HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2022 April 06.

Published in final edited form as:

Mol Cell. 2022 February 17; 82(4): 833–851.e11. doi:10.1016/j.molcel.2022.01.014.

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

Huacheng Luo[1,12], **Ganqian Zhu**[2,12], **Melanie A. Eshelman**[1,12], **Tsz Kan Fung**[3,11], **Qian Lai**[1,4], **Fei Wang**[5], **Bernd B. Zeisig**[3,11], **Julia Lesperance**[1], **Xiaoyan Ma**[1,5], **Shi Chen**[2], **Nicholas Cesari**[1], **Christopher Cogle**[6], **Baoan Chen**[5], **Bing Xu**[4], **Feng-Chun Yang**[7,8], **Chi Wai Eric So**[3,11,*], **Yi Qiu**[9,10,*], **Mingjiang Xu**[2,7,*], **Suming Huang**[1,10,13,*]

[1]Division of Pediatric Hematology/Oncology, Department of Pediatrics, Pennsylvania State University College of Medicine, Hershey, PA 17033, USA

[2]Department of Molecular Medicine, the University of Texas Health Science Center at San Antonio, San Antonio, TX 78229-3904, USA

[3]School of Cancer and Pharmaceutical Science, King's College London, London SE5 9NU, UK

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

[5]Department of Hematology and Oncology, The Affiliated Zhongda Hospital, Southeast University Medical School, Nanjing 21009, China

[6]Division of Hematology/Oncology, Department of Medicine, University of Florida College of Medicine, Gainesville, FL 32610, USA

[7]Department of Cell System & Anatomy, the University of Texas Health Science Center at San Antonio, San Antonio, TX 78229-3904, USA

[8]Mays Cancer Center, Joe R. & Teresa Lozano Long School of Medicine, the University of Texas Health Science Center at San Antonio, San Antonio, TX 78229-3904, USA

[9]Department of Cellular and Molecular Physiology, Pennsylvania State University College of Medicine, Hershey, PA 17033, USA

[10]Penn State Cancer Institute, Pennsylvania State University College of Medicine, Hershey, PA 17033, USA

[11]Department of Haematological Medicine, King's College Hospital, London SE5 9RS, UK

[*]Correspondence: eric.so@kcl.ac.uk (C.W.E.S.), yqiu1@pennstatehealth.psu.edu (Y.Q.), xum1@uthscsa.edu (M.X.), shuang4@pennstatehealth.psu.edu (S.H.).

AUTHOR CONTRIBUTIONS

H.L., Y.Q., X.M., C.W.E.S., and S.H. conceived and designed experiments. H.L., M.A.E., Q.L., F.W., X.M., T.K.F., and B.B.Z. performed experiments. H.L. performed bioinformatics/statistical analysis. G.Z., J.L., T.K.F., B.B.Z., S.C., C.W.E.S., and M.X. established transplantation AML mouse models. N.C., B.X., B.C., F.-C.Y., and C.C. provided critical reagents and patient sample analyses. S.H., Y.Q., C.W.E.S., and M.X. wrote/edited the manuscript. S.H. supervised the studies.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

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

[12]These authors contributed equally

[13]Lead contact

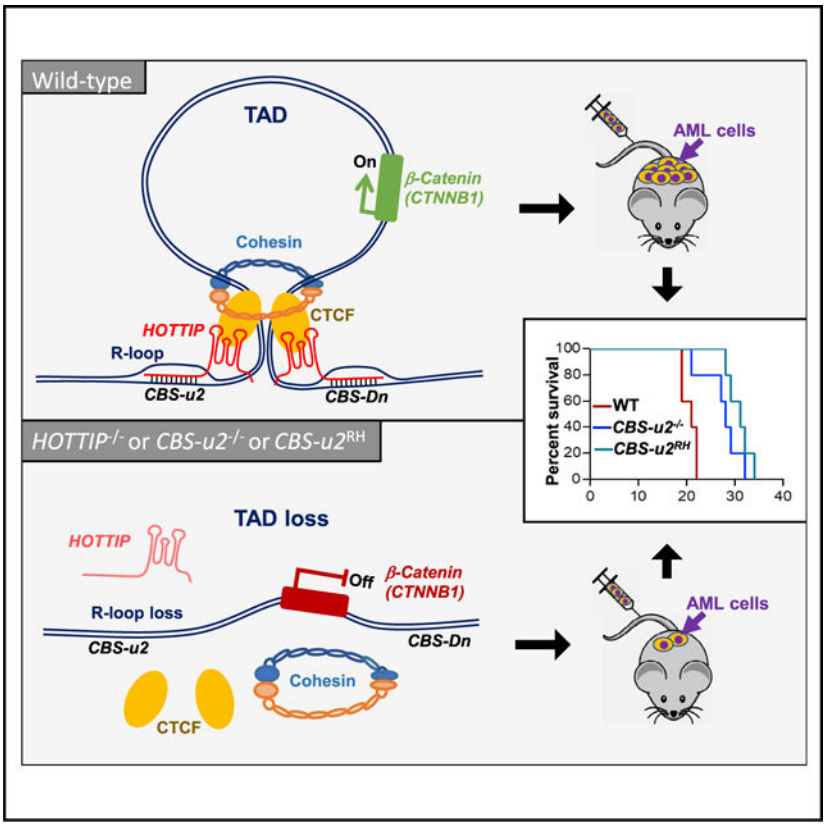
SUMMARY

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 (*CBS*s) 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.

In brief

In this article, Luo et al. discovered that *HOTTIP* lncRNA 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.

Graphical Abstract



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 (Skourti-Stathaki and Proudfoot, 2014; Aguilera and García-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 noncoding 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 (Cristini et al., 2018). However, the detailed mechanism by which R-loops affect genome topology and transcriptional regulation remains unknown.

HOTTIP is a HOXA-locus-associated lncRNA regulating transcription of the 5′ tip of HOXA genes (Wang et al., 2011; Luo et al., 2019). *HOTTIP* is specifically upregulated in

mixed-lineage leukemia (*MLL*) rearranged (*MLLr*$^{+}$) or nucleophosmin 1 (*NPM1*)-mutated (*NPM1*$^{C+}$) acute myeloid leukemia (AML), which correlates with poor survival in AML patients (Luo et al., 2019). Transgenic overexpression of *Hottip* increased hematopoietic stem cell (HSC) self-renewal leading to AML-like disease in mice by altering the HOXA topologically associated domain (TAD) and its 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; Jeannet 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 (*CBS*s) 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 lncRNAs to these CTCF functions remain unknown.

Here, we demonstrate that *HOTTIP* lncRNA 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 lncRNA associates with CTCF/cohesin complex and R-loop-associated proteins in AML cells

HOTTIP facilitates the posterior HOXA TAD to drive a homeotic transcription program partly by binding to annotated *CBS*s (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.

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 *CBS*s. 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 *CBS*s. 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-quad (G-4) sequence in the 5′ region of the *HOTTIP* lncRNA 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; Dumelie 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 *CBS*s and regulate CTCF boundary activity. DNA-RNA immunoprecipitation sequencings (DRIP-seq/DRIPc-seq) were 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 (≥1.5-fold, p ≤ 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 (≥1.5-fold, p ≤ 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 *CBS*s to mediate CTCF binding, which is exemplified at the MDS1 and EVI1 complex locus (*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 *CBS*s 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). 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^{-}Sca1^{+}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*, *Cd44*, *Nonog*, *Cdx4*, *Met*, *Twist*, *Runx1*, and *Wnt3/5a* (Figure 4A). Indeed, one of the primary 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 *CBS*s (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 HSC 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 helixes play critical roles in transcriptional regulation (Skourti-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 *CBS*s, *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 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 *CBS*s (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 *CBS*s 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–S5G). Notably, R-loops at these *CBS*s are not a result of nascent transcription, as there is no active transcription indicated by global run-on sequencing (GRO-seq) (Figures 5E, S5D, and S5F). Consistently, *HOTTIP*$^{-/-;}$ decreased H3K4me3, particularly in the promoters, and increased H3K27me3 within CTCF-defined TAD boundaries of these Wnt pathway loci (Figures 5F, S5E, and S5G). Thus, *HOTTIP*-mediated R-loops play a critical role in establishing and/or maintaining CTCF-defined TAD boundaries at Wnt target gene loci.

The *CBS-u2* boundary at the *CTNNB1* locus is critical to maintain TAD structure, transcription, and β-catenin-driven leukemogenesis

To investigate the role of the upstream *CTNNB1 CBS*s in leukemogenesis, we deleted *CBS-u1* and *CBS-u2* using CRISPR-Cas9 (Figures 6A, S6A, and S6B). Deletion of the boundary *CBS*s reduced CTCF binding to the respective *CBS* (Figure 6B). Strikingly, deletion of *CBS-u2*, which is in convergent orientation with *CBS-Dn*, but not *CBS-u1*, which is in tandem orientation, reduced *CTNNB1* expression (Figure 6C). Consistently, *CBS-u2*$^{-/-;}$ decreased β-catenin target gene expression, while leaving 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). Consequentially, *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* IL2Rgamma^{null} (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 chimerism 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 *NPM1*$^{C+}$;*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 *NPM1*$^{C+}$;*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 chimerism 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.

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^{D210N}, 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*$^{RH\text{-}WT}$, but not *CBS-u2*$^{RH\text{-}Mut}$, reduced *CTNNB1* levels and β-catenin target gene expression without affecting control genes, *LDB1* and *LMO2* (Figure 7B). Additionally, *CBS-u2*$^{RH\text{-}WT}$ impaired cell proliferation (Figure 7C). Hydrolyzing the R-loop reduced CTCF and *HOTTIP* binding 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*$^{RH\text{-}Mut}$ did not significantly affect CTCF or *HOTTIP* binding or R-loop formation any-where in the *CTNNB1* locus (Figures 7D–7F). Importantly, while *CBS-u2*$^{RH\text{-}WT}$ 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*$^{RH\text{-}WT}$ 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*$^{RH\text{-}WT}$ 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*$^{RH\text{-}WT}$ or *CBS-u2*$^{RH\text{-}Mut}$ MOLM13 cells were transplanted into NBSGW mice. WT and *CBS-u2*$^{RH\text{-}Mut}$ MOLM13 transplanted mice died 28–40 days post-transplantation; however, the *CBS-u2*$^{RH\text{-}WT}$ 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*$^{RH\text{-}Mut}$ cells (Figure S7H). Likewise, *CBS-u2*$^{RH\text{-}WT}$ or *CBS-u2*$^{RH\text{-}Mut}$ OCI-AML3 cells were transplanted into NSG mice. While *CBS-u2*$^{RH\text{-}Mut}$ OCI-AML3 transplanted mice died 24 days post-transplantation, *CBS-u2*$^{RH\text{-}WT}$ transplanted mice survived to 29 days and exhibited reduced hCD45^{+} engraftment (Figures 7I and S7I). Importantly, targeting *CBS-u2*$^{RH\text{-}WT}$, but not *CBS-u2*$^{RH\text{-}Mut}$, 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 lncRNA-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-splicing 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* decoys 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 lncRNAs 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 lncRNAs 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 Feldsenfeld'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 lncRNA action on CTCF-mediated TAD boundaries. In light of the discovery that *HOTTIP* interacts with *CBS* 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 boundaries and TAD integrity, particularly in the AML genome. It remains to be determined whether the same R-loop mechanism is employed by other lncRNAs to modulate CTCF-mediated genome organization for gene regulation. Furthermore, it is also important to determine whether lncRNA-driven R-loops at CTCF boundaries can be properly regulated to avoid persistent R-loop-conferred vulnerability to genomic insults. Association of R-loop sensors and regulators with *HOTTIP* (Figure 1) suggests that *HOTTIP* or other lncRNAs 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 *CBS*s 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 *CBS*s to induce R-loop structure in boundary sites; next, lncRNA-dependent R-loops at TAD *CBS*s 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 *CBS*s 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 coregulation 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 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 $NPM1^{C+}$, 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 *CBS*s 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 *CBS*s 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, lncRNA-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 (Angelbello 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 $NPM1^{C+}$-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

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, *NPM1*$^{C+}$*/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% CO_2 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.tRFP 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 pHR-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 pHR-dCas9 vector (Addgene #46911) as a Sbf1-BamHI restriction fragment. sgRNA targeting the *CTNNB1 CBS-u2* was subcloned into the pLKO5.sgRNA.EFS.tRFP 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 linage, 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 $MgCl_2$ 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, 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 $NaHCO_3$, 1% SDS), reverse crosslinked with 2 μL of 10 mg/mL proteinase 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 1mL 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 fresh) and hybridized with 100 pmole of biotinylated DNA probes targeting *HOTTIP* or LacZ, as a negative control. DNA/RNA hybrids were precipitated using 100μL 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 mm^3 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 μg/ml 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 overrepresented 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, 2×10^7 MOLM13 AML cells were collected, washed with cold 1x PBS, and then incubated with freshly prepared nuclear isolation buffer (1.28M sucrose, 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 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 pH7, 150 mM KCl, 5 mM $MgCl_2$, 5 mM EDTA, 0.5% IGEPAL-CA-630, 0.5 mM dithiothreitol, 0.2 mg/mL Heparin, 100 U/mL RNase OUT, 100 U/mL Superase IN, protease inhibitor tablet) 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 $MgCl_2$, 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, HOTAIRM1* 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 Hepes/KOH at pH 7.5; 0.14 M NaCl; 5 mM EDTA; 1% Triton X-100; 0.1% Na-Deoxycholate, ddH_2O) 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 Hepes/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 Hepes/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, treated with RNaseH at 37°C for 30 min, and analyzed by qPCR. DRIP-DNA libraries were prepared using the KAPA Hyper Prep Kit according to the manufacturer's instructions (KAPA, Cat # KK8500). Libraries were subjected to paired-end sequencing to 100 bp length on an Illumina HiSeq 2500. All genomics datasets were deposited in the NCBI GEO: GSE165049.

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 coverages for control and experimental datasets, was generated with the bedGraphToBigWig program (https://www.encodeproject.org/software/bedgraphtobigwig/). Sequencing tracks were viewed using the Integrated Genomic Viewer (IGV) (Robinson et al., 2011). Peak annotation, distribution, and *de novo* motif analysis were performed using HOMER (v4.10) (Heinz et al., 2010). Gene Ontology analysis was performed using DAVID (https://david.ncifcrf.gov, Version 6.8) (Huang et al., 2009a, 2009b). Differential peaks analysis was performed using the DiffBind (v3.2.5) and DESeq2 packages (Cut-off: ≥1.5 fold change; p value ≤ 0.05) in R (Ross-Innes et al., 2012). Overlapping peak analysis was carried out using the "bedtools intersect" program from bedtools (v2.29.2) (Quinlan and Hall, 2010). All datasets were deposited in the NCBI GEO: GSE165049.

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 (NEB) 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 RNA-se 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, 5×10^4 cells were washed with PBS twice and lysed in buffer containing 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 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 –nomodel –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/bedgraphtobigwig/). 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 (Schep et al., 2017, Rubin et al., 2019). Reproducibility between duplicateswas assessed using Pearson's correlation coefficient and Pearson's χ^2-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, 2×10^6 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 ddH_2O. 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 mg 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 oligo, and 1000 pmol Nimblegen HE Index specific blocking oligo 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 ultramers) 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 μl M270 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 (×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-Prkdc*^{scid}*Il2rg*^{tm1Wjl}/SzJ (NSG) mice (6–8 weeks old) were pretreated with 280cGy total body irradiation. Subsequently, irradiated NSG or untreated

adult NOD.Cg-*Kit*$^{W-41J}$*Tyr*$^{+}$*Prkdc*^{scid}*Il2rg*^{tm1Wjl}/ThomJ (NBSGW) mice (6–8 weeks old) were transplanted with 5×10^5 MOLM13, 5×10^5 OCI-AML3 or 1.8×10^5 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, 5×10^6 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://jgi.doe.gov/data-and-tools/bbtools/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 correlation 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) (Durand et al., 2016b), and 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 (5Cy5/CGC AAC CAG GCG GGG AGG GGA GGT GGG CGC GCG), a mutant sequence (Mut) (5Cy5/CGC AAC CAG GCG TTT ATT TGA GGT GGG CGC GCG) and the *HOTTIP* 3′UTR region (*HOTTIP*-3′UTR) (/5Cy5/AAA GTT GCA GTC CAG CCA GCA TGA GAA CTG CCA TG), and Cy3-labeled duplexed DNA oligonucleotides, including the *CBS-u2* region (/5Cy3/CCT CAG CAT GGT CTT CAG CAG CTG GGG GAG GGG CCA GCC TTG CAT CTT C), the *MYC-3′CBS* region (/5Cy3/TGG GTA GCA GGA TGG TGA CTG CTC AGG GGA GGG GCA GAC CCC ACT CAG ACT GGA TGA AGA), the *HOXA-CBS4/5* region (/5Cy3/CAT TTG TTA GGC GAT GAC AGA CTT CAC CTC CAG CAA GGG CTG CTT CAC A), and the *ACTB-CBS* region (/5Cy3/GGA GCC GTT GTC GAC GAC GAG CGC GGC GAT ATC ATC ATC CAT GGT GAG C). 2.5 pmole of each Cy3-labeled duplex DNA oligo was incubated with 5 pmole of each individual Cy5-labeled single strand *HOTTIP* RNA oligo in 10 μl of R-loop binding buffer (90 mM Tris-acetate pH 6.5, 10 mM $MgCl_2$, 100 μg/mL yeast tRNA and 0.5 mM spermine tetrahydrochloride) at room temperature for 2 hrs. Subsequently, dsDNA/ssRNA hybridization mixtures were treated with 7 U RNase H supplemented with 1 mM DTT, 7 U DNaseI, or 7 U RNaseA at room temperature for 30 min, followed by digestion with 10 μg proteinase K at room temperature for 1 hour. Hybrid were resolved on 8% native polyacrylamide gels containing 90 mM Tris-acetate pH 6.5 (or pH 5.0 for triplexes formed at acidic pH) and 1 mM $MgCl_2$ and visualized using the Odyssey CLx Imaging System (LI-COR). For gel shift assays, dsDNA/ssRNA hybridization products were incubated with S9.6 or IgG antibody at 4°C overnight, precipitated with protein A/G agarose beads at 4°C for 2hr, washed 3 times with wash buffer (10 mM Tris-HCl pH 8, 150 mM NaCl, 0.1% Triton X-100, 0.5% Na-Deoxycholate, 1 mM EDTA pH 8), and eluted in fresh elution buffer (1%SDS/0.1M $NaHCO_3$ pH8.0) at room temperature for 30 min with gentle rotation. Eluates were centrifuged at 13000 rpm for 5 min and the supernatant was collected and treated as indicated with RNaseH, DNaseI, or RNaseA. Finally complexes were treated with proteinase K and analyzed by electrophoresis using 8% native polyacrylamide gels containing 90 mM Tris-acetate pH 6.5 and 1 mM MgCl2. Samples were resolved at room temperature and visualized by fluorescence imaging technology using the Odyssey CLx Imaging System.

Global Run-On sequencing (GRO-seq)—GRO-seq experiment was carried out as previously described (Gardini, 2017, Barbieri et al., 2020). Briefly, 1×10^7-1×10^8 cells were washed twice with ice-cold PBS, swelled in swelling buffer (10 mM Tris-HCL pH 7.5, 2mM $MgCl_2$, 3 mM $CaCl_2$, 2U/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 $MgCl_2$, 3 mM $CaCl_2$, 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 $MgCl_2$, 0.1 mM 0.5M EDTA, 50 mM Tris-HCL pH 8, 2 U/ml Superase-in). Nuclei were counted, snap frozen in 100μL aliquots at a concentration of 1×10^6 nuclei/10μL, 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 mM $MgCl_2$, 300 mM KCl, 1 mM DTT, 500 μM ATP, 500 μM GTP, 500 μM Br-UTP, 2 μM CTP, 200 U/ml 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, 1mM EDTA pH 8.0, 0.05 % Tween-20, 37.5 mM NaCl) along with 5mM EDTA and heated to 65°C for 5m. 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 (0.2X SSPE, 0.05% Tween-20, 1mM EDTA pH 8), high salt buffer (0.2X SSPE, 0.05% Tween-20, 1mM EDTA pH 8, 137.5mM 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, 1mM EDTA pH 8.0, 0.1 % SDS, 20mM DTT) by shaking at 42°C for 10 min and purified using TRIzol LS and ethanol precipitation. End repair was conducted by incubating RNA with 5U RppH (NEB) in Thermopol buffer for 1 hr at 37°C, adding 10U PNK (NEB) and 10mM $MgCl_2$, incubating for 15 min at 37°C, adding 1.1X PNK buffer and 10U PNK, incubating for 15 min at 37°C, adding 1mM ATP and 10U PNK, and incubating for 30 min at 37°C. The reactions were stopped with 0.3M NaCl and 8.3mM 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, 1×10^4 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 ± 54.92 (mean ± SD) UMIs per cell and an average of 2785 ± 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 = "pca" and dims = 1:30, followed by running FindClusters function. For visualization, a Uniform Manifold Approximation and Projection (UMAP) cell embedding was generated using the RunUMAP function with the following parameters: 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 20002 (C.W.E.S.), a Cancer Prevention/Research Institute of Texas grant (RP200242, to M.X.), and the Four Diamonds Fund (S.H.).

REFERENCES

Agrotis A, Pengo N, Burden JJ, and Ketteler R. (2019). Redundancy of human ATG4 protease isoforms in autophagy and LC3/GABARAP processing revealed in cells. Autophagy 15, 976–997. [PubMed: 30661429]

Aguilera A, and García-Muse T. (2012). R loops: from transcription by products to threats to genome stability. Mol. Cell 46, 115–124. [PubMed: 22541554]

Angelbello AJ, Chen JL, Childs-Disney JL, Zhang P, Wang ZF, and Disney MD (2018). Using genome sequence to enable the design of medicines and chemical probes. Chem. Rev 118, 1599–1663. [PubMed: 29322778]

Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, Mariappan K, Latrasse D, Blein T, Liu C, et al. (2020). R-loop mediated trans action of the APOLO long noncoding RNA. Mol. Cell 77, 1055–1065.e4. [PubMed: 31952990]

Barbieri E, Hill C, Quesnel-Vallières M, Zucco AJ, 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 Guil S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc. Natl. Acad. Sci. USA 112, 5785–5790. [PubMed: 25902512]

Buckle A, Gilbert N, Marenduzzo D, and Brackley CA (2019). capC-MAP: software for analysis of Capture-C data. Bioinformatics 35, 4773–4775. [PubMed: 31173058]

Buenrostro JD, Wu B, Chang HY, and Greenleaf WJ (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. [PubMed: 29608179]

Chakraborty P, Huang JTJ, and Hiom K. (2018). DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. Nat. Commun 9, 4346. [PubMed: 30341290]

Chen L, Chen JY, Huang YJ, Gu Y, Qiu J, Qian 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. [PubMed: 29395063]

Chen L, Chen JY, 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. [PubMed: 29104020]

Chen PB, Chen HV, Acharya D, Rando OJ, and Fazzio TG (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat. Struct. Mol. Biol 22, 999–1007. [PubMed: 26551076]

Chu C, Qu K, Zhong FL, Artandi SE, and Chang HY (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell 44, 667–678. [PubMed: 21963238]

Chu C, Zhang QC, Da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, and Chang HY (2015). Systematic discovery of Xist RNA binding proteins. Cell 161, 404–416. [PubMed: 25843628]

Cobas M, Wilson A, Ernst B, Mancini SJ, Macdonald HR, Kemler R, and Radtke F. (2004). Beta-catenin is dispensable for hematopoiesis and lymphopoiesis. J. Exp. Med 199, 221–229. [PubMed: 14718516]

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962. [PubMed: 28846090]

Cristini A, Groh M, Kristiansen MS, and Gromak N. (2018). RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. Cell Rep. 23, 1891–1905. [PubMed: 29742442]

Cuddapah S, Jothi R, Schones DE, Roh TY, 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. [PubMed: 19056695]

Davies JO, Telenius JM, Mcgowan SJ, Roberts NA, Taylor S, Higgs DR, and Hughes JR (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods 13, 74–80. [PubMed: 26595209]

De Hoon MJ, Imoto S, Nolan J, and Miyano S. (2004). Open source clustering software. Bioinformatics 20, 1453–1454. [PubMed: 14871861]

De Magis A, Manzo SG, Russo M, Marinello J, Morigi 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, 816–825. [PubMed: 30591567]

De Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, and De Laat W(2015). CTCF binding polarity determines chromatin looping. Mol. Cell 60, 676–684. [PubMed: 26527277]

Deininger MWN, Tyner JW, 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, Bungert J, Qiu Y, and Huang S. (2016). HoxBlinc RNA recruits Set1/MLL complexes to activate hox gene expression patterns and mesoderm lineage development. Cell Rep. 14, 103–114. [PubMed: 26725110]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, and Ren B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380. [PubMed: 22495300]

Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN,Weintraub AS, Schujiers J, Lee TI, Zhao K, and Young RA (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159, 374–387. [PubMed: 25303531]

Dumelie JG, and Jaffrey SR (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. Elife 6, e28306.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, and Aiden EL (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101. [PubMed: 27467250]

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, and Aiden EL (2016b). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98. [PubMed: 27467249]

Eden E, Navon R, Steinfeld I, Lipson D, and Yakhini Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48. [PubMed: 19192299]

Eng JK, Mccormack AL, and Yates JR (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. [PubMed: 24226387]

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, and Mirny LA (2016). Formation of chromosomal domains by loop extrusion. Cell Rep. 15, 2038–2049. [PubMed: 27210764]

Gardini A. (2017). Global run-on sequencing (GRO-seq). Methods Mol. Biol 1468, 111–120. [PubMed: 27662873]

Giladi A, Paul F, Herzog Y, Lubling Y, Weiner A, Yofe I, Jaitin D, Cabezas-Wallscheid N, Dress R, Ginhoux F, et al. (2018). Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. Nat. Cell Biol 20, 836–846. [PubMed: 29915358]

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451. [PubMed: 23849981]

Gröschel S, Schlenk RF, Engelmann J, Rockova V, Teleanu V, Kühn MW, Eiwen K, Erpelinck 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. [PubMed: 23008312]

Grunseich C, Wang IX, Watts JA, Burdick JT, Guber RD, Zhu Z, Bruzel A, Lanman T, Chen K, Schindler AB, et al. (2018). Senataxin mutation reveals how R-loops promote transcription by blocking DNA methylation at gene promoters. Mol. Cell 69, 426–437.e7. [PubMed: 29395064]

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell 162, 900–910. [PubMed: 26276636]

Halász L, Karányi Z, Boros-Oláh B, Kuik-Rózsa T, Sipos É, Nagy É, Mosolygó-L Á, Mázló A, Rajnavölgyi É, Halmos G, and Szekvolgyi L. (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. Genome Res. 27, 1063–1073. [PubMed: 28341774]

Hansen AS, Hsieh TS, Cattoglio C, Pustova I, Saldaña-Meyer R, Reinberg D, Darzacq X, and Tjian R. (2019). Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. Mol. Cell 76, 395–411.e13. [PubMed: 31522987]

Heckl D, Kowalczyk MS, Yudovich D, Belizaire R, Puram RV, Mcconkey ME, Thielke A, Aster JC, Regev A, and Ebert BL (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. Nat. Biotechnol 32, 941–946. [PubMed: 24952903]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]

Huang DW, Sherman BT, and Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13. [PubMed: 19033363]

Huang DW, Sherman BT, and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc 4, 44–57. [PubMed: 19131956]

Hughes JR, Roberts N, Mcgowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R, and Higgs DR (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet 46, 205–212. [PubMed: 24413732]

Izzo F, Lee SC, Poran A, Chaligne R, Gaiti F, Gross B, Murali RR, Deochand SD, Ang C, Jones PW, et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat. Genet 52, 378–387. [PubMed: 32203468]

Jeannet 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. [PubMed: 17906078]

Jenjaroenpun P, Wongsurawat T, Sutheeworapong S, and Kuznetsov VA (2017). R-loopDB: a database for R-loop forming sequences (RLFS) and R-loops. Nucleic Acids Res. 45, D119–D127. [PubMed: 27899586]

Jenjaroenpun P, Wongsurawat T, Yenamandra SP, and Kuznetsov VA (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. [PubMed: 26400173]

Kirstetter P, Anderson K, Porse BT, Jacobsen SE, 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. [PubMed: 16951689]

Kolligs FT, Hu G, Dang CV, and Fearon ER (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. [PubMed: 10409758]

Kung JT, Kesner B, An JY, Ahn JY, Cifuentes-Rojas C, Colognori D, Jeon Y, Szanto A, Del Rosario BC, Pinter SF, et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol. Cell 57, 361–375. [PubMed: 25578877]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. [PubMed: 19261174]

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. [PubMed: 19505943]

Li Y, Liao Z, Luo H, Benyoucef A, Kang Y, Lai Q, Dovat S, Miller B, 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. [PubMed: 32086528]

Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, 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. [PubMed: 23064439]

Lu WT, Hawley BR, Skalka GL, Baldock RA, Smith EM, Bader AS, Malewicz M, Watts FZ, 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. [PubMed: 29416038]

Luis TC, Naber BA, Roozen PP, Brugman MH, De Haas EF, Ghazvini M, Fibbe WE, Van Dongen JJ, Fodde R, and Staal FJ (2011). Canonical wnt signaling regulates hematopoiesis in a dosage-dependent fashion. Cell Stem Cell 9, 345–356. [PubMed: 21982234]

Luo H, Wang F, Zha J, Li H, Yan B, Du Q, Yang F, Sobh 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. [PubMed: 29760161]

Luo H, Zhu G, Xu J, Lai Q, Yan B, Guo Y, Fung TK, Zeisig BB, Cui Y, Zha J, et al. (2019). HOTTIP lncRNA promotes hematopoietic stem cell self-renewal leading to AML-like disease in mice. Cancer Cell 36, 645–659.e8. [PubMed: 31786140]

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012–1025. [PubMed: 25959774]

Manachai N, Saito Y, Nakahata S, Bahirvani AG, Osato M, and Morishita K. (2017). Activation of EVI1 transcription by the LEF1/β-catenin complex with p53-alteration in myeloid blast crisis of chronic myeloid leukemia. Biochem. Biophys. Res. Commun 482, 994–1000. [PubMed: 27908728]

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10–12.

Mcdonald CD, and Maher LJ 3rd. (1995). Recognition of duplex DNA by RNA polynucleotides. Nucleic Acids Res. 23, 500–506. [PubMed: 7533903]

Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, and Reinberg D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 347, 1017–1021. [PubMed: 25722416]

Nguyen HD, Leong WY, Li W, Reddy PNG, Sullivan JD, Walter MJ, Zou L, and Graubert TA (2018). Spliceosome mutations induce R loop-associated sensitivity to ATR inhibition in myelodysplastic syndromes. Cancer Res. 78, 5363–5374. [PubMed: 30054334]

Noordermeer SM, Sanders MA, Gilissen C, Tönnissen E, Van Der Heijden A, Döhner K, Bullinger L, Jansen JH, Valk PJ, and Van Der Reijden BA (2011). High BRE expression predicts favorable outcome in adult acute myeloid leukemia, in particular among MLL-AF9-positive patients. Blood 118, 5613–5621. [PubMed: 21937695]

Park HJ, Li J, Hannah R, Biddie S, Leal-Cervantes AI, Kirschner K, Flores Santa Cruz D, Sexl V, Göttgens B, and Green AR (2016). Cytokine-induced megakaryocytic differentiation is regulated by genome-wide loss of a uSTAT transcriptional program. EMBO J. 35, 580–594. [PubMed: 26702099]

Patel B, Kang Y, Cui K, Litt M, Riberio MS, Deng C, Salz T, Casada S, Fu X, Qiu Y, et al. (2014). Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia. Leukemia 28, 349–361. [PubMed: 23698277]

Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, et al. (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell 163, 1663–1677. [PubMed: 26627738]

Peng J, and Gygi SP (2001). Proteomics: the move to mixtures. J. Mass Spectrom 36, 1083–1091. [PubMed: 11747101]

Perry JM, He XC, Sugimura R, Grindley JC, Haug JS, Ding S, and Li L. (2011). Cooperation between both Wnt/{beta}-catenin and PTEN/PI3K/Akt signaling promotes primitive hematopoietic stem cell self-renewal and expansion. Genes Dev. 25, 1928–1942. [PubMed: 21890648]

Phillips JE, and Corces VG (2009). CTCF: master weaver of the genome. Cell 137, 1194–1211. [PubMed: 19563753]

Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L, Schmitt N, Dold A, Ginsberg D, and Grummt I. (2015). LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. Mol. Cell 60, 626–636. [PubMed: 26590717]

Qi Q, Cheng L, Tang X, He Y, Li Y, Yee T, Shrestha D, Feng R, Xu P, Zhou X, et al. (2021). Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. Blood 137, 1327–1339. [PubMed: 33512425]

Qiu Y, and Huang S. (2020). CTCF-mediated genome organization and leukemogenesis. Leukemia 34, 2295–2304. [PubMed: 32518417]

Qiu Y, Xu M, and Huang S. (2021). Long noncoding RNAs: emerging regulators of normal and malignant hematopoiesis. Blood 138, 2327–2336. [PubMed: 34482397]

Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. [PubMed: 20110278]

Ramírez F, Dündar F, Diehl S, Grüning BA, and Manke T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191. [PubMed: 24799436]

Ren G, Jin W, Cui K, Rodrigez J, Hu G, Zhang Z, Larson DR, and Zhao K. (2017). CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. Mol. Cell 67, 1049–1058.e6. [PubMed: 28938092]

Reya T, Duncan AW, Ailles L, Domen J, Scherer DC, Willert K, Hintz L, Nusse R, and Weissman IL (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature 423, 409–414. [PubMed: 12717450]

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, and Mesirov JP (2011). Integrative genomics viewer. Nat. Biotechnol 29, 24–26. [PubMed: 21221095]

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393. [PubMed: 22217937]

Rowley MJ, and Corces VG (2018). Organizational principles of 3D genome architecture. Nat. Rev. Genet 19, 789–800. [PubMed: 30367165]

Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, Mumbach MR, Ji AL, Kim DS, Cho SW, et al. (2019). Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. Cell 176, 361–376.e17. [PubMed: 30580963]

Saldaña-Meyer R, Rodriguez-Hernaez J, Escobar T, Nishana M, Jácome-López K, Nora EP, Bruneau BG, Tsirigos A, Furlan-Magaril M, Skok J, and Reinberg D. (2019). RNA interactions are essential for CTCF-mediated genome organization. Mol. Cell 76, 412–422.e5. [PubMed: 31522988]

Saldanha AJ (2004). Java TreeView—extensible visualization of microarray data. Bioinformatics 20, 3246–3248. [PubMed: 15180930]

Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA 112, E6456–E6465. [PubMed: 26499245]

Sanjana NE, Shalem O, and Zhang F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784. [PubMed: 25075903]

Sanz LA, and Chédin F. (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat. Protoc 14, 1734–1755. [PubMed: 31053798]

Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, and Chédin F. (2016). Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol. Cell 63, 167–178. [PubMed: 27373332]

Scheller M, Huelsken J, Rosenbauer F, Taketo MM, Birchmeier W, Tenen DG, and Leutz A. (2006). Hematopoietic stem cell and multilineage defects generated by constitutive beta-catenin activation. Nat. Immunol 7, 1037–1047. [PubMed: 16951686]

Schep AN, Wu B, Buenrostro JD, and Greenleaf WJ (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods 14, 975–978. [PubMed: 28825706]

Shevchenko A, Wilm M, Vorm O, and Mann M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem 68, 850–858. [PubMed: 8779443]

Singh S, Ahmed D, Dolatshad H, Tatwavedi D, Schulze U, Sanchi A, Ryley S, Dhir A, Carpenter L, Watt SM, et al. (2020). SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications. Leukemia 34, 2525–2530. [PubMed: 32076118]

Siriboonpiputtana T, Zeisig BB, Zarowiecki M, Fung TK, Mallardo M, Tsai CT, Lau PNI, Hoang QC, 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. [PubMed: 28978671]

Skourti-Stathaki K, and Proudfoot NJ (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 28, 1384–1396. [PubMed: 24990962]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. [PubMed: 16199517]

Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, and Dean C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science 340, 619–621. [PubMed: 23641115]

Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B, et al. (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell 163, 1611–1627. [PubMed: 26686651]

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol 32, 381–386. [PubMed: 24658644]

Trapnell C, Pachter L, and Salzberg SL (2009). TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25, 1105–1111. [PubMed: 19289445]

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, and Pachter L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc 7, 562–578. [PubMed: 22383036]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, 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. [PubMed: 20436464]

Tummala H, Dokal AD, Walne A, Ellison A, Cardoso S, Amirthasigamanipillai S, Kirwan M, Browne I, Sidhu JK, 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. [PubMed: 29987015]

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. [PubMed: 24255920]

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120–124. [PubMed: 21423168]

Wang Y, Krivtsov AV, Sinha AU, North TE, Goessling W, Feng Z, Zon LI, and Armstrong SA (2010). The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. Science 327, 1650–1653. [PubMed: 20339075]

Wingett SW, and Andrews S. (2018). FastQ Screen: a tool for multi-genome mapping and quality control. F1000Res 7, 1338. [PubMed: 30254741]

Wolf FA, Angerer P, and Theis FJ (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15. [PubMed: 29409532]

Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, and Theis FJ (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 20, 59. [PubMed: 30890159]

Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, and Grüning BA (2020). Galaxy HiCExplorer 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. [PubMed: 32301980]

Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, 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. [PubMed: 20966046]

Yeung J, Esposito MT, Gandillet A, Zeisig BB, Griessinger E, Bonnet D, and So CW (2010). Beta-catenin mediates the establishment and drug resistance of MLL leukemic stem cells. Cancer Cell 18, 606–618. [PubMed: 21156284]

Yusufzai TM, 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. [PubMed: 14759373]

Zeisig BB, Fung TK, Zarowiecki M, Tsai CT, Luo H, Stanojevic B, Lynn C, Leung AYH, Zuna J, Zaliova 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, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]

Zhou Z, Giles KE, 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. [PubMed: 30867287]

Zlatanova J, and Caiafa P. (2009). CCCTC-binding factor: to loop or to bridge. Cell. Mol. Life Sci 66, 1647–1660. [PubMed: 19137260]

Highlights

- ***HOTTIP*** **recruits CTCF/cohesin and R-loop regulators to form R-loops at TAD boundaries**
- R-loops reinforce *HOTTIP*-associated CTCF boundaries and facilitate TAD formation
- R-loop-dependent CTCF boundaries coordinate oncogenic Wnt pathway in AML
- Loss of *HOTTIP*-driven R-loops impairs TAD topology, leading to mitigating AML severity

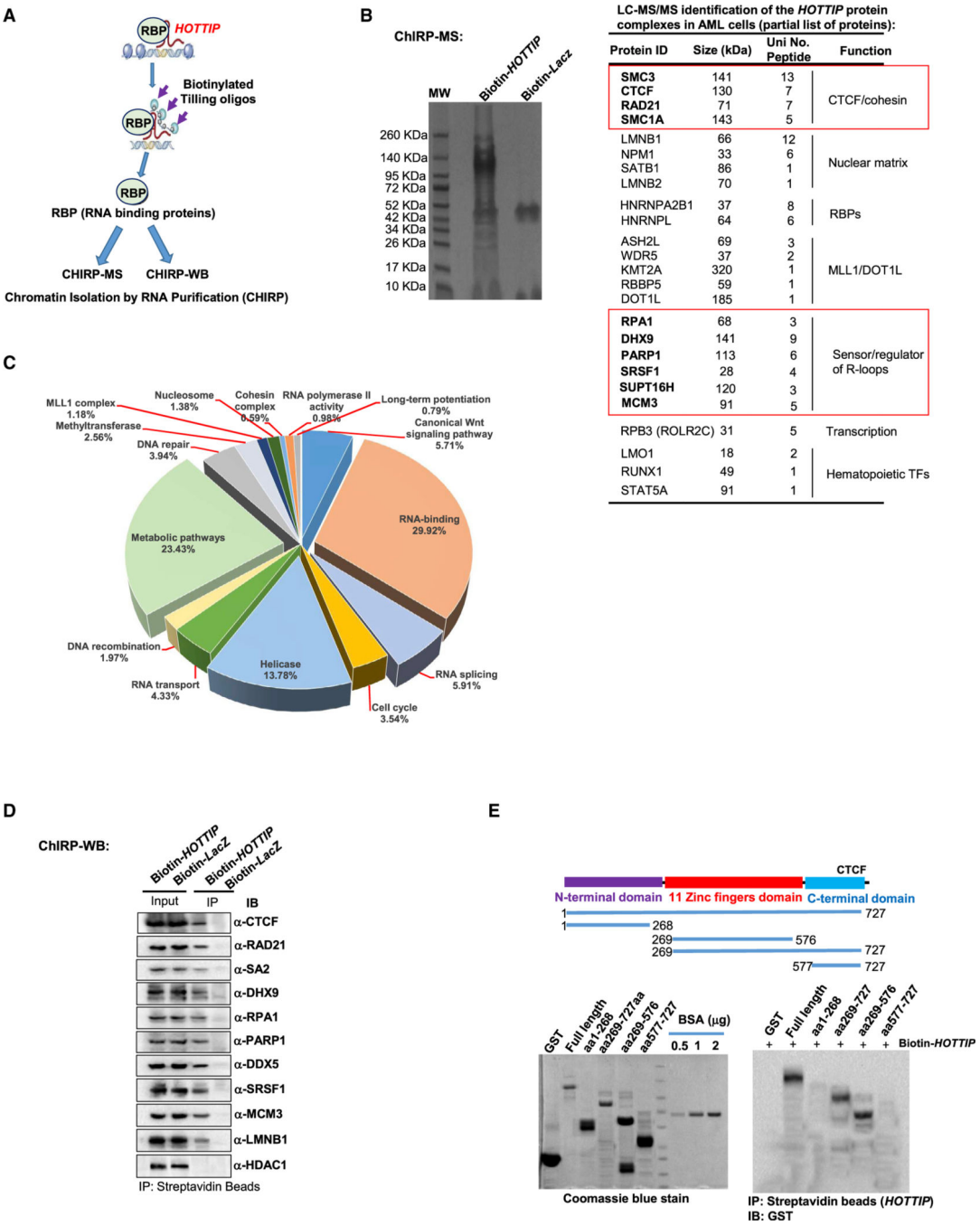


Figure 1. *HOTTIP* interactome isolated from AML cells contains CTCF/cohesin complex and R-loop-associated proteins

(A) Schematic of *HOTTIP* ChIRP-LC-MS/MS/WB workflow.

(B) Left: one-step blue protein gel staining of proteins isolated from MOLM13 cells by ChIRP using *HOTTIP* or control LacZ titling probes. Right: partial list of unique polypeptides identified in the *HOTTIP*, but not in the LacZ ChIRP-LC-MS/MS.

(C) Overrepresentation analysis of enriched *HOTTIP*-interacting protein classes using the DAVID database. The ratio of each protein class represented in the *HOTTIP*-associated

proteome was calculated and statistical significance was ranked according to the Benjamini-Hochberg-corrected p value ($p \leq 0.05$).

(D) ChIRP-WB validation of *HOTTIP*-associated proteins.

(E) Biotin pull-down of CTCF mutants by biotinylated *HOTTIP*.

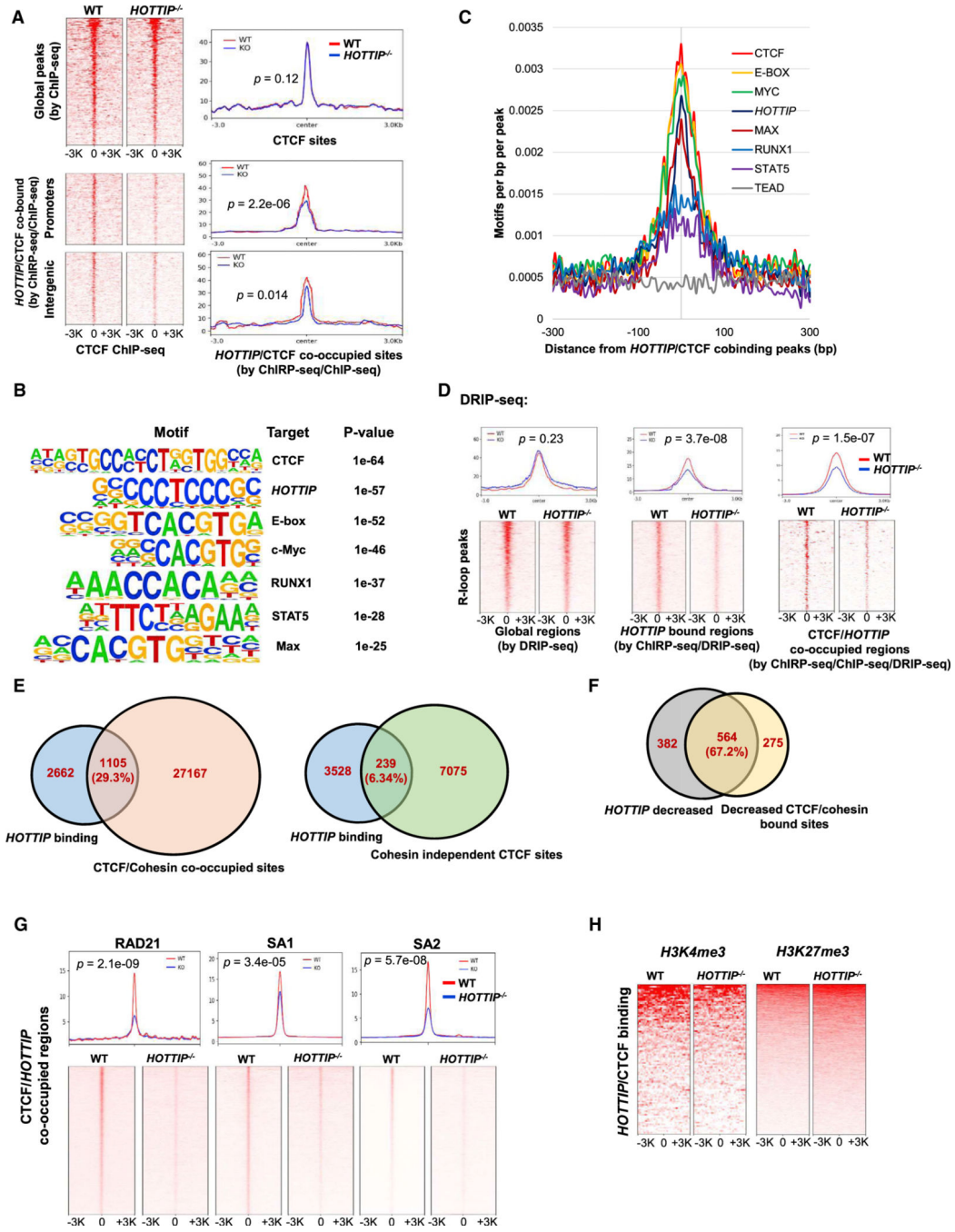


Figure 2. *HOTTIP* directly binds a subset of CTCF boundaries by formation of R-loops in AML genome

(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 ChIRP-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.

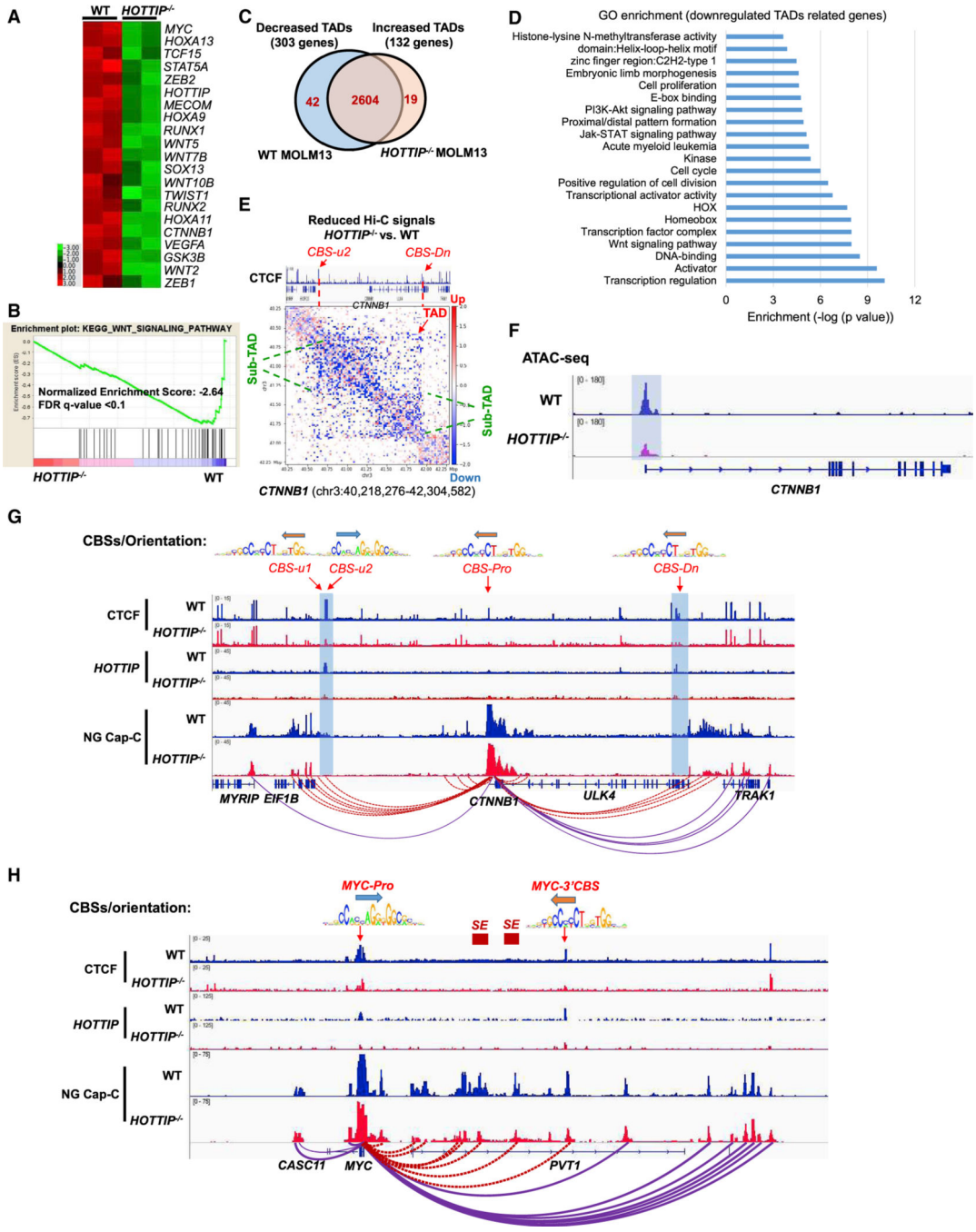


Figure 3. *HOTTIP* regulates CTCF-defined TADs and transcription at canonical Wnt loci in AML genome

(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* ChIRP-seq in WT and *HOTTIP*$^{-/-}$ MOLM13 cells. Solid purple and dashed red lines indicate unchanged and reduced interactions, respectively.

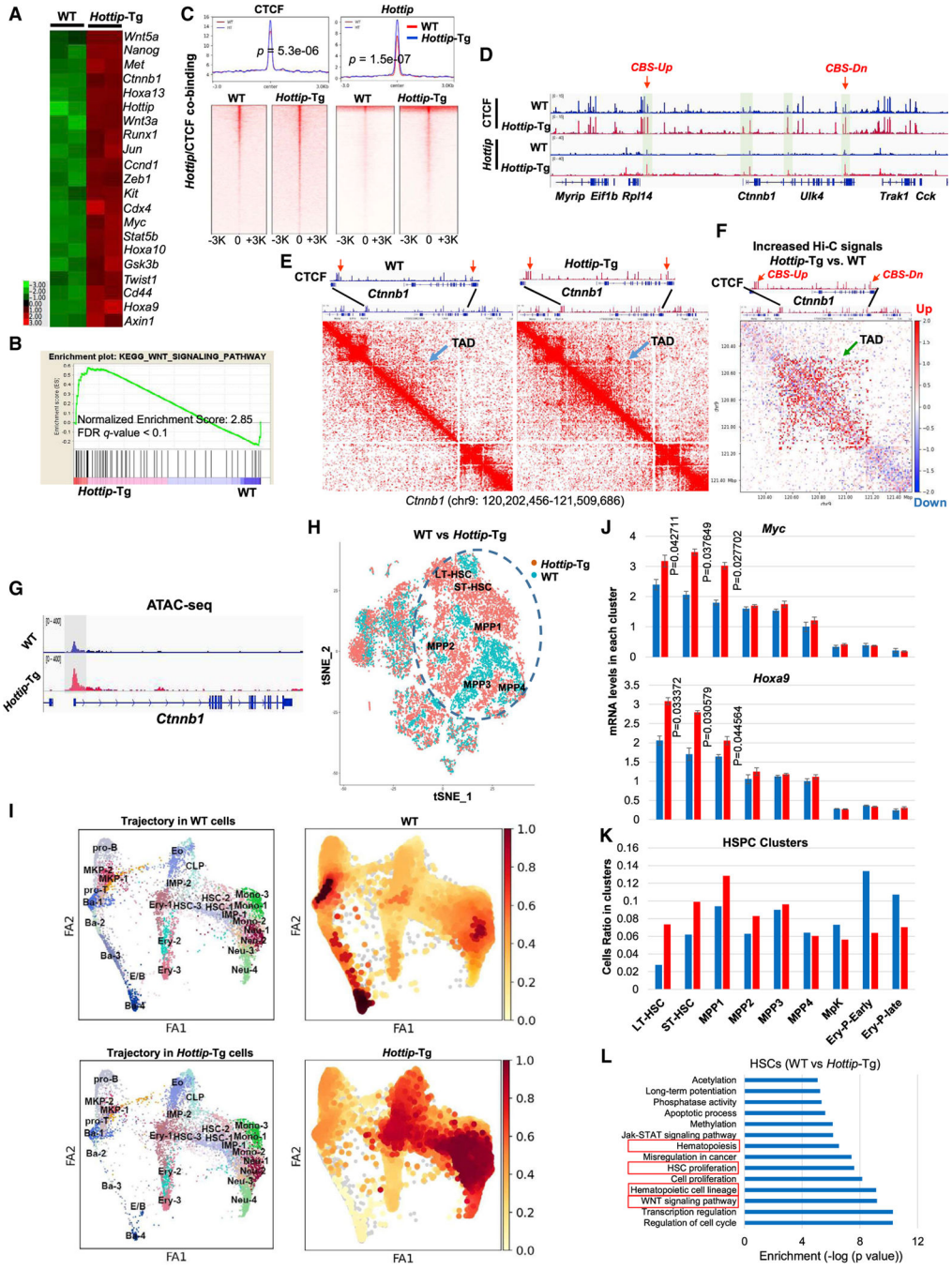


Figure 4. *Hottip* activation perturbs β-catenin and its target TADs leading to aberrant HSC activities

(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 *Ctnnb1* locus in WT and *Hottip*-Tg LK cells.

(E) Hi-C interaction maps at the *Ctnnb1* 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 *Ctnnb1* 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.

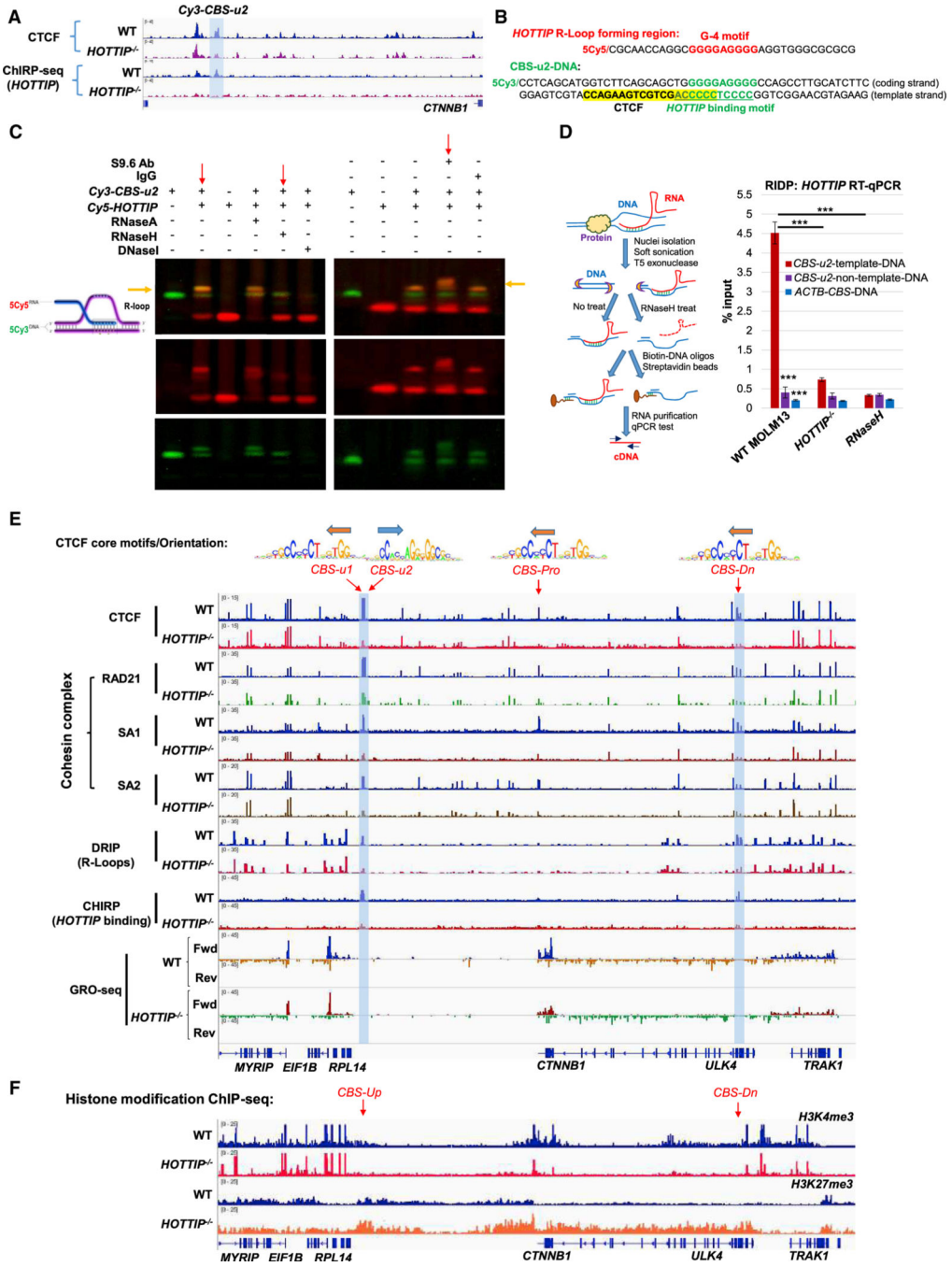


Figure 5. *HOTTIP* recognizes and accesses specific *HOTTIP*/CTCF co-occupied motifs via sequence complementarity to form R-loop structure

(A) CTCF ChIP-seq and *HOTTIP* ChIRP-seq-binding profiles at the upstream *CBS*s of *CTNNB1* in WT and *HOTTIP*$^{-/-}$ MOLM13 cells.

(B) Schematic of probes used in EMSA. The G-4 sequence in *HOTTIP*, the C-rich *HOTTIP*-binding motif in the *CBS-u2* sequence, and the CTCF-binding motif are shown in red, green, and yellow highlighting, respectively.

(C) EMSA of a Cy5-labeled *HOTTIP* RNA probe (red) and a Cy3-labeled *CTNNB1 CBS-u2* probe (green). Orange indicates R-loop formation. RNase A, RNase H, DNase I, and S9.6 antibody added as indicated.

(D) Schematic of RIDP procedure (left). RIDP RT-qPCR of *HOTTIP* precipitated by probes targeting the template or non-template strand of the *CBS-u2* site (in relation to the direction of *CTNNB1* transcription) or targeting the negative control *CBS* at *ACTB* in WT, *HOTTIP*$^{-/-}$; or RNase H treatment (right). Data presented as mean ± SD; ***$p \leq 0.001$.

(E) CTCF and cohesin-binding ChIP-seq, R-loop DRIP-seq, *HOTTIP*-binding ChIRP-seq and nascent RNA GRO-seq profiles at the *CTNNB1* locus in WT and *HOTTIP*$^{-/-}$; MOLM13 cells.

(F) H3K4me3 and H3K27me3 ChIP-seq enrichment profiles at the *CTNNB1* locus in WT and *HOTTIP*$^{-/-}$; MOLM13 cells.

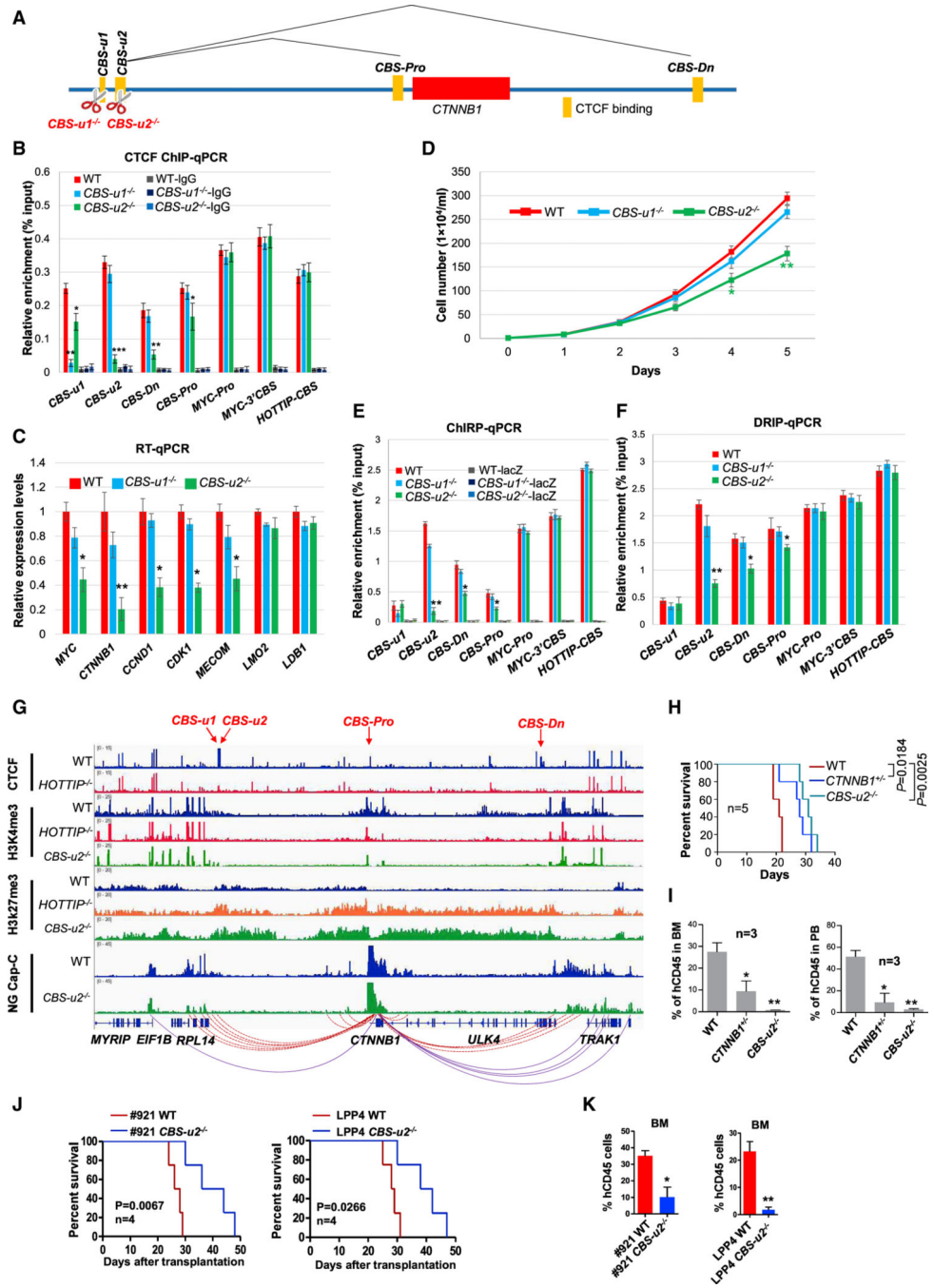


Figure 6. The *CBS-u2* boundary at the *CTNNB1* locus is critical to maintain TAD structure, transcription, and β-catenin-driven leukemogenesis

(A) Schematic of the *CTNNB1* locus showing the locations of *CBS*s, 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* ChIRP-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 chimerism 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.

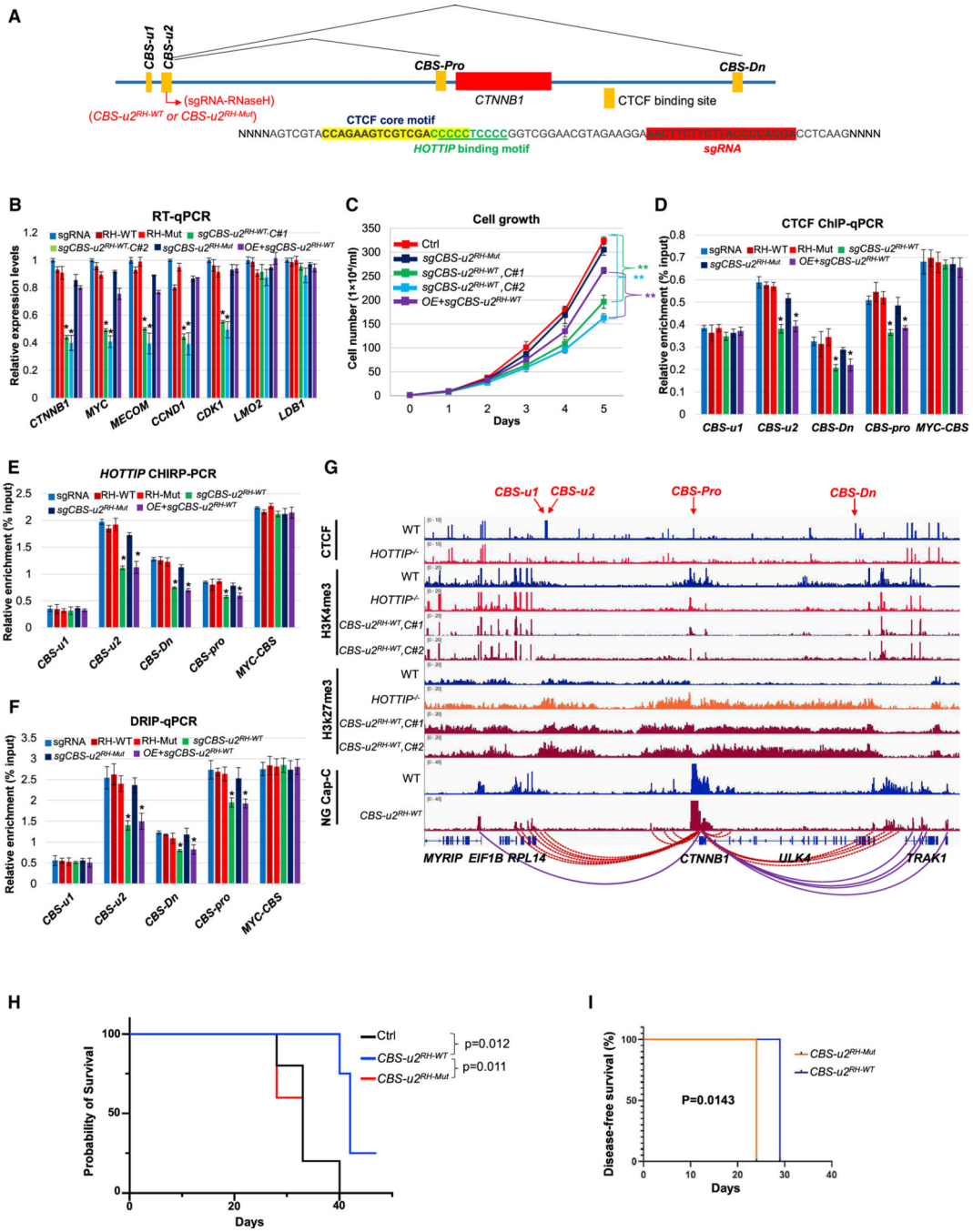


Figure 7. *HOTTIP*-mediated R-loop is required for maintaining CTCF boundary, TAD integrity, and AML pathogenesis

(A) Schematic of the *CTNNB1* 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^{D210N} with or without *CBS-u2*-targeted sgRNA or exogenous β-catenin expression in MOLM13 cells.

(C) Proliferation of WT, $CBS\text{-}u2^{RH\text{-}WT}$, $CBS\text{-}u2^{RH\text{-}Mut}$, and β-catenin-rescued $CBS\text{-}u2^{RH\text{-}WT}$ MOLM13 cells.

(D) CTCF ChIP-qPCR analysis of the indicated sites in WT, *CBS-u2*$^{RH\text{-}WT}$, *CBS-u2*$^{RH\text{-}Mut}$, and β-catenin-rescued *CBS-u2*$^{RH\text{-}WT}$ MOLM13 cells.

(E) *HOTTIP* ChIRP qPCR analysis of the indicated sites in WT, *CBS-u2*$^{RH\text{-}WT}$, *CBS-u2*$^{RH\text{-}Mut}$, and β-catenin-rescued *CBS-u2*$^{RH\text{-}WT}$ MOLM13 cells.

(F) DRIP-qPCR analysis of the indicated sites in WT, *CBS-u2*$^{RH\text{-}WT}$, *CBS-u2*$^{RH\text{-}Mut}$, and β-catenin-rescued *CBS-u2*$^{RH\text{-}WT}$ MOLM13 cells.

(G) NG Capture-C analysis of *CTNNB1* promoter interactions upon expression of *CBS-u2*$^{RH\text{-}WT}$. 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*$^{RH\text{-}WT}$ MOLM13 cells.

(H) Kaplan-Meier survival curves of NBSGW mice transplanted with WT, *CBS-u2*$^{RH\text{-}Mut}$, or *CBS-u2*$^{RH\text{-}WT}$ MOLM13 cells.

(I) Kaplan-Meier survival curves of NSG mice transplanted with *CBS-u2*$^{RH\text{-}WT}$ or *CBS-u2*$^{RH\text{-}Mut}$ 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.

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_313348
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 polyclonal	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

REAGENT or RESOURCE	SOURCE	IDENTIFIER
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
T5 exonuclease	New England Biolabs	Cat#M0363
TRI Reagent	Sigma-Aldrich	Cat#93289
Superase-in	Thermo Fisher Scientific	Cat#AM2696
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		
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
QIAprep Spin Miniprep Kit	QIAGEN	Cat#27106
QIAGEN Plasmid Plus 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 # E7335S
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

REAGENT or RESOURCE	SOURCE	IDENTIFIER
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#KK4824
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
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*^{RH} 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*^{RH} MOLM13 cells	This paper	GEO: GSE165049
NG-Capture-C-seq of WT, *HOTTIP*$^{-/-}$, *CBS-u2*$^{-/-}$, *CBS-u2*^{RH} MOLM13 cells	This paper	GEO: GSE165049
Experimental Models: Cell Lines		
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		
Hottip-Transgenic mouse	(Luo et al., 2019)	N/A
Xenograft AML mouse model	This paper	N/A
Oligonucleotides		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
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		
pL-CRISPR.EFS.GFP	(Heckl et al., 2014)	Addgene Plasmid #57818; RRID:Addgene_57818
pLKO5.sgRNA.EFS.tRFP	(Heckl et al., 2014)	Addgene Plasmid #57824; RRID:Addgene_57824
lentiCRISPR v2	(Sanjana et al., 2014)	Addgene Plasmid # 52961; RRID:Addgene_52961
pHR-SFFV-dCas9-BFP-KRAB	(Gilbert et al., 2013)	Addgene Plasmid #46911; RRID:Addgene_46911
pFRT-TODestGFP_RNAseH1	(Lu et al., 2018)	Addgene Plasmid # 65784; RRID:Addgene_65784
ppyCAG_RNaseH1_D210N	(Chen et al., 2017)	Addgene Plasmid # 111904; RRID:Addgene_ 111904
human beta-catenin pcDNA3	(Kolligs et al., 1999)	Addgene Plasmid # 16828; RRID:Addgene_ 16828
pLenti-CMV-FLAG-puro	(Agrotis et al., 2019)	Addgene Plasmid # 123223; RRID:Addgene_ 123223
pMD2.G	Addgene	Addgene Plasmid # 12259; RRID:Addgene_12259
pBlueScript II SK	Addgene	Addgene Plasmid #212205; RRID:Addgene_212205
HS321/45-*vav-HOTTIP* vector	(Luo et al., 2019)	N/A
pTre-Tight-luciferase-*HOTTIP*	Addgene	Addgene Plasmid #65490; RRID:Addgene_65490
psPAX2	Addgene	Addgene Plasmid #12260; RRID:Addgene_12260
pGEX-5X-1	Addgene	Addgene Plasmid # 27-4584-01; RRID:Addgene_ 27-4584-01
Software and Algorithms		
TopHat/2.0.13	(Trapnell et al., 2012)	https://ccb.jhu.edu/software/tophat/
Bowtie2/2.2.9	(Langmead and Salzberg, 2012)	http://bowtiebio.sourceforge.net/bowtie2/
R/3.6.1	N/A	https://www.r-project.org/
Cufflinks/2.2.1	(Trapnell et al., 2010)	http://cole-trapnell-lab.github.io/cufflinks/
Cuffdiff/2.2.1	(Trapnell et al., 2010)	http://cole-trapnell-lab.github.io/cufflinks/
Integrated Genomic Viewer (version 2.4.19)	(Robinson et al., 2011)	http://software.broadinstitute.org/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deeptools/3.1.3	(Ramirez et al., 2014)	https://deeptools.github.io/
Gene Set Enrichment Analysis (GSEA) (4.0.0)	(Subramanian et al., 2005)	http://software.broadinstitute.org/gsea/
chromVAR(3.10)	(Schep et al., 2017)	https://bioconductor.org/packages/release/bioc/html/chromVAR.html
Homer/4.10	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/index.html
Juicer (version 1.6)	(Durand et al., 2016b)	https://github.com/aidenlab/juicer/
Juicebox(version 2.1.4)	(Durand et al., 2016a)	http://aidenlab.org/juicebox/
HiCExplorer (3.5.3)	(Wolff et al., 2020)	https://hicexplorer.readthedocs.io/en/latest/
QmRLFS-finder (version v1.4)	(Jenjaroenpun et al., 2015)	http://rloop.bii.a-star.edu.sg/?pg=qmrlfs
capC-MAP(1.1.3)	(Hughes et al., 2014, Buckle et al., 2019)	https://capc-map.readthedocs.io/en/latest/