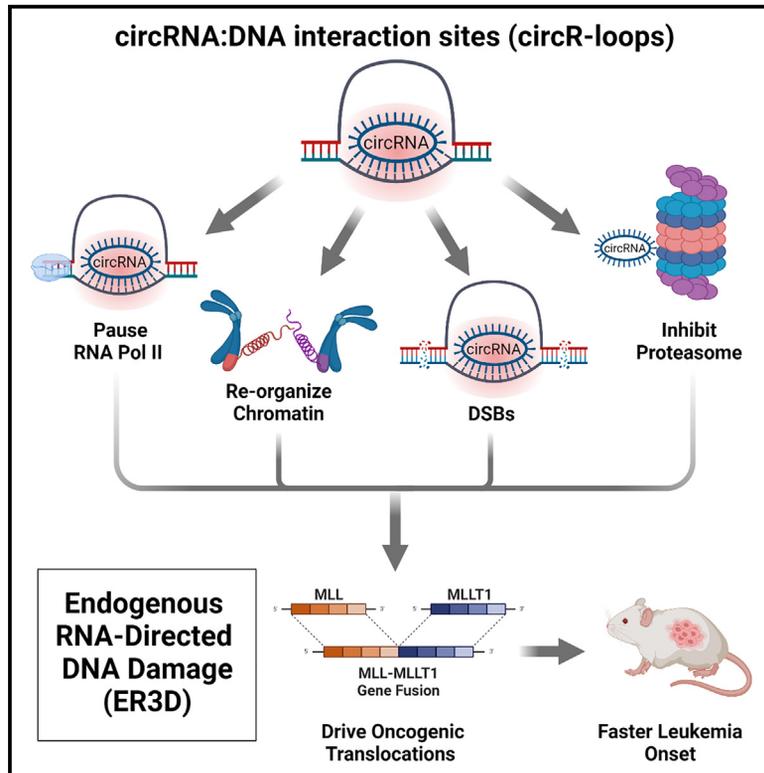# Circular RNAs drive oncogenic chromosomal translocations within the MLL recombinome in leukemia

Graphical abstract

Authors

Vanessa M. Conn, Marta Gabryelska, John Toubia, ..., Stuart M. Pitson, Gregory J. Goodall, Simon J. Conn

Correspondence

simon.conn@flinders.edu.au

In brief

Conn et al. identify circular RNA:DNA interactions concentrated within the MLL recombinome, commonly translocated in acute myeloid leukemia. Both *in vitro* and *in vivo* experiments demonstrate that endogenous circular RNAs inhibit the proteasome, promote DNA double-stranded breaks, reorganize chromatin, and promote oncogenic gene translocations which drive leukemia.

Highlights

- CircRNA:DNA interactions found genome wide, concentrating within the MLL recombinome

- Circ*MLL*(9,10) abundant in infant leukemia and binds MLL breakpoint cluster region

- Circ*MLL*(9,10) inhibits proteasome to promote DNA breaks and chromosomal translocations

- CircRNAs can be endogenous RNA carcinogens driving leukemogenic oncogenes

CellPress

Article

Circular RNAs drive oncogenic chromosomal translocations within the MLL recombinome in leukemia

Vanessa M. Conn,[1,2] Marta Gabryelska,[1] John Toubia,[2,3] Kirsty Kirk,[1] Laura Gantley,[1] Jason A. Powell,[2,4] Gökhan Cildir,[2] Shashikanth Marri,[1] Ryan Liu,[1] Brett W. Stringer,[1] Scott Townley,[1] Stuart T. Webb,[1] He Lin,[1] Saumya E. Samaraweera,[2] Sheree Bailey,[5] Andrew S. Moore,[6,7] Mellissa Maybury,[6] Dawei Liu,[2] Alex D. Colella,[1,8] Timothy Chataway,[1,8] Craig T. Wallington-Gates,[1,2,4,9] Lucie Walters,[10] Jane Sibbons,[11] Luke A. Selth,[1,12] Vinay Tergaonkar,[13] Richard J. D'Andrea,[2] Stuart M. Pitson,[2,4] Gregory J. Goodall,[2,4] and Simon J. Conn[1,2,14,*]

[1]Flinders Health and Medical Research Institute, College of Medicine & Public Health, Flinders University, Bedford Park, SA 5042, Australia
[2]Centre for Cancer Biology, SA Pathology & University of South Australia, Adelaide, SA 5000, Australia
[3]ACRF Cancer Genomics Facility, SA Pathology, Adelaide, SA 5000, Australia
[4]Adelaide Medical School, Faculty of Health and Medical Sciences, the University of Adelaide, Adelaide, SA 5000, Australia
[5]Health and Biomedical Innovation, Clinical and Health Sciences, University of South Australia, Adelaide, SA 5000, Australia
[6]Child Health Research Centre, the University of Queensland, Brisbane, QLD 4101, Australia
[7]Oncology Service, Children's Health Queensland Hospital and Health Service, Brisbane, QLD 4101, Australia
[8]Flinders Omics Facility, College of Medicine & Public Health, Flinders University, Bedford Park, SA 5042, Australia
[9]Flinders Medical Centre, Bedford Park, SA 5042, Australia
[10]Adelaide Rural Clinical School, Faculty of Health and Medical Sciences, the University of Adelaide, Adelaide, SA 5000, Australia
[11]Adelaide Microscopy, Division of Research and Innovation, University of Adelaide, Adelaide, SA 5000, Australia
[12]Freemasons Centre for Male Health and Wellbeing, Flinders University, Bedford Park, SA 5042, Australia
[13]Laboratory of NFκB Signalling, Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), 61 Biopolis Drive, Proteos, Singapore 138673, Republic of Singapore
[14]Lead contact
*Correspondence: simon.conn@flinders.edu.au
https://doi.org/10.1016/j.ccell.2023.05.002

SUMMARY

The first step of oncogenesis is the acquisition of a repertoire of genetic mutations to initiate and sustain the malignancy. An important example of this initiation phase in acute leukemias is the formation of a potent oncogene by chromosomal translocations between the mixed lineage leukemia (*MLL*) gene and one of 100 translocation partners, known as the MLL recombinome. Here, we show that circular RNAs (circRNAs)—a family of covalently closed, alternatively spliced RNA molecules—are enriched within the MLL recombinome and can bind DNA, forming circRNA:DNA hybrids (circR loops) at their cognate loci. These circR loops promote transcriptional pausing, proteasome inhibition, chromatin re-organization, and DNA breakage. Importantly, overexpressing circRNAs in mouse leukemia xenograft models results in co-localization of genomic loci, *de novo* generation of clinically relevant chromosomal translocations mimicking the *MLL* recombinome, and hastening of disease onset. Our findings provide fundamental insight into the acquisition of chromosomal translocations by endogenous RNA carcinogens in leukemia.

INTRODUCTION

RNA-DNA hybrids are commonplace throughout the mammalian genome, including those formed during DNA replication (11 bp hybrid from Okazaki fragments), and transcription (8 bp hybrid within the RNA polymerase active site).[1] However, longer tracts of such three-stranded nucleic acid structures, known as R loops, form when RNA base pairs with its cognate genomic DNA locus, displacing a loop of single-stranded DNA (ssDNA). The exposed ssDNA tracts are sites of genome instability and are susceptible to mutagenesis, which can manifest as double-strand DNA breaks (DSBs) from base excision repair.[2] This R

loop-mediated genome instability can manifest advantageously, including their indispensable role in immunoglobulin class switch recombination (CSR) in activated B cells.[3] CSR requires the mutagenic enzyme, activation-induced cytidine deaminase (AID),[4,5] which belongs to the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of proteins. These APOBEC proteins deaminate deoxycytidine in ssDNA to deoxyuridine, a process that can cause DSBs. Since APOBEC proteins are expressed beyond B cells alone,[6,7] there is potential for widespread, R loop-mediated DNA mutagenesis. Conceivably, aberrant R loop formation could stimulate oncogenic driver mutations, critical in the initiation of cancer.[8]

Acute leukemias carry the lowest mutational burden of any cancer type, on average <0.5 mutations per megabase.[9] Conserved *de novo* chromosomal translocations between the H3K4 histone methyltransferase mixed lineage leukemia (*MLL*, also called *KMT2A*) gene and one of >100 known partner genes are found in about 5%–10% of all acute leukemia patients, but >70% of infant leukemias.[10,11] The spectrum of translocation partner genes, called the MLL recombinome, varies between infant, pediatric, and adult cohorts and between acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). However, common translocation partner genes include *MLLT1-ENL*, *MLLT2-AFF1-AF4*, and *MLLT3-AF9*, which account for approximately 50% of AML and 88% of ALL translocations.[11,12] Rearrangements of the *MLL* gene also occur frequently in therapy-related acute leukemia, accounting for 15% of AMLs and are associated with particular anti-cancer treatments such as topoisomerase inhibitors, which promote DNA breaks.[10] In the background of a paucity of additional gene mutations, this family of translocations can efficiently transform hematopoietic cells into leukemic stem cells[13] and initiate aggressive forms of ALL and AML that are associated with particularly poor outcomes with 5-year survival rates of 5%–27%.[14,15]

Most *MLL* translocations map to an 8.3 kb breakpoint cluster region (bcr) comprising introns 5–12 of *MLL* and common sites within their partner gene introns.[12,16] Some chromosome features have been suggested to play a role in delineating the *MLL* bcr, including transcriptionally active *Alu* elements in intron 9, and a DNase I hypersensitive site and internal RNA polymerase II (RNAPII)-driven promoter in intron 11.[11,17,18] Despite these observations, the mechanisms underlying vulnerability of the bcr to translocation events and how it fuses with conserved sites in partner genes remain to be fully elucidated. Understanding the fundamental basis of these potent *MLL* rearrangements by focusing on their genomic landscape is critical for the prevention of leukemia.[11]

It is clear that R loops can form co-transcriptionally (*cis* R loops), but may also target loci of similar sequence composition in *trans*.[1] Immunoprecipitation with the R loop-specific S9.6 monoclonal antibody coupled to high-throughput sequencing of the resident DNA (DNA:RNA immunoprecipitation sequencing, or DRIP-seq) or the bound RNA (DRIPc-seq) has shown that *cis* R loops comprise the vast majority of R loops genome wide.[19] Defects in termination and elongation factors are thought to stall the RNA polymerase, increasing the time in which nascent RNAs are proximal to cognate DNA sequence and prolonging the unwound state of the DNA and promoting recruitment of AID through SPT5.[20]

The Fanconi anemia (FA) pathway is able to resolve R loops and prevent genome instability,[21] with the mutation of the FA genes common in cancer, particularly AML.[22] In addition, key roles in DNA repair are played by the ubiquitin-proteasome system including through the activation of the FA pathway[23] and through recruitment of both 19S and 20S proteasome subunits to sites of DSBs.[24] Defects in topoisomerases, or chemotherapeutic intervention with topoisomerase inhibitors (e.g., etoposide) can prevent relaxation of the negative supercoils ultimately leading to leukemogenic chromosomal translocations for, as yet, unidentified reasons.[10,25] Furthermore, deficiencies in splicing, or non-canonical splicing, may unmask the RNA, making it more accessible to hybridize with DNA.[26] Therefore, it is apparent that R loops can be augmented by non-canonical splicing and demarcate sites of genome instability. Through impairment of DNA repair, including but not limited to FA pathway inactivation and/or proteasome inhibition, a hallmark of cancer, it is conceivable that R loops could yield oncogenes by driving locus-specific mutagenesis.

It has recently become evident that another class of RNA, circular RNAs (circRNAs), are abundant and ubiquitous across eukaryotes but have been overlooked because they were difficult to detect by traditional methods.[27–29] CircRNAs are covalently closed circles of single-stranded RNA that arise from non-canonical, back-splicing of pre-mRNA. CircRNAs can be formed from RNA-binding proteins such as Quaking, which can be misregulated in cancer.[30,31] Consequently, circRNAs have been identified as cancer biomarkers and function by promoting hallmarks of cancer.[32–35] One report identified circRNAs arising following the oncogenic fusion of *MLL* to known partner genes, and these fusion circRNAs (f-circRNAs) contributed to cellular transformation in leukemia.[36] Critically, there has been no evidence for the capacity of circRNAs to drive oncogenic mutations.

By investigating the functional consequence of circRNA:DNA hybrids (circR loops) in human cancer circRNAs are shown to demarcate the sites of chromosomal translocations and drive oncogenic gene fusions via endogenous RNA-directed DNA damage (ER3D).

RESULTS

circR loops are prevalent across the human genome

To assess the contribution of circRNAs to R loops (Figure 1A) in human cells genome wide, we performed DRIP-seq on HEK293T cells. In accordance with previous studies in mammalian cells,[38] we identified 20,467 discrete R loops (Figure 1B, red bars; Table S1). We assessed which of these peaks overlap with circRNAs at their cognate locus in HEK293T cells by cross-referencing R loops with matched, nuclear-fractionated circRNA sequencing (circRNA-seq) (Figure 1B, blue bars). To facilitate quantitative normalization of circRNAs between RNA samples, we spiked-in three 400-nt *in vitro* synthesized circRNAs[39] before RNase R treatment. This approach identified 11,035 unique circRNAs in the nuclear fraction (and 60,009 circRNAs in the cytosolic fraction), 1,279 of which overlapped with R loop peaks, hereafter called circR loops (Figure 1B, purple bars; Table S1). The panel of circR loops accounts for approximately 6% of all detected R loops, concordant with our previous estimate of circR loops in plants.[40]

We confirmed by DRIP-qRT-PCR that these were bona fide circR loops as, irrespective of whether the chromatin was cross-linked, both the R loops and the circRNAs underlying them were RNase R-resistant and RNase H-sensitive (Figures 1A–1C, S1A, and S1B). Furthermore, as there was no appreciable decrease in R loop intensity for *MLL* and *SSR1* following RNase R digestion the linear RNA from these loci likely do not contribute to the signal, as opposed to the *APOE* and *NOP58* R loops, which were lost following RNase R digestion (Figure 1C). Eight additional putative circR loops from different loci were confirmed as bona fide circR loops using this strategy (Figure S1C).
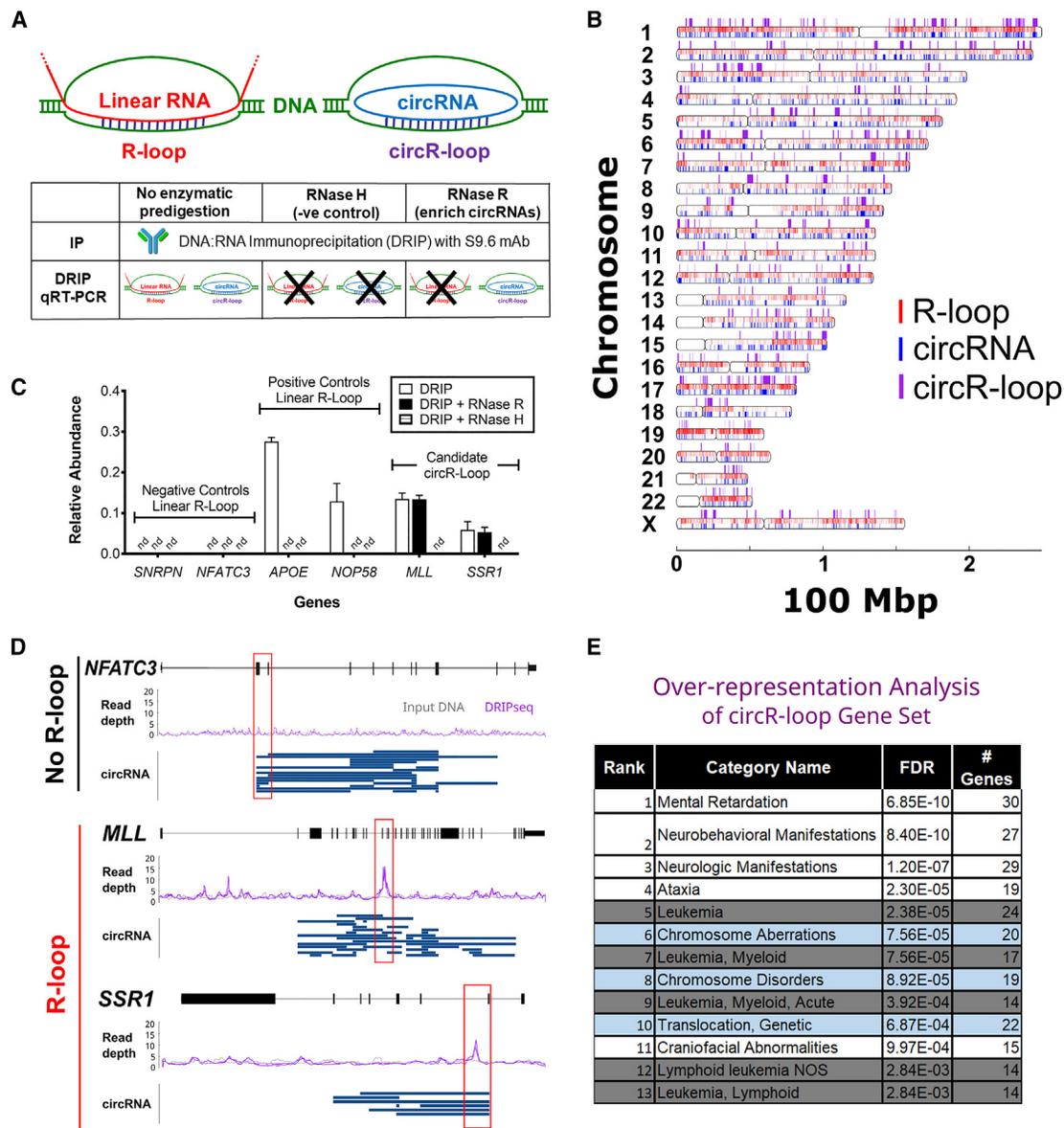
Figure 1. CircR loops are enriched within genes associated with leukemia and chromosomal instability

(A) R loops arising from linear RNA and circRNA (circR loops). DRIP method culminating in DRIP-qRT-PCR. Chromatin was either undigested or digested with RNase H (degrades R loops) or RNase R (digests linear RNAs) before IP with monoclonal antibody S9.6.

(B) Chromosome mapping of genome-wide R loops (red), nuclear circRNAs (blue), and circR loops (purple) in HEK293T cells.

(C) qRT-PCR on DRIP DNA material, showing known loci negative (*SNRPN*, *NFATC3*) and positive (*APOE*, *NOP58*) for linear R loops[37] with RNase R or RNase H digestion. Performed in biological and technical triplicate. nd, not detected. Data shown as mean ± SD.

(D) R loops from duplicate DRIP-seq experiments showing input DNA as background (gray trace), DRIP-seq signal (two purple traces) with mapped circRNAs (blue bars). *NFATC3* (top) has a highly abundant circRNA at exons 2–3, but lacks R loops. Two genes with bona fide circR loops are also shown (*MLL*, *SSR1*), with circR loop region boxed in red. R loop traces shown as moving average (50 nt window).

(E) Over-representation enrichment analysis (ORA) using circR loops geneset (WebGestalt) from HEK293T cells. Top 13 ranked disease categories shown, with leukemias (gray shading) and chromosomal changes (teal shading) highlighted. FDR, FDR-corrected p value. See also Figure S1 and Table S1.

We identified circR loops in 504 unique genes, involving a minimum of 939 circRNAs, as multiple circRNAs can overlap each R loop (Table S1). Interestingly, highly abundant circRNAs detected in the nucleus, including *circCDYL* (the most abundant circRNA in our dataset) and *circNFATC3*, did not form R loops (Figures 1D and S1D). When comparing the 500 most abundant nuclear circR-NAs, there were only 27 circRNAs whose nuclear abundance exceeded their cytoplasmic abundance (Figure S1E). In all cases, these did not overlap R loops, suggesting nuclear abundance alone is not sufficient for circR loop formation. Supporting this concept, circR loops exist for less abundant circRNAs within other genes, including *circMLL* and *circSSR1* (Figure 1D).
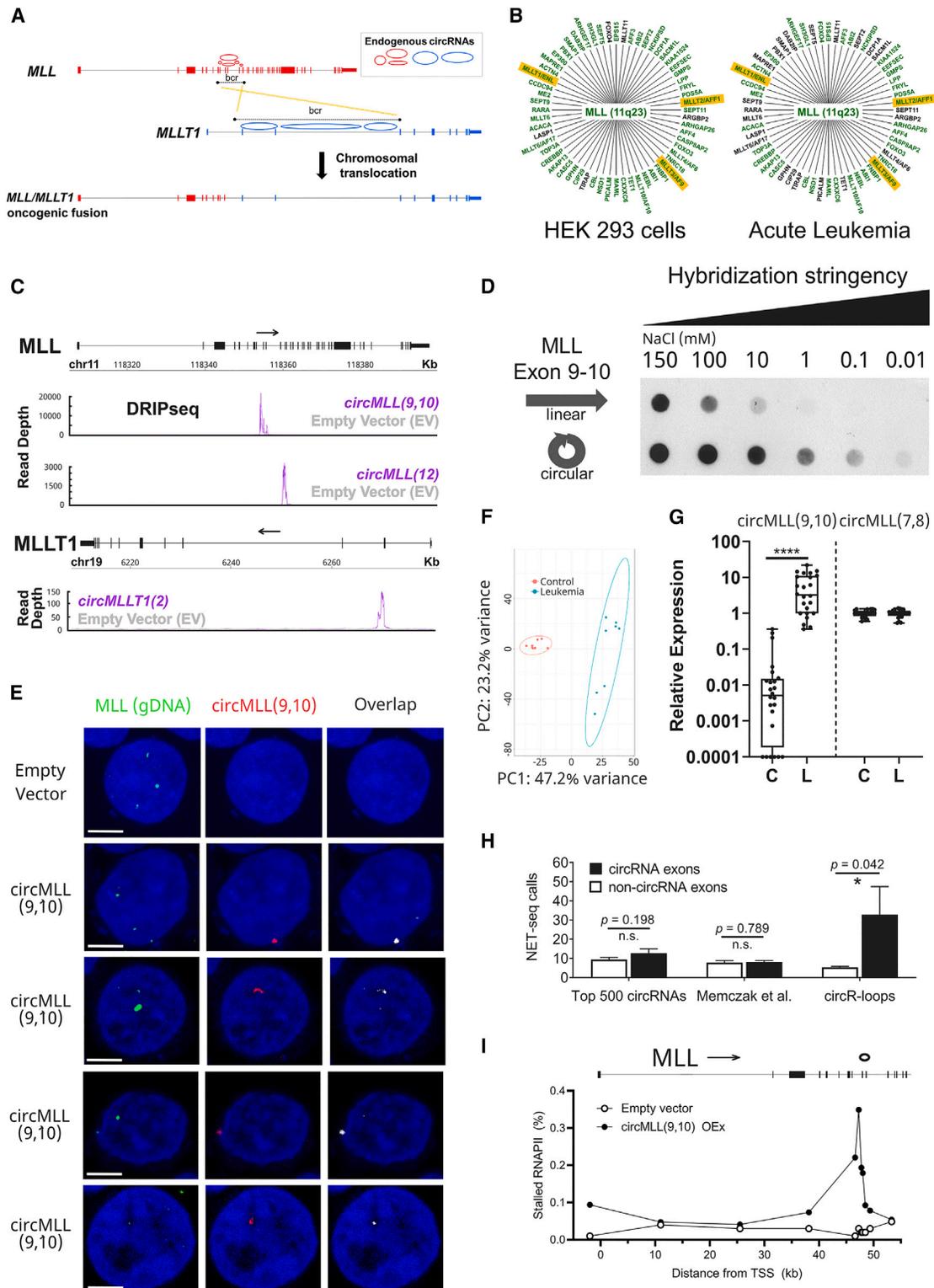
Figure 2. circRNAs form R loops with cognate DNA locus and promote transcriptional pausing

(A) Stylized image of *MLL* and *MLLT1* (*ENL*) genes and circRNAs showing breakpoint cluster sites (bcrs) that concentrate chromosomal translocations, resulting in a genomic *MLL-MLLT1* fusion.

(B) Enrichment of circRNAs in bcrs of *MLL* recombinome. Green indicates circRNA(s) adjacent to known bcr, black indicates circRNAs not adjacent to known bcr from HEK293T (this study) or human acute promyelocytic leukemic (APL) cell datasets.[46]

(legend continued on next page)

CircRNAs are overwhelmingly enriched in the cytoplasm.[28] Our validated cellular fractionation data (Table S1) also supports this observation, with only a minor increase in nuclear retention for the circRNAs forming circR loops compared with the global circRNA average in HEK293T cells (cytosol:nuclear ratio of 8.4 versus 9.7). However, specific examples—including circMLL(12), circMLLT1(2), and circMLLT2(3,4)—showed higher nuclear abundance by qRT-PCR (Figure S1F). While intron-retentive circRNAs have been shown to favor nuclear localization,[41] including circSSR1 which forms a circR loop, we also found no evidence for overall enrichment in intron retention for circRNAs that form circR loops (data not shown). Further analysis found that circRNAs forming circR loops were significantly increased (Tukey's multiple comparison of means, p < 0.0001) in size (nucleotide length), exon number, and GC content compared with circRNAs present in the nucleus that do not form circR loops (Figures S1G–S1I). Certainly elevated GC content has been noted as a weak predictor of R loop regions.[42,43] Therefore, it is likely that many factors, including the circRNA size, sequence, and secondary structure, rather than simply its abundance, are critical factors in potentiating circR loop formation.

circR loops are enriched within the MLL recombinome

To investigate biological contexts for the 504 genes containing circR loops, we performed gene over-representation enrichment analysis using WebGestalt.[44] Within the top 13 significantly enriched disease categories were 5 leukemia, 4 neurological/neurobehavioral manifestations, and 3 genome instability categories (Figure 1E; Table S1). Some circR loop-containing genes were common to both the leukemia and genome instability categories, including MLL (KMT2A), MLLT1 (ENL), MLLT2 (AFF1/AF4), and MLLT3 (AF9). Taken together, these three circR loop containing genes comprise up to 88% of the known leukemogenic chromosomal translocations in AML and ALL with the MLL gene, which is itself replete with circRNAs and circR loops (Figure 1D).[11] Since R loops have been implicated in genome instability,[2,45] we hypothesized that this enrichment may represent a mechanistic insight linking circR loops with chromosomal translocations in leukemogenesis.

Despite being made co-transcriptionally, circRNAs are not pervasive among spliced genes, being detected in 17%–31% of transcribed genes from our data. However, focusing on the MLL recombinome, we found circRNA-containing genes were highly enriched in HEK293T cells (56 circRNA-containing genes from 61 expressed genes = 91.8%), acute promyelocytic leukemia cells (71.4%), ALL (60.5%), and non-leukemic mesHMLE cells (80%), suggesting that these are circRNA-active genes (Figures 2A and 2B). Even more convincingly, we co-localized the circRNAs from these genes and the circR loops to within the bcr fusion hotspots, with 90% of the circRNAs overlapping or immediately flanking these hotspots as annotated in the Catalog of Somatic Mutations in Cancer (COSMIC) database[49] (Table S2). These data implicate circRNAs in the translocation process within the MLL recombinome.

Quantification and visualization of circMLL R loops

To assess the feasibility of circR loops underlying the MLL recombinome and causing chromosomal translocations, we overexpressed circRNAs entirely contained within the bcr of MLL (circMLL(9,10) and circMLL(12)) and its common partner gene MLLT1 (circMLLT1(2)) in HEK293T cells and performed DRIP-seq. For all circRNAs, DRIP peaks were seen spanning the cognate exons within the bcrs of MLL and MLLT1 (Figure 2C). Since circMLL(9,10) (hsa_CIRCpedia_342868) formed the strongest circR loop within the bcr of the MLL recombinome master regulator, all further experiments focused on this circRNA. We next assessed the capacity of in vitro synthesized linear RNA and circRNA (circMLL(9,10)) to bind its cognate DNA, exons 9–10 of MLL, encoded on a plasmid. This was performed as per Conn et al.[40] under increasing hybridization stringency and quantified by dot-blot using the R loop-specific S9.6 antibody. While both linear RNA and circRNA from this site could form R loops at low hybridization stringency, the circR loop had stronger binding as it was detectable at 100-fold higher stringency (Figure 2D). This result demonstrates the high avidity of circRNA for cognate DNA.

To visually validate the formation of circR loops, we undertook three-dimensional RNA-DNA fluorescence in situ hybridization (3D RNA/DNA-FISH) with a single RNA probe targeting the back-splice junction of circMLL(9,10) and DNA probe for the MLL genomic DNA locus. HEK293T cells were transfected with empty vector (pcDNA3.1) or the circMLL(9,10) overexpression construct before 3D RNA/DNA-FISH. A maximum of three discrete signals for the MLL gDNA probe were observed in

---

(C) DRIP-seq traces (purple) following overexpression of circMLL(9,10) (upper), circMLL(12) (middle), and circMLLT1(2) (lower) in HEK293T cells compared with empty vector (gray).

(D) In vitro R loop dot-blot between a plasmid encoding MLL exons 9–10 (50 ng, equating to 15 fmol) and circular or linear forms of MLL exon 9–10 RNA (15 fmol) stained with S9.6 primary antibody. Hybridization stringency increased left-to-right by decreasing ionic strength, with sodium chloride (NaCl) concentrations shown in mM.

(E) Three-dimensional RNA/DNA-FISH in HEK293T cells transfected with empty vector (upper panels) and circMLL(9,10) overexpression construct. MLL genomic DNA (TAMRA, green), circMLL(9,10) (AlexaFluor 647, red), and DAPI nuclear stain (blue), with overlap between MLL gDNA and circRNA shown (gray). Magnification, 40×. Scale bars, 5 μm.

(F) PCA plot of mRNA-seq from Guthrie cards of patients who developed infant leukemia with MLL rearrangements (teal) and normal cohort without blood cancers (orange). N = 8 patients per cohort.

(G) qRT-PCR of circMLL(9,10) and circMLL(7,8) between normal (C) and MLL-rearranged leukemia patients (L), using GAPDH as normalization control. N = 8 patients per cohort, with all technical triplicates plotted. Significance calculated by one-way ANOVA, ****p < 0.0001 (circMLL(9,10)) and p = 0.8741 (circMLL(7,8)). Box shows upper/lower quartile and median with whiskers showing minimum and maximum values.

(H) Enrichment for RNA polymerase II pausing with circR loops using published HEK293T NET-seq data.[47] Counts within 1 kb upstream of circRNA exons in three circRNA populations (top 500 most abundant nuclear circRNAs, circRNAs overlapping circR loops, HEK293T circRNAs from Memczak et al.[48]) compared with counts for non-circRNA exons from the same genes. Mean ± SEM. Statistical analysis performed by two-way ANOVA.

(I) ChIP-qPCR for paused RNA polymerase II in MLL from HEK293T cells with (filled circles) and without (unfilled circles) overexpression of circMLL(9,10). Data presented and stalled RNAPII (%) by subtracting RNAPII pSer2 signal from RNAPII pSer5 signal and dividing by total RNAPII. See also Figure S2 and Table S2.

each cell nucleus, which is expected as HEK293T is a hypotriploid line. Demonstrating that the RNA probe was circRNA specific, no signal was found in the EV control cell line, which does not express *circMLL(9,10)*, but commonly one to two discrete signals were found in the *circMLL(9,10)* overexpression line delimited by the nuclear DAPI stain (Figures 2E and S2A). Critically, the *circMLL(9,10)* signal(s) were always found to co-localize with these *MLL* gDNA signals (Figure 2E), demonstrating physical, nuclear co-localization of *circMLL(9,10)* with its cognate genomic locus.

CircMLL(9,10) is more highly expressed in pediatric leukemia samples

To assess the clinical relevance of *circMLL(9,10)* in *MLL*-rearranged (MLLr) leukemia before the onset of disease, we harvested RNA from neonatal blood spots for eight patients who developed MLLr leukemia before the age of 14 (Table S2). The control cohort comprised eight Guthrie cards from patients who did not develop any blood cancer. RNA was harvested and amplified from the Guthrie cards and underwent RNA-seq (Table S2). Principal-component analysis appears to stratify normal and MLLr leukemia cohorts by transcript profiling even at birth and before detection of gene fusions or disease manifestation (Figure 2F; Table S2). The remaining RNA was reverse transcribed and underwent qRT-PCR to quantify circRNAs. While *circMLL(9,10)* was detected in 14 of the 16 patients (absent in 2 control samples), there was >100-fold higher abundance of this circRNA in the MLLr leukemia patient samples compared with the control cohort (Figure 2G; one-way ANOVA, $p < 0.0001$). The most abundant circRNA from *MLL* in published datasets, *circMLL(7,8)*, was detected in all samples and did not differ between the normal and MLLr patient cohorts ($p = 0.8741$).

Overexpression of circMLL(9,10) does not alter the global transcriptome or alternative splicing of linear MLL

CircRNAs can regulate the abundance of RNA transcripts through microRNA sponging or by directly influencing gene transcription.[41,48,50,51] Therefore, to assess if overexpression of *circMLL(9,10)* in HEK293T cells could impact gene expression we performed ribosomal RNA-depleted RNA-seq and differential expression analysis compared with empty vector control (Table S2). Of over 17,000 expressed transcripts only 8 were significantly differentially expressed (FDR < 0.05) and only one of these, LOC100505841, was found to be more highly expressed with $\log_2$ fold-change >2 after 48 h. Furthermore, *MLL* expression was not altered by overexpression of *circMLL(9,10)*. Importantly, this suggests that *circMLL(9,10)* did not impact gene transcription en masse within this time frame. Furthermore, we were unable to detect any exon-skipped, alternative spliced variants in *MLL* exons 8–11 following overexpression of this circRNA by targeted RT-PCR (Figure S2B). Moreover, no reads crossing the exon 8-exon 11 junction were detected in our total RNA-seq datasets, demonstrating that effects mediated by *circMLL(9,10)* are not caused by altering *MLL* gene expression or alternative splicing.

CircMLL(9,10) can pause RNAPII within the MLL bcr

R loops have been shown to increase transcriptional pausing by providing a physical barrier for RNAPII elongation.[52,53] To assess whether circR loops promote transcriptional pausing, we examined RNAPII occupancy from published HEK293T NET-seq data,[47] which globally maps RNAPII density at single-nucleotide resolution. This analysis found enrichment for RNAPII pausing adjacent (both upstream and downstream introns) to circR loops compared with other exons within the same genes (Figure 2H; Table S2), but not for the 500 most abundant circRNAs, or at sites corresponding to a previously published circRNA dataset from HEK293T cells.[48] This suggests that RNAPII pausing is enriched adjacent to circR loops and is not a universal feature of circRNA transcription.

To further probe this, ChIP-qPCR for total RNAPII, phosphoserine 2 (pSer2) RNAPII, and phosphoserine 5 (pSer5) RNAPII was performed in HEK293T cells with overexpression of *circMLL(9,10)*. As the overexpression construct contains the exons and fragments of the flanking introns from the locus of interest, we avoided amplifying regions that overlapped these sites. Increased occupancy was seen for paused RNAPII (phosphorylated on serine 5 of YSPTSPS repeat in the RBP1 subunit) in the introns both upstream and downstream of exons 9 and 10 of *MLL* (Figure 2I), correlating with the site of the previously identified circR loop (Figure 2C), but not appreciably elsewhere in the gene.

circR loops cause DSBs

Paused RNAPII is capable of recruiting the hematopoietic lineage-specific mutagenic protein, AID, through the action of the adaptor protein, SPT5, to drive chromosomal translocations at various loci.[20] Furthermore, there is a general correlation of DSB sites with RNAPII pausing[54] and RNAPII release.[55] This provides a tantalizing putative link between circR loops, transcriptional pausing, APOBEC family proteins, and chromosomal translocations.

As a first step in assessing whether circR loops are associated with chromosomal translocations, we correlated circR loop proximity with genome-wide DSBs. Exploiting publicly available BLESS (breaks labeling, enrichment on streptavidin, and next-generation sequencing) data from wild-type HEK293T cells,[56,57] which provides a genome-wide snapshot of DSBs at single-nucleotide resolution, we found an increased prevalence of DSBs in proximity to circR loops compared with the highest-abundance, nuclear-localized circRNAs (Figure 3A; Table S3). This finding suggests that circR loops can promote DSBs.

To assess this experimentally, we analyzed the capacity of candidate circRNAs to drive specific DSBs at their cognate locus by expressing individually and in combination *circMLL(9,10)* and *circMLLT1(2)*, which both form precise, strong circR loops (Figure 2C), with FLAG-tagged AID or a catalytically dead, negative control AID mutant protein lacking the cytidine deaminase domain ($AID^{\Delta 58-94}$, or mutAID) in HEK293T cells.[59] The exogenous AID proteins were targeted to the nucleus with a strong, constitutive nuclear localization signal since endogenous AID in HEK293T cells is only translocated to the nucleus under stress conditions.[60,61] We confirmed with anti-FLAG antisera that the transgenic AID and mutAID were both enriched in nuclear fractions and did not affect the abundance of either *MLL* or *MLLT1* mRNA (Figures S2C and S2D).

By performing the more scalable version of BLESS, called BLISS (breaks labeling *in situ* and sequencing),[62] we discovered
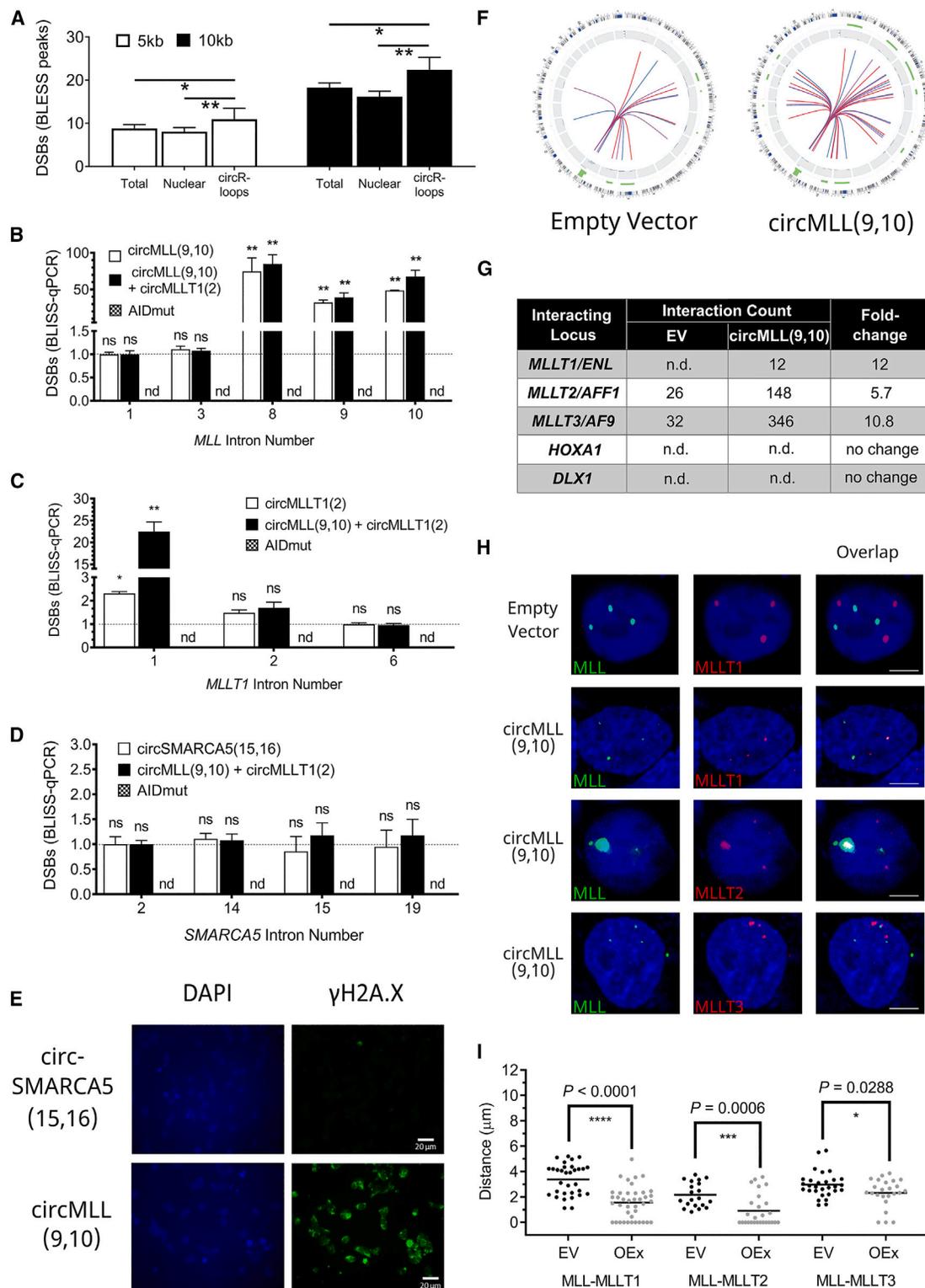
Figure 3. circRNAs drive DSBs and promote co-localization within *MLL* recombinome

(A) Enrichment of DSBs from published HEK293T BLESS data[56] within circR loops. BLESS peak count within 5 kb (white) or 10 kb (black) of circRNA exons for three circRNA populations from this study. Mean ± SD. *p < 0.05, **p < 0.01 Student's t test, two-tailed.

(B–D) BLISS-qRT-PCR used to assess DSB frequency in HEK293T cells overexpressing uniquely (B) *circMLL(9,10)*, (C) *circMLLT1(2)*, (D) *circSMARCA5(15,16)* (unfilled columns), or co-expressing *circMLL(9,10)* + *circMLLT1(2)* (filled columns). Co-expression of circRNA with nuclear-targeted mutant AID (checked columns) or wild-type AID (filled and unfilled columns). Mean ± SD. *p < 0.05, **p < 0.01, Student's t test, two-tailed. nd, none detected; ns, not significant.

increased DSB formation in the introns adjacent to the circR loops in *MLL* and *MLLT1* in the presence of AID, but no significant DSBs when co-expressed with mutAID (Figures 3B and 3C), or when overexpressing *circSMARCA5(15,16)*, despite a circR loop previously being identified at that site (Figure 3D).[63] We also observed an increase in nuclear γH2A.X signal, a known marker of DNA damage, following transient overexpression with the *circMLL(9,10)* (average foci/cell = 7.2), but not *circSMARCA5(15,16)* (average foci/cell = 0.2) (Figure 3E). This suggests effects mediated by *circMLL(9,10)* are not *MLL* locus specific, but impact the genome more broadly. Interestingly, the amount of DSBs in intron 1 of *MLLT1* was increased 10-fold, with co-expression of *circMLL(9,10)*, further suggesting a symbiotic effect in *trans* from *circMLL(9,10)* expression (Figure 3C). These experiments were repeated in the human leukemic cell line, HL-60, chosen as it is known to express high amounts of members of the APOBEC family in abundance and lacks *MLL* translocations.[64] Overexpression of *circMLL(9,10)* in HL-60 cells resulted in an increase in DSBs at both introns adjacent to the circR loop, with no additional increase in DSBs observed with co-expression of transgenic AID (Figure S2E). This suggests that HL-60 cells, unlike HEK293T cells, have sufficient capacity to drive DSBs without supplemental APOBEC protein expression.

CircMLL(9,10) alters chromosome organization and increases contact between loci within the MLL recombinome

Spatial organization of the genome is influenced by a plethora of chemical and environmental factors and influences the ability of two loci to undergo translocation events.[65] A recent publication identified subtype-specific 3D genome alterations in AML.[66] While the "contact-first" and "breakage-first" hypotheses regarding chromosomal translocations differ in the order of events, they both necessitate DSBs at co-localized loci. To assess whether circRNAs can alter the 3D proximity of *MLL* with its known translocation partners we transiently overexpressed *circMLL(9,10)* in HEK293T cells and performed circular chromosome conformation capture (4C).[67] In empty-vector (pcDNA3.1) transfected cells the bcr of *MLL* interacts with 10,173 coding loci, which equated to approximately 47% of the expressed coding loci genome wide (Table S3). However, the *MLL* recombinome was found to be under-represented with only 14 interactions detected among the 61 expressed loci (~23%). Following overexpression of *circMLL(9,10)*, the number of interactions increased ~3.6-fold (51 interactions, rep-

resenting ~84% of the expressed MLL recombinome (Figure 3F). The three most common *MLL* translocation partners were in closer proximity to *MLL* following overexpression of *circMLL(9,10)*: *MLLT1/ENL* (12-fold increase), *MLLT2/AFF1* (5.7-fold increase), and *MLLT3/AF9* (10.8-fold increase) (Figure 3G; Table S3). No interaction was detected in either library between *MLL* and other loci previously shown to dynamically co-localize by 4C in stem cells, including the HOX gene, *HoxA1*, and *Dlx1*,[58] suggesting that *circMLL(9,10)* is not causing promiscuous relocation of chromatin. On the contrary, the data suggest that, only in the presence of *circMLL(9,10)*, does the MLL recombinome co-localize, a necessary precursor for *MLL* translocations. Interestingly, by performing the assay for transposase-accessible chromatin with sequencing on HEK293T cells overexpressing *circMLL(9,10)*, there was no detectable increase in chromatin accessibility at the *MLL* locus, suggesting that the altered chromatin architecture and RNAPII pausing does not result from this phenomenon (Figure S2F).

The physical co-localization of *MLL* and each of the three major partner loci—*MLLT1*, *MLLT2*, and *MLLT3*—was assessed by 3D FISH in the hypotriploid HEK293T cell line (Figure 3H). Assessing the interaction frequencies and the likelihood of translocations between any two loci, it only requires one pair of these interacting loci to be in closer proximity. Therefore, we quantified the shortest 3D distance between any *MLL* and partner locus within the nucleus. Supporting the 4C results, the average distance between loci was found to be significantly less between MLL and all three partner genes compared with their empty vector controls (Figure 3I).

CircMLL(9,10) interacts with and inhibits proteasome components

To provide insight into the mechanism underlying the abilities of *circMLL(9,10)* to pause RNAPII and cause DSBs at cognate and partner loci we performed circRNA immunoprecipitation to profile the protein interactome of *circMLL(9,10)*. Following immunoprecipitation without fixation in biological triplicate, the protein eluate underwent tryptic digestion and mass spectrometry (Figure 4A). KEGG gene ontology analysis was performed on the 121 proteins that satisfied the stringent selection criteria (Table S4). The most enriched functional category, represented by a pull-down of 9 members out of a total of 46 (FDR = 7.49×10^{-9}), was the proteasome (Figure 4B). While each protein is a member of the larger 26S proteasome, 5 of the 14 subunits of the catalytic core subunit, 20S proteasome, were bound by *circMLL(9,10)* (Figure 4C).

(E) Representative immunofluorescence images of HEK293T cells using gamma H2A.X antisera and DAPI (blue) with transient overexpression of *circSMARCA5(15,16)*, or *circMLL(9,10)* in combination with nuclear AID. Magnification, 20×. Scale bars, 20 μm.

(F) Circos plots showing confirmed interactions between *MLL* (bait) and genes within the *MLL* recombinome in control (empty vector, left panel) and *circMLL(9,10)* overexpressor (right panel) HEK293T lines obtained by Circular chromosome conformation capture (4C-seq).

(G) Quantitative increase in common *MLL* fusion partner interactions assessed by 4C and absence of interaction with known HOX genes and partners found in other 4C experiments.[58] Interaction count between MLL bait region and interacting locus after normalizing for read depth, fold-change represented as ratio between *circMLL(9,10)* overexpression line and empty vector (EV) line.

(H) Three-dimensional DNA-FISH (3D FISH) performed on HEK293T cells transfected with empty vector (upper panel set) or *circMLL(9,10)* (lower three panel sets). DNA-FISH probes sets targeting *MLL-MLLT1*, *MLL-MLLT2*, or *MLL-MLLT3* and counterstained with DAPI (blue). Representative collapsed z stack is presented per line captured at 40× magnification. Scale bars, 3 μm.

(I) Distance between nuclear *MLL* and partner signals from 3D FISH experiment (n = minimum 20 cells per condition) for EV and *circMLL(9,10)* overexpressors. Paired graphs comparing distances between the nearest loci in each cell (overlapping loci given a nominal value of 0). Mean ± SD. Repeated measures one-way ANOVA analysis, p values shown. See also Figure S2 and Table S3.
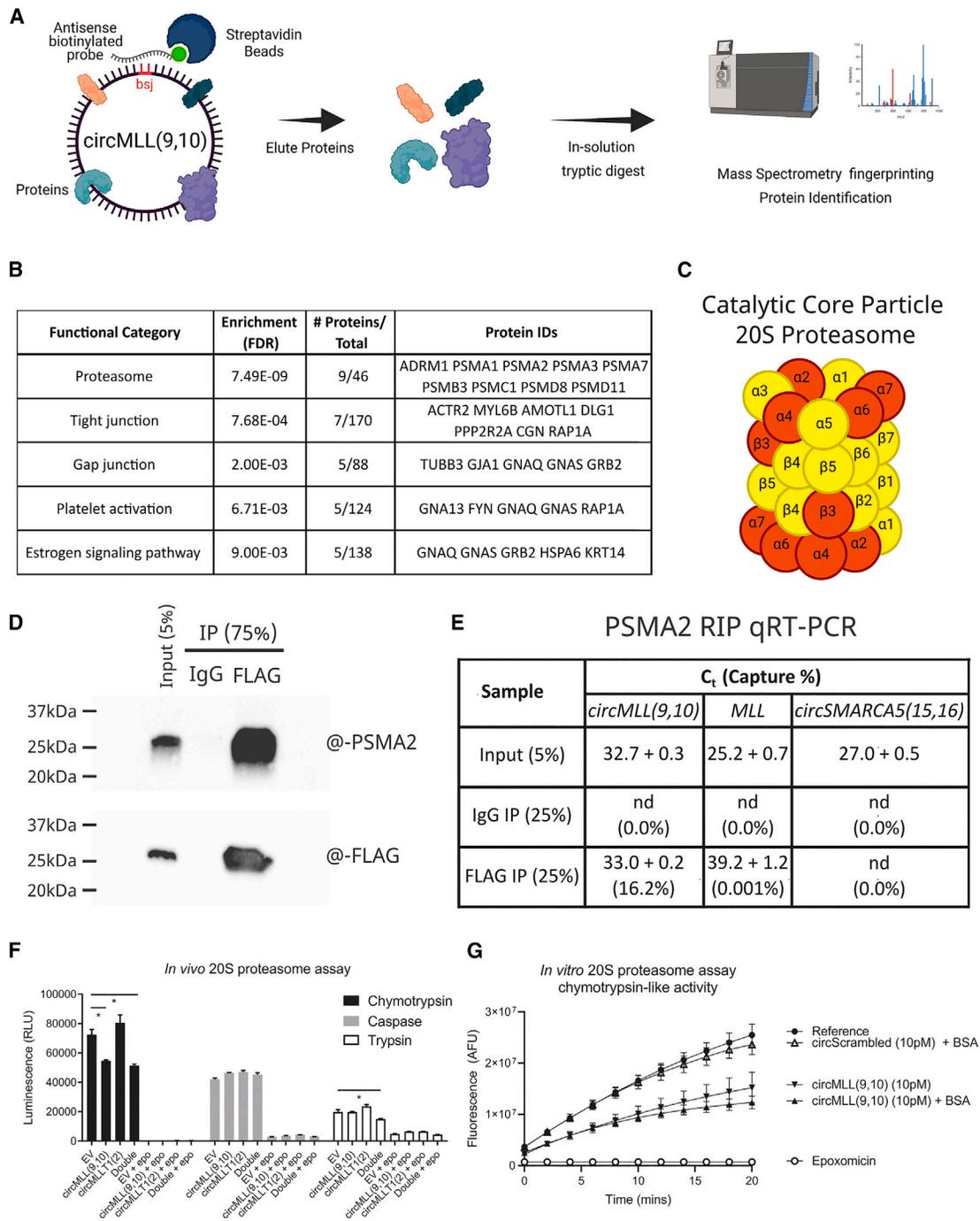
Figure 4. *CircMLL(9,10)* interacts with multiple members of the 20S proteasome and inhibits chymotrypsin-like activity

(A) Schematic of circRNA-IP using an antisense biotinylated probe (and sense probe as control) spanning the back-splice junction (bsj) followed by mass spectrometry to identify interacting proteins for *circMLL(9,10)*. Created with BioRender.com.

(B) Gene ontology (KEGG) analysis of proteins identified in *circMLL(9,10)* RIP-MS in groups with five or more proteins.

(C) The eukaryotic catalytic core particle (20S proteasome). Alpha and beta subunits shown, with members bound by *circMLL(9,10)* colored red. Created with BioRender.com.

(D and E) RNA Immunoprecipitation of FLAG-tagged PSMA2 protein using mouse anti-FLAG antibody. (D) Western blot probed with rabbit anti-FLAG (lower panel) or mouse anti-PSMA2 (upper panel) and (E) qRT-PCR for *MLL* mRNA, *circMLL(9,10)*, *circSMARCA5(15,16)* showing Ct (mean ± SD, three technical replicates) and capture (%) of total starting abundance from input. nd, not detected.

(legend continued on next page)

Given that inhibition of the 20S proteasome has been shown to be needed for removal of stalled RNAPII[68] and increase DSBs,[23,24] showing that *circMLL(9,10)* effects proteasome activity would provide compelling mechanistic insight for its ability to promote DSBs. To validate this interaction, a FLAG-tagged form of one of the hits from the mass spectrometry, proteasome subunit alpha type 2 (PSMA2), was co-expressed in HEK293T cells with *circMLL(9,10)*. We demonstrated successful enrichment of PSMA2-FLAG by western blotting following RNA immunoprecipitation (RIP) (Figure 4D). Importantly, qRT-PCR for *circMLL(9,10)* demonstrated that it was only detected in the FLAG RIP and not in the IgG pull-down, confirming the interaction between PSMA2 and *circMLL(9,10)* (Figure 4E). Specificity for this interaction was seen with the *MLL* mRNA detected at very low levels in the FLAG IP (1,000-fold less than *circMLL(9,10)*), while a highly expressed circRNA, *circSMARCA5(15,16)*, was undetected.

The 20S proteasome comprises catalytic (β type subunits) and non-catalytic (α type subunits) proteins, required for the assembly, ingress/egress of protein substrates, and protease activity of the complex.[69] Despite the circRNA binding exclusively to non-catalytic subunits of this complex, we assessed whether catalytic activity was affected by *circMLL(9,10)*. To achieve this, we generated stable HL-60 cell lines overexpressing *circMLL(9,10)* and *circMLLT1(2)*, individually or in combination. These overexpression constructs, containing GFP as a selectable marker, were generated as described previously[31,50] with the entire plasmid construct sequences verified by Nanopore sequencing and EPI2ME analysis (Figure S3A). After making stable cell lines via lentiviral transduction and sorting GFP-positive cells by FACS (Figure S3B), we confirmed by qRT-PCR the increase in expression of circRNAs in each line compared with empty vector control (Figure S3C). qRT-PCR demonstrated that each circRNA was increased up to 9-fold compared with the control HL-60 cells, with a circularization efficiency of ∼60% (Figure S3D) conforming to best practice standards for circRNA overexpression studies.[70] We verified by qRT-PCR and Northern blotting that the circRNAs were resistant to RNase R and not contaminated by large quantities of linear or circRNA concatamers (Figures S3E and S3F).

We first assessed the three major proteasome activities—chymotrypsin-like, trypsin-like, and caspase-like—using specific luminogenic substrates for each specific protease activity on each of the HL-60 cell lines. As these are suspension cells, there was no need to trypsinize cells before the assay, which afforded greater confidence in the results obtained. Epoxomicin (5 μM) was used as a control proteasome inhibitor, which significantly reduced all protease activities (Figure 4F). Overexpression of *circMLL(9,10)*, either alone or in combination with *circMLLT1(2)*, resulted in a statistically significant and specific decrease in chymotrypsin-like activity compared with the empty vector control cell line (Figure 4F). However, overexpression of *circMLLT1(2)* alone did not alter chymotrypsin-like, trypsin-like,

or caspase-like proteasome activity. Despite a 10% reduction in trypsin-like activity seen with co-expression of the two circRNAs, there were no other statistically significant alterations in proteasome activity. This demonstrates that cellular proteasome activity is reduced in the presence of *circMLL(9,10)*.

To assess whether *circMLL(9,10)* was directly capable of impacting the 20S proteasome itself, we performed an *in vitro* activity assay using a purified recombinant 20S proteasome and incubated this with *in vitro* synthesized *circMLL(9,10)* or a scrambled circRNA control. The specific activity of the purified 20S proteasome (100 ng protein) in this assay was 165 ± 14 pmol min^{-1} mg^{-1} protein. As expected, 5 μM epoxomicin completely suppressed 20S activity (Figure 4G). In the presence of 10 pM *circMLL(9,10)*, 20S protease activity was reduced by over 40% but was unaltered in the scrambled circRNA control (Figure 4G). To account for the inherent capacity of naked RNA to bind proteins non-specifically, we precoated *circMLL(9,10)* with BSA and noted that the inhibition caused by *circMLL(9,10)* was not significantly attenuated.

CircRNAs can drive oncogenic translocations *in vitro* and *in vivo* and hasten the onset of leukemia in mice

Since *circMLL(9,10)* can promote co-localization of the *MLL* recombinome and drive DSBs at these sites and inhibit the proteasome needed for repair of DSBs, we wanted to assess whether circRNAs could be capable of driving *de novo* chromosomal translocations. To achieve this, we utilized the stable HL-60 cell lines overexpressing *circMLL(9,10)* and *circMLLT1(2)*, individually and in combination. To assess the capacity of these circRNAs to induce DSBs genome wide, the neutral comet assay was employed. This assay demonstrated that there was a significantly higher amount of DSBs in *circMLL(9,10)*, *circMLL(9,10)/circMLLT1(2)*, and *circMLLT1(2)* lines compared with empty vector based on tail moment measurements, with *circMLLT1(2)* cells showing the smallest increase (Figures 5A and 5B) compared with the *circMLL(9,10)*-expressing lines. UV-C irradiation on HL-60 EV cells was used as a positive control and this displayed the greatest tail moment (approximately 14-fold higher than EV).

RNA-seq was performed on these cell lines and the data underwent detection of fusion transcripts using the software STAR-fusion and Clinker.[71] Within a background of very few translocations in the empty vector HL-60 lines, chromosomal translocations involving the *MLL* gene were seen uniquely in *circMLL(9,10)*-expressing lines (Figure 5C). Specifically, precise *MLL/MLLT1* fusion mRNAs were detected in all three replicates of the dual overexpression line between exon 10 of *MLL* and exon 2 of *MLLT1*. Interestingly, another translocation event was noted in a single replicate between exon 27 of *MLL* and exon 2 of *MLLT1* in the *circMLL(9,10)*-overexpressing line only, an event that is outside the known bcr of MLL and not found in patients. Overall, these experiments demonstrated the ability

(F) Cell-based 20S proteasome activity assay performed on HL-60 cells overexpressing circRNAs for three biological replicates (n = 3 technical replicates). Chymotrypsin-like (black), caspase-like (gray), and trypsin-like (unfilled columns) activities assayed. Double: *circMLL(9,10)/circMLLT1(2)*. Epoxomicin (0.5 μM) used as irreversible inhibitor of chymotrypsin activity. Mean +SD. Student's t test, *p < 0.05.

(G) *In vitro* 20S chymotrypsin-like activity of purified human 20S proteasome in the presence of *in vitro* synthesized circRNAs, *circMLL(9,10)*, or scrambled circRNA (10 pM) with and without BSA blocking. Time-resolved fluorescence assay over 20 min. Epoxomicin (0.5 μM) used as irreversible inhibitor of chymotrypsin activity. Mean ± SD for n = 3 biological replicates. See also Figure S3 and Table S4.
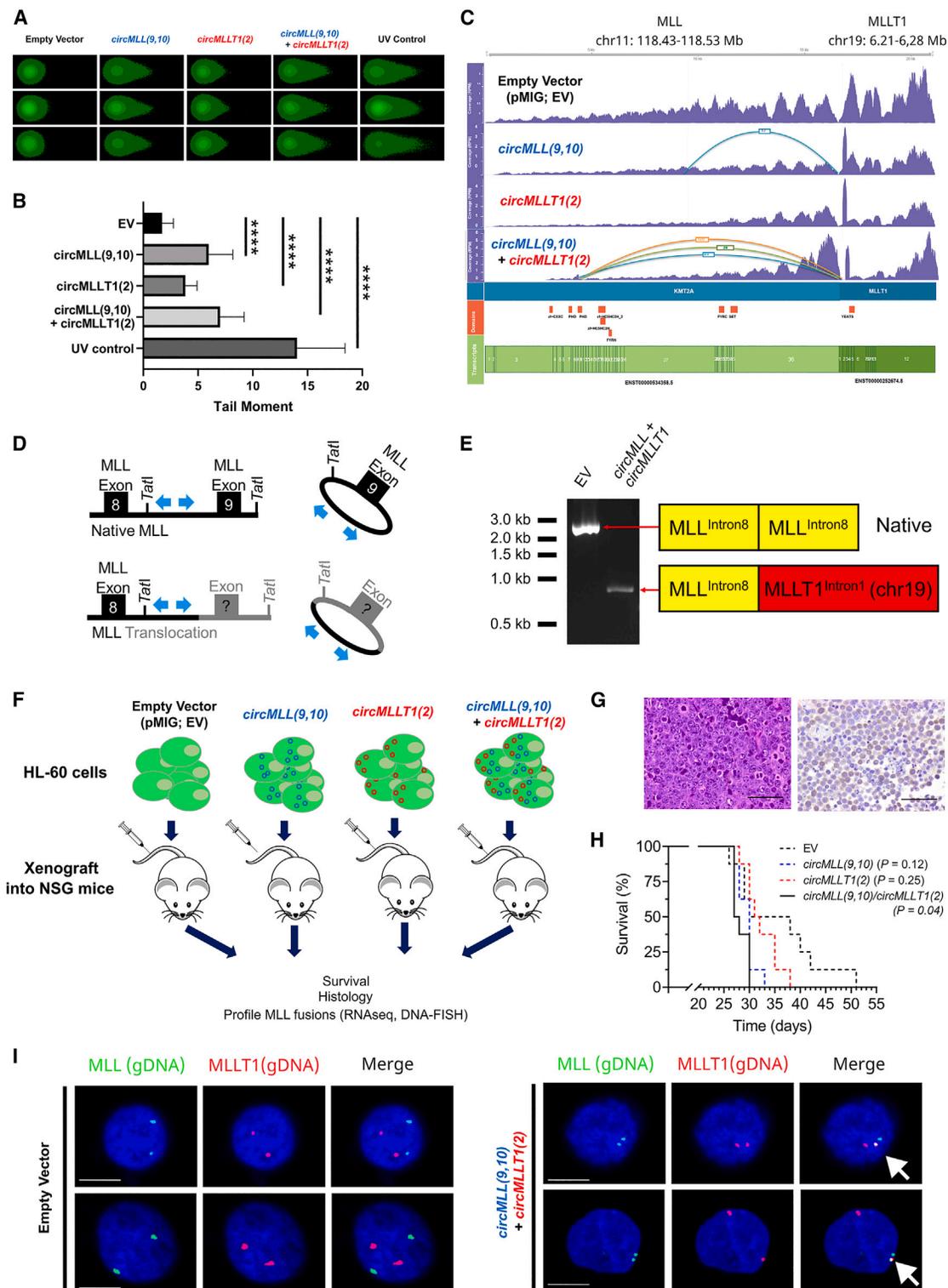
Figure 5. CircRNAs drive oncogenic chromosomal translocations and hasten onset of leukemia

(A and B) Neutral comet assay performed on HL-60 cells expressing empty vector (EV), *circMLL(9,10)*, *circMLLT1(2)*, *circMLL(9,10)* + *circMLLT1(2)*, and UV control (EV dosed with 10 J/m^2). (A) Three representative images of cells from each group stained with EvaGreen and (B) quantification of tail moment from n = 50 cells, showing mean + standard deviation. Student's t test, ****p < 0.0001.

(C) *MLL-MLLT1* gene fusion events across four HL-60 cell lines that underwent RNA-seq. Clinker output (mapped reads, fusion superTranscript, and annotation track) for the four cell lines with sashimi plot showing fusion breakpoint counts and location in three biological replicates (orange, green, and blue lines).

(D) Schematic for detection of translocation events at the DNA level by custom inverse PCR method targeting intron 8 of *MLL*.

(legend continued on next page)

of circRNAs to drive precise oncogenic translocations within the MLL recombinome.

It is possible that low-frequency, poorly transcribed or low-stability fusions, occurring at other sites, were undetected by RNA-seq. Dictated by the highest rate of DSBs in intron 8 of *MLL* from our BLISS experiments (Figure 3B), we developed a custom inverse PCR assay based on the widely used inverse PCR method targeting intron 11 of MLL[72] to identify chromosomal translocations at the DNA level (Figure 5D). In short, it involves digestion with the 4 bp restriction endonuclease *Tat*I, followed by ligation and subsequent digestion with *Bgl*II to mitigate the amplification of the native *MLL* sequence and encourage detection of the translocated loci, necessary for polyclonal populations of cells. In the absence of *MLL* translocations, this inverse PCR method produced a wild-type product of 2,273 bp (Figure 5E).

No translocations were detected by the inverse PCR method in the empty vector or single overexpression lines at 2 weeks post-selection (Figure S4A). However, inverse PCR performed on cells co-expressing *circMLL(9,10)* and *circMLLT1(2)* yielded a single, prominent amplicon differing in size from that of wild-type *MLL* seen in the EV control cells (Figure 5E). This was confirmed by Sanger sequencing as a specific translocation between intron 8 of *MLL* and intron 1 of *MLLT1/ENL* (Figure S4B), adjacent to the cognate circRNAs. This rearrangement matches the breakpoints seen by BLISS and is the second most common fusion site seen between these loci in AML patients.[11,49] Additional fusions were confirmed by repeating the inverse PCR across five additional stable polyclonal populations of HL-60 cells (Figure S4C). In each instance, the inverse PCR analysis on the *circMLL(9,10)/circMLLT1(2)* transgenic dual overexpression lines yielded additional novel fusions comprising a distinct site in *MLL* and *MLLT1*. In five of these six fusions, the breakpoint in *MLL* immediately precedes a cytidine in the native sequence (Table S5). Therefore, the mutational signature of these fusions is consistent with cytidine deamination by APOBEC proteins followed by base excision repair and therefore consistent with our earlier data. In addition, despite the low number of fusions assessed herein, we can speculate that one likely DNA repair mechanism contributing to the fusions is non-homologous end-joining (NHEJ) as four of these fusions show examples of microhomology between the two fused introns and one showed the insertion of a 4 bp sequence not present in either loci, both known signatures of NHEJ.[73] This presents an example of ER3D manifesting as oncogenic chromosomal translocations.

Performing gene-specific RT-PCR on GFP-sorted HL-60 cells we found *de novo* gene fusion transcripts at precisely the ex-pected sites between *MLL* exon 8 and *MLLT1* exon 2 (Figures S4D and S4E). Critically, these fusion mRNAs were undetectable in empty vector and single circRNA overexpression lines. Guarnerio et al.[36] showed that, following the detection of linear RNA fusions for *MLL*, unique f-circRNAs between *MLL* and the translocation partner were detectable. By utilizing circRNA-specific divergent primers, with a single primer in *MLL* exon 8 and a second primer at the *MLLT1* exon 2, on RNase R-treated RNA from each cell line, we could reliably detect f-circRNAs only in the dual circRNA overexpression line that contained the fusion mRNA, which was confirmed by Sanger sequencing (Figures S4D and S4F).

Next, we assessed the impact of circRNA overexpression *in vivo* with empty vector control, individual *circMLL(9,10)* and *circMLLT1(2)*, or dual circRNA overexpression (*circMLL(9,10)/circMLLT1(2)*) cells xenografted into immunocompromised NOD.Cg-*Prkdc*[scid] *Il2rg*[tm1Wjl]/SzJ (NOD *scid* gamma) mice (Figure 5F). Leukemic progression was tracked and animals were humanely killed when they reached ethical welfare endpoints resulting from leukemic burden. Consistent with previous reports utilizing HL-60 cells, tumors excised from each group of mice showed a homogeneous population of medium-size primitive-appearing cells with moderately high nucleus to cytoplasm ratio, fine to moderately clumped chromatin, and one or more prominent nucleoli, consistent with HL-60 morphology. Mitotic figures were abundant in all sections and shown to be GFP positive by immunohistochemistry (Figure 5G) and flow cytometry (Figure S3B). Kaplan-Meier survival analysis demonstrated that overexpression of individual circRNAs did not significantly impact the onset of disease in these mice compared with the empty vector control; however, the dual overexpression line showed a significant hastening of disease (Figure 5H; p = 0.04, log rank Mantel-Cox test). RNA-seq and fusion detection by STAR-fusion was performed from GFP-positive HL-60 cells purified from two mice from each group. This confirmed the expression of *MLL-MLLT1* fusion transcripts in only the mice xenografted with *circMLL(9,10)/circMLLT1(2)* (Table S5). Critically, we confirmed by DNA-FISH using *MLL-MLLT1* genomic DNA probes that fusions were seen uniquely in the *circMLL(9,10)/circMLLT1(2)* lines (Figure 5I).

To ensure that our results were not restricted to a specific cell line, the xenograft experiments were repeated using another leukemic cell line, K562, a model of chronic myelogenous leukemia. As opposed to HL-60 cells, these cells carry known chromosomal translocations (including *BCR-ABL*). Specific translocations between *MLL* and *MLLT1* were generated only in the dual circRNA overexpression line (Figure S5), replicating our

(E) Identification of translocation events in HL-60 cells stably transduced with empty vector or co-expressing *circMLL(9,10)* and *circMLLT1(2)*. Native *MLL* locus inverse PCR product seen exclusively in stable EV (pMIG) line. Prominent product seen in circRNA overexpressing line confirmed as *MLL-MLLT1* translocation.

(F) Schematic of HL-60 mouse xenotransplantation experiments into NSG mice. EV, pMIG, overexpressing *circMLL(9,10)*, *circMLLT1(2)*, or both circRNAs (*circMLL(9,10) + circMLLT1(2)*) xenografted into NSG mice.

(G) Left panel: paraffin-embedded, hematoxylin and eosin-stained section of representative tumor excised from xenotransplanted mice. Right panel: representative GFP expression by immunohistochemistry. Magnification 40× mag. Scale bars, 100 μm.

(H) Kaplan-Meier survival plot for four mouse experimental groups xenotransplanted with HL-60 cells carrying empty vector (dashed black line, pMIG), *circMLL(9,10)* (dashed blue line), *circMLLT1(2)* (dashed red line), or both *circMLL(9,10) + circMLLT1(2)* (solid black line) (n = 8 mice per group). p value from log rank Mantel-Cox test.

(I) Representative DNA-FISH images of HL-60 cells expressing empty vector or *circMLL(9,10) + circMLLT1(2)* using *MLL* (green) and *MLLT1* (red) translocation dual fusion probes to visualize gDNA loci. Overlapping signals (white arrow) indicate *MLL-MLLT1* translocations. Nuclei stained with DAPI (blue). Scale bars 5 μm. Magnification, 60×. See also Figures S4, S5 and Table S5.

results with the HL-60 cells. RNA-seq revealed that the normalized fusion transcript counts were higher in all three biological replicates of the K562 line compared with the HL-60 line (Figure S5A), but the fusion mRNA transcript between exon 10 of *MLL* and exon 2 of *MLLT1* was again found uniquely in the dual circRNA overexpression line. We also assessed the prevalence of an additional fusion transcript between *MLL* exon 8 and *MLLT1* exon 2 via gene-specific RT-PCR on GFP-sorted K562 cells purified from euthanized mice and found *de novo MLL* gene fusion transcripts at the expected sites only in the dual overexpression lines (Figures S5B–S5D). Again, critically, the dual circRNA overexpression line was associated with significantly faster onset of disease in a mouse xenograft model compared with empty vector or individual circRNAs (Figure S5E). Together, our results represent *in vivo* evidence of ER3D manifesting as chromosomal translocations that drive oncogenesis and disease aggressiveness in leukemia.

DISCUSSION

CircRNAs are increasingly shown to play functional roles in eukaryotic cells, including known hallmarks of cancer, mediated via their interactions with microRNAs, proteins, and more recently mRNA[74] and DNA.[40] Here, we show that circRNAs, which are formed co-transcriptionally,[75] can form circRNA:DNA hybrids (circR loops) and cause DNA breaks resulting in chromosomal translocations. R loops formed during transcription have been shown to be dynamically resolved with an average half-life of ~10–20 min, theorized to enable cells to mitigate the negative effects of R loop formation on genomic stability.[19] However, we found that circR loops are stronger than R loops involving cognate linear RNAs, suggesting that they could persist for longer and encourage genome instability.

While two circRNAs underlying circR loops within the *MLL* recombinome can generate chromosomal rearrangements, not every circRNA may be a threat to genome integrity. This is demonstrated by many circRNAs not forming circR loops and circR loops residing in genes not known to enrich for somatic or oncogenic mutations. Interestingly, a number of circRNAs have also been shown to regulate transcription at their cognate locus,[76] including exon-intron circRNAs that retain introns in their mature form and are enriched in the nucleus.[41] Concordantly, R loops have been linked with gene-specific and expansive (chromatin condensation, CpG islands) transcriptional control.[38,77–79]

While the circRNAs underlying circR loops are lowly abundant and mildly enriched in the nucleus compared with all circRNAs, they are not exclusively nuclear localized. However, factors that increase the biogenesis of these circRNAs, or their retention in the nucleus, could promote translocations in hematopoietic cells and other cell types. Mutations in pre-mRNA splicing machinery have been observed in leukemia and other tumors and may alter circRNA abundance.[80] Furthermore, two dead-box RNA helicases, DDX39A and DDX39B, were shown to control circRNA nuclear export in a size-dependent manner.[81] Concordantly, loss of these helicases precipitates genome instability,[82] while a number of these helicases have also been associated with acute and chronic leukemias.[83] CircRNAs could act symbiotically and encourage translocations in the presence of additional destabilizers of genome integrity in hematopoietic stem cells—including alterations in the DNA damage response proteins, including *ATM* mutations shown to play a pathogenic role in MLL rearranged leukemia.[84] An intriguing interplay is also found between the DNA damage response pathway and RNA splicing,[85] suggesting potential co-regulation of these pathways, which could protect somatic cells from ER3D by circRNAs, while tumor cells may exhibit chaotic dysregulation manifesting as chromosomal translocations. Analysis of patient samples would go some way toward addressing these links and identifying the biogenesis factor(s) for these oncogenic circRNAs.

Intriguingly, the overexpression of *circMLL(9,10)* was able to promote co-localization of its cognate locus with other loci in the MLL recombinome in HEK293T cells. While we combined proximity ligation (4C) and visualization experiments (3D-FISH) in this study to demonstrate closer association of loci before fusion, follow-up experiments could achieve greater spatial resolution to illuminate the 3D genome architecture of the *MLL* locus by incorporating shorter FISH probes (1–10 kb in size) or molecular beacons.[86] The ability of particular circRNAs to form DSBs at their cognate loci, in combination with promoting co-localization, increases the likelihood of resultant chromosomal translocations through APOBEC-directed base excision and non-homologous DNA repair, for which we provided supportive evidence of herein.

The increase in DSBs within the *MLLT1* bcr and the circR loop, increased RNAPII pausing at the MLL locus, and the formation of chromosomal translocations in the presence of *circMLL(9,10)* implicated this circRNA as pivotal in functional targeting of networks that impact cellular DNA repair. Specifically, the demonstration of proteasome components bound to this circRNA and its ability to inhibit chymotrypsin-like proteasome activity *in vivo* and *in vitro* all point to this as a likely mechanism underlying these processes. Targeting of the proteasome, including the use of the reversible inhibitor of the 26S proteasome, bortezomib, has emerged as a promising therapy in a number of cancers, including multiple myeloma, lymphoma, and AML.[87] As distinct from its role in DNA repair, the mechanism for this effect is believed to be via other targets of the proteasome, including NF-κB in leukemic stem cells.[88] In apparent contrast to this report, increased proteasome abundance is found in the plasma of AML and ALL patients.[89] However, when normalized, and in agreement with our findings herein, the chymotrypsin activity of the 20S proteasome was uniquely lower in AML.[89] Furthermore, lower proteasome levels were a strong predictor of poor survival in AML and ALL.[89] In the Guthrie card RNA-seq data herein, it was noted by gene ontology analysis that 22 of the 46 members of the proteasome were reduced by 2-fold or more in the leukemia patients, before onset of disease symptoms, compared with controls. Another indication of the impact of reduced proteasome activity in cancer is that cancer stem cells show lower levels of proteasome activity.[90] The important role of leukemic stem cells in development of leukemia and the fact that they are refractory to standard-of-care therapies, suggests that there may be more roles for proteasome inhibition to be illuminated in leukemia and cancer, more broadly.

Guarnerio et al.[36] demonstrated that fusion circRNAs, arising after chromosomal translocations from the *MLL* gene, are critical

for progression of leukemia. However, this report did not address how transgenic MLL fusion cDNAs, which lack the capacity to form fusion circRNAs, can transform cells in numerous reports, or how these potent MLL fusion oncogenes arose as the key nucleating factor in leukemia.[91,92] An *in silico* analysis of published circRNAs in public databases supports our experimental observation that the *MLL* recombinome is enriched for circRNA-producing genes.[93] However, here we provide the evidence to show that circRNAs are involved in initiating translocations among the *MLL* recombinome, forming leukemic oncogenes. This is an example of ER3D, mediated by circRNAs and manifesting as oncogenic chromosomal translocations. The focus on AML, known to have the lowest mutation burden of any cancer,[9] has illuminated the oncogenic role of circRNAs in driving these common, yet poorly prognostic, *MLL* translocations upstream of their already demonstrated role in progression of *MLL*-rearranged leukemias.[36] ER3D represents a previously unknown, but fundamental, mechanism of genome instability, whose prevalence is likely more widespread and may be critical in understanding genomic mutations more broadly and potentially provide targets for therapy.

Limitations of the study

The ability to stably express circRNAs in primary hematopoietic cells would have allowed further interrogation of their oncogenic capacity. However, a limitation of this and other circRNA studies in hematological disease is the inability to establish expression of circRNAs in primary cells, including CD34[+] or leukemic stem cells. Furthermore, while our finding of the ability to stratify leukemia and normal patients based on RNA-seq at birth from their Guthrie cards offers potential for clinical translation, we recognize that increased sample sizes and controls to eliminate potential confounding variables would further strengthen this claim.

STAR★METHODS

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

- ● KEY RESOURCES TABLE
- ● RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- ● EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - ○ Cell lines
  - ○ Guthrie cards
  - ○ Mice
- ● METHOD DETAILS
  - ○ Constructs
  - ○ Retroviral transduction of circRNA overexpression constructs
  - ○ R-loop assay (*in vitro*)
  - ○ DNA:RNA immunoprecipitation (DRIP)
  - ○ DRIP-sequencing
  - ○ DRIP sequencing and alignment statistics
  - ○ Cell fractionation and reverse transcription
  - ○ RNA sequencing
  - ○ RNA sequencing and alignment statistics

- ○ RNA sequencing of human patient guthrie cards
- ○ Identification of fusion transcripts in RNAseq data
- ○ Circular chromosome conformation capture (4C)
- ○ 4C DNA sequencing and alignment statistics
- ○ Northern blotting
- ○ Three-dimensional DNA fluorescence *in situ* hybridization (3D FISH)
- ○ 3-Dimensional RNA/DNA fluorescence *in situ* hybridization (3D RNA/DNA FISH)
- ○ Immunofluorescence/immunohistochemistry
- ○ Chromatin immunoprecipitation qRT-PCR (ChIP-qRT-PCR)
- ○ Long-range inverse PCR
- ○ Comet assay
- ○ Western blotting
- ○ Over-representation enrichment analysis
- ○ Mouse xenotransplantation
- ○ BLISS and BLESS
- ○ NETSeq analysis
- ○ CircRNA pulldown
- ○ MS sample preparation and analysis
- ○ Cell-based proteasome assays
- ○ *In vitro* proteasome inhibition assays
- ○ RNA immunoprecipitation
- ○ ATAC-seq
- ○ Code availability
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

AUTHOR CONTRIBUTIONS

Conceptualization, S.J.C. and V.M.C.; methodology, V.M.C., M.G., J.T., K.K., J.A.P., G.C., S.M., S.T., S.T.W., R.L., S.B., J.S., S.E.S., D.L., A.D.C., T.C., C.T.W.-G., S.J.C.; formal analysis, V.M.C., J.T., B.W.S., J.A.P., S.M., and S.J.C.; investigation, V.M.C., M.G., J.T., K.K., L.G., J.A.P., R.L., S.B., J.S., G.C., S.M., S.T., S.T.W., L.H., S.E.S., D.L., A.D.C., T.C., C.T.W.-G., S.B.,

and S.J.C.; formal analysis, V.M.C., M.G., J.T., L.G., G.C., S.M., R.L., A.D.C., B.W.S., J.A.P., and S.J.C.; data curation, J.T., S.M., and S.J.C.; writing – original draft, S.J.C.; writing – review & editing, V.M.C., M.G., J.T., K.K., L.G., J.A.P., G.C., S.M., R.L., S.B., B.W.S., S.T., J.S., S.T.W., L.H., S.E.S., A.S.M., M.M., D.L., A.D.C., T.C., C.T.W.-G., L.W., L.A.S., V.T., R.J.D., S.M.P., G.J.G., and S.J.C.; visualization, V.M.C., S.J.C., H.L., J.T., S.M., and K.K.; supervision, T.C., L.A.S., V.T., R.J.D., S.M.P., G.J.G., and S.J.C.; funding, V.M.C., L.A.S., V.T., R.J.D., S.M.P., G.J.G., and S.J.C.

REFERENCES

1. Costantino, L., and Koshland, D. (2015). The Yin and Yang of R-loop biology. Curr. Opin. Cell Biol. 34, 39–45. https://doi.org/10.1016/j.ceb.2015.04.008.

2. Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nat. Rev. Genet. 16, 583–597. https://doi.org/10.1038/nrg3961.

3. Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat. Immunol. 4, 442–451. https://doi.org/10.1038/ni919.

4. Gómez-González, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc. Natl. Acad. Sci. USA 104, 8409–8414. https://doi.org/10.1073/pnas.0702836104.

5. Rada, C. (2015). Mutagenesis by AID: being in the right place at the right time. PLoS Genet. 11, e1005489. https://doi.org/10.1371/journal.pgen.1005489.

6. Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature 494, 366–370. https://doi.org/10.1038/nature11881.

7. Refsland, E.W., and Harris, R.S. (2013). The APOBEC3 family of retroelement restriction factors. Curr. Top. Microbiol. Immunol. 371, 1–27. https://doi.org/10.1007/978-3-642-37765-5_1.

8. Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol. 25, 514–522. https://doi.org/10.1016/j.tcb.2015.05.003.

9. Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. Science 349, 1483–1489. https://doi.org/10.1126/science.aab4082.

10. Cowell, I.G., Sondka, Z., Smith, K., Lee, K.C., Manville, C.M., Sidorczuk-Lesthuruge, M., Rance, H.A., Padget, K., Jackson, G.H., Adachi, N., et al. (2012). Model for MLL translocations in therapy-related leukemia involving topoisomerase IIβ-mediated DNA strand breaks and gene proximity. Proc. Natl. Acad. Sci. USA 109, 8989–8994. https://doi.org/10.1073/pnas.1204406109.

11. Meyer, C., Burmeister, T., Gröger, D., Tsaur, G., Fechina, L., Renneville, A., Sutton, R., Venn, N.C., Emerenciano, M., Pombo-de-Oliveira, M.S., et al. (2018). The MLL recombinome of acute leukemias in 2017. Leukemia 32, 273–284.

12. Marschalek, R. (2016). Systematic classification of mixed-lineage leukemia fusion partners predicts additional cancer pathways. Ann. Lab. Med. 36, 85–100. https://doi.org/10.3343/alm.2016.36.2.85.

13. Krivtsov, A.V., and Armstrong, S.A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat. Rev. Cancer 7, 823–833. https://doi.org/10.1038/nrc2253.

14. Chen, Y., Kantarjian, H., Pierce, S., Faderl, S., O'Brien, S., Qiao, W., Abruzzo, L., de Lima, M., Kebriaei, P., Jabbour, E., et al. (2013). Prognostic significance of 11q23 aberrations in adult acute myeloid leukemia and the role of allogeneic stem cell transplantation. Leukemia 27, 836–842. https://doi.org/10.1038/leu.2012.319.

15. Martens, J.H.A., and Stunnenberg, H.G. (2010). The molecular signature of oncofusion proteins in acute myeloid leukemia. FEBS Lett. 584, 2662–2669. https://doi.org/10.1016/j.febslet.2010.04.002.

16. Harper, D.P., and Aplan, P.D. (2008). Chromosomal rearrangements leading to MLL gene fusions: clinical and biological aspects. Cancer Res. 68, 10024–10027. https://doi.org/10.1158/0008-5472.CAN-08-2208.

17. Scharf, S., Zech, J., Bursen, A., Schraets, D., Oliver, P.L., Kliem, S., Pfitzner, E., Gillert, E., Dingermann, T., and Marschalek, R. (2007). Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene. Oncogene 26, 1361–1371. https://doi.org/10.1038/sj.onc.1209948.

18. Strissel, P.L., Strick, R., Rowley, J.D., and Zeleznik-Le, N.J. (1998). An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region. Blood 92, 3793–3803.

19. Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., 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. https://doi.org/10.1016/j.molcel.2016.05.032.

20. Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell 143, 122–133. https://doi.org/10.1016/j.cell.2010.09.017.

21. Schwab, R.A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C.-C., Cohn, M.A., Gibbons, R.J., Deans, A.J., and Niedzwiedz, W. (2015). The fanconi anemia pathway maintains genome stability by coordinating replication and transcription. Mol. Cell 60, 351–361. https://doi.org/10.1016/j.molcel.2015.09.012.

22. Alter, B.P. (2014). Fanconi anemia and the development of leukemia. Best Pract. Res. Clin. Haematol. 27, 214–221. https://doi.org/10.1016/j.beha.2014.10.002.

23. Jacquemont, C., and Taniguchi, T. (2007). Proteasome function is required for DNA damage response and fanconi anemia pathway activation. Cancer Res. 67, 7395–7405. https://doi.org/10.1158/0008-5472.CAN-07-1015.

24. Krogan, N.J., Lam, M.H.Y., Fillingham, J., Keogh, M.-C., Gebbia, M., Li, J., Datta, N., Cagney, G., Buratowski, S., Emili, A., et al. (2004). Proteasome involvement in the repair of DNA double-strand breaks. Mol. Cell 16, 1027–1034. https://doi.org/10.1016/j.molcel.2004.11.033.

25. Mistry, A.R., Felix, C.A., Whitmarsh, R.J., Mason, A., Reiter, A., Cassinat, B., Parry, A., Walz, C., Wiemels, J.L., Segal, M.R., et al. (2005). DNA topoisomerase II in therapy-related acute promyelocytic leukemia. N. Engl. J. Med. 352, 1529–1538. https://doi.org/10.1056/NEJMoa042715.

26. Han, J., Xiong, J., Wang, D., and Fu, X.-D. (2011). Pre-mRNA splicing: where and when in the nucleus. Trends Cell Biol. 21, 336–343. https://doi.org/10.1016/j.tcb.2011.03.003.

27. Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA 19, 141–157. https://doi.org/10.1261/rna.035667.112.

28. Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One 7, e30733. https://doi.org/10.1371/journal.pone.0030733.

29. Wilusz, J.E., and Sharp, P.A. (2013). Molecular biology. A circuitous route to noncoding RNA. Science *340*, 440–441. https://doi.org/10.1126/science.1238522.

30. Pillman, K.A., Phillips, C.A., Roslan, S., Toubia, J., Dredge, B.K., Bert, A.G., Lumb, R., Neumann, D.P., Li, X., Conn, S.J., et al. (2018). miR-200/375 control epithelial plasticity-associated alternative splicing by repressing the RNA-binding protein Quaking. EMBO J. *37*, e99016. https://doi.org/10.15252/embj.201899016.

31. Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. Cell *160*, 1125–1134.

32. Chen, S., Huang, V., Xu, X., Livingstone, J., Soares, F., Jeon, J., Zeng, Y., Hua, J.T., Petricca, J., Guo, H., et al. (2019). Widespread and functional RNA circularization in localized prostate cancer. Cell *176*, 831–843.e22. https://doi.org/10.1016/j.cell.2019.01.025.

33. Vo, J.N., Cieslik, M., Zhang, Y., Shukla, S., Xiao, L., Zhang, Y., Wu, Y.-M., Dhanasekaran, S.M., Engelke, C.G., Cao, X., et al. (2019). The landscape of circular RNA in cancer. Cell *176*, 869–881.e13. https://doi.org/10.1016/j.cell.2018.12.021.

34. Kristensen, L.S., Hansen, T.B., Venø, M.T., and Kjems, J. (2018). Circular RNAs in cancer: opportunities and challenges in the field. Oncogene *37*, 555–565. https://doi.org/10.1038/onc.2017.361.

35. Gabryelska, M.M., and Conn, S.J. (2023). The RNA interactome in the hallmarks of cancer. Wiley Interdiscip. Rev. RNA, e1786. https://doi.org/10.1002/wrna.1786.

36. Guarnerio, J., Bezzi, M., Jeong, J.C., Paffenholz, S.V., Berry, K., Naldini, M.M., Lo-Coco, F., Tay, Y., Beck, A.H., and Pandolfi, P.P. (2016). Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. Cell *165*, 289–302. https://doi.org/10.1016/j.cell.2016.03.020.

37. Powell, W.T., Coulson, R.L., Gonzales, M.L., Crary, F.K., Wong, S.S., Adams, S., Ach, R.A., Tsang, P., Yamada, N.A., Yasui, D.H., et al. (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc. Natl. Acad. Sci. USA *110*, 13938–13943. https://doi.org/10.1073/pnas.1305426110.

38. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell *45*, 814–825. https://doi.org/10.1016/j.molcel.2012.01.017.

39. Conn, V., and Conn, S.J. (2019). SplintQuant: a method for accurately quantifying circular RNA transcript abundance without reverse transcription bias. RNA N. Y. *25*, 1202–1210. https://doi.org/10.1261/rna.070953.119.

40. Conn, V.M., Hugouvieux, V., Nayak, A., Conos, S.A., Capovilla, G., Cildir, G., Jourdain, A., Tergaonkar, V., Schmid, M., Zubieta, C., et al. (2017). A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. Nat. Plants *3*, 17053. https://doi.org/10.1038/nplants.2017.53.

41. Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L., et al. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. Nat. Struct. Mol. Biol. *22*, 256–264. https://doi.org/10.1038/nsmb.2959.

42. El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014). Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. PLoS Genet. *10*, e1004716. https://doi.org/10.1371/journal.pgen.1004716.

43. Li, X., Zhang, J.-L., Lei, Y.-N., Liu, X.-Q., Xue, W., Zhang, Y., Nan, F., Gao, X., Zhang, J., Wei, J., et al. (2021). Linking circular intronic RNA degradation and function in transcription by RNase H1. Sci. China Life Sci. *64*, 1795–1809. https://doi.org/10.1007/s11427-021-1993-6.

44. Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-Based GEne SeT AnaLysis toolkit (WebGestalt): update 2013. Nucleic Acids Res. *41*, W77–W83. https://doi.org/10.1093/nar/gkt439.

45. Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. *28*, 1384–1396. https://doi.org/10.1101/gad.242990.114.

46. Li, S., Ma, Y., Tan, Y., Ma, X., Zhao, M., Chen, B., Zhang, R., Chen, Z., and Wang, K. (2018). Profiling and functional analysis of circular RNAs in acute promyelocytic leukemia and their dynamic regulation during all-trans retinoic acid treatment. Cell Death Dis. *9*, 651. https://doi.org/10.1038/s41419-018-0699-2.

47. Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554. https://doi.org/10.1016/j.cell.2015.03.010.

48. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338. https://doi.org/10.1038/nature11928.

49. Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. *45*, D777–D783. https://doi.org/10.1093/nar/gkw1121.

50. Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388. https://doi.org/10.1038/nature11993.

51. Taulli, R., Loretelli, C., and Pandolfi, P.P. (2013). From pseudo-ceRNAs to circ-ceRNAs: a tale of cross-talk and competition. Nat. Struct. Mol. Biol. *20*, 541–543. https://doi.org/10.1038/nsmb.2580.

52. Chen, L., Chen, J.-Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. Mol. Cell *68*, 745–757.e5. https://doi.org/10.1016/j.molcel.2017.10.008.

53. Dumelie, J.G., and Jaffrey, S.R. (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. Elife *6*, e28306. https://doi.org/10.7554/eLife.28306.

54. Burger, K., Schlackow, M., and Gullerova, M. (2019). Tyrosine kinase c-Abl couples RNA polymerase II transcription to DNA double-strand breaks. Nucleic Acids Res. *47*, 3467–3484. https://doi.org/10.1093/nar/gkz024.

55. Dellino, G.I., Palluzzi, F., Chiariello, A.M., Piccioni, R., Bianco, S., Furia, L., De Conti, G., Bouwman, B.A.M., Melloni, G., Guido, D., et al. (2019). Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. Nat. Genet. *51*, 1011–1023. https://doi.org/10.1038/s41588-019-0421-z.

56. Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191. https://doi.org/10.1038/nature14299.

57. Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., et al. (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat. Methods *10*, 361–365. https://doi.org/10.1038/nmeth.2408.

58. Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proc. Natl. Acad. Sci. USA *112*, 4672–4677. https://doi.org/10.1073/pnas.1504783112.

59. Fagarasan, S., Kinoshita, K., Muramatsu, M., Ikuta, K., and Honjo, T. (2001). In situ class switching and differentiation to IgA-producing cells in the gut lamina propria. Nature *413*, 639–643. https://doi.org/10.1038/35098100.

60. Lambert, L.J., Walker, S., Feltham, J., Lee, H.J., Reik, W., and Houseley, J. (2013). Etoposide induces nuclear re-localisation of AID. PLoS One *8*, e82110. https://doi.org/10.1371/journal.pone.0082110.

61. Patenaude, A.-M., and Di Noia, J.M. (2010). The mechanisms regulating the subcellular localization of AID. Nucleus *1*, 325–331. https://doi.org/10.4161/nucl.1.4.12107.

62. Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L., et al. (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat. Commun. *8*, 15058. https://doi.org/10.1038/ncomms15058.

63. Xu, X., Zhang, J., Tian, Y., Gao, Y., Dong, X., Chen, W., Yuan, X., Yin, W., Xu, J., Chen, K., et al. (2020). CircRNA inhibits DNA damage repair by interacting with host gene. Mol. Cancer *19*, 128. https://doi.org/10.1186/s12943-020-01246-x.

64. Ferrucci, P.F., Grignani, F., Pearson, M., Fagioli, M., Nicoletti, I., and Pelicci, P.G. (1997). Cell death induction by the acute promyelocytic leukemia-specific PML/RARalpha fusion protein. Proc. Natl. Acad. Sci. USA *94*, 10901–10906. https://doi.org/10.1073/pnas.94.20.10901.

65. Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678. https://doi.org/10.1038/nrg.2016.112.

66. Xu, J., Song, F., Lyu, H., Kobayashi, M., Zhang, B., Zhao, Z., Hou, Y., Wang, X., Luan, Y., Jia, B., et al. (2022). Subtype-specific 3D genome alteration in acute myeloid leukaemia. Nature *611*, 387–398. https://doi.org/10.1038/s41586-022-05365-x.

67. van de Werken, H.J.G., de Vree, P.J.P., Splinter, E., Holwerda, S.J.B., Klous, P., de Wit, E., and de Laat, W. (2012). 4C technology: protocols and data analysis. Methods Enzymol. *513*, 89–112. https://doi.org/10.1016/B978-0-12-391938-0.00004-5.

68. Verma, R., Oania, R., Fang, R., Smith, G.T., and Deshaies, R.J. (2011). Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. Mol. Cell *41*, 82–92. https://doi.org/10.1016/j.molcel.2010.12.017.

69. Nandi, D., Tahiliani, P., Kumar, A., and Chandu, D. (2006). The ubiquitin-proteasome system. J. Biosci. *31*, 137–155. https://doi.org/10.1007/BF02705243.

70. Nielsen, A.F., Bindereif, A., Bozzoni, I., Hanan, M., Hansen, T.B., Irimia, M., Kadener, S., Kristensen, L.S., Legnini, I., Morlando, M., et al. (2022). Best practice standards for circular RNA research. Nat. Methods *19*, 1208–1220. https://doi.org/10.1038/s41592-022-01487-2.

71. Schmidt, B.M., Davidson, N.M., Hawkins, A.D.K., Bartolo, R., Majewski, I.J., Ekert, P.G., and Oshlack, A. (2018). Clinker: visualizing fusion genes detected in RNA-seq data. Gigascience *7*, giy079. https://doi.org/10.1093/gigascience/giy079.

72. Libura, J., Slater, D.J., Felix, C.A., and Richardson, C. (2005). Therapy-related acute myeloid leukemia-like MLL rearrangements are induced by etoposide in primary human CD34+ cells and remain stable after clonal expansion. Blood *105*, 2124–2131. https://doi.org/10.1182/blood-2004-07-2683.

73. Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. Nat. Rev. Genet. *15*, 585–598. https://doi.org/10.1038/nrg3729.

74. Zhu, P., Zhu, X., Wu, J., He, L., Lu, T., Wang, Y., Liu, B., Ye, B., Sun, L., Fan, D., et al. (2019). IL-13 secreted by ILC2s promotes the self-renewal of intestinal stem cells through circular RNA circPan3. Nat. Immunol. *20*, 183–194. https://doi.org/10.1038/s41590-018-0297-6.

75. Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. Mol. Cell *56*, 55–66. https://doi.org/10.1016/j.molcel.2014.08.019.

76. Huang, S., Yang, B., Chen, B.J., Bliim, N., Ueberham, U., Arendt, T., and Janitz, M. (2017). The emerging role of circular RNAs in transcriptome regulation. Genomics *109*, 401–407. https://doi.org/10.1016/j.ygeno.2017.06.005.

77. Castellano-Pozo, M., Santos-Pereira, J.M., Rondón, A.G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., and Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol. Cell *52*, 583–590. https://doi.org/10.1016/j.molcel.2013.10.006.

78. Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439. https://doi.org/10.1038/nature13787.

79. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science *340*, 619–621. https://doi.org/10.1126/science.1234848.

80. Scott, L.M., and Rebel, V.I. (2013). Acquired mutations that affect pre-mRNA splicing in hematologic malignancies and solid tumors. J. Natl. Cancer Inst. *105*, 1540–1549. https://doi.org/10.1093/jnci/djt257.

81. Huang, C., Liang, D., Tatomer, D.C., and Wilusz, J.E. (2018). A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. Genes Dev. *32*, 639–644. https://doi.org/10.1101/gad.314856.118.

82. Yoo, H.H., and Chung, I.K. (2011). Requirement of DDX39 DEAD box RNA helicase for genome integrity and telomere protection. Aging Cell *10*, 557–571. https://doi.org/10.1111/j.1474-9726.2011.00696.x.

83. Cai, W., Xiong Chen, Z., Rane, G., Satendra Singh, S., Choo, Z., Wang, C., Yuan, Y., Zea Tan, T., Arfuso, F., Yap, C.T., et al. (2017). Wanted DEAD/H or alive: helicases winding up in cancers. J. Natl. Cancer Inst. *109*, djw278. https://doi.org/10.1093/jnci/djw278.

84. Oguchi, K., Takagi, M., Tsuchida, R., Taya, Y., Ito, E., Isoyama, K., Ishii, E., Zannini, L., Delia, D., and Mizutani, S. (2003). Missense mutation and defective function of ATM in a childhood acute leukemia patient with MLL gene rearrangement. Blood *101*, 3622–3627. https://doi.org/10.1182/blood-2002-02-0570.

85. Shkreta, L., and Chabot, B. (2015). The RNA splicing response to DNA damage. Biomolecules *5*, 2935–2977. https://doi.org/10.3390/biom5042935.

86. Ma, T., Chen, L., Shi, M., Niu, J., Zhang, X., Yang, X., Zhanghao, K., Wang, M., Xi, P., Jin, D., et al. (2018). Developing novel methods to image and visualize 3D genomes. Cell Biol. Toxicol. *34*, 367–380. https://doi.org/10.1007/s10565-018-9427-z.

87. Csizmar, C.M., Kim, D.-H., and Sachs, Z. (2016). The role of the proteasome in AML. Blood Cancer J. *6*, e503. https://doi.org/10.1038/bcj.2016.112.

88. Antony, M.L., Noble-Orcutt, K., Jensen, J.L., He, F., and Sachs, Z. (2021). Proteasome inhibition attenuates self-renewal in human acute myeloid leukemia by targeting NF-Kappa B in leukemia stem cells. Blood *138*, 3347. https://doi.org/10.1182/blood-2021-153338.

89. Ma, W., Kantarjian, H., Zhang, X., Wang, X., Estrov, Z., O'Brien, S., and Albitar, M. (2011). Ubiquitin-proteasome system profiling in acute leukemias and its clinical relevance. Leuk. Res. *35*, 526–533. https://doi.org/10.1016/j.leukres.2010.09.009.

90. Voutsadakis, I.A. (2017). Proteasome expression and activity in cancer and cancer stem cells. Tumor Biol. *39*. 1010428317692248. https://doi.org/10.1177/1010428317692248.

91. Krivtsov, A.V., Feng, Z., Lemieux, M.E., Faber, J., Vempati, S., Sinha, A.U., Xia, X., Jesneck, J., Bracken, A.P., Silverman, L.B., et al. (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell *14*, 355–368. https://doi.org/10.1016/j.ccr.2008.10.001.

92. Lavau, C., Luo, R.T., Du, C., and Thirman, M.J. (2000). Retrovirus-mediated gene transfer of MLL-ELL transforms primary myeloid progenitors and causes acute myeloid leukemias in mice. Proc. Natl. Acad. Sci. USA *97*, 10984–10989. https://doi.org/10.1073/pnas.190167297.

93. Dal Molin, A., Bresolin, S., Gaffo, E., Tretti, C., Boldrin, E., Meyer, L.H., Guglielmelli, P., Vannucchi, A.M., te Kronnie, G., and Bortoluzzi, S.

(2019). CircRNAs are here to stay: a perspective on the MLL recombinome. Front. Genet. *10*, 88. https://doi.org/10.3389/fgene.2019.00088.

94. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

95. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

96. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192. https://doi.org/10.1093/bib/bbs017.

97. Phanstiel, D.H., Boyle, A.P., Araya, C.L., and Snyder, M.P. (2014). Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinforma. Oxf. Engl. *30*, 2808–2810. https://doi.org/10.1093/bioinformatics/btu379.

98. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

99. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

100. Zhang, X.-O., Dong, R., Zhang, Y., Zhang, J.-L., Luo, Z., Zhang, J., Chen, L.-L., and Yang, L. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. Genome Res. *26*, 1277–1287. https://doi.org/10.1101/gr.202895.115.

101. Haas, B.J., Dobin, A., Li, B., Stransky, N., Pochet, N., and Regev, A. (2019). Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol. *20*, 213. https://doi.org/10.1186/s13059-019-1842-9.

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

103. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma. Oxf. Engl. *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

104. Kocanova, S., Goiffon, I., and Bystricky, K. (2018). 3D FISH to analyse gene domain-specific chromatin re-modeling in human cancer cell lines. Methods *142*, 3–15. https://doi.org/10.1016/j.ymeth.2018.02.013.

105. Zhao, Q., Liu, J., Deng, H., Ma, R., Liao, J.-Y., Liang, H., Hu, J., Li, J., Guo, Z., Cai, J., et al. (2020). Targeting mitochondria-located circRNA SCAR alleviates NASH via reducing mROS output. Cell *183*, 76–93.e22. https://doi.org/10.1016/j.cell.2020.08.009.

106. Attema, J.L., Bert, A.G., Lim, Y.-Y., Kolesnikoff, N., Lawrence, D.M., Pillman, K.A., Smith, E., Drew, P.A., Khew-Goodall, Y., Shannon, F., et al. (2013). Identification of an enhancer that increases miR-200b~200a~429 gene expression in breast cancer cells. PLoS One *8*, e75517. https://doi.org/10.1371/journal.pone.0075517.

107. Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. *45*, W130–W137. https://doi.org/10.1093/nar/gkx356.

108. Du, W.W., Yang, W., Chen, Y., Wu, Z.-K., Foster, F.S., Yang, Z., Li, X., and Yang, B.B. (2017). Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. Eur. Heart J. *38*, 1402–1412. https://doi.org/10.1093/eurheartj/ehw001.

109. Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Couté, Y., Conn, S., Kadlec, J., Sachidanandam, R., et al. (2014). RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell *157*, 1698–1711.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal S9.6	Stephen Leppla (NIH, Bethesda)	N/A
Mouse monoclonal RNA Pol II S2p (ChIP-seq grade)	Diagenode	Cat# C15200005-50 RRID: AB_2713925
Mouse monoclonal RNA Pol II S5p (ChIP-seq grade)	Diagenode	Cat# C15200007-50 RRID: AB_2713926
Mouse monoclonal RNA Pol II (ChIP-seq grade)	Diagenode	Cat# C15200004-50
Mouse monoclonal anti-FLAG M2	Sigma Aldrich	Cat# F3165 RRID:AB_259529
Rabbit monoclonal anti-FLAG antibody	Sigma Aldrich	Cat# SAB4301135 RRID:AB_2811010
Rabbit polyclonal anti-PSMA2	Sigma Aldrich	Cat# SAB2101893 RRID: AB_10605648
Rabbit polyclonal Histone H3 (H3pan, 1B1B2)	Diagenode	Cat# C15410324-50
Rabbit polyclonal anti-GFP	Cell Signaling Technology	Cat# 2555 RRID:AB_10692764
Rabbit polyclonal gammaH2A.X	Novusbio	Cat# NB100-384
Horse Anti-Rabbit IgG Antibody (H + L), Biotinylated	Vector Laboratories	Cat# BA-1100
IRDye® 800CW Goat anti-Mouse IgG Secondary Antibody	LI-COR Biosciences	Cat# 926-32210
Goat anti-Mouse IgG (H + L) Secondary Antibody, HRP	ThermoFisher Scientific	Cat# 31430
Goat anti-Rabbit IgG (H + L) Secondary Antibody, HRP	ThermoFisher Scientific	Cat# 31460
Goat anti-Rabbit IgG (H + L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	ThermoFisher Scientific	Cat# A11008
Normal mouse IgG	Santa Cruz	Cat# sc-2025
Bacterial and virus strains		
One Shot™ TOP10 Chemically Competent *E. coli*	ThermoFisher Scientific	Cat# C404010
Biological samples		
Neonatal Guthrie cards	Queensland Child Tumor Bank, ANZCHOG (https://anzchog.org/clinical-trials-research/anzchog-biobanking-network/queensland-childrens-tumour-bank-qld/)	De-identified
Chemicals, peptides and recombinant proteins		
Antibiotic/antimycotic (100X)	Sigma Aldrich	Cat# A5955
TrypLE™ Express Enzyme (1X), no phenol red	ThermoFisher Scientific	Cat# 12604013
DNase I, rec., RNase-free, GMP Grade	Roche	Cat# 09483195103
Ribonuclease R	Epicentre	Cat# RNR07250
Ribonuclease H	New England Biolabs	Cat# M0297S
*Nla*III	New England Biolabs	Cat# R0125L
*Bgl*II	New England Biolabs	Cat# R0144L
*Tat*I FastDigest	ThermoFisher Scientific	FD1294
*Kpn*I-HF	New England Biolabs	Cat# R3142L
*Eco*RI-HF	New England Biolabs	Cat# R3101L
*Hind*III-HF	New England Biolabs	Cat# R3104L
*Not*I-HF	New England Biolabs	Cat# R3189L
TRIzol™ Reagent	ThermoFisher Scientific	Cat# 15596018

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TrioRNA-seq kit	Tecan Biosciences	Cat# 0506
Custom AnyDeplete (rRNA, human adult and fetal globin)	Tecan Biosciences	N/A
VECTASHIELD® PLUS Antifade Mounting Medium,	Vector Laboratories	Cat# VEH19002
T4 DNA ligase	New England Biolabs	Cat# M0202L
Phusion® High-Fidelity DNA Polymerase	New England Biolabs	Cat# M0530L
ATAC-Seq kit	Active Motif	Cat# ATAC-Seq kit
Critical commercial assays		
Proteasome 20S assay kit	Enzo Life Sciences	Cat# BML-AK740-0001
Proteasome-Glo 3 substrate system	Promega	Cat# G1180
NEBNext® rRNA Depletion Kit (Human/Mouse/Rat)	New England Biolabs	Cat# E6310L
NEBNext® Ultra™ II DNA Library Prep kit	New England Biolabs	Cat# E7645S
KAPA RNA Hyper-Prep Kit for illumina®	KAPA Biosystems	Cat# KK8540
QuantiTect® reverse transcription kit	Qiagen	Cat# 205314
QuantiTect® SYBR Green PCR kit	Qiagen	Cat# 204145
DIG Northern Starter Kit	Roche	Cat# 12039672910
Translocation/Dual fusion probes KMT2A/MLLT1	Metasystems	Cat# D-5136-100-OG
MLL FISH probe (TAMRA)	Empire Genomics	Cat# MLL-20-OR
MLLT2 FISH probe (fluorescein)	Empire Genomics	Cat# AFF1-20-GR
Translocation/Dual fusion probes KMT2A/MLLT3	Metasystems	Cat# D-5133-100-OG
CometAssay® Single Cell Gel Electrophoresis Assay	R&D systems	Cat# 4250-050-K
FISH Tag™ RNA Multicolour Kit, Alexa Fluor™ dye combination	ThermoFisher Scientific	F32956
Deposited data		
DNA:RNA Immunoprecipitation (DRIPseq), HEK293T cells	This paper	GEO: GSE125985
Nuclear/cytosolic fractionation (circRNAseq), HEK293T cells	This paper	GEO: GSE125984
circMLL(9,10) overexpression (mRNAseq), HEK293T cells	This paper	GEO: GSE125984
Guthrie cards (mRNAseq), patient material	This paper	GEO: GSE217126
Circular Chromosome Confirmation Capture (4Cseq), HEK293T cells	This paper	GEO: GSE135872
CircRNA overexpression (ATACseq), HL-60 cells	This paper	GEO: GSE216874
Overexpression cell lines and tumor cell lines (mRNAseq), HL-60 cells	This paper	GEO: GSE216873
Overexpression cell lines and tumor cell lines (mRNAseq) K562 cells	This paper	GEO: GSE216873
Uniprot human database	https://www.uniprot.org/	v2017-10-25
NET-seq data	Mayer et al.[47]	GEO: GSE61332
Bless data	NCBI SRA repository	SRR1949772, SRR1949773, SRR1949774, SRR1949775
Experimental models: Cell lines		
Human: K562 cells	Richard D'Andrea	ATCC® CCL-243™ RRID:CVCL_0004
Human: HL-60 cells	Sigma Aldrich	Cat# 98070106 RRID:CVCL_0002
Human: HEK293T cells	ATCC	ATCC® CRL-3216™ RRID:CVCL_0063
Human: Human Mammary Epithelial Cells (immortalized)	Gregory Goodall	N/A
Human: MDA-MB-231	Gregory Goodall	N/A
Experimental models: Organisms/strains		
NSG mice (NOD.Cg-Prkdc scid Il2rg tm1Wjl/SzJ/Arc)	Animal Resource Center, Perth	Cat# 005557

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc)	N/A	N/A
BWA alignment algorithm	Li,[94]	ver 0.7.9a-r786
MACS2 peak caller	Zhang et al.[95]	ver 2.1.1.20160309
Integrative Genomics Viewer	Thorvaldsdóttir et al.[96]	v2.3.80
'Sushi'	Phanstiel et al.[97]	N/A
BEDTools	Quinlan and Hall,[98]	N/A
STAR spliced alignment algorithm	Dobin et al.[99]	N/A
CIRCexplorer2	Zhang et al.[100]	N/A
STAR-Fusion	Haas et al.[101]	N/A
FusionInspector tool	https://github.com/FusionInspector/FusionInspector/wiki	N/A
Clinker	Schmidt et al.[71]	N/A
cutadapt	Martin,[102]	Ver 1.8
featureCounts	Liao et al.[103]	N/A
IMARIS software	BitPlane	Ver 9.3
WebGestalt program	Ref.[107]	N/A
ProteomeDiscoverer	ThermoFisher Scientific	Ver2.4
Oligonucleotides		
Oligonucleotide sequences	N/A	See Table S6
Recombinant DNA		
pMIG (MSCV-IRES-GFP)	Richard D'Andrea	N/A
pEQ-PAM3	Richard D'Andrea	N/A
Synthetic gene fragments	N/A	See Table S6
Other		
Streptavidin-conjugated peroxidase	Pierce	Cat# 21127
SuperSignal™ West Pico PLUS chemiluminescent substrate	ThermoFisher Scientific	Cat# 34578
Clarity Max™ Western ECL substrate	Bio-Rad	Cat# 1705062
Fetal Bovine Serum, NZ origin, filtered, sterile	Bovogen	Cat# FBSAU-2007A (Lot1710A)
Protein A - Sepharose™ 4B	ThermoFisher Scientific	Cat# 101041
PARIS™ kit	ThermoFisher Scientific	Cat# AM1921
HiScribe™ T7 RNA High Yield RNA Synthesis kit	New England Biolabs	Cat# E2040S
Proteinase K, Molecular Biology Grade	New England Biolabs	Cat# P8107S
MEGAScript™ T7 transcription kit	ThermoFisher Scientific	Cat# AMB13345
Lipofectamine® 2000 reagent	ThermoFisher Scientific	Cat#11668019

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Simon Conn (simon.conn@flinders.edu.au).

Materials availability
Plasmids and cell lines generated in this study can be requested from the lead contact.

Data and code availability
RNA-seq, circRNA-seq, DRIP-seq, 4C-seq and ATAC-seq data have been deposited at GEO and are publicly available. Accession numbers are listed in the key resources table. Microscopy data reported in this paper will be shared by the lead contact upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Cell lines

All human cell lines were incubated at 37°C with 5% CO$_2$. Human embryonic kidney cells (HEK293T, ATCC CRL-3216) were grown in DMEM (GIBCO BRL) with 10% fetal bovine serum (Thermo Fisher Scientific), 1x antibiotic/antimycotic (Sigma Aldrich). Human leukemic cell lines HL-60 (ATCC CCL-240) and K562 (ATCC CCL-243) were grown in RPMI 1640 with 2 mM L-glutamine (Thermo Fisher Scientific), 10% fetal bovine serum and 1x antibiotic/antimycotic. Human mammary epithelial cells (HMLE) were treated with 2 ng/mL TGF-β1 for 21 days to become mesenchymal HMLE (mesHMLE) cells, as per Conn et al.[31]

Guthrie cards

Neonatal blood spots (Guthrie cards) were obtained for eight patients who developed acute leukemia prior to 14 years of age with MLL rearrangements confirmed by cytogenetics and eight control Guthrie cards (children showed no hematological disease). Use of human neonatal guthrie cards was approved by the Central Adelaide Local Health Network Human Research Ethics Committee (Approval # HREC/17/RAH/284) and parental informed consent was obtained for all human subjects.

Mice

Eight-week-old, female NSG (NOD.Cg-Prkdc scid Il2rg tm1Wjl/SzJ/Arc) mice were obtained from the Animal Resource Center (Cat#005557) and acclimatized for one week prior to xenografting. Mouse were housed in individually ventilated cages with a maximum of five animals per cage, with sterile bedding and fed with sterilized food pellets and water. All animal studies were approved by the Flinders University and Southern Adelaide Local Health Network Animal Welfare Committee (Approval #AEM1634).

METHOD DETAILS

Constructs

Overexpression of circRNAs was achieved as per Conn et al.,[31] using pcDNA3.1 (Life Technologies) for transient overexpression in HEK293T cells, or the retroviral MSCV-IRES-GFP (pMIG) vector for stable overexpression in HL-60 and K562 cells. Oligonucleotides are listed in Table S6.

Retroviral transduction of circRNA overexpression constructs

Retroviral particles were generated in HEK293T cells with pEQ-PAM3 packaging plasmid. Target cells (HL-60) were exposed to virus-containing supernatant and selected by fluorescence-activated cell sorting (FACS) for GFP 48 h after transduction. Cultures with GFP proportions exceeding 97% were used for experiments as assessed by the TN Gallios Flow Cytometer (Beckman Coulter). Cells were checked at weekly intervals to track circRNA expression and assess chromosomal translocations.

R-loop assay (*in vitro*)

In vitro R-loop assays were performed as described in Conn et al.,[40] with the following modifications. Synthetic circular RNAs were made by *in vitro* transcribing exons 9–10 of *MLL* with HiScribe T7 RNA High Yield RNA Synthesis kit (New England Biolabs) from gBlock synthetic gene fragment (Integrated DNA Technologies, Table S6). Circularization was performed by depleting linear RNA from the circular RNA preparation with the use of Ribonuclease R as described.[40] Linear and circular RNAs for exons 9–10 of *MLL* were hybridized to pcDNA3.1 plasmid containing the cognate genomic locus spanning the 5′ intron to the 3′ intron. Probing with S9.6 mouse mAb (kind gift from Stephen Leppla) followed by secondary goat anti-mouse-HRP (ThermoFisher Scientific, Cat#31430) and chemiluminescent detection was performed with Clarity Max Western ECL substrate (Bio-Rad).

DNA:RNA immunoprecipitation (DRIP)

Plasmid DNA (6 μg) was transfected into HEK293T cells at 60–70% confluency in a 10-cm dish with Lipofectamine 2000 reagent (Thermo Fisher Scientific) according to manufacturer's instructions. Three days post-transfection, cells were washed with PBS twice and resuspended in 1 mL lysis buffer (10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM EDTA, 0.5% SDS) supplemented with 8 U Proteinase K (New England Biolabs) and incubated at 37°C for 18 h. Genomic DNA was isolated with GenElute Mammalian Genomic DNA Miniprep Kit (Merck) and eluted in 200 μL TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Genomic DNA in TE buffer was directly used for sonication in S220 focused ultra-sonicator (Covaris) with 200 bp target peak option (130 μL sample volume, Intensity 5, Duty cycle 10%, 200 cycles per burst, 180 s treatment time). Sonicated samples were checked on 1% agarose gel to confirm the sonication efficiency. Some samples were treated with either 40 U Ribonuclease R (RNase R, Epicentre) or 10U RNase H and incubated at 37°C for 30 min. These were phenol:chloroform extracted and subsequently 60 μg genomic DNA was added into DRIP buffer (50 mM HEPES/KOH at pH 7.5, 140 mM NaCl, 5 mM EDTA, 1% Triton X-100) and incubated with 20 μg S9.6 antibody for 18 h in 400 μL total volume on a rotator at 4°C. Subsequently, protein A Sepharose beads (Thermo Fisher Scientific) were blocked with 5 mg/mL Bovine Serum Albumin (BSA) in PBS for 2 h and 40 μL of blocked-beads (50% slurry) was incubated with DNA-antibody complexes for 2 h on a rotator at 4°C. Samples were centrifuged for 2 min at 800 *g*, and beads were washed sequentially with 1 mL TSE buffer-1 (0.1% NP-40, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH: 8.0), 150 mm NaCl), 1 mL TSE buffer-2 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH: 8.0), 500 mM NaCl), 1 mL TSE buffer-3 (0.25 M LiCl, 1% NP-40, 1% deoxycholate,

1 mM EDTA, 10 mM Tris-HCl (pH: 8.0)) and 1 mL TE washing buffer (2 mM EDTA, 10 mM Tris-HCl (pH: 8.0)). Finally, R-loops were eluted in 100 μL of elution buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS) for 15 min at 65°C and further purified with QIAquick PCR Purification Kit (Qiagen). The concentration and size of the complexes were determined using Qubit Fluorometer and Agilent Bioanalyzer, respectively.

DRIP-sequencing

Following DRIP, 2 ng DNA was used to generate DRIP-seq libraries with NEBNext Ultra II DNA Library Prep kit (New England Biolabs) according to manufacturer's protocol. DNA-RNA immunoprecipitation followed by high-throughput sequencing (DRIP-seq) samples were multiplexed and sequenced on the Illumina NextSeq 500 using the stranded single end protocol with a read length of 75 (high output kit). Raw reads were adaptor trimmed and filtered for short sequences using cutadapt v1.8.1,[102] using the same parameter set as described in the RNA Sequencing methods section. The resulting FASTQ files averaging 52.7 million reads per sample were analyzed and quality checked using the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

The filtered reads were mapped against the human reference genome (Hg19) using the BWA alignment algorithm[94] (version 0.7.9a-r786 with default parameters) returning an average alignment rate of 92% (table below). Enriched regions of the genome were identified using the MACS2 peak caller[95] (version 2.1.1.20160309) using default parameters and reporting only peaks with an FDR cut-off (-q) < 0.05. Alignments were again visualized and interrogated using the Integrative Genomics Viewer v2.3.80.[96] The R package 'Sushi'[97] was used to visualize the read depth/coverage of samples across a specific gene. For this, BEDTools[98] were used to calculate the gene coverage using 'genomeCoverageBed' command and stored in bedgraph format. The plots were generated using the calculated log2 fold read coverage across 100nt window sizes.

DRIP sequencing and alignment statistics

ID	Replicate #	library	# input reads	# uniquely mapped reads	% uniquely mapped reads
Input_DNA	1	DRIP-seq	77,618,516	74,136,108	95.51
circMLL(9,10) 1	1	DRIP-seq	24,236,934	21,265,086	87.74
circMLL(9,10) 2	2	DRIP-seq	9,561,297	8,384,347	87.69
circMLL(12) 1	1	DRIP-seq	66,811,746	62,763,049	93.94
circMLL(12) 2	2	DRIP-seq	54,371,615	51,157,686	94.09
pcDNA3.1 EV1	1	DRIP-seq	68,818,892	64,390,627	93.57
pcDNA31 EV2	2	DRIP-seq	68,020,155	63,510,097	93.37
		Average	52,777,022	49,372,429	92

Cell fractionation and reverse transcription

Nuclear and cytoplasmic RNA were fractionated from human cell lines (HEK293T, HL-60) using the PARIS system (Ambion), according to manufacturer's instructions. Purity of cellular fractionation was confirmed by qRT-PCR using nuclear (*45S*, *MALAT1* ncRNA) and cytoplasmic (*GAPDH, 18S rRNA*) markers. Equivalent cellular input of cytoplasmic RNA and nuclear RNA from each sample was treated with DNase I (Roche) at 37°C for 15 min, the digested with RNase R, at 2 U/μg RNA, then cleaned up with RNeasy micro kit (Qiagen). RNA was then depleted for rRNA using the NEBNext rRNA Depletion Kit (Human/Mouse/Rat) (New England Biolabs) before libraries were constructed with the KAPA RNA Hyper-Prep Kit for illumina (KAPA Biosystems). For qPCR, prior to RNase R digestion, RNA was reverse transcribed with QuantiTect reverse transcription kit (Qiagen). After RT, QuantiTect SYBR Green PCR Kit (Qiagen) was used for qRT-PCR as per Conn et al.[31] For detecting fusion mRNAs, reverse transcription was specially primed using a gene specific reverse primer, rather than oligodT and random hexamer master mix.

RNA sequencing

Whole transcriptome RNA, or nuclear and cytoplasmic-fractionated RNA from ribo-depleted (total RNA-seq) or RNase R treated (circular RNA-seq) samples, respectively, were multiplexed and sequenced separately on the Illumina NextSeq 500 platform (high output mode) using the stranded, single-end protocol with a read length of 150. Raw reads were adaptor trimmed and filtered for short sequences using cutadapt v1.8.1,[102] setting minimum-length option to 18, error-rate 0.2 and overlap 5. The resulting FASTQ files averaging 67.5 and 40.4 million reads per sample for total RNAseq and circRNAseq respectively were analyzed and quality checked using the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

Reads were mapped against the human reference genome (Hg19) using the STAR spliced alignment algorithm[99] (version 2.5.3a with default parameters and –chimSegmentMin 20) returning an average unique alignment rate of 79% and 81% for total RNAseq and circRNAseq respectively (Table below). The resulting STAR produced Chimeric.out.junction file for each sample was parsed and annotated for circRNA prediction and back-splice abundance using CIRCexplorer2.[100] Alignments were visualized and interrogated using the Integrative Genomics Viewer v2.3.80.[96]

RNA sequencing and alignment statistics

ID	Replicate #	library	# input reads	# uniquely mapped reads	% uniquely mapped reads
MLL_e9-10_R1	1	riboMinus	73,666,457	58,060,262	78.82
MLL_e9-10_R2	2	riboMinus	69,103,134	53,975,124	78.11
pcDNA31_R1	1	riboMinus	61,293,184	48,857,812	79.71
pcDNA31_R2	2	riboMinus	65,566,376	51,829,336	79.05
Nuc_Rep1	1	RNase R	37,834,224	30,386,037	80.31
Nuc_Rep2	2	RNase R	30,436,146	24,101,861	79.19
Cyto_Rep1	1	RNase R	33,842,915	28,242,889	83.45
Cyto_Rep2	2	RNase R	59,379,873	47,520,026	80.03
		Average	40,373,290	32,562,703	80.75

RNA sequencing of human patient guthrie cards

RNA was extracted from one-half of a bloodspot using TRIzol Reagent (ThermoFisher Scientific) from these and 8 control Guthrie cards for children who did not develop cancer. Informed parental consent was obtained for all participants. DNAse I (Roche) treatment was performed, and RNA purified with RNeasy MinElute cleanup kit (Qiagen). The RNA quality was assessed by LabChip and 50 ng was used as the input material for TrioRNA-seq kit according to manufacturer's instructions (Tecan Biosciences). A custom AnyDeplete module was employed to remove ribosomal RNA, adult and fetal globin transcripts. The RNAseq libraries were sequenced with the NextSeq 500 (Illumina) and mapped as above. RNA was also made into cDNA using the Quantitect Reverse Transcriptase kit (Qiagen) and used in qRT-PCR to detect mRNA and circRNA transcripts. Oligonucleotides are listed in Table S6.

Identification of fusion transcripts in RNAseq data

STAR-Fusion[101] was used to identify the fusion transcripts from RNAseq data and the output generated was used by the STAR aligner to map junction reads during the circRNA detection workflow and Trinity Cancer Transcriptome Analysis Toolkit (CTAT) libraries. The FusionInspector tool (https://github.com/FusionInspector/FusionInspector/wiki) was applied for *in silico* characterization, and interpretation of the candidate fusion transcripts from STAR-Fusion and exploration of their sequence and expression characteristics. Fusion plots were generated from RNA-seq data using Clinker.[71]

Circular chromosome conformation capture (4C)

The 4C experiments were performed according to van de Werken et al.[67] Briefly, *circMLL(9,10)* overexpression constructs (in pcDNA3.1 vector backbone) or the empty pcDNA3.1 vector were transfected into HEK293T cells using lipofectamine 2000, with 1×10^7 cells fixed at 24 h post-transfection. Overexpression was confirmed by RT-PCR. The *MLL* genomic region targeted as the bait was chr11:118,354,788-118,355,290 (hg19), exploiting primary and secondary restriction enzyme *Nla*III and *Bgl*II, respectively. Oligonucleotides are listed in Table S6. High-throughput sequencing was performed using Illumina NextSeq (SE150). The unindexed reads were sorted according to the barcode and sorted reads were trimmed and filtered for short sequences using cutadapt v1.8,[102] setting minimum-length option to 15, error-rate 0.2 and overlap 5. The quality of the data was checked using the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The reads mapped against the human reference genome (hg19) using the STAR spliced alignment algorithm[99] (version 2.7.0e with default parameters) and resulted in 80% of average unique alignment. The featureCounts program was used to assign the sequence reads to genomic features and counted the reads assigned to each gene.[103]

4C DNA sequencing and alignment statistics

ID	Barcode # and sequence	library	# input reads	# uniquely mapped reads	% uniquely mapped reads
pcDNA3.1_R1	3 (ATG)	DNAseq	6,763,703	5,612,462	82.98
pcDNA3.1_R2	5 (AGCA)	DNAseq	12,575,305	10,278,182	81.73
*circMLL(9,10)*_R1	4 (CCG)	DNAseq	8,585,484	6,645,830	77.41
*circMLL(9,10)*_R2	6 (TACG)	DNAseq	13,939,501	11,085,291	79.52
		Average	9,308,164	8,405,441	80.75

Northern blotting

A single, 200 nt, DIG-labeled RNA probe was generated with DIG Northern Starter Kit (Roche) to target the back-splice junction of *circMLL(9,10)*. Template for the *in vitro* transcription reaction was a gBlock (Integrated DNA Technologies) with a T7 promoter

sequence at the 5′ end (Table S6). The Northern was run on a 1% Agarose formaldehyde MOPS gel, with transfer, cross-linking, pre-hybridization, hybridization, washing and development done as per manufacturer's instructions.

Three-dimensional DNA fluorescence *in situ* hybridization (3D FISH)

3D FISH was performed as per Kocanova et al.[104] on HEK293T cells grown on poly L-lysine-coated glass coverslips transfected with pcDNA3.1 (empty vector) or pcDNA3.1*circMLL(9,10)* 48 h prior to fixation. Fluorescently labeled BAC DNA FISH probe sets were purchased for *MLL* (*KMT2A,* 5-TAMRA label) and *MLLT2* (*AFF1*, 5-fluorescein label)(Empire Genomics). Translocation/Dual Fusion probe sets were obtained for KMT2A-MLLT1 (Metasystems, D-5136-100-OG) and KMT2A-MLLT3 (Metasystems, D-5130-100-OG). Nuclei were counterstained with DAPI (1 μg/mL) in Vectashield mounting medium. Coverslips were imaged using the Olympus FV3000 confocal microscope (Olympus, Tokyo, Japan) and z-stacks obtained using optimized intervals (z-voxel approximately 0.41 μm). Images were obtained with Laser excitation: 405 nm and emission detection wavelength: 430–470 nm for DAPI; laser excitation: 488 nm and emission detection wavelength: 500–540 nm for *MLLT2* (Empire Genomics) or MLL (Metasystems) and laser excitation: 561 nm, and emission detection wavelength: 570–620 nm for *MLL* (Empire Genomics) or *MLLT1*, *MLLT3* (Metasystems) using the 40× objective and 4.63X zoom. Measurement of distances between signals in the nucleus was achieved on 3-dimensional reconstructions using IMARIS software ver9.3 (BitPlane) for each treatment (n = 20 cells for each cell population).

3-Dimensional RNA/DNA fluorescence *in situ* hybridization (3D RNA/DNA FISH)

3D RNA/DNA FISH was performed on HEK293T cells grown on poly L-lysine-coated glass coverslips transfected with pcDNA3.1 (empty vector) or pcDNA3.1*circMLL(9,10)* 48 h prior to fixation. DNA-FISH for *MLL* (*KMT2A*) with 5-TAMRA label was performed as above. Localization of circRNA was achieved as per Zhao et al.[105] using an *in vitro* transcribed RNA targeting the back-splice junction of *circMLL(9,10)* and labeled with Alexa Fluor 647 using the FISH Tag RNA Multicolor kit (ThermoFisher Scientific) (Table S6). Nuclei were counterstained with DAPI in Vectashield mounting medium. Slides were imaged using the Zeiss LSM880 confocal microscope (Zeiss, Oberkochen, Germany) and z-stacks obtained using optimized intervals (z-voxel approximately 0.41 μm). Images were obtained with Laser excitation: 405 nm and emission detection wavelength: 430-470nm for DAPI; laser excitation: 650nm and emission detection wavelength: 650-690nm for *circMLL(9,10)* (Alexa Fluor 647) and laser excitation: 561nm, and emission detection wavelength: 570-620nm for *MLL* (5-TAMRA) using the 40× objective and 4.63X zoom. Signal overlap in the nucleus was achieved on 3-dimensional reconstructions using IMARIS software ver9.3 (BitPlane) for each treatment.

Immunofluorescence/immunohistochemistry

Immunofluorescence was performed on HEK293T cells transiently transfected with pcDNA3.1*circSMARCA5(15,16)* or pcDNA3.1-*circMLL(9,10)* for 36 h and grown on 13 mm diameter poly-L-lysine (Sigma Aldrich, 1% solution) coated glass coverslips. These were permeabilized and probed with gammaH2A.X (Novusbio, NB100-384 @ 1:500), with secondary antibody (Goat anti-Rabbit IgG (H + L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488, ThermoFisher Scientific A11008, @ 4 μg/mL) according to manufacturer's protocol. Imaging was performed using an SP5 Spectral Confocal Microscope (Leica).

For the immunohistochemical detection cells of human origin from mouse tumors, 10 μm sections of FFPE tumor were probed with anti-GFP rabbit polyclonal antisera (Cell Signaling Technology, MA, USA, Code # 2555) using a standard streptavidin-biotinylated immunoperoxidase technique. Sections were dewaxed using xylene and then rehydrated through alcohols. Sections were then treated with methanol/H₂0₂ for 30 min. The sections were then rinsed twice in PBS (pH 7.4) for a further 5 min each wash. Antigen retrieval was performed using Citrate buffer (pH 6.0). Slides were allowed to cool and washed twice in PBS (pH 7.4). Non-specific proteins were blocked using normal horse serum for 30 min. The antibody was applied at a dilution of 1:500 at room temperature overnight. Slides minus the primary antibody were also run in parallel. The following day, the sections were given two washes in PBS then a biotinylated anti-rabbit secondary (Vector Laboratories, USA, Cat # BA-1100) was applied for 30 min at room temperature. Following two PBS washes, the slides were incubated for 1 h at room temperature with a streptavidin-conjugated peroxidase tertiary (Pierce, USA, Cat # 21127). Sections were then visualized using diaminobenzidinetetrahydrochloride (DAB), washed, counterstained with haematoxylin, dehydrated, cleared and mounted on glass slides.

Chromatin immunoprecipitation qRT-PCR (ChIP-qRT-PCR)

ChIP was performed as per Attema et al.[106] on HEK293T cells, with and without overexpression of *circMLL(9,10)*, using rabbit polyclonal, ChIP-grade antisera targeting total RNAPII, phosphoSerine2 (pSer2) RNAPII and phosphoSerine5 (pSer5) RNAPII (Diagenode). Oligonucleotides are listed in Table S6.

Long-range inverse PCR

Genomic DNA was harvested from transduced HL-60 cells to screen for translocations by inverse PCR. Custom inverse PCR method targeting Intron 8 was performed based on Libura et al.[72] Digestion of purified genomic DNA (1 μg) was performed for 18 h with 20 U *Tat*I (ThermoFisher Scientific) and purified by Qiaquick PCR purification kit (Qiagen). Purified DNA was ligated with 40 U T4 DNA ligase (New England Biolabs) at 4°C for 18 h in a 250 μL volume (1 ng/μL DNA concentration) to promote intramolecular circularization. After column purification, the circularized DNA was then digested with *Bgl*II (6U; New England Biolabs) prior to PCR amplification to deplete native *MLL* circularized DNA molecules.

PCR amplification was performed with Phusion High-Fidelity DNA Polymerase (NEB) with 1x Q-solution (Qiagen) in the first round (F1: GCAGTTGCGTGTAGTCCCAG; R1: GCCCAGCTAAAAGCCTAAAC) and the second round (F2: CCTGGGTGACAAAGCAAAAC; R2: CCCTGAGAAATGGCAGAGAA) with the following PCR conditions: initial denaturation for 3 min s at 95°C, 35 cycles of 95°C (15 s), 57°C (15 s) and 72°C (2 min 30 s), final extension at 72°C for 5 min. All PCR products were gel purified and underwent Sanger sequencing.

Comet assay
Comet Assay Single Cell Gel Electrophoresis Assay (R&D systems) was performed on HL-60 cells expressing circRNAs according to manufacturer's instructions. Neutral assay was performed to detect the extent of DSBs genome-wide. UV-C irradiation (peak wavelength 253.7 nm) of HL-60 EV cells in DPBS was carried out at room temperature using a UV Stratalinker 1800 (Stratagene) with an irradiation dose of 10 J/cm^2.

Western blotting
Western blotting was performed as per Conn et al[31] using the following antibodies, primary: Monoclonal anti-FLAG M2 antibody (F3165, Sigma Aldrich), rabbit anti-FLAG antibody (SAB4301135, Sigma Aldrich), rabbit anti-PSMA2 (SAB2101893, Sigma Aldrich) and Histone H3 (H3pan) monoclonal antibody (1B1B2, Diagenode) and secondary: Goat anti-mouse IRDye800 (LI-COR, 926–32210; 1:20,000), Goat Anti-Mouse IgG HRP (cat. #31430, ThermoFisher Scientific) and goat Anti-Rabbit IgG HRP (cat. #31460, ThermoFisher Scientific). Chemiluminescent detection was performed with the SuperSignal West Pico PLUS chemiluminescent substrate (cat. #34578, ThermoFisher Scientific). Blots were imaged using the Odyssey 9120 imaging system (Li-Cor Biosciences) or ChemiDoc instrument (Bio-Rad).

Over-representation enrichment analysis
Analysis of gene sets were performed for genes containing the circR-loops, using the WebGestalt program.[107] We searched the *Homo sapiens*, disease functional database (GLAD4U). We utilized the genome_protein-coding reference genome and restricted the output to those categories with >200 genes.

Mouse xenotransplantation
HL-60 cells or K562 cells stably expressing pMSCV-IRES-GFP (pMIG; empty vector), *circMLL(9,10)* and *circMLLT1(2)*, or *circMLL(9,10)* and *circMLLT1(2)* were sorted for GFP expression (FACSAria II, Becton Dickinson) prior to intravenous injection of 5 × 10^6 cells into 6–8 week old NSG mice (8 mice per group). Mice were monitored daily and humanely culled when they showed overt signs of morbidity due to the tumor burden. The experiment was terminated at day 94. Tumors were excised, homogenized through 70 μm cell strainers into IMDM containing 10% FBS, then GFP sorted (FACSAria II, Becton Dickinson) to purify HL-60 cells, with RNA extracted by TRIzol (ThermoFisher Scientific).

BLISS and BLESS
Breaks labeling *in situ* sequencing (BLISS) was performed as described,[62] with qRT-PCR performed, rather than NGS, on HEK293T and HL-60 cells. Mammalian cells were transfected with overexpression vectors encoding circRNAs (*circMLL(9,10)*, *circMLLT1(2)*, *circMLL(9,10)* + *circMLLT1(2), circSMARCA5(15,16)*) along with AID, or a catalytically dead mutant version (mutAID) using Lipofectamine 2000 (ThermoFisher). After 48 h, these were trypsinized (for HEK293T cells), counted and 4 × 10^5 cells allowed to adhere to poly-lysine (0.1% w/v, Sigma-Aldrich) coated 12mm glass coverslips. All remaining steps until the ligation of adapters to DSBs were completed on slides following cell permeabilization. *In vitro* transcription was performed using the MEGAScript T7 transcription kit (Thermo Fisher Scientific), Reverse transcription and qRT-PCR was performed as described above. Oligonucleotides are listed in Table S6.

The raw data (fastq files) of BLESS for HEK293T cells was downloaded from the NCBI SRA repository (SRR1949772, SRR1949773, SRR1949774, SRR1949775) and processed to generate the DSB (DNA double-stranded breaks) clusters as described.[56] The DSB cluster information converted into BED file format and collected overlapping regions to the circular RNA using bedtools intersect and closest options. A custom PERL script was developed to process the data to find the DSB clusters in upstream regions (5 kb and 10 kb) from the first exon of circular RNAs and averaged the counts of replicates.

NETSeq analysis
The bedgraph files of NET-seq data for HEK293T cells were downloaded from NCBI Accession Number GSE61332.[47] These files were converted to BED file format and overlapping regions of circular RNA were calculated using bedtools in a strand-specific manner. Custom scripts were developed to count the number of peaks in the 1 kb region upstream of the first exon of circular RNAs, and downstream of circular RNAs, based on the hg19 human genome annotation.

CircRNA pulldown
Biotinylated (5′TEG-biotin), 30 nt-long DNA oligonucleotides were designed to span the back-splice junction of *circMLL(9,10)* (Integrated DNA Technologies) - antisense strand 5′- CGGAGGTGGTTTTTCCTCTACATGCCCACT-3′, and sense strand (negative control) 5′-AGTGGGCATGTAGAGGAAAAACCACCTCCG-3′. BLAST analysis of probes revealed no significant matches and secondary

structure folding assessment with RNAfold showed no strong secondary structure. CircRNA pulldown was performed as per Du et al.[108] using unfixed HEK293T cells, transiently transfected (40 h) with pcDNA3.1circMLL(9,10). Three biological replicates were prepared for both the antisense and sense (control) probes.

MS sample preparation and analysis

Protein eluate from the circRNA pulldown was concentrated, and protein was quantified with Bio-Rad Protein Assay Dye reagent concentrate. One microgram of total protein was subjected to in solution trypsin digestion by incubating for 30 min at 56°C in the following conditions: 111 mM Tris pH 8.0, 1.1 mM $CaCl_2$ (BDH Chemicals), 5.6 mM DTT (ThermoFisher Scientific) in 18 μL reduction reaction. Subsequently, 1 μL of 200 mM 2-Chloroacetamide (Sigma) was added to above reaction after it has cooled to room temperature and incubated in the dark at room temperature for 30 min. Protein samples were digested at a 1:50 Enzyme:Protein ratio reaction, containing 19 μL of the above reaction mixture and 1 μL Trypsin (Pierce) solution to the final concentration of 1 ng/μL. The protein sample was vortexed and incubated at 37°C overnight, transferred to MS vials (ThermoFisher) and analyzed by MS within 24 h. MS was performed using a Thermo Exploris 480 coupled to an Ultimate 3000 uHPLC. Peptides were loaded onto a PepMap 100 trap (300 μm × 5 mm, 5 μm C18 beads) and separated on a custom 40 cm × 75 μm pulled spray tip column packed with 1.9 μm beads (Dr Maisch) using a 55 min 3-25% and 6 min 25-50% acetonitrile gradients. The MS instrument was operated in Top20 DDA mode with a 1.4 m/z quad isolation width, 30% HCD collision energy, 60,000 MS1/15,000 MS2 resolution and dynamic exclusion of 20 s. Spectra were analyzed with ProteomeDiscoverer 2.4 (ThermoFisher) using the Uniprot human database (v2017-10-25), with Dynamic modifications of methionine oxidation (+15.995 Da), deamidated Asp/Gln (+0.984 Da) and N-terminal modification of Gln to pyro-Glu (−17.027 Da Q) and static modification of carbamidomethylation to cysteine with (+57.021 da). Protein hits were identified by the following filtering criteria (1) rejecting proteins with only a single peptide, (2) having to be present in at least 2 of 3 biological replicates for the antisense strand pulldowns and (3) absent from all three sense (negative control) pulldowns.

Cell-based proteasome assays

Cell-based proteasome assays were performed using the Proteasome-Glo 3 substrate system (Promega, G1180) to quantify chymotrypsin-like, caspase-like and trypsin-like proteasome activity as per manufacturer's instructions. Briefly, 20,000 HL-60 cells in biological triplicate were obtained following cell counting with CellDrop automated cell counter (De-Novix) and underwent lysis. Assay activity was measured with luminescence with the ClarioStar plus spectrophotometer (BMG Labtech). Epoxomicin (5 μM) was utilized as a potent, irreversible inhibitor of the 20S proteasome, with increased effect on the chymotrypsin-like activity.

In vitro proteasome inhibition assays

The Proteasome 20S assay kit (Enzo Life Sciences, BML-AK740-0001) was utilized to quantify effects of co-incubation with circular RNAs specifically on chymotrypsin-like activity using purified human 20S proteasome. The Suc-LLVY-AMC fluorogenic peptide substrate permits measurement the chymotrypsin-like activity by the release of the free AMC fluorophore after proteolysis. This fluorescence was measured using the ClarioStar plus spectrophotometer (BMG Labtech) with Excitation at 360/10 nm and emission at 450/10 nm. The AMC fluorophore was serially diluted in biological triplicate to establish a standard curve to allow calculation of specific activity with line of best fit ($R^2 = 1.00$) as below:

Arbitrary fluorescence units (AFU) = 3,444,608.46* AMC concentration (μM) + 75,728.46.

Ten nanograms of *in vitro* circularized, PAGE-purified *circMLL(9,10)* or synthetic, scrambled control (circScrambled) was incubated with 3x molar excess of purified BSA for 1h on ice, prior to use in this assay. Experiments were performed in triplicate and data presented as mean ± SD over 20 min. As recommended, specific activity was calculated between 6 and 12 min after assay commencement to target the linear phase of the assay.

RNA immunoprecipitation

Human PSMA2 without a stop codon was ordered as a double stranded DNA fragment (gBlock, Integrated DNA technologies) and cloned into pcDNA3.1-FLAG using *Kpn*I and *Eco*RI restriction sites in-frame with the C-terminal FLAG tag sequence. The construct was verified by Sanger sequencing. This construct was co-transfected with pcDNA3.1circMLL(9,10) into HEK293T cells and cell pellets were collected 36 h later with RNA immunoprecipitation (RIP) performed as per.[109] RNA was purified from 25% of the IgG IP (control) and FLAG IP and qRT-PCR was performed on input (5%) in technical triplicate. Oligonucleotides are listed in Table S6. Capture (%) was calculated based on mean Ct values of RNA target in each IP sample and compared with input abundance.

ATAC-seq

The ATAC-Seq kit (Active Motif) was utilized to assess regions of open chromatin in the empty vector HL-60 cell line, and cell lines overexpressing either *circMLL(9,10)* or *circMLLT1(2)* as per the manufacturer's protocol with 100,000 cells. The resultant three libraries were multiplexed and sequenced on the Illumina NovaSeq SP (100bp, single end), with libraries ranging from 81 to 156 million reads (76–80% alignment to the hg19 genome assembly with removal of mitochondrial genes and duplicate reads). Peaks were called with Genrich and MAC2. Traces were produced as per DRIP-seq methods above.

Code availability
Custom PERL scripts were used to process the data for analysis of the public BLESS and NET-seq data as mentioned. These are available from the corresponding author upon reasonable request.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis was performed using GraphPad PRISM ver9.1.1. Alpha (type I error rate) was set at 0.05 throughout all analyses as the measure of statistical significance. N refers to the number of biological replicates in the experiment, while technical replication is detailed in the figure legend. No subjects/datapoints were excluded from analysis. Sample sizes for animal experimental groups were calculated using GPower with power of 0.85 and sigma of 20, giving sample size of 8 animal per group.