. *Nat Genet.* 2020;52(9):931-938.

16. Rawat VP, Humphries RK, Buske C. Beyond Hox: the role of ParaHox genes in normal and malignant hematopoiesis. *Blood.* 2012; 120(3):519-527.

17. Rinn JL, Kertesz M, Wang JK, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007;129(7): 1311-1323.

18. Wang KC, Yang YW, Liu B, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature.* 2011;472(7341):120-124.

19. Tseng YY, Moriarity BS, Gong W, et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature.* 2014;512(7512): 82-86.

20. Delás MJ, Sabin LR, Dolzhenko E, et al. lncRNA requirements for mouse acute myeloid leukemia and normal differentiation. *eLife.* 2017;6:e25607.

21. Yeh E, Cunningham M, Arnold H, et al. A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat Cell Biol.* 2004;6(4): 308-318.

22. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal.* 2010; 3(107):ra8.

23. Melo CA, Drost J, Wijchers PJ, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell.* 2013;49(3):524-535.

24. Tan SH, Leong WZ, Ngoc PCT, et al. The enhancer RNA *ARIEL* activates the oncogenic transcriptional program in T-cell acute lymphoblastic leukemia. *Blood.* 2019; 134(3):239-251.

25. Ariel F, Lucero L, Christ A, et al. R-loop mediated trans action of the APOLO long noncoding RNA. *Mol Cell.* 2020;77(5): 1055-1065.e4.

26. Cloutier SC, Wang S, Ma WK, et al. Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation [published correction appears in *Mol Cell*; 2016;62(1):148]. *Mol Cell.* 2016;61(3):393-404.

27. Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol.* 2015; 22(12):999-1007.

28. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science.* 2013; 340(6132):619-621.

29. Boque-Sastre R, Soler M, Oliveira-Mateos C, et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci USA.* 2015;112(18):5785-5790.

30. Chakraborty P, Huang JTJ, Hiom K. DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat Commun.* 2018;9(1):4346.

31. Zhou Z, Giles KE, Felsenfeld G. DNA·RNA triple helix formation can function as a *cis*-acting regulatory mechanism at the human β-globin locus. *Proc Natl Acad Sci USA.* 2019;116(13):6130-6139.

32. Kung JT, Kesner B, An JY, et al. Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Mol Cell.* 2015;57(2):361-375.

33. Minajigi A, Froberg J, Wei C, et al. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science.* 2015;349(6245):10.1106.

34. Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 2010;24(22): 2543-2555.

35. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell.* 2013;152(6): 1308-1323.

36. Saldaña-Meyer R, Rodriguez-Hernaez J, Escobar T, et al. RNA interactions are essential for CTCF-mediated genome organization. *Mol Cell.* 2019;76(3):412-422.e5.

37. Hansen AS, Hsieh TS, Cattoglio C, et al. Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. *Mol Cell.* 2019;76(3):395-411.e13.

38. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011;25(18):1915-1927.

39. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature.* 2011;477(7364):295-300.

40. Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* 2014;24(4): 616-628.

41. Cabezas-Wallscheid N, Klimmeck D, Hansson J, et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell.* 2014;15(4): 507-522.

42. Wu Z, Gao S, Zhao X, et al. Long noncoding RNAs of single hematopoietic stem and progenitor cells in healthy and dysplastic human bone marrow. *Haematologica.* 2019; 104(5):894-906.

43. Luo M, Jeong M, Sun D, et al. Long non-coding RNAs control hematopoietic stem cell function. *Cell Stem Cell.* 2015;16(4): 426-438.

44. Zhu G, Luo H, Feng Y, et al. HOXBLINC long non-coding RNA activation promotes leukemogenesis in NPM1-mutant acute myeloid leukemia. *Nat Commun.* 2021;12(1): 1956.

45. Schwarzer A, Emmrich S, Schmidt F, et al. The non-coding RNA landscape of human hematopoiesis and leukemia. *Nat Commun.* 2017;8(1):218.

46. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell.* 2004; 13(2):291-298.

47. Drabkin HA, Parsy C, Ferguson K, et al. Quantitative HOX expression in chromosomally defined subsets of acute myelogenous leukemia. *Leukemia.* 2002; 16(2):186-195.

48. Andreeff M, Ruvolo V, Gadgil S, et al. HOX expression patterns identify a common signature for favorable AML. *Leukemia.* 2008;22(11):2041-2047.

49. Dou DR, Calvanese V, Sierra MI, et al. Medial HOXA genes demarcate haematopoietic stem cell fate during human development. *Nat Cell Biol.* 2016;18(6): 595-606.

50. Lawrence HJ, Christensen J, Fong S, et al. Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. *Blood.* 2005;106(12):3988-3994.

51. Deng C, Li Y, Liang S, et al. USF1 and hSET1A mediated epigenetic modifications regulate lineage differentiation and HoxB4 transcription. *PLoS Genet.* 2013;9(6): e1003524.

52. Alharbi RA, Pettengell R, Pandha HS, Morgan R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia.* 2013;27(5):1000-1008.

53. Luo H, Wang F, Zha J, et al. CTCF boundary remodels chromatin domain and drives aberrant *HOX* gene transcription in acute myeloid leukemia. *Blood.* 2018;132(8): 837-848.

54. Spencer DH, Young MA, Lamprecht TL, et al. Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. *Leukemia.* 2015; 29(6):1279-1289.

55. Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci USA.* 2014;111(52):18679-18684.

56. Mer AS, Lindberg J, Nilsson C, et al. Expression levels of long non-coding RNAs

are prognostic for AML outcome. *J Hematol Oncol.* 2018;11(1):52.

57. Wang Y, Krivtsov AV, Sinha AU, et al. The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. *Science.* 2010;327(5973):1650-1653.

58. Yeung J, Esposito MT, Gandillet A, et al. β-Catenin mediates the establishment and drug resistance of MLL leukemic stem cells. *Cancer Cell.* 2010;18(6):606-618.

59. Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell.* 2009;137(7): 1194-1211.

60. Qiu Y, Huang S. CTCF-mediated genome organization and leukemogenesis. *Leukemia.* 2020;34(9):2295-2304.

61. Fang C, Rao S, Crispino JD, Ntziachristos P. Determinants and role of chromatin organization in acute leukemia. *Leukemia.* 2020;34(10):2561-2575.

62. Katainen R, Dave K, Pitkänen E, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet.* 2015; 47(7):818-821.

63. Thol F, Bollin R, Gehlhaar M, et al. Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood.* 2014;123(6):914-920.

64. Chen L, Chen JY, Zhang X, et al. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol Cell.* 2017; 68(4):745-757.e5.

65. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell.* 2011;44(4): 667-678.

66. Chu C, Chang HYCIRP-MS. ChIRP-MS: RNA-directed proteomic discovery. *Methods Mol Biol.* 2018;1861:37-45.

67. Hacisuleyman E, Goff LA, Trapnell C, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol.* 2014;21(2):198-206.

68. Verma-Gaur J, Torkamani A, Schaffer L, Head SR, Schork NJ, Feeney AJ. Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. *Proc Natl Acad Sci USA.* 2012;109(42): 17004-17009.

69. Kajino T, Shimamura T, Gong S, et al. Divergent lncRNA MYMLR regulates MYC by eliciting DNA looping and promoter-enhancer interaction. *EMBO J.* 2019;38(17): e98441.

70. Pachera E, Assassi S, Salazar GA, et al. Long noncoding RNA H19X is a key mediator of TGF-β-driven fibrosis. *J Clin Invest.* 2020;130(9): 4888-4905.

QIU et al