In this article, Luo et al. discovered that HOTTIP IncRNA and CTCF/cohesin co-occupied a subset of the AML genome driven by HOTTIP aberration. HOTTIP regulates the CTCF boundary and TAD topology of β-catenin and its target loci by binding to complementary sequences to form R-loop structures that drive leukemic transcription programs and leukemogenesis.
**SUMMARY**

*HOTTIP*-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia

Hottip lncRNA is highly expressed in acute myeloid leukemia (AML) driven by MLL rearrangements or NPM1 mutations to mediate HOXA topologically associated domain (TAD) formation and drive aberrant transcription. However, the mechanism through which Hottip accesses CCCTC-binding factor (CTCF) chromatin boundaries and regulates CTCF-mediated genome topology remains unknown. Here, we show that Hottip directly interacts with and regulates a fraction of CTCF-binding sites (CBSs) in the AML genome by recruiting CTCF/cohesin complex and R-loop-associated regulators to form R-loops. Hottip-mediated R-loops reinforce the CTCF boundary and facilitate formation of TADs to drive gene transcription. Either deleting CBS or targeting RNase H to eliminate R-loops in the boundary CBS of β-catenin TAD impaired CTCF boundary activity, inhibited promoter/enhancer interactions, reduced β-catenin target expression, and mitigated leukemogenesis in xenograft mouse models with aberrant Hottip expression. Thus, Hottip-mediated R-loop formation directly reinforces CTCF chromatin boundary activity and TAD integrity to drive oncogene transcription and leukemia development.

**INTRODUCTION**

R-loops are DNA-RNA hybrid structures that occur naturally during DNA replication and gene transcription mediated by all three families of RNA polymerases (Skouri-Stathaki and Proudfoot, 2014; Aguilar and Garcia-Muse, 2012). Passive formation of transcription-dependent R-loops in promoters leaves the G-rich non-template DNA strand displaced, suggesting that R-loops can modulate promoter accessibility, RNA polymerase II (RNA Pol II) recruitment, transcriptional pausing, and RNA splicing (Chen et al., 2015, 2017; Sun et al., 2013; Boque-Sastre et al., 2015; Chakraborty et al., 2018). The auxin-regulated promoter loop (Apolo) long non-coding RNA (lncRNA) recognizes promoters and regulates transcription by forming R-loops (Ariel et al., 2020), suggesting a role for R-loops in lncRNA-mediated transcriptional regulation. However, increased unprogrammed R-loop formation due to mutation of the R-loop resolving protein, ERCC6L2, resulted in transcription-associated genome instability and bone marrow failure (Tummala et al., 2018). Thus, proper regulation of R-loop formation is critical for genome transcription and stability. Isolation of R-loop-associated complexes has identified DHX9, PARP1, LMNB1, DDX5, MCM3, SRSF1, and SUPT16H as R-loop associated proteins that may regulate R-loops.
A. Chromatin isolation by RNA Purification (ChIRP)

B. LC-MS/MS identification of the HOTTIP protein complexes in AML cells (partial list of proteins):

| Protein ID | Size (kDa) | Uni. No. | Function                          |
|-----------|------------|----------|-----------------------------------|
| SMC3      | 141        | 12       | CTCF/locatin                      |
| CTCF      | 130        | 7        |                                   |
| RAD21     | 71         | 7        |                                   |
| SMCSA     | 143        | 5        |                                   |
| LMNB1     | 66         | 12       |                                   |
| NPM1      | 33         | 6        | Nuclear matrix                    |
| SATB1     | 86         | 1        |                                   |
| LMNB2     | 70         | 1        |                                   |
| HNRNPA2B1 | 37         | 8        | RBP                              |
| HNRNP1L   | 64         | 6        |                                   |
| ASH2L     | 69         | 3        |                                   |
| WDR5      | 37         | 2        |                                   |
| KMT2A     | 320        | 1        | MLL1/DOT1L                       |
| RBBP5     | 59         | 1        |                                   |
| DOT1L     | 185        | 1        |                                   |
| RPA1      | 68         | 3        |                                    |
| DHX9      | 141        | 9        |                                    |
| PARP1     | 113        | 6        |                                    |
| SRSF1     | 28         | 4        |                                    |
| SUPT16H   | 120        | 3        |                                    |
| MCM3      | 91         | 5        |                                    |
| RPB3 (ROL2RC) | 31 | 5 | Transcription |
| LMO1      | 18         | 2        |                                    |
| RUNX1     | 49         | 1        | Hematopoietic TFs                |
| STAT5A    | 91         | 1        |                                    |

C. Metabolic pathways

D. ChIRP-WB:

E. N-terminal domain 11 Zinc fingers domain C-terminal domain

(legend on next page)
transcription, suggesting HOTTIP may contribute to oncogenic genome organization and leukemogenesis (Luo et al., 2019). However, it remains unclear whether and how HOTTIP aberration impacts CCCTC-binding factor (CTCF)-defined TADs to regulate AML genome topology and transcription.

The canonical Wnt pathway is integral to the self-renewal and long-term (LT) reconstitution capacity of HSCs (Reya et al., 2003; Perry et al., 2011; Kirstetter et al., 2006; Scheller et al., 2006; Luis et al., 2011). However, its core component, β-catenin is largely dispensable for adult HSC development (Cobas et al., 2004; Jeannot et al., 2008). In contrast, depletion of β-catenin in leukemia suppressed target gene expression, resulting in reduced leukemia incidence (Wang et al., 2010) and resensitization to GSK3β inhibitor treatment (Yeung et al., 2010). In particular, AML driven by MLLr- requires intact Wnt/β-catenin activity for leukemia stem cell (LSC) self-renewal (Wang et al., 2010; Yeung et al., 2010), suggesting that β-catenin is a therapeutic target for leukemia. Thus, elucidating the mechanisms regulating β-catenin in AML could lead to novel therapeutic strategies. Canonical Wnt/β-catenin transcription was activated by Hottip transgene-driven leukemia (Luo et al., 2019); therefore, this pathway may contribute to Hottip-driven AML. However, the mechanism by which HOTTIP modulates oncogenic β-catenin remains elusive.

CTCF is a master regulator of mammalian genome organization (Phillips and Corces, 2009; Rowley and Corces, 2018; Dowen et al., 2014; Tang et al., 2015). CTCF/cohesin complexes define TAD boundaries and constrain long-range interactions within TADs (Dixon et al., 2012; Phillips and Corces, 2009). Impaired CTCF boundaries perturb the associated TAD and alter gene expression programs, demonstrating that CTCF-derived TADs are structural and functional units (Narendra et al., 2015; de Wit et al., 2015; Guo et al., 2015; Lupiáñez et al., 2015; Li et al., 2020; Luo et al., 2018). Additionally, CTCF-binding sites (CBSs) located within TADs facilitate and stabilize enhancer/promoter interactions to control-lineage-specific transcription and cell identity (Ren et al., 2017). Intriguingly, CTCF-mediated enhancer/promoter interactions and genome organization depend on RNA-interacting domains within CTCF (Saldaña-Meyer et al., 2019; Hansen et al., 2019). However, the contributions of IncRNAs to these CTCF functions remain unknown.

Here, we demonstrate that HOTTIP IncRNA forms R-loop structures to localize CTCF at TAD boundaries of Wnt/β-catenin gene loci, which promotes TAD formation, enhancer/promoter interactions, and β-catenin transcriptional activation, resulting in AML transformation.

RESULTS

HOTTIP IncRNA associates with CTCF/cohesin complex and R-loop-associated proteins in AML cells

HOTTIP facilitates the posterior HOXA TAD to drive a homoeotic transcription program partly by binding to annotated CBSs (Luo et al., 2019), suggesting that HOTTIP can regulate CTCF boundary activity, perhaps by interacting with CTCF complexes. We performed chromatin isolation by RNA purification in conjunction with liquid chromatography tandem mass spectrometry (ChIRP-MS) to unbiasedly identify protein complexes associated with HOTTIP in MLLr+ MOLM13 cells (Figure 1A). Compared with control LacZ probes, HOTTIP-specific probes precipitated known HOTTIP-interacting MLL1/DOT1L complexes (Figure 1B; Luo et al., 2019), indicating the specificity of the ChIRP-enriched protein interactome. Within the HOTTIP-enriched proteome, proteins involved in chromatin structure and organization, the cohesin complex, transcription and RNA splicing, DNA repair, the nuclear matrix, DNA/RNA binding, and canonical Wnt signaling pathway were overrepresented (Figures 1B, 1C, and S1A). Notably, HOTTIP ChIRP-LC-MS/MS specifically identified CTCF, the cohesin complex, and proteins implicated in R-loop formation and regulation, e.g., RPA1 and DHX9, as HOTTIP-interacting partners (Figure 1B).

To confirm the specificity, we carried out ChIRP-western blot and verified that HOTTIP specifically interacts with CTCF, RAD21, stromal antigen 2 (SA2/STAG2), replication protein A1 (RPA1), DHX9, PARP1, SRSF1, LMNB1, DDX5, and MCM3, but not with negative control, HDAC1 (Figure 1D). Furthermore, RNA immunoprecipitation (RIP) assay in MOLM13 cells revealed that HOTTIP, but not HOTAIRM1, interacts with CTCF, RAD21, SA2, RPA1, PARP1, DDX5, LMNB1, SRSF1, and DHX9 (Figure S1B). As CTCF contains RNA-interacting domains (Saldaña-Meyer et al., 2019; Hansen et al., 2019), we performed an in vitro biotinylated RNA pull-down assay by incubating biotinylated HOTTIP with GST-CTCF fusions encompassing full-length and truncated mutant proteins. Biotinylated HOTTIP precipitated the RNA-binding zinc-finger domains of CTCF (Figure 1E), supporting a direct interaction between HOTTIP and CTCF. Thus, HOTTIP associates with CTCF/cohesin complexes and R-loop-associated proteins.
A

Global peaks (by ChIP-seq)

CTCF sites

-3K 0 +3K -3K 0 +3K

WT HOTTIP^+

\[ p = 0.12 \]

HOTTIP^+ cotagged promoters

HOTTIP^+ctagged intergenic

CTCF ChIP-seq

HOTTIP^+ctagged sites (by ChIP-seq)

-3K 0 +3K -3K 0 +3K

B

Motif Target P-value

CTCF 1e-64

HOTTIP 1e-57

E-box 1e-52

Myc 1e-46

RUNX1 1e-37

STAT5 1e-28

Max 1e-25

C

Motifs per bp per peak

Distance from HOTTIP/CTCF co-binding peaks (bp)

D

DRIP-seq:

\[ p = 0.23 \]

\[ p = 3.7e-08 \]

\[ p = 1.5e-07 \]

HOTTIP^+

WT

E

HOTTIP binding

CTCF/Cohesin co-occupied sites

2662 (29.3%)

27167

3528 (6.34%)

7675

Cohesin independent CTCF sites

HOTTIP binding

HOTTIP decreased

Decreased CTCF/cohesin bound sites

F

HOTTIP binding

HOTTIP decreased

Decreased CTCF/cohesin bound sites

G

\[ p = 2.1e-09 \]

\[ p = 3.4e-05 \]

\[ p = 5.7e-09 \]

H3K4me3

H3K27me3

HOTTIP binding

HOTTIP^+

CTCF/CTCF

cotagged regions

-3K 0 +3K -3K 0 +3K -3K 0 +3K -3K 0 +3K -3K 0 +3K
**HOTTIP directly regulates a subset of CTCF boundaries in the AML genome by forming R-loop structures**

Despite an interaction between HOTTIP and CTCF/cohesin complexes, the mechanism by which HOTTIP accesses and regulates the CTCF boundary remains unknown. Given the interaction of CTCF and HOTTIP (Figure 1), we posited that HOTTIP may regulate CTCF binding by binding to CBSSs. While HOTTIP did not affect global CTCF binding, it significantly reduced CTCF occupancy at HOTTIP and CTCF co-occupied sites, especially in promoters and intergenic regions (Figure 2A). Only 1,325 of 35,586 (3.72%) CTCF peaks in MOLM13 cells were co-occupied by HOTTIP (Figure S2A), suggesting that HOTTIP regulates CTCF boundary function of a subset of CBSSs. Motif analysis of HOTTIP/CTCF co-bound peaks identified the CTCF motif ($p = 1e-64$), a C-rich motif ($p = 1e-57$), and hematopoietic TF-binding motifs, including E-box, MYC, RUNX1, STAT5, and Max motifs ($p < 1e-25$), as top enriched motifs (Figure 2B). These motifs were not enriched in HOTTIP-independent CTCF peaks (Figure S2B; Table S1). The top C-rich motif identified was complementary to a partial G-quadruplex (G-4) sequence in the 5’ region of the HOTTIP IncRNA and defined as the HOTTIP-binding motif (Figure 2B). These motifs resided within 100 bp of the HOTTIP/CTCF co-bound peaks, indicating that this subset of CTCF peaks was enriched for E-box, MYC, MAX, and HOTTIP binding within the leukemic genome (Figure 2C).

R-loops are a mechanism by which RNAs regulate transcription (Qiu et al., 2021; Ariel et al., 2020; Skourti-Stathaki and Proudfoot, 2014; Sun et al., 2013; Dumelle and Jaffrey, 2017; Postepska-Igielska et al., 2015). Given the HOTTIP association with R-loop-associated proteins and CTCF/cohesin complexes (Figure 1), we determined whether HOTTIP forms DNA-RNA hybrid complexes to access CBSSs and regulate CTCF boundary activity. DNA-RNA immunoprecipitation sequencing (DRIP-seq/DRIPcseq) was performed to identify locations of R-loops formed across the genome (Chen et al., 2017) in WT and HOTTIP$^{-/-}$ MOLM13 cells. Both DRIP-seq and DRIPc-seq provided reproducible profiles of R-loops across the AML genome, especially in promoters, exons, and intergenic regions (Figures S2C and S2D). HOTTIP$^{-/-}$ greatly reduced R-loop structures at promoters and intergenic regions, but not at exonic, intronic, and UTR regions (Figures S2C and S2D). Although HOTTIP$^{-/-}$ did not affect global R-loop formation, HOTTIP$^{-/-}$ significantly reduced R-loops at HOTTIP$^{-/-}$-binding sites, especially where it is co-bound with CTCF (Figure 2D). Of 1,325 CTCF/HOTTIP co-bound sites, 812 (61.3%) had a significant decrease in CTCF binding in HOTTIP$^{-/-}$ cells ($\geq 1.5$-fold, $p \leq 0.05$; Figure S2E). Notably, 38.8% (315) of the CTCF-decreased, HOTTIP/CTCF co-bound sites had a concurrent decrease in R-loop formation in HOTTIP$^{-/-}$ cells ($\geq 1.5$-fold, $p \leq 0.05$; Figure S2E). These HOTTIP-dependent, R-loop-containing CTCF sites were predominantly located in promoter (20.43%) and intergenic regions (49.33%) (Figure S2F). These results suggest that HOTTIP forms R-loop structures at a subset of CBSSs to mediate CTCF binding, which is exemplified at the MDS1 and EVI1 complex loci (MECOM, MYC, and CTNNB1 loci, where the TAD boundaries exhibit HOTTIP-dependent R-loop formation and CTCF binding (Figures S2G–S2I).

Cohesin and CTCF coordinate interphase folding of the genome into 3D TAD domains through a loop extrusion mechanism, in which CTCF-bound CBSSs act as barriers to extrusion (Qiu and Huang, 2020; Sanborn et al., 2015; Fudenberg et al., 2016). Similar to CTCF, cohesin complex components were overrepresented in the HOTTIP ChIRP-LC-MS/MS (Figure 1). Therefore, we profiled cohesin (RAD21, SA1, and SA2) binding at HOTTIP-bound sites in WT and HOTTIP$^{-/-}$ cells. Of all HOTTIP-bound sites, only 6.34% were bound solely by CTCF, whereas 29.3% were co-occupied by CTCF and cohesin (RAD21) (Figure 2E). Furthermore, 67.2% of the CTCF/cohesin co-bound sites that decreased upon HOTTIP$^{-/-}$ also exhibited decreased HOTTIP binding (Figure 2F). Notably, HOTTIP$^{-/-}$ did not alter global CTCF (Figure 2A) or cohesin complex binding (Figure S2J). The recruitment of RAD21, SA1, and SA2 was significantly decreased at HOTTIP/CTCF co-bound sites upon HOTTIP$^{-/-}$ (Figure 2G). Consistently, H3K27me3 levels increased and H3K4me3 levels decreased in proximity of HOTTIP/CTCF co-bound sites upon HOTTIP$^{-/-}$ (Figure 2H). Together, these data indicate that HOTTIP forms R-loops to regulate CTCF/cohesin binding genome-wide and coordinate chromatin boundary function.

**HOTTIP regulates CTCF-defined TADs and transcription at canonical Wnt/β-catenin loci in AML genome**

HOTTIP regulates the canonical Wnt pathway in AML as HOTTIP$^{-/-}$ consistently reduced the expression levels of many Wnt/β-catenin targets and dysregulated the Wnt signaling pathway (Figures 3A and 3B). We reasoned that HOTTIP could be modulating the CTCF/cohesin-mediated genome organization at Wnt gene loci. Although HOTTIP$^{-/-}$ did not dramatically affect global TAD organization (Figure S3A), HOTTIP$^{-/-}$ led to a decrease of 42 TADs and increase of 19 TADs (Figure 3C).

**Figure 2. HOTTIP directly binds a subset of CTCF boundaries by formation of R-loops in AML genome**

(A) Heatmap of CTCF binding globally (top) and at HOTTIP/CTCF co-bound promoters (middle) and intergenic regions (bottom) in WT and HOTTIP$^{-/-}$ MOLM13 cells. p value calculated by Kolmogorov-Smirnov (K-S) test.

(B) Top enriched TF-binding motifs in HOTTIP/CTCF co-occupied peaks according to de novo motif analysis.

(C) Histogram of the distribution of HOTTIP and TF motifs within HOTTIP/CTCF co-occupied peaks.

(D) Heatmap of R-loop peaks identified by DRIP-seq globally (left), at HOTTIP-bound regions (middle) and at HOTTIP/CTCF co-bound regions (right) in WT and HOTTIP$^{-/-}$ MOLM13 cells. p value calculated by K-S test.

(E) Overlap of HOTTIP-binding peaks identified by ChIP-seq and CTCF/cohesin co-occupied sites (left) or cohesin independent CTCF sites (right) identified by CTCF and RAD21 ChIP-seq in MOLM13 cells.

(F) Overlap of total reduced HOTTIP peaks and all decreased CTCF/cohesin co-bound sites in the genome comparing WT and HOTTIP$^{-/-}$ MOLM13 cells.

(G) Heatmap of RAD21, SA1, and SA2 binding identified by ChIP-seq at HOTTIP/CTCF co-bound sites in WT and HOTTIP$^{-/-}$ MOLM13 cells. p value calculated by K-S test.

(H) Heatmap of H3K27me3 and H3K4me3 profiles identified by ChIP-seq at HOTTIP/CTCF co-bound sites in WT and HOTTIP$^{-/-}$ MOLM13 cells.
The decreased TADs encompassed 303 genes primarily involved in transcription regulation, Wnt signaling, HOX gene regulation, AML, and cell-cycle processes (Figure 3D). Notably, HOTTIP−/− impaired TAD topology in Wnt target loci including Ctnnb1, Myc, and Mecom, without affecting neighboring TADs (Figures 3E, S3B, and S3D). The Ctnnb1 TAD is flanked by an upstream (CBS-u2) and downstream (CBS-Dn) boundary and divided by the CTCF-bound Ctnnb1 promoter into two sub-TADs, both of which are impaired in HOTTIP−/− cells (Figure 3E). Concurrent with the structural alterations, chromatin accessibility was reduced in the promoters of these target genes (Figures 3F, S3C, and S3E). Next-generation (NG) capture-C of the Ctnnb1, Myc, and Mecom promoters demonstrated that TAD disruption influences promoter-looping at these loci. Indeed, HOTTIP−/− led to markedly decreased long-range interactions between the Ctnnb1 promoter and both upstream and downstream chromatin regions (Figure 3G) and the interactions of both Myc and Mecom promoters with their downstream chromatin regulatory regions (Figures 3H and S3F). Similar results were observed upon shRNA-mediated knockdown of HOTTIP (Figures S3G–S3L), indicating that HOTTIP is directly responsible for the perturbed genome topology. Collectively, HOTTIP regulates β-catenin and its target gene TADs, enhancer/promoter interactions, and transcription.

**Hottip overexpression reinforces CTCF boundaries and enhances Wnt/β-catenin TADs**

To complement our findings in HOTTIP−/− cells, we explored Hottip-dependent, CTCF-mediated genome topology using Hottip transgenic (Hottip-Tg) mice (Luo et al., 2019). Transcriptional profiling of Lin−/−c−Kit+ (LSK) cells from the bone marrow of these mice demonstrated elevated levels of the Hoxa9-a13 genes and several Wnt/β-catenin targets, including Ctnnb1, Myc, Axin, Kit, Cdx4, Nonog, Cdx4, Met, Twist, Runx1, and Wnt3/Sa (Figure 4A). Indeed, one of the pathways affected by Hottip aberration was the Wnt signaling pathway (Figures 4B and S4A).

We examined whether Hottip modulates the Wnt/β-catenin pathway by facilitating CTCF boundaries and TAD topology. Hottip activation resulted in significant enrichment of Hottip binding to the CTCF core motifs in Lin+c−Kit+ (LK) cells (Figure S4B). Although Hottip overexpression did not alter global or Hottip-independent CTCF binding (Figure S4C), activation of Hottip in LK cells significantly enhanced Hottip and CTCF binding at Hottip/CTCF co-occupied CBSs (Figure 4C), including at the TAD boundaries of the Ctnnb1, Myc, and Hoxa loci (Figures 4D, S4D, and S4E). Thus, Hottip facilitates CTCF binding at TAD boundaries of Wnt target loci, suggesting that it may also remodel CTCF-defined TADs. Indeed, although Hottip activation did not alter global TAD formation (Figure S4F), it increased 49 TADs and decreased 26 TADs, encompassing 287 and 125 genes, respectively (Figure S4G). The TADs in the Ctnnb1 and Myc loci were among the increased TADs (Figures 4E, 4F, and S4H), which was consistent with changes in promoter accessibility (Figures 4G and S4I), transcription (Figures 4A and 4B), and enhanced Hottip and CTCF binding at Wnt/β-catenin loci (Figures 4D and S4D).

The hematopoietic stem and progenitor cell (HS/PC) population of origin for Hottip-driven leukemia is unknown. Therefore, single-cell RNA-seq (scRNA-seq) was used to determine the HS/PC subpopulations impacted by Hottip activation. Hottip-Tg mice exhibited expansion of the LT-HSC, short-term (ST)-HSC, and MPP1 populations (Figures 4H–4K and S4J) with a myeloid-biased trajectory (Figure 4I). Oncogenic Hoxa9 and Myc levels were specifically elevated in these expanded HSC subsets compared with more differentiated progenitors and lineages (Figures 4J, 4K, and S4K). Furthermore, the most significantly altered pathways in HSCs included the Wnt signaling pathway and pathways controlling HSC proliferation and hematopoiesis (Figures 4L and S4L). Thus, Hottip aberration drives leukemic transcription in HSC subpopulations, in part by regulating the CTCF-defined TADs of Hoxa and Wnt pathway genes.

**HOTTIP binds to the Ctnnb1 TAD boundary and forms R-loops by a base-pairing mechanism to regulate CTCF boundary function**

DNA-RNA triple-strand hybrid complexes known as R-loops and triple helices play critical roles in transcriptional regulation (Skourtis-Stathaki and Proudfoot, 2014; Sun et al., 2013; Dumelie and Jaffrey, 2017; Postepska-Igielska et al., 2015). While HOTTIP contains a G-4 motif predicted to form R-loops by the QmRLFS-finder algorithm (Figures 5A and 5B), it is unknown whether HOTTIP and the β-catenin CBS sequences can form these structures. Using an in vitro electrophoretic mobility shift assay (EMSA), we found that a Cy3-labeled DNA probe that encompassed CBS-u2 formed a DNA-RNA triple-strand structure with a Cy5-labeled HOTTIP RNA probe (Figure 5C). In contrast, probes encompassing HOTTIP-independent CBSs, CBS4/5 of HOXA and promoter CBS of ACTB, do not complex with HOTTIP (Figure S5A). The DNA-RNA complex was sensitive to RNase H, which hydrolyzes RNA in RNA/DNA hybrids, and DNase I, but not RNase A, which digests single-strand RNA (Figure 5C), indicating that HOTTIP and CBS-u2 form an R-loop structure, not a DNA-RNA triplex. Indeed, the DNA-RNA hybrid complex was super-shifted by an S9.6 antibody that specifically recognizes R-loop structures.

![Figure 3. HOTTIP regulates CTCF-defined TADs and transcription at canonical Wnt loci in AML genome](https://example.com/image1)

(A) Heatmap of ≥2-fold downregulated genes in MOLM13 cells upon HOTTIP−/− as determined by RNA-seq.
(B) GSEA of downregulated genes after HOTTIP−/−.
(C) Overlap of TADs identified by Hi-C in WT and HOTTIP−/− MOLM13 cells. The domain score of an altered TAD was normalized (quantile-normalization) by subtracting the mean of all TAD Hi-C signals. ANOVA was used to identify significantly altered TADs (Bonferroni-corrected p value < 0.05).
(D) GO analysis of genes encompassed by the decreased TADs upon HOTTIP−/−.
(E) Hi-C interaction map at the Ctnnb1 locus comparing WT and HOTTIP−/− MOLM13 cells.
(F) ATAC-seq analysis of Ctnnb1 in WT and HOTTIP−/− MOLM13 cells.
(G and H) NG Capture-C analysis of Ctnnb1 (G) or MYC (H) promoter interactions, CTCF ChIP-seq and HOTTIP ChIP-seq in WT and HOTTIP−/− MOLM13 cells. Solid purple and dashed red lines indicate unchanged and reduced interactions, respectively.
(Figure 5C). R-loop formation was also observed between HOTTIP and the 3' CBS of MYC (Figure S5B).

At the CTCF/HOTTIP co-bound sites in the CTNNB1 and MYC loci, the C-rich HOTTIP-binding motif that is complementary to the G-4 sequence in the 5' terminus of HOTTIP is overlapping or in proximity to the CBSs (Figures 5B and S5C). Mutation of the G-4 sequence in HOTTIP abolished its ability to form an R-loop with CBS-u2 (Figure S5A), indicating that HOTTIP binds and forms an R-loop with CBS-u2 through complementary sequences. Furthermore, RNA isolation by DNA purification (RIDP) of the template and non-template of CBS-u2 demonstrated that HOTTIP only forms an RNase H-sensitive R-loop with the template strand, with respect to CTNNB1 transcription, of CBS-u2 (Figure 5D).

Thus, HOTTIP accesses CTCF-defined TAD boundary CBSs to form R-loops in a base-pairing-dependent mechanism.

We next examined whether HOTTIP+/− reduces R-loop formation and cohesin complex binding at the CTCF boundaries in the CTNNB1, MYC, and MECOM loci in MOLM13 cells. HOTTIP+/− not only reduced HOTTIP binding and R-loop formation at these TAD boundaries but also disrupted CTCF and cohesin complex binding (Figures 5E and S5D–S5F). Notably, R-loops at these CBSs are not a result of nascent transcription, as there is no active transcription indicated by global run-on sequencing (GRO-seq) or CTCF ChIP-seq and transcription of non-target genes, LDB1 and LMO2, and HOTTIP unaltered (Figures 6C and S6C). CBS-u2+/− also impaired MOLM13 cell proliferation (Figure 6D). Furthermore, CBS-u2+/− reduced HOTTIP binding and R-loop formation at CBS-u2, CBS-Dn, and the CTNNB1 promoter, without affecting control regions, CBS-u1 or MYC (Figures 6E and 6F). Consequently, CBS-u2+/− impaired the CBS-u2 and CBS-Dn interaction (Figure S6D) and disrupted promoter long-range interactions within two sub-TADs in the CTNNB1 locus (Figure 6G), leading to decreased H3K4me3 and increased H3K27me3 within the CTNNB1 TAD (Figures 6G and S6E–S6G). As a control, CBS-u2+/− did not affect global H3K4me3 and H3K27me3 levels (Figures S6H and S6I). Notably, CBS-u1−/− did not have any effect on these phenotypes (Figures 6B–6F). Thus, CBS-u2 boundary controls the integrity of the CTNNB1 TAD.

Given the requirement for β-catenin in LSC self-renewal in MLLr+ AML (Wang et al., 2010; Yeung et al., 2010), we examined the role of CBS-u2 in leukemogenesis by transplanting WT, CTNNB1+/− or CBS-u2+/− MOLM13 cells into sub-lethally irradiated NOD-scid IL2Rγnull (NSG) mice. Mice receiving WT cells died 18–21 days after transplantation, while those receiving CTNNB1+/− (positive control) or CBS-u2+/− cells survived significantly longer (21–34 days) (Figure 6H). Consistently, WT recipient mice had significantly higher human CD45+ cell chimera in the BM and PB after 15 days than mice receiving CTNNB1+/− or CBS-u2+/− cells (Figures 6I and S6J). Thus, the CBS-u2 TAD boundary is critical for oncogenic genome topology and transcription in AML.

To extend our findings to AML patient samples, we deleted CBS-u2 in primary AML cells from an MLLr+ (#LPP4) or an NPM1C+/−;FLT-ITD+ (#921) patient with elevated expression of HOTTIP by CRISPR-Cas9 (Figure S6K). Upon transplantation of control or CBS-u2+/− primary AML cells into NSG mice, mice receiving control MLLr+ (#LPP4) or NPM1C+/−;FLT-ITD+ (#921) cells died around 30 days after transplantation, whereas mice receiving CBS-u2+/− AML cells survived up to 49 days (Figure 6J). Consistently, CBS-u2+/− dramatically decreased the hCD45+ cell chimera in the BM 24 days post-transplantation (Figures 6K and S6L). Together, these data suggest that the CBS-u2 boundary is required to maintain the CTNNB1 TAD and potentiate leukemogenesis in HOTTIP-driven AML.

Figure 4. Hottip activation perturbs β-catenin and its target TADs leading to aberrant HSC activities

(A) Heatmap of ≥2-fold upregulated genes in BM LSK cells from Hottip-Tg mice.

(B) GSEA of upregulated genes in LSK cells upon Hottip activation.

(C) Heatmap of CTCF binding, from ChiP-seq, and Hottip binding, from ChiRP-seq, at CTCF/Hottip co-occupied sites in BM LK cells from WT and Hottip-Tg mice.

(D) CTCF ChiP-seq and Hottip ChiRP-seq binding profiles at the Cttnb1 locus in WT and Hottip-Tg LK cells.

(E) Hi-C interaction maps at the Cttnb1 locus in WT and Hottip-Tg BM LK cells. CTCF-bound TAD boundaries indicated by red arrows.

(F) Overlap of WT and Hottip-Tg Hi-C signals from (E).

(G) ATAC-seq analysis of Cttnb1 in WT and Hottip-Tg BM LSKs.

(H) tSNE visualization of BM LK cell subsets from Hottip-Tg (red) and WT (blue) mice by scRNA-seq. LT-HSC, ST-HSC, and MPP populations encompassed by blue circle.

(i) Trajectory inference branches/clusters were generated based on the expression levels of lineage-associated genes in cell clusters (left) from WT and Hottip-Tg BM LK cells. Sub-population cell density analysis (right) correlated with the enriched cell number of each population. Higher cell densities shown in dark red.

(j) The levels of Myc and Hoxa9 in each cell subset along HSC to MEP differentiation in WT and Hottip-Tg BM LK cells by scRNA-seq. The FDR-corrected p value ≤0.05 by binomial and hypergeometric test.

(k) Relative cell numbers in each cell subset along HSC to MEP differentiation in WT and Hottip-Tg BM LK cells by scRNA-seq. The FDR-corrected p value ≤0.05 by binomial and hypergeometric test.

(l) GO analysis of upregulated genes in LT- and ST-HSC populations upon Hottip activation by scRNA-seq.
**HOTTIP-mediated R-loop formation directly contributes to CTCF boundary activity and TAD integrity**

To determine the role for the HOTTIP-mediated R-loop at the CBS-u2 boundary in leukemogenesis, we targeted dCas9-RNase H, which can hydrolyze the local R-loop, or dCas9-RNase H^{Δ31N}, a control catalytically dead mutant, to the CBS-u2 boundary (Figures 7A and S7A–S7D). Importantly, the sgRNA was targeted downstream of the CTCF core motif to avoid direct competition with the CTCF protein (Figure 7A). CBS-u2^{RH-Mut}, but not CBS-u2^{RH-WT}, reduced CTNNB1 levels and β-catenin target gene expression without affecting control genes, LDB1 and LMO2 (Figure 7B). Additionally, CBS-u2^{RH-WT} impaired cell proliferation (Figure 7C). Hydrolyzing the R-loop reduced cotranscriptional and R-loop formation in the CBS-u2 boundary and, to a lesser extent, the CBS-Dn boundary and CTNNB1 promoter (Figures 7D–7F). CBS-u2^{RH-Mut} did not significantly affect CTCF or HOTTIP binding or R-loop formation anywhere in the CTNNB1 locus (Figures 7D–7F). Importantly, while CBS-u2^{RH-WT} did not affect global H3K4me3 and H3K27me3 levels, H3K4me3 levels were decreased and H3K27me3 levels were elevated within the CTNNB1 TAD (Figures 7G, S6H, S6I, S7E, and S7F), indicating a loss of TAD boundary activity and spread of repressive H3K27me3 into the active CTNNB1 TAD. Supporting this, CBS-u2^{RH-WT} impaired long-range interactions between the CTNNB1 promoter and upstream or downstream regulatory regions (Figures 7G and S7G). Critically, the transcriptional and proliferative defects caused by CBS-u2^{RH-WT} can largely be rescued by exogenous β-catenin expression (Figures 7B and 7C) despite R-loop formation and HOTTIP/CTCF binding in the CTNNB1 locus remaining impaired (Figures 7D–7F). Thus, the R-loop at the CBS-u2 boundary specifically controls the CTNNB1 TAD to regulate β-catenin expression, which is essential for Wnt pathway activation and AML cell proliferation.

We next assessed the effect of CBS-u2 R-loop hydrolyzation on AML leukemogenesis by transplanting HOTTIP-driven AML cells harboring either MLLr^{+} (MOLM13) or NPM1^{C+} (OCI-AML3) mutations (Luo et al., 2019). First, WT, CBS-u2^{RH-WT} or CBS-u2^{RH-Mut} MOLM13 cells were transplanted into NBSGW mice. WT and CBS-u2^{RH-Mut} MOLM13 transplanted mice died 28–40 days post-transplantation; however, the CBS-u2^{RH-WT} transplanted mice prolonged survival significantly. The mice began to die from disease on day 40, with one surviving past 47 days post-transplantation without obvious symptoms (Figure 7H) with markedly reduced hCD45⁺ chimerism in the BM after 33 days compared with the mice receiving WT or CBS-u2^{RH-Mut} cells (Figure S7H). Likewise, CBS-u2^{RH-WT} or CBS-u2^{RH-Mut} OCI-AML3 cells were transplanted into NSG mice. While CBS-u2^{RH-Mut} OCI-AML3 transplanted mice died 24 days post-transplantation, CBS-u2^{RH-WT} transplanted mice survived to 29 days and exhibited reduced hCD45⁺ engraftment (Figures 7I and S7I). Importantly, targeting CBS-u2^{RH-WT}, but not CBS-u2^{RH-Mut}, in OCI-AML3 cells inhibited CTNNB1 expression and their LT proliferative potential (Figures S7J–S7L). Together, these data provide evidence that HOTTIP-mediated R-loop formation in CTCF-defined TAD boundaries controls the boundary activity, genome topology, and transcriptional programs required for AML leukemogenesis.

**DISCUSSION**

The mechanism underlying IncRNA-driven genome organization

R-loops have been widely implicated in transcription, DNA damage, and genome instability (De Magis et al., 2019; Aguilera and García-Muse, 2012). R-loops were induced upon mutation of pre-mRNA-spooling factors to contribute to DNA damage in myelodysplastic syndrome (Chen et al., 2018; Singh et al., 2020; Nguyen et al., 2018). R-loops are thought to be universal products of transcription events in the mammalian genome (Sanz et al., 2016). Persistence of R-loops confer vulnerability to genomic insults leading to genome instability (Aguilera and García-Muse, 2012). However, R-loops were also demonstrated to block promoter methylation and therefore promote transcription (Grunseich et al., 2018), suggesting an active role in transcriptional activation. Several recent studies also correlated R-loop formation with transcriptional activation and RNA Pol II pausing (Boque-Sastre et al., 2015; Chen et al., 2017). Interestingly, APOLO decays polycomb proteins from promoters for gene activation by forming R-loops in Arabidopsis (Ariel et al., 2020). Yet, it remains unknown if R-loops are the direct cause or a consequence of transcription events and how R-loops contribute to genome regulation and organization. Combining HOTTIP KO and dCas9-mediated targeting RNase H to a specific CTCF boundary, we showed that HOTTIP-dependent R-loop formation is required for establishing CTCF-mediated TAD boundaries to maintain TAD integrity and drive β-catenin and its target gene expression that functionally promotes AML leukemogenesis. Consistently, two recent studies suggested that...
RNA-interacting domains of CTCF indeed play a critical role in CTCF-mediated long-range interactions (Saldaña-Meyer et al., 2019; Hansen et al., 2019). Thus, it is conceivable that IncRNAs and their associated R-loops may widely contribute to genome organization and transcriptional regulation.

**Regulation of CTCF-mediated TAD boundary activity**

The role of CTCF in mammalian genome regulation is dependent on CTCF’s ability to homodimerize with itself and to heterodimerize with other proteins including the cohesin complex (Zlatanova and Caiafa, 2009; Yusufzai et al., 2004). Although CTCF mostly interacts with the same DNA sites in different cell types, it often functions as a chromatin barrier in one cell type but does not in the other (Cuddapah et al., 2009). Thus, the remaining question is how boundary activities and TAD formation are regulated across the genome to participate in cell-type-specific fashion. Given that many IncRNAs are associated with CTCF (Kung et al., 2015) and expressed in a hematopoietic/leukemic-cell-specific manner (Qiu et al., 2021), it is possible that RNA is required for proper chromatin boundary function. Additional evidence supporting this notion stemmed from a report by Feldsenfield’s group showing that DEAD-box RNA helicase p68 and steroid receptor RNA activator (SRA) interact with both CTCF and cohesin to stabilize their interaction (Yao et al., 2010). However, little is known about the mode of IncRNA action on CTCF-mediated TAD boundaries. In light of the discovery that HOTTIP interacts with CTCF boundaries to form R-loops that maintain the integrity of the TAD boundary and TAD topology, these results imply that HOTTIP-mediated R-loops indeed play a regulatory role in controlling CTCF boundary activities and TAD integrity, particularly in the AML genome. It remains to be determined whether the same R-loop mechanism is employed by other IncRNAs to modulate CTCF-mediated genome organization for gene regulation. Furthermore, it is also important to determine whether IncRNA-driven R-loops at CTCF boundaries can be properly regulated to avoid persistent R-loop-conferring vulnerability to genomic insults. Association of R-loop sensors and regulators with HOTTIP (Figure 1) suggests that HOTTIP or other IncRNAs may regulate CTCF boundary-mediated genome organization and stability by modulating R-loop formation.

**Mechanism of HOTTIP-mediated R-loops in boundary activity and TAD integrity**

R-loops were thought to passively associate with enhancer and insulator states due to transcription activity (Sanz et al., 2016). Here, we showed that mechanistically HOTTIP interacts with CTCF/cohesin complex and R-loop-associated proteins by binding to CBSs to form R-loops via sequence complementarity, which regulates TAD boundary activity and maintains TAD integrity for transcription. Our results suggest two regulatory steps exerted by HOTTIP to control CTCF boundary activities. First, HOTTIP binds to CBSs to induce R-loop structure in boundary sites; next, IncRNA-dependent R-loops at TAD CBSs are required for stabilizing CTCF and cohesin binding, as loss of HOTTIP or disruption of R-loop at specific CBS impairs chromatin boundary activity and TAD formation. Given that CTCF contains RNA-binding domains (Saldaña-Meyer et al., 2019; Hansen et al., 2019), it is also possible that HOTTIP is brought to the particular TAD CBSs by CTCF and then the recruited HOTTIP in turn stabilizes CTCF and cohesin complex binding via formation of R-loop. Mechanistically, HOTTIP employs a partial G-4 motif located at its 5’ terminus to recognize the C-rich HOTTIP-binding motif at the HOTTIP/CTCF co-occupied boundaries by sequence complementarity.

The **HOTTIP-HOXA9-β-catenin axis in HSC/HPC function and regulation**

Aberrant self-renewal is an indispensable feature of LSCs that sustain the malignant phenotypes (Deininger et al., 2017). Although Hoxa9 and β-catenin functionally complement each other in mediating self-renewal of MLLr-AML stem cells derived from early myeloid progenitors but not HSCs (Siriboonpiputtana et al., 2017), a novel cross-talk between β-catenin and Hoxa9 is required for self-renewal of MLLr-AML stem cells in both mouse (Siriboonpiputtana et al., 2017) and human (Zeisig et al., 2021) AML of HSC origins. Intriguingly, our studies also revealed that the co-regulation of the posterior HOXA locus and the canonical Wnt pathway by HOTTIP-mediated genome topology coordinates HS/PC cluster formation, which provides a molecular link between these two previously unrelated key self-renewal pathways (Luo et al., 2019). In support of this notion, we showed that HOTTIP binds and forms R-loops specifically in the TAD.

Figure 6. The CBS-u2 boundary at the CTNNB1 locus is critical to maintain TAD structure, transcription, and β-catenin-driven leukogenesis

(A) Schematic of the CTNNB1 locus showing the locations of CBSs, sub-TADs, and TAD.
(B) CTCF ChiP-qPCR analysis of the indicated sites in WT, CBS-u1+/− and CBS-u2+/− MOLM13 cells.
(C) RT-qPCR analysis of the indicated transcripts in WT, CBS-u1+/−, and CBS-u2+/− MOLM13 cells.
(D) Proliferation of WT, CBS-u1+/−, and CBS-u2+/− MOLM13 cells.
(E) HOTTIP ChiRIP-qPCR analysis of the indicated sites in WT, CBS-u1+/−, and CBS-u2+/− MOLM13 cells.
(F) DRIP-qPCR analysis of the indicated sites in WT, CBS-u1+/−, and CBS-u2+/− MOLM13 cells.
(G) NG Capture-C analysis of CTNNB1 promoter interactions in WT and CBS-u2+/− MOLM13 cells. Solid purple and dashed red lines indicate unchanged and reduced interactions, respectively. Capture-C data were aligned with CTCF, H3K4me3, and H3K27me3 Chip-seq profiles in the CTNNB1 locus in WT, HOTTIP+/−, and CBS-u2+/− MOLM13 cells.
(H) Kaplan-Meier survival curves of NSG mice transplanted with WT, CTNNB1+/−, and CBS-u2+/− MOLM13 cells. n = 5.
(I) HCD45+ cells chimera in the BM and PB of NSG mice transplanted with WT, CTNNB1+/−, and CBS-u2+/− MOLM13 cells. n = 3.
(J) Kaplan-Meier survival curves of NSG mice transplanted with WT or CBS-u2+/− primary AML cells carrying MLL* (LPP4) or NPM1(C)-FLT3-ITD* (#974) mutations. n = 4.
(K) HCD45+ cell chimerism in the BM of NSG mice transplanted with WT or CBS-u2+/− primary AML cells. Data in (B)-(F) and (H)-(K) are presented as mean ± SD. *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001.
boundaries to activate transcription of β-catenin and Hoxa9, which are essential for self-renewal of MLlR+ AML LSCs (Wang et al., 2010; Yeung et al., 2010). Dysregulation of β-catenin target oncogene, EVI1 (encoded by MECOM gene) (Manachai et al., 2017) is also a signature of HSC-derived MLlR+ AML and serves as a clinical prognosis factor associated with inferior survival in AML patients carrying MLlR+ (Gröschel et al., 2013; Noordermeer et al., 2011), which apparently is also topologically modulated by HOTTIP and its associated R-loop formation (Figures S2G and S3D–S3F). It is conceivable that R-loops add another layer of fine-tuned regulation of AML genome organization in the cells carrying MLLR+ or NPM1C+, perhaps initiated by regulatory lncRNAs. In this way, regulatory R-loops can be directed in a developmental stage or lineage-specific fashion by lncRNA.

Interestingly, motif analysis of the HOTTIP/CTCF co-occupied peaks in AML cells also revealed that HOTTIP-dependent CBSs were highly co-enriched with E-box, MYC/MAX, RUNX1, and STAT5-binding motifs. Since cell-type-specific TF-binding sites are often in proximity to CBSs in the genome (Qi et al., 2021; Park et al., 2016), the likely scenario is that the proximity of these hematopoietic TF-binding sites cooperates with HOTTIP and CTCF/cohesin complex to reorganize or stabilize the HOTTIP-dependent CTCF TAD boundary to drive hematopoietic/leukemic-cell-specific transcription. Besides its role as biomarker, IncRNA-mediated R-loops also play important regulatory roles in transcription and AML genome organization. Thus, R-loops are potential druggable targets for anti-leukemic therapy (Angelibello et al., 2018).

Limitations of the study
In this study, we demonstrated that HOTTIP lncRNA regulates CTCF chromatin boundaries via formation of R-loops at the β-catenin target gene TAD boundaries to drive the leukemic transcription program in MLLR- or NPM1C-mutated AML. However, AML is a heterogeneous disease with different genetic and epigenetic alterations. Thus, further exploration of the role of HOTTIP-mediated R-loops in the regulation of genome organization in other HOTTIP aberration-driven leukemias could further strengthen the relevance of this mechanism in leukemogenesis. Since HOTTIP interacts with the CTCF/cohesin complex (Figure 1) and CTCF possesses putative RNA-interacting domains (Saldaña-Meyer et al., 2019; Hansen et al., 2019), whether other lncRNAs or architectural RNAs are generally required for CTCF-mediated boundary function and genome organization remains unexplored and warrants further investigation. In terms of dCas9-RNase H-mediated site-specific hydrolysis of R-loop, the targeted dCas9-RNase H reduced 30%–50% of R-loops and HOTTIP/CTCF binding at the sites but did not eliminate them (Figure 7). The target sites of which sgRNA localizes the fusion protein could be more carefully titrated to efficiently hydrolyze R-loops while avoiding direct competition with CTCF binding. Such an approach would help to address the direct mechanisms underlying R-loop-driven regulation of gene transcription, splicing, DNA replication, as well as DNA damage and repair.

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - AML patient samples
  - Cell lines
  - Hottip transgenic (Tg) mouse model
- **METHOD DETAILS**
  - Generation of CRISPR-Cas9 mediated knockout AML cells and WT or catalytically dead dCas9-RNaseH AML cells
  - Hematopoietic stem and progenitor cell (HSPCs) sorting
  - In vitro pull-down assay with biotinylated RNA
  - RNA isolation, quantitative RT-PCR, and RNA-sequencing (RNA-seq)
  - Chromatin immunoprecipitation (ChIP) assay
  - Chromatin Isolation by RNA Immunoprecipitation (ChIRP), ChIRP-WB, and ChIRP-LC-MS/MS assays
  - RNA Immunoprecipitation (RIP)
  - DNA:RNA immunoprecipitation sequencing (DRIP-seq) assay
  - ChIP-seq, ChIRP-seq, and DRIP-seq data analysis

Figure 7. HOTTIP-mediated R-loop is required for maintaining CTCF boundary, TAD integrity, and AML pathogenesis
(A) Schematic of the CTCFNB1 locus. Shown is the CBS-u2 sequence with CTCF (yellow) and HOTTIP (green) motifs and the sgRNA target site (red) indicated.
(B) RT-qPCR analysis of β-catenin and its target gene expression upon expression of dCas9-RNase H or dCas9-RNase H(320R)E with or without CBS-u2-targeted sgRNA or exogenous β-catenin expression in MOLM13 cells.
(C) Proliferation of WT, CBS-u2HWT, CBS-u2HMut, and β-catenin-rescued CBS-u2HWT MOLM13 cells.
(D) CTCF ChIP-qPCR analysis of the indicated sites in WT, CBS-u2HWT, CBS-u2HMut, and β-catenin-rescued CBS-u2HWT MOLM13 cells.
(E) HOTTIP ChIRP qPCR analysis of the indicated sites in WT, CBS-u2HWT, CBS-u2HMut, and β-catenin-rescued CBS-u2HWT MOLM13 cells.
(F) DRIP-qPCR analysis of the indicated sites in WT, CBS-u2HWT, CBS-u2HMut, and β-catenin-rescued CBS-u2HWT MOLM13 cells.
(G) NG Capture-C analysis of CTCFNB1 promoter interactions upon expression of CBS-u2HWT. Solid purple and dashed red lines indicate unchanged and reduced interactions, respectively. Capture-C data were aligned with CTCF, H3K4me3, and H3K27me3 ChIP-seq profiles in the CTCFNB1 locus in WT, HOTTIP+/−, and CBS-u2HWT MOLM13 cells.
(H) Kaplan-Meier survival curves of NBSGW mice transplanted with WT, CBS-u2HMut, or CBS-u2HWT MOLM13 cells.
(I) Kaplan-Meier survival curves of NSG mice transplanted with CBS-u2HWT or CBS-u2HMut OCI-AML3 cells. Mice were sacrificed when they were paralyzed due to the disease.

Data in (B)–(F) and (H)–(I) are presented as mean ± SD. *p < 0.05; **p < 0.01; ***p < 0.001.
RNA isolation by DNA Purification (RIDP)

Transposase-Accessible Chromatin using sequencing (ATAC-seq) assay

ATAC-seq analysis

Chromosome conformation capture (3C) assay

Next generation (NG)-Capture-C sequencing assay

NG-Capture-C sequencing analysis

Xenotransplantation of human leukemic cells and Patient-Derived Xenografts (PDX)

The Next-Gen Chromosome Conformation Capture (Hi-C) Assay

Hi-C sequencing data analysis

In vitro R-loop structure detection and gel shift assay

Global Run-On sequencing (GRO-seq)

Analysis of GRO-seq data

Single cell RNA-seq analysis

QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2022.01.014.

ACKNOWLEDGMENTS

We thank the Penn State College of Medicine Genome Science Facility for sequencing and the UTHSA and the Penn State Hershey College of Medicine Flow cytometry Cores. This work was supported by grants from the National Institutes of Health (S.H. and M.X., R01CA260729 and R01HL141950; S.H., R01CA204044; M.X., R01CA172408, R01HL145883; Y.Q., R01HL144712), CRUK program grant 29213 and Blood Cancer UK program continuity grant R01CA204044; M.X., R01CA172408, R01HL145883; Y.Q., R01HL144712), and the Four Diamonds Fund (S.H.).

REFERENCES

Agrotis, A., Pengo, N., Burden, J.J., and Ketteler, R. (2019). Redundancy of human ATG4 protease isoforms in autophagy and LC3/GABARAP processing revealed in cells. Autophagy 15, 976–987.

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell 46, 115–124.

Angelibello, A.J., Chen, J.J., Childs-Disney, J.L., Zhang, P., Wang, Z.F., and Disney, M.D. (2018). Using genome sequence to enable the design of medicines and chemical probes. Chem. Rev. 118, 1599–1663.

Ariel, F., Lucero, L., Christ, A., Mammary, F.M., Jugu, T., Veluchamy, A., Mariappan, K., Latrasse, D., Blein, T., Liu, C., et al. (2020). R-loop mediated trans action of the APOL1 long noncoding RNA. Mol. Cell 77, 1055–1065.e4.

Barbier, E., Hill, C., Quesnel-Vallières, M., Zucco, A.J., Barash, Y., and Gardini, A. (2020). Rapid and scalable profiling of nascent RNA with fastGRO. Cell Rep 33, 108373.

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Gui, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc. Natl. Acad. Sci. USA 112, 5785–5790.

Buckley, A., Gilbert, N., Marenduzzo, D., and Brackley, C.A. (2019). capC-MAP software for analysis of Capture-C data. Bioinformatics 35, 4773–4775.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 109, 21.29.1–21.29.9.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420.

Chakraborty, P., Huang, J.T.J., and Hiom, K. (2018). DHX9 helicase promotes R-loop formation in cells impaired RNA splicing. Nat. Commun. 9, 4348.

Chen, L., Cher, J.Y., Huang, Y.J., Gu, Y., Qiu, J., Gian, H., Shao, C., Zhang, X., Hu, J., Li, H., et al. (2018). The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. Mol. Cell 69, 412–425.e6.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. Mol. Cell 68, 745–757.e5.

Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat. Struct. Mol. Biol. 22, 999–1007.

Chu, C., Qiu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic long read mapping RNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell 44, 667–678.

Chu, C., Zhang, Q.C., Da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell 161, 404–416.

Cobas, M., Wilson, A., Ernst, B., Mancini, S.J., Macdonald, H.R., Kemler, R., and Radtke, F. (2004). Beta-catenin is dispensable for hematopoiesis and lymphopoiesis. J. Exp. Med. 199, 221–229.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962.

Cristini, A., Groh, M., Kristiansen, M.S., and Gromak, N. (2018). RNA/DNA hybrid interactome identifies DHX9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. Cell Rep. 23, 1891–1905.

Cuddapah, S., Johri, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. 19, 24–32.

Davies, J.O., Telenius, J.M., Mougou, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2018). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods 15, 74–80.

De Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics 20, 1453–1454.

De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morici, R., Sordet, O., and Capranico, G. (2019). DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. Proc. Natl. Acad. Sci. USA 116, 816–826.

De Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quesada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Kriger, P.H., and De Laat, W.
(2015). CTCF binding polarity determines chromatin looping. Mol. Cell 60, 678–684.

Deininger, M.W.N., Tyner, J.W., and Solary, E. (2017). Turning the tide in myelodysplastic/myeloproliferative neoplasms. Nat. Rev. Cancer 17, 425–440.

Deng, C., Li, Y., Zhou, L., Cho, J., Patel, B., Terada, N., Li, Y., Burgnet, J., Qi, Y., and Huang, S. (2016). HoxB11c RNA recruits Set1/MLL complexes to activate hox gene expression patterns and mesoderm lineage development. Cell Rep. 14, 103–114.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 487, 376–380.

Dowen, J.M., Fan, Z.P., Hinisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuriers, J., Lee, T.I., Zhao, K., and Young, R.A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159, 374–387.

Dumelle, J.G., and Jaffrey, S.R. (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. Elife 6, e28306.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOriilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.

Eng, J.K., Mccormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Golubovskiy, A. (2015). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 162, 442–451.

Gröschel, S., Schlenk, R.F., Engelmann, J., Rockova, V., Tealeau, V., Kuhn, M.W., Eiwen, K., Erpelinc, C., Havermans, M., Lübbert, M., et al. (2013). Deregulated expression of EVI1 defines a poor prognostic subset of MLL-rearranged acute myeloid leukemias: a study of the German-Austrian Acute myeloid leukemia study group and the Dutch-Belgian-Swiss HOVON/SAKK cooperative group. J. Clin. Oncol. 31, 95–103.

Gruneich, C., Wang, I.X., Watts, J.A., Burdick, J.T., Guber, R.D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A.B., et al. (2018). Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. Nat. Cell Biol. 20, 836–846.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudrina, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 175, 442–451.

Halaš, L., Karányi, Z., Boros-Öláh, B., Kuik-Rózsás, T., Sipos, É., Nagy, M., Mosolygó-L.-A., Mázó, A., Najafí, E., Halmos, G., and Szekvölgyi, L. (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. Genome Res. 27, 1063–1073.

Hansen, A.S., Hsieh, T.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct classes of chromatin loops revealed by deletion of a RNA-binding region in CTCF. Mol. Cell 76, 395–411.e13.

Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R.V., Mcconkey, M.E., Thielke, A., Aster, J.C., Regew, A., and Ebert, B.L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. Nat. Biotechnol. 32, 941–948.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Hughes, J.R., Roberts, N., Mcowgan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. 46, 205–212.

Izzo, F., Lee, S.C., Poran, A., Chaliligne, R., Galti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat. Genet. 52, 378–387.

Jeannot, G., Scheller, M., Scarpellino, L., Duboux, S., Gardiol, N., Back, J., Kuttler, F., Malanchi, I., Birchmeier, W., Leutz, A., et al. (2008). Long-term, multilineage hematopoiesis occurs in the combined absence of beta-catenin and gamma-catenin. Blood 111, 142–149.

Jenjareenpun, P., Wongsurawat, T., Sudheerawongan, S., and Kuznetsov, V.A. (2017). R-loopDB: a database for R-loop forming sequences (R-loops) and R-loops. Nucleic Acids Res. 45, D119–D127.

Jenjareenpun, P., Wongsurawat, T., Yanamadara, S.P., and Kuznetsov, V.A. (2015). QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. Nucleic Acids Res. 43, 10081.

Kirstetter, P., Anderson, K., Porse, B.T., Jacobson, S.E., and Nerlov, C. (2006). Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. Nat. Immunol. 7, 1048–1056.

Kolligs, F.T., Hu, G., Dang, C.V., and Fearon, E.R. (1999). Neoplastic transformation of RK3E by mutant beta-catenin requires deregulation of Tcf/Lef transcription but not activation of c-myc expression. Mol. Cell Biol. 19, 5696–5706.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., Del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol. Cell 57, 361–375.

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

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, Y., Xiao, Z., Luo, H., Benyoucef, A., Kang, Y., Lai, Q., Dovat, S., Miller, C., Chepelev, I., Li, Y., et al. (2020). Alteration of CTCF-associated chromatin neighborhood inhibits TAL1-driven oncogenic transcription program and leukemogenesis. Nucleic Acids Res. 48, 3119–3133.

Liu, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes...
in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. 13, 1196–1204.

Lu, W.T., Hawley, B.R., Skalka, G.L., Baldock, R.A., Smith, E.M., Bader, A.S., Malewicz, M., Watts, F.Z., Wilczynska, A., and Bushell, M. (2018). Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. Nat. Commun. 9, 532.

Luo, H., Wang, F., Zha, J., Li, H., Yan, B., Du, Q., Yang, F., Sofh, A., Vulpe, C., Drusbosky, L., et al. (2018). CTCF boundary remodels chromatin domain and drives aberrant HOX gene transcription in acute myeloid leukemia. Blood 132, 837–848.

Luo, H., Zhu, G., Xu, J., Lai, Q., Yang, G., Guo, Y., Zhou, X., et al. (2019). HOTTIP IncRNA promotes hematopoietic stem cell self-renewal leading to AML-like disease in mice. Cancer Cell 36, 645–659.e8.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J.

Martin, M., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Paul, F., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell 138, 653–663.

Molecular Cell

in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. 13, 1196–1204.

Lu, W.T., Hawley, B.R., Skalka, G.L., Baldock, R.A., Smith, E.M., Bader, A.S., Malewicz, M., Watts, F.Z., Wilczynska, A., and Bushell, M. (2018). Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. Nat. Commun. 9, 532.

Luo, H., Wang, F., Zha, J., Li, H., Yan, B., Du, Q., Yang, F., Sofh, A., Vulpe, C., Drusbosky, L., et al. (2018). CTCF boundary remodels chromatin domain and drives aberrant HOX gene transcription in acute myeloid leukemia. Blood 132, 837–848.
Molecular Cell

Acad. Sci. USA

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. 68, 850–858.

Singh, S., Ahmed, D., Dolatshad, H., Tatwavedi, D., Schulze, U., Sanchi, A., Ryley, S., Dhir, A., Carpenter, L., Watt, S.M., et al. (2020). SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications. Leukemia 34, 2525–2530.

Siriboonpiputtana, T., Zeisig, B.B., Zarfowiecki, M., Fung, T.K., Mallardo, M., Tsai, C.T., Lau, P.N.I., Hoang, Q.C., Veiga, P., Barnes, J., et al. (2017). Transcriptional memory of cells of origin overrides β-catenin requirement of MLL cancer stem cells. EMBO J. 36, 3139–3155.

Skoult-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 28, 1384–1396.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550.

Sun, Q., Cserba, T., Skoult-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science 340, 619–621.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Włodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell 163, 1611–1627.

Trapnell, C., Caciarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelson, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386.

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

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578.

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

Tummala, H., Dokal, A.D., Walne, A., Ellison, A., Cardoso, S., Amirthasiganamipilla, S., Kirwan, M., Browne, I., Sidhu, J.K., Rajeeve, V., et al. (2018). Genome instability is a consequence of transcription deficiency in patients with bone marrow failure harboring biallelic ERCC6L2 variants. Proc. Natl. Acad. Sci. USA 115, 7777–7782.

Wang, J., Li, Z., He, Y., Pan, F., Chen, S., Rhodes, S., Nguyen, L., Yuan, J., Jiang, L., Yang, X., et al. (2014). Loss of Asxl1 leads to myelodysplastic syndrome-like disease in mice. Blood 123, 541–553.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120–124.

Wang, Y., Krivtsov, A.V., Sinha, A.U., North, T.E., Goessling, W., Feng, Z., Zon, L.I., and Armstrong, S.A. (2010). The Wnt/β-catenin pathway is required for the development of leukemia stem cells in AML. Science 327, 1650–1653.

Wingett, S.W., and Andrews, S. (2018). FastQ Screen: a tool for multi-genome mapping and quality control. F1000Res 7, 1338.

Wolff, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15.

Wolff, F.A., Hamye, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction recovers clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 20, 59.

Wolff, J., Rabbani, L., Gilisbach, R., Richard, G., Manke, T., Backofen, R., and Grünig, B.A. (2020). Galaxy HiExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res. 48, W177–W184.

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

Yeung, J., Esposito, M.T., Gandillet, A., Zeisig, B.B., Griessinger, E., Bonnet, D., and So, C.W. (2010). Beta-catenin mediates the establishment and drug resistance of MLL leukemic stem cells. Cancer Cell 18, 606–618.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell 13, 291–298.

Zeisig, B.B., Fung, T.K., Zarfowiecki, M., Tsai, C.T., Luo, H., Stanojevic, B., Lynn, C., Leung, A.Y.H., Zuna, J., Zalova, M., et al. (2021). Functional reconstruction of human AML reveals stem cell origin and vulnerability of treatment-resistant MLL-rearranged leukemia. Sci. Transl. Med. 13, eabc4822.

Zhang, Y., Krivtsov, A.V., Sinha, A.U., North, T.E., Goessling, W., Feng, Z., Zon, L.I., and Armstrong, S.A. (2010). The Wnt/β-catenin pathway is required for the development of leukemia stem cells in AML. Science 327, 1650–1653.

Zhou, Z., Giles, K.E., and Felsenfeld, G. (2019). DNA.RNA triple helix formation can function as a cis-acting regulatory mechanism at the human beta-globin locus. Proc. Natl. Acad. Sci. USA 116, 6130–6139.

Zlatanova, J., and Caiafa, P. (2009). CCCTC-binding factor: to loop or to bridge. Cell. Mol. Life Sci. 66, 1647–1660.
## STAR METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Antibodies**      |        |            |
| Anti-H3K4me3 antibody, rabbit monoclonal | Millipore | Cat#04-745; RRID:AB_1163444 |
| Anti-H3K27me3 antibody, rabbit polyclonal | Millipore | Cat#07-449; RRID:AB_310624 |
| Anti-DNA-RNA Hybrid [S9.6] Antibody | Kerafast | Cat# ENH001; RRID:AB_2687463 |
| Anti-CD45 antibody, rabbit polyclonal | Abcam | Cat#ab10559; RRID:AB_442811 |
| Anti-Ly-6A/E (Sca-1) antibody, mouse monoclonal | BioLegend | Cat#108111, RRID:AB_313948 |
| Anti-CD117 (c-kit) antibody, mouse monoclonal | BioLegend | Cat#135135, RRID:AB_2632808 |
| Anti-GST antibody, rabbit polyclonal | Cell Signaling Technology | Cat# 2622, RRID:AB_331670 |
| Anti-CTCF antibody, rabbit polyclonal | Cell Signaling Technology | Cat# 2899, RRID:AB_2086794 |
| Anti-SA1 antibody, goat polyclonal | Abcam | Cat#ab4457; RRID:AB_2286589 |
| Anti-SA2 antibody, goat polyclonal | Abcam | Cat#ab4464; RRID:AB_304472 |
| Anti-RAD21 antibody, rabbit polyclonal | Abcam | Cat#ab992; RRID:AB_2176601 |
| Anti-DHX9 antibody, mouse monoclonal | Sigma-Aldrich | Cat# WH0001660M1, RRID:AB_1841284 |
| Anti-RPA1 antibody, rabbit monoclonal | Abcam | Cat#ab79398; RRID:AB_1603759 |
| Anti-PARP1 antibody, rabbit polyclonal | Sigma-Aldrich | Cat#AV33754, RRID:AB_1854978 |
| Anti-DDX5 antibody, rabbit polyclonal | Cell Signaling Technology | Cat# 4387, RRID:AB_2090733 |
| Anti-Lamin B1 antibody, rabbit monoclonal | Cell Signaling Technology | Cat# 13435, RRID:AB_2737428 |
| Anti-SRSF1 antibody, rabbit monoclonal | Thermo Fisher Scientific | Cat# PA5-30220, RRID:AB_2547694 |
| Anti-MCM3 antibody, rabbit polyclonal | Cell Signaling Technology | Cat# 4012, RRID:AB_2235150 |
| Anti-NPM1 Antibody, mouse monoclonal | Novus Biologicals | Cat#NB600-1030, RRID:AB_10001674 |
| Mouse IgG Isotype Control antibody | Thermo Fisher Scientific | Cat# 31903, RRID:AB_10959891 |
| Normal Rabbit IgG antibody | Cell Signaling Technology | Cat# 2729, RRID:AB_1031062 |
| Anti-HDAC1, rabbit polyclonal | Abcam | Cat#ab7028, RRID:AB_305705 |
| **Chemicals, peptides, and recombinant proteins** |        |            |
| Lipofectamine 3000 reagent | Thermo Fisher Scientific | Cat#L3000-008 |
| Proteinase K | Thermo Fisher Scientific | Cat#25530049 |
| RNase A | Thermo Fisher Scientific | Cat#EN0531 |
| RNase H | New England Biolabs | Cat#M0297L |
| DpnII | New England Biolabs | Cat#R0543L |
| Alt-R S.p. Cas9 Nuclease 3NLS | Integrated DNA Technologies | Cat#1074181 |
| Alt-R CRISPR-Cas9 tracrRNA | Integrated DNA Technologies | Cat#1072532 |
| Alt-R Cas9 Electroporation Enhancer | Integrated DNA Technologies | Cat#1075915 |
| Protease inhibitor Cocktail | Abcam | Cat#ab65621 |
| Dynabeads Protein G | Thermo Fisher Scientific | Cat#10003D |
| Dynabeads Protein A | Thermo Fisher Scientific | Cat#10001D |
| Dynabeads MyOne Streptavidin C1 | Thermo Fisher Scientific | Cat#65001 |
| Dynabeads M-280 Streptavidin | Thermo Fisher Scientific | Cat#11206D |
| SUPERase In RNase Inhibitor | Thermo Fisher Scientific | Cat#AM2694 |
| RNaseOUT Ribonuclease Inhibitor | Thermo Fisher Scientific | Cat#10777019 |
| Pierce Protease Inhibitor Tablets | Thermo Fisher Scientific | Cat#A32963 |
| TURBO DNase | Thermo Fisher Scientific | Cat#AM2238 |
| Pierce Glutathione Agarose | Thermo Fisher Scientific | Cat#16101 |
| D-Biotin | Thermo Fisher Scientific | Cat#B20656 |

(Continued on next page)
### Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| T5 exonuclease      | New England Biolabs | Cat#M0363 |
| TRI Reagent         | Sigma-Aldrich | Cat#93289 |
| Nonidet P40         | Millipore | Cat#11332473001 |
| IGEPAL CA-630       | Millipore | Cat#I8896 |
| laemmli sample buffer (4X) | Thermo Fisher Scientific | Cat#84788 |
| BrUTP               | Biotium | Cat#40026 |
| One-Step Blue Protein Gel Stain | Biotium | Cat#21003 |
| AMPure XP beads     | Beckman Coulter | Cat#A63881 |

### Critical commercial assays

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| RNeasy mini-isolation kit | QIAGEN | Cat#74106 |
| Neon Transfection System Kit | Thermo Fisher Scientific | Cat#MPK1025 |
| QIAquick Gel Extract kit | QIAGEN | Cat#28706 |
| Alt-R CRISPR-Cas9 Control Kit | Integrated DNA Technologies | Cat#1072554 |
| mirVana PARIS kit | Thermo Fisher Scientific | Cat#AM1556 |
| Superscript II reverse Transcriptase | Thermo Fisher Scientific | Cat#18064014 |
| AmpliScribe T7-Flash Biotin-RNA Transcription Kit | Lucigen | Cat#ASB71110 |
| QIAquick PCR purification kit | QIAGEN | Cat#28106 |
| QIAGEN Spin Miniprep Kit | QIAGEN | Cat#27106 |
| QIAGEN Plasmid Maxi Kit | QIAGEN | Cat#12965 |
| Nextera DNA Library Preparation Kit | Illumina | Cat#FC-121-1030 |
| TruSeq Stranded mRNA Library Prep | Illumina | Cat#20020594 |
| TruSeq CHIP Library Preparation Kit | Illumina | Cat#IP-202-1012 |
| Arima-HIC Kit | Arima | Cat#A410030 |
| KAPA Hyper Prep Kit | KAPA | Cat # KK8500, KK4824 and KK8502 |
| NEBNext DNA Library Prep Master Mix Set for Illumina | New England Biolabs | Cat # E6040 |
| NEBNext Ultra II | New England Biolabs | Cat # 7645S |
| NEBNext Multiplex Oligos for Illumina Primer set 1 | New England Biolabs | Cat # E7355S |
| NEBNext Multiplex Oligos for Illumina Primer set 2 | New England Biolabs | Cat # E7500S |
| SureSelectXT Mouse All Exon | Agilent | Cat # 5190–4641 |
| pGEM-T Easy Vector Systems | Promega | Cat#A137A |
| Herculase II Fusion Polymerase kit | Agilent | Cat#600677 |
| Nimblegen SeqCap EZ HE-oligo kit A | Roche | Cat#06777287001 |
| Nimblegen SeqCap EZ HE-oligo kit B | Roche | Cat#06777317001 |
| Nimblegen SeqCap EZ Accessory kit v2 | Roche | Cat#07145594001 |
| Nimblegen SeqCap EZ Hybridisation and wash kit | Roche | Cat#05634261001 |
| KAPA Library Quantification Complete Kit (Universal) | KAPA | Cat#K4824 |
| Qubit High Sensitivity Assay kit | Thermo Fisher Scientific | Cat#Q32854 |
| NEBNext Ultra II Directional RNA Library Prep kit | New England Biolabs | Cat#E7765 |
| SingleShot SYBR Green One-Step Kit | Bio-Rad Laboratories | Cat#1725095 |

### Deposited data

- **RNA-seq in WT vs HOTTIP−/− MOLM13 cells**: Luo et al., 2019  
  GEO: GSE114981
- **ATAC-seq in WT vs HOTTIP−/− MOLM13 cells**: Luo et al., 2019  
  GEO: GSE114981

(Continued on next page)
### REAGENT or RESOURCE SOURCE IDENTIFIER

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| CHIRP-seq of WT vs HOTTIP^- MOLM13 cells | Luo et al., 2019 | GEO: GSE114981 |
| CHIRP-seq of WT vs Hottip-Tg LK cells | This paper | GEO: GSE165049 |
| H3K4me3/H3K27me3 ChIP-seq of WT vs HOTTIP^- MOLM13 cells | Luo et al., 2019 | GEO: GSE114981 |
| H3K4me3/H3K27me3 ChIP-seq of WT, CBS-u2^-/-, CBS-u2^RHI MOLM13 cells | This paper | GEO: GSE165049 |
| RNA-seq of WT vs Hottip-Tg mice LSK cells | This paper | GEO: GSE114981 |
| ATAC-seq of WT vs Hottip-Tg mice LSK cells | This paper | GEO: GSE114981 |
| CTCF/SA1/SA2/RAD21 ChIP-seq of WT vs HOTTIP^- MOLM13 cells | This paper | GEO: GSE165049 |
| CTCF ChIP-seq of WT vs Hottip-Tg LK cells | This paper | GEO: GSE165049 |
| DRIP-seq of WT vs HOTTIP^- MOLM13 cells | This paper | GEO: GSE165049 |
| DRIPc-seq of WT vs HOTTIP^- MOLM13 cells | This paper | GEO: GSE165049 |
| Single cell RNA-seq of WT vs Hottip-Tg mice | This paper | GEO: GSE165049 |
| HiC-seq of WT vs HOTTIP^-/- MOLM13 cells | Luo et al., 2019 | GEO: GSE114981 |
| HiC-seq of WT vs Hottip-Tg mice LSK cells | This paper | GEO: GSE165049 |
| GRO-seq of WT vs HOTTIP^-/- MOLM13 cells | This paper | GEO: GSE165049 |
| RNA-seq of WT vs CBS-u2^RHI MOLM13 cells | This paper | GEO: GSE165049 |
| NG-Capture-C-seq of WT, HOTTIP^-/-, CBS-u2^-/-, CBS-u2^RHI MOLM13 cells | This paper | GEO: GSE165049 |

### Experimental models: Cell lines

| Models | Source | Identifier |
|--------|--------|------------|
| MOLM-13 | DSMZ | Cat# ACC-554, RRID:CVCL_2119 |
| HEK293T | ATCC | Cat# CRL-3216, RRID:CVCL_0063 |
| OCI-AML2 | DSMZ | Cat# ACC-99, RRID:CVCL_1619 |
| OCI-AML3 | DSMZ | Cat# ACC-582, RRID:CVCL_1844 |

### Experimental models: Organisms/strains

| Models | Source | Identifier |
|--------|--------|------------|
| Hottip-Transgenic mouse | Luo et al., 2019 | N/A |
| Xenograft AML mouse model | This paper | N/A |

### Oligonucleotides

| Oligonucleotides | Source | Availability |
|------------------|--------|-------------|
| sgRNAs | This paper | Upon request |
| RT-qPCR primers | This paper | Upon request |
| ChIP-qPCR primers | This paper | Upon request |
| ATAC primers | This paper | Upon request |
| CHIRP Probes | This paper | Upon request |
| RiDP Probes | This paper | Upon request |
| NG-Capture-C-biotin oligos | This paper | Upon request |
| Cy3 or Cy5 labelled oligos | This paper | Upon request |
| 3C primers | This paper | Upon request |

### Recombinant DNA

| Recombinant DNA | Source | Identifier |
|-----------------|--------|------------|
| pL-CRISPR.EFS.GFP | Heclk et al., 2014 | Addgene Plasmid #57818; RRID:Addgene_57818 |
| pLKO5.sgRNA.EFS.tRFP | Heclk et al., 2014 | Addgene Plasmid #57824; RRID:Addgene_57824 |
| lentiCRISPR v2 | Sanjana et al., 2014 | Addgene Plasmid #52961; RRID:Addgene_52961 |

(Continued on next page)
**RESOURCE AVAILABILITY**

**Lead contact**
Further information and requests for reagents should be directed to and will be fulfilled by the lead contact, Dr. Suming Huang (shuang4@pennstatehealth.psu.edu)

**Materials availability**
This study did not generate new unique reagents. All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.
**Data and code availability**

- All genomics datasets generated in this study can be accessed at GEO database (GEO: GSE114981, GSE165049), are listed in the key resources table, and are publicly available as of date of publication.
- All original western blots, gel images, and MS proteomic data have been deposited at Mendeley and are publicly available as of the date of publication.
- This paper does not report original code since it is based on implementation of publicly available software.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

**AML patient samples**

The acquisition of primary AML patient samples, #921, NPM1C+/FLT-ITD+; #LPP4, MLLr+, was approved by the Institutional Review Boards of the University of Florida and the Cancer Institute of the Pennsylvania State University College of Medicine in accordance with the Declaration of Helsinki.

**Cell lines**

HEK293T cells were cultured in DMEM supplemented with 10% bovine serum (FBS) and 1% penicillin/streptomycin, MOLM13 AML cells were cultured in RPMI 1640 supplemented with 10% FBS and 1% penicillin/streptomycin, and OCI-AML2 and OCI-AML3 cells were cultured in alpha-MEM supplemented with 10% FBS and 1% penicillin/streptomycin. All cells were maintained at 37°C with 5% CO2 in a humidified incubator.

**Hottip transgenic (Tg) mouse model**

The complete coding region of murine Hottip was cloned downstream of the Vav1 promoter (HS321/45-vav vector) and enhancer to ensure restricted transgene expression in hematopoiesis (Luo et al., 2019). All studies were approved by the Institutional Animal Care and Use Committees (IACUC) at Penn State Hershey Medical Center and The University of Texas Health Science Center at San Antonio and performed in accordance with relevant guidelines and regulations.

**METHOD DETAILS**

**Generation of CRISPR-Cas9 mediated knockout AML cells and WT or catalytically dead dCas9-RNaseH AML cells**

All guide RNAs were subcloned into the pLKO5.sgRNA.EFS.tRFP (Addgene #57824), pL-CRISPR.EFS.GFP (Addgene #57818), or lentiCRISPR v2 (Addgene #52961) vector, confirmed by sequencing, and transduced into MOLM13 cells. To generate the CTNNB1 CBS-u2 knockout, two sgRNAs flanking the CBS site upstream of CTNNB1 were designed. One sgRNA was subcloned into the pLKO5.sgRNA.EFS.tRF vector, and another into the pL-CRISPR.EFS.GFP vector, and then single GFP/RFP double positive clones were selected by FACS. For generation of the dCas9-RNaseH fusion, GFP-tagged RNaseH (Addgene #65784) was subcloned into the pH-R-dCas9 vector (Addgene #46911) as a Sbf1-BamHI restriction fragment. For generation of the catalytically dead RNaseH-fused dCas9, mutated RNaseH (D210N) (Addgene #111904) was subcloned into the pH-R-dCas9 vector (Addgene #46911) as a Sbf1-BamHI restriction fragment. sgRNA targeting the CTNNB1 CBS-u2 was subcloned into the pLKO5.sgRNA.EFS.tRF vector (Addgene #57824) and transduced along with the dCas9-RNaseH fusion vector into AML cells, and then GFP/RFP double positive cells were selected by FACS. For overexpression of β-catenin in RNaseH-dCas9 clone, firstly, CTNNB1 cDNA (Addgene #16828) was subcloned into the lentiviral-puro backbone vector (Addgene #123223), and then transduced into RNaseH-dCas9 cells. Finally, positive clones were selected with puromycin (2μg/mL).

**Hematopoietic stem and progenitor cell (HSPCs) sorting**

Hematopoietic stem and progenitor cells (HSPCs) were sorted from WT and Hottip transgenic (Tg) mice as previously described (Wang et al., 2014). Briefly, Lin− BM cells were depleted from the BM of 6-8 week old mice using Miltenyi Biotec magnetic beads (130-110-470). Subsequently, Lin− BM cells were stained with lineage, Sca-1, and c-kit antibodies. LK (Lin− Kit+) or LSK (Lin− Sca1+ Kit+) cells were sorted using a BD FACSAriaII flow cytometer. The purity of selected LSK cells was routinely over 98%. Data were analyzed with FlowJo-V10 software.

**In vitro pull-down assay with biotinylated RNA**

Full-length CTCF (FL-CTCF, 727aa) and truncated CTCF (1-268aa, 269-576aa, 269-727aa, 577-727aa) cDNAs were subcloned into the pGEX-5X-1 vector (Promega). GST fusion proteins were generated in BL21 bacteria cells and purified using glutathione sepharose beads (GE Healthcare Life Sciences). Biotinylated HOTTIP RNAs were generated in vitro using the AmpliScribe T7-flash Biotin-RNA Transcription Kit according to the manufacturer’s instructions. In vitro transcribed RNAs were denatured and analyzed by agarose gel electrophoresis. 50 pmol of biotinylated RNAs were refolded by heating at 65°C for 5 min and cooling to room temperature in 1X refolding buffer (100 mM KCl, 10 mM MgCl2, and 10 mM Tris-HCl pH7.0). GST-fusion proteins and in vitro transcribed
HOTTIP RNAs were incubated together at room temperature (RT) for 1 hr, then equilibrated streptavidin-coupled Dynabeads were added for 1 hr at RT. The complexes were precipitated, washed four times, and analyzed by western blotting.

**RNA isolation, quantitative RT-PCR, and RNA-sequencing (RNA-seq)**

Total RNA was isolated from the indicated cells or cell lines using the RNeasy mini kit (Qiagen) according to the manufacturer’s instructions. The Superscript II Reverse Transcriptase (Invitrogen) was used to convert 2 μg RNA into cDNA, which was analyzed by quantitative PCR, using the primers upon request.

RNA libraries were prepared for next generation sequencing using the Illumina TruSeq mRNA sample preparation kit (Illumina, Cat# 20020594). Briefly, mRNA was purified using poly-T oligo beads and fragmented. Subsequently, first- and second- strand cDNA synthesis was performed, libraries were amplified, and samples were indexed. Library quality was assessed by Qubit and Agilent Bioanalyzer. Libraries were subjected to paired-end sequencing at a 50 bp length on an Illumina HiSeq 3000. Sequence reads were trimmed and quality filtered using cutadapt (http://cutadapt.readthedocs.io/, version 1.2.0) program (Martin, 2011) and aligned to the mouse (mm9) or human (hg19) genome using TopHat (version 2.0) and Bowtie2 (Trapnell et al., 2012, Langmead et al., 2009, Trapnell et al., 2009). FPKM (paired-end fragments per kilobase of exon model per million mapped reads) values were calculated using Cufflinks v2.2.1 and differential expression analysis was performed using Cuffdiff (Trapnell et al., 2010). The heatmap was generated using cluster3.0 and treeview based on log_2 transformation of the FPKM values (de Hoon et al., 2004). Gene set enrichment analysis (GSEA) was performed with gene sets obtained from the Molecular Signatures Database (Subramanian et al., 2005). GO analysis of differentially expressed (greater than 2 fold) genes was generated using Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources (Huang et al., 2009b) and Gorilla (Eden et al., 2009). Sequence reads have been deposited in the NCBI GEO: GSE114981.

**Chromatin immunoprecipitation (ChIP) assay**

ChIP assay was completed as described previously (Luo et al., 2018). Briefly, 5 × 10^6 cells were crosslinked in 1% formaldehyde and/or EGS (for TFs) for 10 min at room temperature, quenched by 125 mM glycine for 5 min at room temperature, and washed with ice-cold 1X PBS twice. Cells were then lysed in ChIP lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) and DNA was fragmented by sonication with the Bioruptor UCD200. Lysates were centrifuged at 14,000 rpm for 10 min at 4°C and diluted in ChIP buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 1% Triton X-100, 150 mM NaCl). 10% of the lysate was retained as an input. Chromatin was immunoprecipitated with 5μg CTCF, 2.5μg H3K4me3, 2.5μg H3K27me3, 5μg SA1, 5μg SA2, and 5μg RAD21 antibody overnight at 4°C. Each sample was then incubated with 50 μL Protein G Dynabeads (Thermo Fisher Scientific) for 2 hrs at 4°C and washed sequentially with low salt washing buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), high salt washing buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), lithium chloride washing buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1% Triton X-100, 250 mM LiCl, 1% sodium deoxycholate), and TE buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA). DNA was eluted for 30 min at room temperature in elution buffer (100 mM NaHCO₃, 1% SDS), reverse crosslinked with 2 μL of 10 mg/mL protease K overnight at 65°C, purified, and analyzed by quantitative PCR. Results represent percentage of input and error bars indicate standard deviations (S.D.) through triplicate experiments.

ChIP-DNA libraries were prepared using Illumina’s TruSeq ChIP Sample Preparation Kit according to the manufacturer’s instructions (Cat# IP-202-1012). Briefly, 10 ng ChIP DNA fragments underwent end repair and were purified with AMPure XP beads. The 3′ ends were adenylated. Fragments were ligated with adapter indices and amplified with adapter primers. Quality control was done using Qubit and Agilent Bioanalyzer and libraries were subjected to paired-end sequencing at 100 bp length on an Illumina NovaSeq 6000.

**Chromatin Isolation by RNA Immunoprecipitation (CHIRP), ChIRP-WB, and ChIRP-LC-MS/MS assays**

The Chromatin Isolation by RNA Immunoprecipitation (CHIRP) assay was performed as described previously (Chu et al., 2011) with modifications. Briefly, 20 million cells were cross-linked in PBS with 1% glutaraldehyde for 10 min at room temperature with shaking, quenched by 125 mM glycine at room temperature for 5 min, and washed with chilled PBS twice. Cells were then lysed in 1 ml of cell lysis buffer (50 mM Tris-HCl pH 7.0, 10 mM EDTA, 1% SDS, add PMSF, DTT, P.I. and Superase-in in fresh) per 100mg of cells and DNA was fragmented by sonication with a Bioruptor UCD200 (Diagenode). Chromatin was diluted in hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris-HCl pH 7.0, 1 mM EDTA, 15% formamide, add DTT, PMSF, P.I. and Superase-in in fresh) and hybridized with 100 pmole of biotinylated DNA probes targeting HOTTIP or LacZ, as a negative control. DNA/RNA hybrids were precipitated using 100 μL of Streptavidin-magnetic C1 beads (Invitrogen) and washed 5 times with washing buffer (2X SSC, 0.5% SDS). RNA was isolated using TRIzol reagent and HOTTIP enrichment was confirmed by RT-qPCR, with β-actin as a negative control. DNA was isolated using Phenol/Chloroform extraction and ethanol precipitation. CHIRP-seq library was prepared for sequencing as described for ChIP-seq library.

ChIRP-WB or mass spectrometry (MS) was performed as previous described (Chu et al., 2015). Briefly, 500 million MOLM13 cells were cross-linked, lysed, sonicated, and hybridized with biotinylated tiling probes as described above. Total proteins isolated by the Streptavidin-magnetic C1 beads were eluted from the beads by shaking for 20 min at room temperature followed by incubation at 65°C for 10 min. Then, proteins were precipitated by adding 50 μL TCA (25% v/v) and vortexing overnight at 4°C. Proteins were pelleted at 16,000 g at 4°C for 30 min, washed once with cold acetone, air-dried for 1 min, and solubilized in 1× laemmli sample...
buffer. Proteins were reverse cross-linked at 95 °C for 30 min with occasional mixing and resolved on bis-tris SDS-PAGE gels for western blot, using the indicated antibodies or mass-spectrometry. ChIRP-LC/MS/MS (LC/MS/MS), experiments were carried out at Taplin Mass Spectrometry Facility at Harvard Medical School (Shevchenko et al., 1996, Eng et al., 1994, Peng and Gygi, 2001). Briefly, Coomassie brilliant blue gel bands were cut into approximately 1 mm² pieces and subjected to a modified in-gel trypsin digestion procedure. Gel pieces were washed and dehydrated with acetonitrile for 10 min, the acetonitrile was removed, and they were completely dried in a SpeedVac Vacuum. Gel pieces were rehydrated with 50 mM ammonium bicarbonate solution containing 12.5 ng/μl modified sequencing-grade trypsin (Promega) for 45 min at 4 °C. Subsequently, excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate solution to cover the gel pieces overnight at 37 °C. Peptides were then extracted by removing the ammonium bicarbonate solution, washing once with a 50% acetonitrile/1% formic acid solution, and drying in a speed-vac. Next, peptides were subjected to electrospray ionization and entered in an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences were determined by matching protein databases. Cellular compartment analysis was based on GO term cellular component analysis, and overrepresentation analysis of protein classes (protein enrichment) was performed using DAVID database (https://david.ncifcrf.gov, Version 6.8) (Huang et al., 2009a, 2009b). Protein classes over-represented in the HOTTIP-ChIRP were ranked according to their Benjamini-Hochberg-corrected p values (p value threshold 0.05).

RNA Immunoprecipitation (RIP)
The RNA-IP assay was performed as the previously described (Deng et al., 2016; Tsai et al., 2010). Briefly, 2×10⁷ MOLM13 AML cells were collected, washed with cold 1x PBS, and then incubated with freshly prepared nuclear isolator buffer (1.28M sucrose, 40 mM Tris-HCl pH 7.5, 20 mM MgCl₂, 4% Triton X-100) for 20 min on ice with frequent mixing. Nuclei were precipitated by centrifugation at 2,500 g for 15 min at 4 °C, and then resuspended with freshly prepared lysis buffer (10 mM HEPES-KOH pH 7.5, 150 mM KCl, 5 mM MgCl₂, 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) on the ice, and immediately sonicated briefly using the Bioruptor. The nuclear membrane and other debris were pelleted by centrifugation for 10 min at 14,000 g at 4 °C and the supernatant was collected. Then, antibodies (2-10 μg) targeting CTCF, RAD21, SA2, PARP1, DHX9, DDX5, LMNB1, SRSF1, RPA1, HDAC1 or IgG were added to the supernatant and incubated overnight at 4 °C with rotation. Next, complexes were precipitated using 40μl equilibrated Protein A/G magnetic beads at 4 °C with gentle rotation, and washed three times with ice-cold wash buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM MgCl₂, 0.05% IGEPAL-CA-630) supplemented with 0.02 mg/mL heparin, and then eluted with 500 μL SDS-EDTA (50 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA, 1% SDS) for 10 min at 65 °C. Co-precipitated RNA was isolated using TRIzol Reagent, precipitated with isopropanol, washed with 70% ethanol, and eluted in nuclease-free water. RNA was subsequently DNase treated and reverse transcribed according to the manufacturer instructions to generate cDNA, which was measured by qPCR using primers targeting HOTTIP, HOTAIR1 and GAPDH.

DNA:RNA immunoprecipitation sequencing (DRIP-seq) assay
DNA:RNA immunoprecipitation sequencing (DRIP-seq) was performed as described previously (Sanz and Chedin, 2019, Halasz et al., 2017). Briefly, MOLM13 cells were crosslinked in 1% formaldehyde for 10 min at room temperature, quenched by 125 mM glycine for 5 min at room temperature, and washed with ice-cold 1x PBS twice. Cells were then lysed in lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 1% Triton X-100, 0.1% Na-Deoxycholate, 1% SDS) and fragmented by sonication (Bioruptor) to 300bp. The fragmented chromatin was treated with RNase A (10 mg/ml) in TE buffer (10 mM Tris-HCl pH 8, 10 mM EDTA pH 8) at 37 °C for 1 hr, followed by Proteinase K (10 mg/ml) at 65 °C overnight. Nucleic acids were extracted by phenol-chloroform extraction, ethanol precipitation, and resuspension in 5 mM Tris-HCl pH 8.5. Two percent of each sample was retained for input DNA. Half of the samples were treated with RNaseH at 37 °C overnight. 50 μL of BSA-blocked Protein A Dynabeads (Thermo Fisher Scientific) were incubated with 10 μg of S9.6 antibody in IP buffer (50 mM Heps/KOH at pH 7.5; 0.14 M NaCl; 5 mM EDTA; 1% Triton X-100; 0.1% Na-Deoxycholate, ddH₂O) at 4 °C for 4 h with rotation. Purified genomic DNA was added and rotated at 4 °C overnight. Beads were washed sequentially with low salt buffer (50 mM Heps/KOH pH 7.5, 0.14 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), high salt buffer (50 mM Heps/KOH pH 7.5, 0.5 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), LiCl wash buffer (10 mM Tris-HCl pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% Na-Deoxycholate, 1 mM EDTA pH 8), and TE (100 mM Tris-HCl pH 8, 10 mM EDTA pH 8) twice. Complexes were eluted in 100 μL of elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) for 15 min at 65 °C. Subsequently, excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate solution to cover the gel pieces overnight at 37 °C. Peptides were then extracted by removing the ammonium bicarbonate solution, washing once with a 50% acetonitrile/1% formic acid solution, and drying in a speed-vac. Next, peptides were subjected to electrospray ionization and entered in an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences were determined by matching protein databases. Cellular compartment analysis was based on GO term cellular component analysis, and overrepresentation analysis of protein classes (protein enrichment) was performed using DAVID database (https://david.ncifcrf.gov, Version 6.8) (Huang et al., 2009a, 2009b). Protein classes over-represented in the HOTTIP-ChIRP were ranked according to their Benjamini-Hochberg-corrected p values (p value threshold 0.05).

ChIP-seq, ChIRP-seq, and DRIP-seq data analysis
Sequence reads were trimmed and filtered using cutadapt (http://cutadapt.readthedocs.io/, version 1.2.0) (Martin, 2011), underwent quality control by FastQC program (Wingett and Andrews, 2018), and were mapped to the human (hg19) or mouse (mm9) reference genome using Bowtie2 with default parameters (Langmead et al., 2009). SAM files were converted to BAM files and sorted using Samtools (Li et al., 2009). Peak calling was performed using MACS2 (Zhang et al., 2008). A bigWig file, including fragment or read
RNA Isolation by DNA Purification (RiDP)
RNA Isolation by DNA Purification (RiDP) assay was performed as previously described (Ariel et al., 2020). Briefly, nuclei were extracted and purified as for the ChIP assay without crosslinking. Subsequently, nuclear DNA was fragmented by gentle sonication with the Bioruptor UCD200 (3 cycles 30'' ON - 30'' OFF). Sonicated chromatin was treated with 0.6 units of T5 exonuclease for 45 min at 37°C, and digestion was stopped by adding 4 µL of 0.5 M EDTA. 10% of these samples volume was saved for input, and the remainder was treated with or without RNaseH at 37°C for 30 min. Then, samples were incubated with 50 µL Dynabeads MyOne Streptavidin C1 pre-coated with 2 µL of indicated 100 µM biotinylated probe for 30 min at 37°C with rotation. After hybridization, beads were washed three times with ChIRP wash buffer (see ChIRP protocol) and resuspended in 200 µL of RNAase free water. RNA was purified using TRI Reagent (Sigma-Aldrich) and reverse transcribed into cDNA. The RNA components of precipitated complexes were assessed using RT-qPCR. Results are represented as percent of input with triplicate experiments.

Transposase-Accessible Chromatin using sequencing (ATAC-seq) assay
ATAC-seq assay was performed using the Nextera DNA library preparation kit as described previously (Buenrostro et al., 2015). Briefly, 5 x 10^6 cells were washed with PBS twice and lysed in buffer containing 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl2, and 0.1% NP-40. Cells were washed in PBS then treated with Tn5 transposase at 37°C for 30 min. DNA was purified using the MinElute Kit (QiAGEN) and fragments were simultaneously amplified and indexed. Following purification with AMPure XP beads (Beckman Coulter), libraries were quantified using qPCR Kapa Library Quantification Kit for Illumina (Roche). Quality control was done using Qubit and Agilent Bioanalyzer and libraries were subjected to paired-end sequencing at 100 bp length on an Illumina NextSeq 500.

ATAC-seq analysis
Sequence reads were trimmed and filtered using cutadapt (http://cutadapt.readthedocs.io/, version 1.2.0) (Martin, 2011). PCR duplicates were removed using Picard MarkDuplicates (version 2.0.1), mitochondrial reads were removed with samtools (Corces et al., 2017), and sequences underwent quality control by FastQC (Wingett and Andrews, 2018). Reads and were mapped to the human (hg19) or mouse (mm9) reference genome using Bowtie2 with default parameters (Langmead et al., 2009). ENCODE blacklist regions were filtered (https://sites.google.com/site/anshulkundaje/projects/blacklists). SAM files were converted to BAM files and sorted using Samtools (Li et al., 2009). Peak calling was performed using MACS2 with parameters (-g mm (or hs) -p 1e-9 –nolambda -f BAMPE –nomodal –shiftsize 100 –extsize 200) (Zhang et al., 2008). A bigwig file, including fragment or read coverages for control and experimental datasets, was generated with the bedGraphToBigWig program, (https://www.encodeproject.org/software/bedgraphhtobigwig/). Sequencing tracks were viewed using the Integrated Genomic Viewer (Robinson et al., 2011). peak annotation was performed using HOMER (Heinz et al., 2010), DESeq2 (Benjamini-Hochberg adjusted p< 0.05; FoldChange ≥ 2) was used to identify differentially accessible sites (Ross-Innes et al., 2012). For each genomic feature (peaks or chromVAR annotation), the chromatin accessibility median deviation z-score (for chromVAR features) or fragment counts (for peaks) were examined in control and experimental groups with chromVAR package in R language (Schip et al., 2017, Rubin et al., 2019). Reproducibility between duplicates was assessed using Pearson’s correlation coefficient and Pearson’s χ²-test. All genomics datasets were deposited in the NCBI GEO: GSE114981.

Chromosome conformation capture (3C) assay
3C assay was performed as previously described (Patel et al., 2014, Deng et al., 2016, Luo et al., 2018) with minor modifications. Briefly, 2x10^6 cells were crosslinked in 2% formaldehyde for 10 min at room temperature, quenched by 125 mM glycine for 5 min at room temperature, and washed with ice-cold 1x PBS twice. Cells were then suspended in enzymatic digestion buffer containing 0.3% SDS and incubated at 37°C overnight with shaking. The SDS was sequestered by addition of 2% Triton X-100 incubation for 1.5 hours at 37°C with shaking. Chromatin was digested with 800 U of DpnII (NEB) at 37°C overnight. The reaction was stopped using 1.6% SDS buffer for 20 minutes at 65°C. Samples were equilibrated in T4 DNA ligation buffer (NEB) containing 1% Triton X-100 for 1.5 hrs at 37°C with shaking. DNA ends were ligated with 800 U of T4 DNA ligase (NEB) at 16°C for 3 days and 1 hr at room temperature. Crosslinks were reversed with 200 µg of Proteinase K (Invitrogen) and incubation at 65°C overnight and DNA was extracted using phenol:chloroform, purified with ethanol, and dissolved in ddH2O. The 3C-ligated DNA was analyzed with PCR or qPCR.

Next generation (NG)-Capture-C sequencing assay
NG-Capture-C assay was performed as previously described (Davies et al., 2016). Briefly, 3C libraries were prepared as described above and 5 µg were sheared by sonication to 200 bp. Libraries were indexed with Illumina Truseq indexed sequencing adapters using NEBnext reagents (E6000 / E6040 / E7335 / E7500) for end repair, dA labelling, adaptor ligation, and PCR indexing following
the manufactures instructions with the following exceptions: DNA clean up steps were performed with Ampure XP beads and the Herculase II PCR kit (Agilent) was used to add the Truseq indices. Libraries were assessed using an Agilent Bioanalyzer both pre- and post-PCR and addition of sequencing adaptors. 1.5-2 µg of adapter ligated DNA, 5µg COT Human DNA, 1000 pmol Nimblegen HE Universal blocking olio, and 1000 pmol Nimblegen HE Index specific blocking olio were dried using a vacuum centrifuge at 55 °C. DNA was resuspended in 7.5 µl Nimblegen Hybridization Buffer and 3 µl Nimblegen Hybridization Component A and denatured at 95 °C for 10 min. Concurrently, 4.5 µl of 2.9 µM pooled, locus-specific biotinylated capture oligonucleotide (total 13 pmol) (IDT uliters) was heated to 47 °C in a PCR thermocycler. After denaturation, the 3C library and blocking oligonucleotides were added to the biotinylated oligonucleotides without removing them from the heating block and incubated at 47 °C for 64-72 h (with a heated lid at 57 °C). 100 µM270 streptavidin beads per library were brought to room temperature then washed twice with 200 µl pre-warmed Bead Wash Buffer. The hybridization reaction mixture was added to the beads and mixed thoroughly and incubated at 47 °C for 45 min, shaking at 500 rpm. Washing was performed using the SeqCap EZ Hybridization and wash kit (Roche) according to the manufacturer’s directions. Washed beads were resuspended in 40 µl of PCR grade water and the captured material was amplified using the SeqCap EZ Post-Capture LM PCR Master Mix and Post LM-PCR oligos (x 18 cycles). DNA was purified using Ampure-XP beads and analyzed by the Agilent Bioanalyzer. 75% of the amplified captured material was used for a second round of oligonucleotide capture as described above and hybridized for 24 h at 47 °C. Quality control was done using Qubit and Agilent Bioanalyzer and 4 nM libraries were subjected to paired-end sequencing at 150 bp length on an Illumina HiSeq 2500 for 10-20 million reads.

NG-Capture-C sequencing analysis

Sequence reads were trimmed and filtered using cutadapt (http://cutadapt.readthedocs.io, version 1.2.0) (Martin, 2011). The human genome (hg19) index and DpnII fragment list were generated using capC-MAP and bowtie2 (Langmead et al., 2009, Buckle et al., 2019). Interaction profiles were generated using capC-MAP with the following parameters: input files with paired-end fastq.gz files, genome files with human genome (hg19) index and restriction enzyme DpnII fragment files, restriction enzyme option with DpnII, binning parameters with BIN 1000 3000, BIN 3000 6000 and BIN 5,000 10,000. All genomics datasets were deposited in the NCBI GEO: GSE165049.

Xenotransplantation of human leukemic cells and Patient-Derived Xenografts (PDX)

Adult NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ (NSG) mice (6-8 weeks old) were pretreated with 280cGy total body irradiation. Subsequently, irradiated NSG or untreated adult NOD.Cg-KjW-41/Tyrα PrkdcscidIl2rgtm1Wjl/ThomJ (NBSGW) mice (6-8 weeks old) were transplanted with 5x10⁵ MOLM13, 5x10⁵ OCI-AML3 or 1.8x10⁵ primary patient AML cells each by tail-vein injection. At 18-21 days after transplantation, peripheral blood was collected and depleted of red blood cells by ammonium chloride treatment, BM was isolated from the tibias, femurs, and pelvis, and the spleen was processed into a single cell suspension. Human CD45⁺ chimerism in the BM, spleen, liver and PB cells were analyzed by flow cytometry (FACS LSR II–BD Biosciences, San Jose, CA, USA).

The Next-Gen Chromosome Conformation Capture (Hi-C) Assay

Hi-C assay was performed as described previously with Arima-HiC Kit (Cat: A410030) (https://arimagenomics.com/). Briefly, 5x10⁶ cells were washed with PBS twice, crosslinked in 1% formaldehyde for 10 min at room temperature, quenched by 125 mM glycine for 5 min at room temperature, and washed with ice-cold 1x PBS twice. Cells were then resuspended in lysis buffer and incubated at 4 °C for 15 min, conditioning solution was added at 62 °C for 10 min, and the reaction was stopped by stop solution for 15 min at 37 °C. Cell pellets were then digested with the enzymatic digestion reaction buffer and restriction enzyme cocktails (Arima-HiC Kit) overnight at 37 °C with rotation and DNA was purified using AMPure XP Beads. 750 ng of DNA was sheared by sonication (Bioruptor) and size-selected to 200-600 bp. Sequence libraries were then prepared from 250ng of DNA using KAPA Hyper Prep Kit (Catalog # KK8500, KK4824 and KK8502). Libraries were subjected to paired-end sequencing to 100 bp length on an Illumina HiSeq 2500.

Hi-C sequencing data analysis

Adapters and low quality reads were removed using bbmap and bbduk.sh (https://i2b2-tools.org/bb-tools/bb-tools-user-guide/bbduk-guide/). Reads were trimmed using Homer (version 4.10) (Heinz et al., 2010). PCR duplicates were removed using Picard MarkDuplicates (version 2.0.1), and underwent quality control by FastQC program (Wingett and Andrews, 2018). Filtered reads were mapped to the mouse (mm9) or human (hg19) reference genome using Bowtie2 with parameters (‘‘-n 1 -m 1 -p 8’’) (Langmead et al., 2009). Mapped sequencing data were used to generate a contact matrix using Homer (version 4.10). Normalization, generation of Hi-C interaction matrices, principal component analysis (PCA), and identifying significant interactions were performed as previously described (Lin et al., 2012). Briefly, the analyzeHiC program was used to generate a normalized and visualizable interaction matrix with default parameters, and the intra-chromatin interactions of the specific loci were generated by analyzeHiC program in Homer/4.10 with parameters (-res 10,000 -superRes 20,000 -pos chromosome location). The normalized interaction matrix was used to perform principal component analysis (PCA) analysis on full chromosome matrices with runHiCpca.pl program (res 10,000 -cpu 8 -genome mm9 or hg19) in R. ANOVA analysis was performed to identify TADs with significantly different domain scores between WT and HOTTIP-KO MOLM13 AML cells or between WT and Hottip-Tg mice LSK cells (cutoff: Bonferroni-corrected p value < 0.05). Differential chromatin interactions were evaluated through HICExplorer (version 3.5.3) (Wolff et al., 2020). These normalized and visualized chromatin interaction matrices were used to generate the Hi-C heatmap using juicer (version 1.5.5) (Du-
HOTTIP, and was visualized with Juicebox (Durand et al., 2016a) and Java Treeview (Saldanha, 2004). Furthermore, the domain score of the TAD was normalized by subtracting the mean of all TADs, and quantile-normalization was applied on domain scores to facilitate comparison among all Hi-C signals with HOMER program (v4.10) (Lin et al., 2012). All genomics datasets were deposited in the NCBI GEO: GSE114981 and GEO: GSE165049.

**In vitro R-loop structure detection and gel shift assay**

*In vitro* DNA:RNA hybrid (R-loop) formation assay was performed as previously described with modifications (McDonald and Maher, 1995, Zhou et al., 2019). R-loop structure prediction and Cy3/Cy5 labelled oligos design were carried out according to the DRIP-seq peaks and qMRLFS-finder database (Jenjaroenpun et al., 2015, 2017). Briefly, we synthesized Cy5-labeled single stranded HOTTIP RNA oligos, including the WT sequence (5C5/GGC AAC CAG GCG GGG AGG GAA GGT GGG GCG GCG) and the HOTTIP 3’UTR region (HOTTIP 3’UTR) (5C5/GAA GTT GCA TGC CAG CCA GCA TGA GAA CGT CCA TG), and Cy3-labeled duplexed DNA oligonucleotides, including the CBS-u2 region (5Cy3/CGC AAC CAG GCG GGG AGG GGA GGT GGG GCG GCG) and the CBS-u2 region (5Cy3/CGC AAC CAG GCG GGG AGG GGA GGT GGG GCG GCG) and the CBS-u2 region (5Cy3/CGC AAC CAG GCG GGG AGG GGA GGT GGG GCG GCG) and the CBS-u2 region (5Cy3/CGC AAC CAG GCG GGG AGG GGA GGT GGG GCG GCG) and the CBS-u2 region (5Cy3/CGC AAC CAG GCG GGG AGG GGA GGT GGG GCG GCG). In vitro the NCBI GEO: GSE114981 and GEO: GSE165049.

**Global Run-On sequencing (GRO-seq)**

GRO-seq experiment was carried out as previously described (Gardini, 2017, Barbieri et al., 2020). Briefly, 1x10^7–1x10^8 cells were washed twice with ice-cold PBS, swelled in swelling buffer (10 mM Tris-HCl pH 7.5, 2 mM MgCl2, 3 mM CaCl2, 2 U/ml Superase-in (Invitrogen)) for 5 min on ice, resuspended in swelling buffer with 10% glycerol, and lysed in lysis buffer (10 mM Tris-HCl pH 7.5, 2 mM MgCl2, 3 mM CaCl2, 10% glycerol, 1% Igepal NP-40, 2 U/ml Superase-in) for 5 min on ice. Nuclei were pelleted at 600 g for 5 min at 4°C, and resuspended in freezing buffer (40% glycerol, 5 mM MgCl2, 0.1 mM 5’-triphosphate EDTA, 50 mM Tris-HCl pH 8, 2 U/ml Superase-in). Nuclei were counted, snap frozen in 100 µL aliquots at a concentration of 1x10^6 nuclei/100 µL, and stored at -80°C. Nuclei were thawed on ice, an equal volume of pre-warmed nuclear run-on reaction buffer (10 mM Tris-HCl pH 8.5, 5 mM MgCl2, 300 mM KCl, 1 mM DTT, 500 µM ATP, 50 µM TTP, 500 µM UTP, 200 µM Superase-in, 1% Sarkosyl (Nlaurylsarcosine sodium salt solution)) was added, and the reaction was incubated for 7 min at 30°C. RNA was extracted with TRIzol LS reagent (Invitrogen), purified by ethanol precipitation, and dissolved in water. RNA was DNase treated, fragmented using RNA fragmentation reagents (Ambion), and purified Micro Bio-Spin P-30 Gel Columns (BioRad). Fragmented RNA was added to pre-blocked Anti-BrdU-conjugated agarose beads (Santa Cruz Biotechnologies) in binding buffer (0.25× SSPE, 1 mM EDTA pH 8.0, 0.05% Tween-20, 37.5 mM NaCl) along with 5 mM EDTA and heated to 65°C for 5 min. The reaction was cooled on ice for 2 min then rotated 1 hr at room temperature. Beads were washed for 3 minutes with binding buffer, low salt buffer (0.2× SSPE, 0.05% Tween-20, 1 mM EDTA pH 8), high salt buffer (0.2× SSPE, 0.05% Tween-20, 1 mM EDTA pH 8, 137.5 mM NaCl), and TET buffer (0.05% Tween-20 in TE buffer), twice. RNA was eluted twice in preheated elution buffer (0.15 M NaCl, 0.05 M Tris pH 7.5, 1 mM EDTA pH 8.0, 0.1% SDS, 20 mM DTT) by shaking at 42°C for 10 min and purified using TRizol LS and ethanol precipitation. End repair was conducted by incubating with 5 U RppH (NEB) in Thermopol buffer for 1 hr at 37°C, adding 10U PKN (NEB) and 10U MmMgCl2, incubating for 15 min at 37°C, adding 1.1X PKN buffer and 10U PKN, incubating for 15 min at 37°C, adding 1mM ATP and 10U PKN, and incubating for 30 min at 37°C. The reactions were stopped with 0.3M NaCl and 8.3M EDTA and RNA was purified using TRizol LS and ethanol precipitation. RNA was analyzed on the Bioanalyzer and sequencing libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep kit (New England Biolabs). Libraries were pooled and sequenced to 150bp length on the Illumina NovaSeq 6000.

**Analysis of GRO-seq data**

Raw sequence reads were trimmed and filtered using cutadapt (http://cutadapt.readthedocs.io/, version 1.2.0) (Martin, 2011), PCR duplicates were removed using Picard MarkDuplicates (version 2.0.1), and underwent quality control by FastQC program (Wingett and Andrews, 2018), and were mapped to the human (hg19) reference genome using Bowtie2 with default parameters (Langmead
et al., 2009). SAM files were converted to BAM files and sorted using Samtools (Li et al., 2009). Peak calling was performed using MACS2 with default parameters (Zhang et al., 2008). A visualized bigWig file, including fragment or read coverages for control and experimental datasets, was generated with the bedGraphToBigWig program. Sequencing tracks were viewed using the Integrated Genomic Viewer (IGV) (Robinson et al., 2011). Peak annotation and distribution analysis were performed using HOMER (Heinz et al., 2010). Sequence reads have been deposited in the NCBI GEO: GSE165049.

**Single cell RNA-seq analysis**

Single cells were generated using Chromium Controller (10x Genomics), and scRNA-seq libraries were constructed using chromium single cell 3’ reagent kits v2 (10x Genomics, California USA) according to the manufacturer’s recommendations. Briefly, $1 \times 10^4$ LK cells were loaded in each channel. Reverse transcription and library preparation were performed on C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad). Amplified cDNA and final libraries were examined with the Agilent BioAnalyzer using a High Sensitivity DNA Kit (Agilent Technologies). Differentially barcoded libraries were diluted to 4nM and pooled for sequencing with the NovaSeq 6000 Sequencing System (Illumina). Samples were sequenced with an average of 40,000 reads per cell.

Briefly, pooled samples were demultiplexed with cellranger mkfastq program (10x genomics), and demultiplexed FASTQ files were then performed the alignment with reference mm10 genome, filtering, barcode counting, and unique molecular identifier (UMI) counting with cellranger count under the default parameters (10x genomics). QC filtering was performed, and low complexity cell barcodes with number of genes detected were filtered out using the following parameters: percentage of reads mapping to the mitochondrial chromosome < 10%; UMI counts per cell > 500; and number of detected genes per cell (normalized counts > = 1) > 250. After filtering, we obtained a total of 9271 cells from WT LK group, a total of 9455 cells from Hottip-Tg LK group. Additionally, $12,056 \pm 54.92$ (mean ± SD) UMIs per cell and an average of $2785 \pm 60.11$ (mean ± SD) genes per cell were detected. Then, data normalization, integration, clustering and dimensionality reduction T-distributed stochastic neighbor embedding (tSNE) or uniform manifold approximation and projection (UMAP) was performed using ‘Seurat’ package (Butler et al., 2018) in R language. Once data were successfully integrated, principal component analysis (PCA) by running the RunPCA function were performed with default parameters, and then by running the FindNeighbors function with reduction = “umap”, dims = 1:30. Defining the clusters was manually assigned and curated on the basis of expressed genes previously reported (Giladi et al., 2018, Paul et al., 2015, Izzo et al., 2020). Additionally, differentially expressed genes were calculated using Bonferroni corrected Wilcoxon Sum-Rank Test as implemented in FindAllMarkers function (default parameters) of the ‘Seurat’ package with adjusted p values < 0.05. FeaturePlot, CoveragePlot and DotPlot were performed to show the specific gene expression in each cell and cluster using ‘Seurat’ package. Average expression level of the specific genes in different clusters was calculated with AverageExpression function in ‘Seurat’ package. We performed clustering of cells using louvain algorithm in SCANPY (Wolf et al., 2018). In addition, we carried out PAGA analysis using sc.tl.paga function to explain the connectedness of the clusters (Wolf et al., 2019). Trajectory inference analysis was performed with plot_cell_trajectory program in ‘Monocle’ package (Trapnell et al., 2014) and PAGA in Python (Wolf et al., 2019).

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Statistical differences were determined by Student’s t test or analysis of variance (ANOVA) followed by Newman-Keuls multiple comparison tests. * indicates p value less than 0.05, ** indicates p value less than 0.01, and *** indicates a p value less than 0.001. For in vitro experiments, at least three independent experiments with at least three biological replicates for each condition/genotype were carried out. For in vivo experiments, the sample size of at least 3-5 mice/group/genotype was chosen based on the generalized linear model with Bonferroni multiple comparison adjustments, and animals were randomly assigned to each study.