SUPPLEMENTARY METHODS

Isolation of common myeloid progenitors (CMP) and 3'RNA sequencing

GM-CSF mobilized leukapheresis products were obtained from UC San Diego Health Stem Cell Processing Lab (La Jolla, CA) in accordance with a university-approved Institutional Review Board (IRB) protocol. Thawed products were diluted in PBS supplemented with 1mg/mL DNase I (Sigma, #11284932001). Live, mononuclear cells were isolated by density gradient centrifugation using Ficoll-Paque (VWR, #17-1440-02). Magnetic bead CD34-enrichment was performed using MACS Miltenyi Kit 130-046-702 per the manufacturer's instructions. Enriched cells were stained with CD34-APC (BD Biosciences, #555824), CD38-PECy7 (Biolegend, #303516), CD123-PE (Biolegend, #306005), CD45RA-FITC (Biolegend, #304106), and propidium iodide. Common myeloid progenitors (CD34+/CD38+/CD123+/CD45RA-) were sorted on a FACS Aria II (BD Biosciences, San Diego, CA). RNA was extracted from sorted cells using Trizol reagent (ThermoFisher Scientific, #15596026) per the manufacturer's protocol. 3' region extraction and deep sequencing (3'READS) libraries were prepared as previously described[1] and sequenced on an Illumina HiSeq4000 (San Diego, CA). *RUNX1* poly(A) site usage was determined using the bigwig files generated via the standard 3'READS pipeline[1]. Using multiBigwigSummary from DeepTools2.0[2], a count matrix of read coverage at each base pair was produced. Each peak (a defined region of 200bp) was quantified using R software (R version 3.4.4) and calculated as a percentage of all four peaks.

Tandem poly(A) reporters and RNase protection assays

The pGL3 basic empty vector backbone (Promega, E1751) was used to clone the tandem poly(A) reporters for cleavage efficiency assays. First, the synthetic poly(A) site upstream of the luciferase region was removed using the QuikChange II XL Site Mutagenesis Kit (Agilent, #200517) and inserted downstream of luciferase by BamHI/SalI restriction digestion, adding an EcoRI restriction site. The pCMV promoter was inserted upstream of luciferase by XhoI/HindIII restriction enzyme digestion. *RUNX1* poly(A) sites #1 and #4 were individually subcloned between the coding region of luciferase and the synthetic poly(A) site by XbaI/EcoRI restriction digestion. RUNX1 poly(A) fragments were amplified from KG-1a genomic DNA using the following primer sets:

Poly(A) site #1: 5'–CCTGCAGAAATCACTTGATGCAC–3' / 5'–AATAAGGAGACACCGGGGGAA–3'

Poly(A) site #4: 5'–CTCTCCTGGGAGCATTCGTC–3' / 5'–GCCTTCAGAAGCCAGAGTGT–3'

Plasmids for in vitro transcription of antisense biotinylated probes were individually cloned for each *RUNX1* poly(A) site using primers 5'–GCCTGAATGGCGAATGGGA–3' / 5'–TCAGAGAGATCCTCATAAAGGC–3' to amplify a fragment from each of the tandem poly(A) reporter constructs. Amplified fragments were ligated into the pGEM t-easy vector digested with SacII/Sal1. These constructs were linearized by SalI restriction enzyme digestion, purified by gel extraction, and used as a substrate for T7 in vitro transcription using the MEGAscript T7 Transcription Kit (Ambion, AM1334) per the manufacturer's instructions. Reactions were supplemented with biotin-UTP. Probes were purified using the MEGAclear Transcription Clean-up Kit (Ambion, AM1908) following the manufacturer's instructions. Full length probes were further purified by gel extraction (5% polyacrylamide/urea gel).

For RNase protection assays, 293T cells were transfected with the respective tandem poly(A) reporter construct. 1 µg of the appropriate construct was mixed with 4 µL of polyethylenimine (PEI) in 100 µL of Opti-MEM I Reduced Serum Medium (Gibco, #31985-070). 24 hours following transfection, RNA was extracted using Trizol reagent (ThermoFisher Scientific, #15596026). 1 µg of 293T total RNA or 10 µg of control yeast RNA was mixed with excess T7 antisense biotinylated probe and RNase protection was performed as described using the RPA III Ribonuclease Protection Assay Kit (Ambion, AM1414). Protected fragments were run on a polyacrylamide/urea gel and transferred to Hybond-N+ nylon membrane (GE Healthcare, RPN1210B). RNA was crosslinked to the membrane using the Spectrolinker XL-1000 UV Crosslinker set to 120 mJ/cm^2 (Spectronics Corporation; New York, New York). The blot was probed with IRDye 680RD Streptavidin (Li-Cor, 926-68079) and visualized on a Li-Cor Odyssey Classic Infrared Imaging System (Lincoln, Nebraska). Densitometry was performed using the Li-Cor Application Software Version 3.0 and band intensities were normalized based on the number of biotinylated uracil nucleotides in each protected fragment.

Plasmids and additional minigene cloning

Chimeric minigene constructs were generated by amplifying individual intronic / exonic regions from the wild type exon 7a and 7b split GFP minigene constructs by PCR, joining fragments by overlap PCR, and inserting the full product into the split GFP vector by HindIII/KpnI digestion. For exon 7a chimeric constructs, the 3'

splice site was maintained so that splicing can occur normally with the first GFP exon. For exon 7b chimeric constructs, the 3' and 5' splice sites were both maintained so that splicing can occur normally with both flanking GFP exons.

The mCherry-P2A-split GFP *RUNX1* exon 7a minigene (dual fluorescent reporter) was generated by cloning mCherry and a P2A peptide directly upstream of GFP. mCherry was amplified from MSCV-IRES-mCherry, using primers to add the P2A peptide. Exon 1 of the split GFP minigene was also amplified and added to the mCherry-P2A fragment by overlap PCR such that translation of the dual fluorescent reporter was in-frame. The combined fragment was reintroduced into the split GFP minigene vector by NheI/SacI restriction enzyme digestion.

The HNRNPA1 binding site mutant construct was made with the QuikChange II XL Site Mutagenesis Kit (Agilent, #200517) per the manufacturer's protocol, using primers:

5'–CCCTCTCCCCAGCCAGGATCCAGCTATCTTTTCCA–3' /

5'–TGGAAAAGATAGCTGGATCCTGGCTGGGGAGAGGG–3'.

The pEGFP-C1 plasmid was obtained from Clontech (#6084-1).

Cell culture

HEK293T cells were purchased from ATCC and cultured in DMEM supplemented with 10% BCS and 100U/mL penicillin/streptomycin. HL-60, K562, KG-1a, THP-1, and U937 leukemia cell lines were all purchased from ATCC and maintained as low-passage stocks. MDS-L cells were kindly provided by Dr. Daniel Starczynowski (Cincinnati Children's Hospital, Ohio). ME-1 cells were obtained from Dr. Paul Liu (NIH, Maryland). KT-1 cells were generously provided by Dr. Ikuya Sakai (Ehime University School of Medicine, Japan). K562, THP-1, KT-1, HL-60, U937, and MDS-L cells were cultured in RPMI supplemented with 10% FBS and 100U/mL penicillin/streptomycin. MDS-L cells were also supplemented with 10 ng/mL recombinant human IL-3 (Peprotech, #200-03). KG-1a and ME-1 cells were cultured in RPMI supplemented with 20% FBS and 100U/mL penicillin/streptomycin.

Cryopreserved, >95% CD34+ enriched HSPCs from healthy donors were obtained from Fred Hutchinson Cooperative Center for Excellence in Hematology (CCEH) (Seattle, Washington). Cells were thawed and serially diluted in PBS with 2% FBS. They were then cultured in StemSpan SFEM (STEMCELL Technologies, #09600) supplemented with 100U/mL penicillin/streptomycin, 100 ng/mL hTPO, 100 ng/mL hSCF, 100 ng/mL hFLT3L, and 20 ng/mL hIL-6 (Peprotech, #300-07, #300-18, #300-19, #200-06).

All cells were cultured at 37°C and 5% CO_2.

Nucleofection of minigene constructs and siRNAs

For all minigene assays, DNA was nucleofected into KG-1a cells using program V-001 of the AMAXA II Nucleofector (Lonza; Basel, Switzerland). 3 µg of the indicated minigene DNA was mixed with 1 million KG-1a cells in 100 µL of nucleofection buffer (140 mM Na_2HPO_4/NaH_2PO_4 (pH 7.2), 5 mM KCl, 15 mM $MgCl_2$, 48 mM $NaCl_2$). Following nucleofection, cells were transferred to a 12-well plate containing pre-warmed RPMI supplemented with 20% FBS and no penicillin/streptomycin. Nucleofected cells were analyzed by flow cytometry, RT-PCR, or RT-qPCR 24 hours post-nucleofection.

For the siRNA knockdown secondary validation screen, MDS-L cells were nucleofected with 200 pmol of ON-TARGETplus Human SMARTpool siRNA (Dharmacon) targeting the indicated RBP. 1.5 million MDS-L cells were mixed with siRNA in 100 µL of nucleofection buffer (140 mM Na_2HPO_4/NaH_2PO_4 [pH 7.2], 5 mM KCl, 15 mM $MgCl_2$) and nucleofected using program V-001 of the AMAXA II Nucleofector. Cells were carefully transferred to a 12-well plate containing pre-warmed RPMI supplemented with 20% FBS and 10 ng/mL IL-3. Cells were collected 48 hours post-nucleofection for RNA extraction and RT-qPCR analysis. The ON-TARGETplus Human SMARTpool siRNAs purchased from Dharmacon are as follows: siControl (Non-targeting siRNA, D-001810-01-05), siGAPDH (D-001830-10-05), siDHX15 (L-011250-01), siHNRNPA1 (L-008221-00), siLUC7L3 (L-015383-00), siSF3A2 (L-018282-02), siRBM7 (L-017936-02), siDBR1 (L-008290-00), siHNRNPK (L-011692-00), siRBMX2 (L-020763-02), siDDX41 (L-010394-00), siHNRNPC (L-011869-03), siTHOC5 (L-015317-01), siKHDRBS1 (L-020019-00). Lyophilized siRNAs were resuspended in siRNA buffer (B-002000-UB-100) diluted to 1x with RNase-free water and stored in aliquots at -20°C.

Lentivirus and retrovirus production and transduction

For sgRNA CRISPR library lentiviral production, HEK293T cells were seeded in 15 cm tissue culture plates. One hour prior to transfection, media was replaced with 8 mL of pre-warmed Opti-MEM. Cells were then transfected with a mixture consisting of 62.5 µL Lipofectamine 2000, 125 µL Plus reagent, 12.5 µg lentiCRISPR plasmid library, 6.25 µg of pMD2.G, and 9.375 µg psPAX2. 6 hours following transfection, the media was changed to DMEM supplemented with 10% FBS. Supernatant was collected 48 hours later and filtered through a 0.45 µm low protein binding membrane. Virus was ultracentrifuged (24,000 rpm) for 2 hours at 4°C and resuspended in PBS overnight. Virus aliquots were stored in -80°C. For infection of the MDS-L dual reporter line, the volume of virus to achieve a multiplicity of infection ~0.3 was titrated prior to the screen. For the CRISPR screen, virus was added to 6-well dishes containing MDS-L reporter cells seeded at a density of 0.3-0.5 million cells per mL and supplemented with 4 µg/mL polybrene. Cells were transduced by centrifugation (2,000 x g) on two consecutive days for 3 hours at 32°C in an Allegra X-12R centrifuge (Beckman Coulter; Brea, CA). Cells were resuspended in puromycin 24 hours following the second transduction and cultured in the presence of puromycin for 3 days.

For shRNA knockdown experiments, MDS-L, K562, or primary CD34+ HSPCs were infected with lentivirus produced in HEK293T cells. HEK293T cells were transfected with 3 µg of pLKO-based lentiviral vector, 5 µg of psPAX2, 2.5 µg of pMD2.G, and 42 µL of polyethylenimine (PEI) in 1 mL of Opti-MEM I Reduced Serum Medium. 24 hours post-transfection, the cell supernatant was aspirated and replaced with fresh RPMI supplemented with 10% FBS (IMDM supplemented with 2% FBS for CD34+ HSPC experiments). 24 hours following the media replacement, the 293T cell supernatant containing lentiviral particles was collected and passed through a 0.45 µm syringe filter. Virus was added to the cells seeded at 0.3-0.5 million cells per mL and supplemented with 4 µg/mL polybrene. Cells were transduced on two consecutive days in 6-well tissue culture plates by centrifugation (2,000 x g) for 3 hours at 32°C in an Allegra X-12R centrifuge (Beckman Coulter; Brea, CA). 24 hours following the second transduction, cells were resuspended in fresh media with 1 µg/mL puromycin. After 48 hours of puromycin selection, cells were diluted 1:2 and maintained at 0.5 µg/mL puromycin until RNA or protein was collected on day 6 post-transduction. The HNRNPA1 pLKO shRNA clones utilized in this study were: TRCN0000235097 (shRNA #1) and TRCN0000235098 (shRNA #2). The KHDRBS1

pLKO shRNA clones utilized were: TRCN0000000044 (shRNA #1) and TRCN0000000048 (shRNA #2). The pLKO control shRNA was Addgene plasmid #1864.

For KHDRBS1 overexpression experiments, retrovirus was produced in HEK293T cells. 293T cells were transfected with 5 µg of MSCV-IRES-PURO (MIP) empty vector or MIP vector expressing KHDRBS1 cDNA, 5 µg of pCL-10A1, and 40 µL of PEI in 1 mL of Opti-MEM I. Virus was collected and K562 cells were transduced as described above. Puromycin selection of transduced cells was also performed as previously described. The KHDRBS1 cDNA was obtained from transOMIC Technologies (BC019109) and subcloned into the MIP vector.

For HNRNPA1 knockdown and KHDRBS1 overexpression dual treated K562 cells, the same protocol for viral production and transduction was followed as above. Instead of MSCV-IRES-PURO overexpression vectors, MSCV-IRES-NEO plasmids were used. Consequently, double transduced cells were selected with 1 µg/mL puromycin and 1 mg/mL G418. RNA and protein were collected 6 days post-transduction.

MDS-L dual fluorescent reporter cell line and HNRNPA1 CRISPR knockout cell line generation

The clonal MDS-L dual fluorescent reporter cell line used for the CRISPR screen was generated by nucleofection of MDS-L cells as described with linearized mCherry-P2A-spilt GFP RUNX1 exon 7a minigene plasmid DNA. Following two weeks of G418 selection, single mCherry+/GFP+ cells were sorted into 96-well plates using a BD FACSAria II (BD Biosciences, San Diego, CA).

For generating HNRNPA1 knockout MDS-L cells, the HNRNPA1 sgRNA that was most dramatically enriched in the GFP low population (5'–GCCGTCATGTCTAAGTCAG–3') was subcloned into the lentiCRISPR v2 vector (Addgene, #52961). MDS-L cells were then transduced as described with the Cas9/sgRNA vector followed by puromycin selection. On-target cleavage in the bulk cell population was confirmed by performing the T7 endonuclease assay on DNA amplified from transduced cells by a primer pair that spans the target site (5'–ACGACCGAAGGAATGACGTT–3' / 5'–TTACCACACAGTCCGTGAGC–3'). Single cells were sorted into 96-well plates using a FACSAria II. Clonal lines were screened for HNRNPA1 protein by western blotting and alleles of selected lines were TA-cloned into the pGEM t-easy vector (Promega, A1360) followed by Sanger sequencing.

Flow cytometry

GFP mean fluorescent intensity (MFI) was measured in minigene nucleofected KG-1a cells using a BD FACSCanto flow cytometer (San Diego, CA) and BD FACSDiva Acquisition software. Propidium iodide was used to exclude dead cells.

GFP and mCherry fluorescence were measured in dual fluorescent minigene nucleofected KG-1a cells using a BD LSRFortessa and BD FACSDiva Acquisition software. Sytox Blue Dead Cell Stain (ThermoFisher Scientific, #S34857) was used to exclude dead cells. Post-acquisition data analysis was performed using FlowJo software (FlowJo, LLC).

RT-PCR analysis and imaging

RNA was extracted from KG-1a cells nucleofected with the indicated minigene construct using Trizol reagent. cDNA was prepared from 0.5-1µg of RNA by reverse transcription using the First Strand cDNA Synthesis Kit (Molecular Cloning Laboratories, FSCS-200) following the manufacturer's instructions. PCR of the cDNA was performed with KOD Hot Start DNA polymerase (EMD Millipore, 71086-3) using the following primer sets:

GFP: 5'–AGTGCTTCAGCCGCTACCC–3' / 5'–GTTGTACTCCAGCTTGTGCC–3'

Exon 7a spliced and polyadenylated product: 5'–AGTGCTTCAGCCGCTACCC–3' /

5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATAGTG–3'

RT-PCR products were run on 2% agarose gels and imaged with a Biorad Gel Doc XR+ (Hercules, CA) using the Biorad Quantity One 1-D Analysis Software.

RT-qPCR

RNA isolation and cDNA preparation were performed as described above. Quantitative PCR (qPCR) reactions were performed in duplicate or triplicate on a BioRad CFX Connect (Hercules, CA) using KAPA SYBR FAST qPCR Master Mix (2X) Universal (KAPA Biosystems, #KK4618). Data were analyzed using the delta delta Ct method and normalized as described in the figure legends. The following primer pairs were used:

RUNX1 exon 7a minigene product: 5'–AGTGCTTCAGCCGCTACCC–3' / 5'–GAAAGTGTACCGGGATCCATG–3'

GAPDH: 5'–TCGCTCAGACACCATGGGGAAG–3' / 5'–GCCTTGACGGTGCCATGGAATTTG–3'

β-Actin: 5'–TCCCTGGAGAAGAGCTACGA–3' / 5'–AGCACTGTGTTGGCGTACAG–3'

RUNX1a: 5'–CTCCCTGAACCACTCCACTG–3' / 5'–TTAACATCTCCAGGGTGCTGTGTCTTC–3'

RUNX1b/c: 5'–CTCCCTGAACCACTCCACTG–3' / 5'–CAGAGAGGGTTGTCATGCCG–3'

RUNX1 (Total): 5'–CCTCAGGTTTGTCGGTCGAA–3' / 5'–CTGCCGATGTCTTCGAGGTT–3'

mCherry: 5'–CACGAGTTCGAGATCGAGGG–3' / 5'–CAAGTAGTCGGGGATGTCGG–3'

KHDRBS1: 5'–GAAATTTCTAGTACCGGATATGATG–3' / 5'–TGGTGTACCACGTACCAAAG–3'

Western blotting

Cells were lysed on ice in RIPA buffer (50 mM Tris, 1 M NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitors (Roche, # 11873580001). Lysates were cleared by centrifugation (10,000 x g), mixed with loading buffer, and denatured at 95°C. Conventional SDS-PAGE and western blotting techniques were used to analyze the lysates. The following primary antibodies were used in this study, following the manufacturer-recommended dilutions: HNRNPA1 (Santa Cruz, sc-32301), KHDRBS1 (Bethyl Laboratories, A302-110A), RUNX1 (generated by Covance)[3], β-Actin (Sigma, A1978), and α-Tubulin (Developmental Studies Hybridoma Bank, #12G10). The following secondary western antibodies were used in this study: IRDye 800CW goat anti-rabbit IgG (LI-COR, #926-32211), IRDye 800CW goat anti-mouse IgG (LI-COR, # 926-32210), and IRDye 680RD goat anti-mouse IgG (LI-COR, #926-68070). Membranes were scanned on the LI-COR Odyssey Classic Infrared Imaging System and densitometry was performed using the LI-COR Application Software Version 3.0.

Expression of *HNRNPA1* and *KHDRBS1* throughout hematopoiesis and in AML patients

HNRNPA1 and *KHDRBS1* mRNA expression data in various hematopoietic populations and in AML patients compared to healthy bone marrow were obtained from the BloodSpot server (bloodspot.eu)[4]. For

hematopoietic subpopulations, the HemaExplorer dataset (GSE17054, GSE19599, GSE11864, E-MEXP-1242) is shown for each RBP with the following Affymetrix probe IDs: HNRNPA1 (214280_x_at) and KHDRBS1 (200040_at). For AML patients and healthy bone marrow, the Leukemia MILE dataset (GSE13159) is shown for each RBP with the same probe IDs listed above.

Statistical analyses

All statistical analyses were performed using the GraphPad Prism Software (Version 8.3.1). All statistical tests are described in the respective figure legends. P-values are consistently annotated throughout: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Data availability

3'RNA sequencing data from healthy common myeloid progenitors (CMP) and CRISPR RBP screen results were deposited to the Gene Expression Omnibus (GEO), accession GSE145968. All DNA constructs cloned for this study will either be deposited to Addgene for purchase, or available by email request to the corresponding author.

SUPPLEMENTARY REFERENCES

1. Hoque M, Ji Z, Zheng D, et al. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods*. 2013;10(2):133-139.
2. Ramirez F, Ryan DP, Gruning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-165.
3. Yan M, Kanbe E, Peterson LF, et al. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med*. 2006;12(8):945-949.
4. Bagger FO, Kinalis S, Rapin N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res*. 2019;47(D1):D881-d885.
5. Bagger FO, Rapin N, Theilgaard-Mönch K, et al. HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. *Nucleic Acids Res*. 2013;41(Database issue):D1034-1039.
6. Haferlach T, Kohlmann A, Wieczorek L, et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *J Clin Oncol*. 2010;28(15):2529-2537.

SUPPLEMENTARY TABLE 1. Composition of significantly enriched RBPs in GFP low and high populations

The most significantly enriched ($p < 0.01$) RBPs in the GFP low (left) and GFP high (right) populations are listed. EGFP and mCherry targeting sgRNAs, the positive control for each sorted population, were most significantly enriched. 'RNA splicing' (GO: 0008380) / 'Regulation of mRNA splicing' (GO: 0043484) and 'mRNA 3'end processing' (GO:0031124) annotations are indicated for each RBP.

Putative RUNX1a Repressors (GFP Low)				Putative RUNX1a Activators (GFP High)			
RBP	**P-value**	**RNA splicing GO:0008380; Regulation of mRNA splicing GO:0043484**	**mRNA 3'-end processing GO:0031124**	**RBP**	**P-value**	**RNA splicing GO:0008380; Regulation of mRNA splicing GO:0043484**	**mRNA 3'-end processing GO:0031124**
EGFP	0	N/A	N/A	mCherry	0	N/A	N/A
DHX15	0.00163	Yes	No	DDX41	0	Yes	No
HNRNPA1	0.002445	Yes	No	U2SURP	0	Yes	No
LUC7L3	0.002445	Yes	No	NARS	0	No	No
SF3A2	0.002445	Yes	No	GARS	0.000815	No	No
RBM7	0.002445	Yes	No	HNRNPC	0.000815	Yes	No
ADAT2	0.00326	No	No	MTPAP	0.000815	No	Yes
DBR1	0.004075	Yes	No	CARS	0.000815	No	No
RRP15	0.00489	No	No	RNF113A	0.00163	Yes	No
HNRNPK	0.00489	Yes	No	THOC5	0.00163	Yes	Yes
RBMX2	0.00489	Yes	No	KHDRBS1	0.002445	Yes	No
LSM12	0.00489	No	No	DYNC1H1	0.002445	No	No
RPL7	0.005705	No	No	EIF2S3	0.00326	No	No
DDX21	0.005705	No	No	MRPL14	0.00489	No	No
TIAL1	0.005705	No	No	IPO7	0.00652	No	No
RPL30	0.00652	No	No	RPS15	0.007335	No	No
MARS2	0.00652	No	No	DDX49	0.00815	No	No
RPL26	0.00652	No	No				
XRN2	0.00815	No	No				
RPL19	0.00815	No	No				
RPL23	0.00815	No	No				
PSMD1	0.00978	No	No				

SUPPLEMENTARY FIGURE 1. *RUNX1* endogenous poly(A) site usage in common myeloid progenitors (CMPs) via 3'READS

A. Sample fluorescence activated cell sorting (FACS) plots depicting how CMPs were identified and sorted following CD34 bead enrichment of leukapheresis products obtained from healthy donors.
B. Genome browser tracks depicting sequencing reads in the full *RUNX1* gene obtained from 3'READS of sorted common myeloid progenitors (CMPs). The boxed region is expanded in Figure 1A.

SUPPLEMENTARY FIGURE 2. Sequence comparison of major *RUNX1* poly(A) sites

A. mRNA sequences containing the four major *RUNX1* poly(A) sites. The location of each site within the *RUNX1* genomic structure is marked with a red arrow (top). The three major cis-acting elements for polyadenylation are color-coded and underlined. A vertical line indicates the predicted cleavage site.
B. Table summarizing the sequence features of the four major *RUNX1* poly(A) sites. A checkmark indicates presence, whereas an 'X' represents absence of the indicated cis element.

SUPPLEMENTARY FIGURE 3. *RUNX1* poly(A) sites differ in cleavage efficiency

A. Schematic of the RNase protection assay. The *RUNX1* poly(A) site sequences cloned between luciferase and the SV40 poly(A) site are shown in Supplementary Figure 2A.
B. Representative blots showing the protected fragments generated from *RUNX1* poly(A) site #1 and #4 relative to a common synthetic poly(A) site using RNase protection assays (RPA). For each blot, lane 1 is the experimental lane. Lane 2 is a positive control for effective RNase digestion and a negative control for non-specific binding of the probe. Lane 3 is a control of full-length probe integrity. Yeast RNA is mixed with probe for control lanes (2-3). *RUNX1* poly(A) site and synthetic poly(A) site signal intensities were normalized to each other based on the number of uracil nucleotides present in each protected fragment. Example quantifications are listed below the experimental lanes.
C. Table showing sample calculations of cleavage at either *RUNX1* poly(A) site #1 or #4 relative to a common synthetic poly(A) site via RPA. Band intensities were normalized to each other by the biotinylated uracil content of each protected fragment.

SUPPLEMENTARY FIGURE 4. Total *RUNX1* and relative *RUNX1* isoform expression in various leukemia cell lines

A. RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in the indicated leukemia cell lines. mRNA levels were normalized to K562 cells, which had the lowest relative *RUNX1a* transcript. Blue bars represent '*RUNX1a* low,' yellow bars represent '*RUNX1a* intermediate,' and red bars represent '*RUNX1a* high' cell lines. Data are mean +/- s.d. of three independent batches of RNA.
B. RT-qPCR analysis of *RUNX1b/c* mRNA normalized to total *RUNX1* mRNA in the indicated leukemia cell lines. mRNA levels were normalized to K562 cells. Data are mean +/- s.d. of three independent batches of RNA.
C. RT-qPCR analysis of *RUNX1* mRNA normalized to total *GAPDH* mRNA in the indicated leukemia cell lines. mRNA levels were normalized to K562 cells. Data are mean +/- s.d. of three independent batches of RNA.

SUPPLEMENTARY FIGURE 5. CRISPR screen quality control metrics

A. Dot plot showing the correlation between normalized sgRNA read counts in the plasmid library used for lentiviral production and counts in transduced cells on 'day 0.'
B. Bar graphs showing the number of significantly enriched sgRNAs per RBP in the sorted GFP low and GFP high populations relative to the bulk population of cells on day 21. For each RBP, 10 unique sgRNAs were present in the screen. There were fewer sgRNAs targeting the fluorescent protein controls: 3 targeting EGFP and 2 targeting mCherry.

SUPPLEMENTARY FIGURE 6. Secondary siRNA screen reveals regulators of *RUNX1* exon 7a usage

A. RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in MDS-L cells nucleofected with the indicated siRNAs. mRNA levels were normalized to siControl nucleofected cells. The teal bar marks the repressor RBP that significantly increased *RUNX1a* mRNA upon knockdown. Data are mean +/- s.d. of three biological replicates. * $p < 0.05$, *** $p < 0.001$, one-way ANOVA with post-hoc Tukey test.

B. RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* in MDS-L cells nucleofected with the indicated siRNAs. mRNA levels were normalized to siControl nucleofected cells. The teal bars mark activator RBPs that significantly decreased *RUNX1a* mRNA upon knockdown. Data are mean +/- s.d. of three biological replicates. * $p < 0.05$, one-way ANOVA with post-hoc Tukey test. Though U2SURP and RNF113A are involved in splicing (**Supplementary Table 1**), these two RBPs were excluded from the secondary screen because they were also enriched in the GFP low cell population ($0.01 < p < 0.1$).

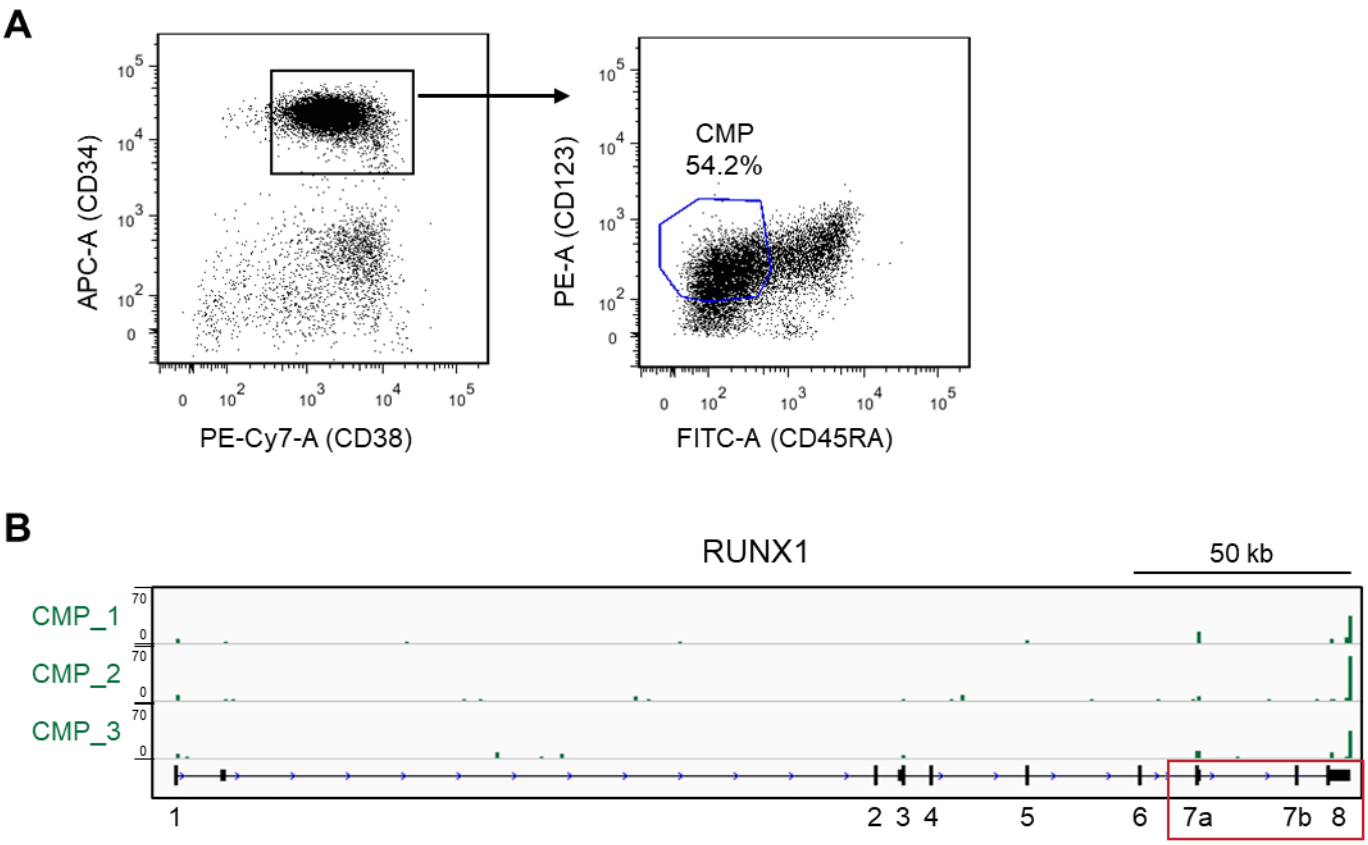
SUPPLEMENTARY FIGURE 7. CRISPR knockout of HNRNPA1 increases endogenous RUNX1a levels

A. (Left) Western blot showing HNRNPA1 and tubulin (loading control) protein in the parental MDS-L cell line, a clonal MDS-L cell line containing empty lenti-v2 Cas9 vector, and individual HNRNPA1 sgRNA-targeted MDS-L clonal cell lines. (Right) Protein was quantified by normalizing HNRNPA1 signal intensity to tubulin signal intensity using LI-COR Image Studio software. Clonal cell line HNRNPA1 protein levels were normalized to the parental MDS-L cell line. Data are mean +/- s.d. of three protein lysates from each clonal line.
B. Table showing the sequence of each HNRNPA1 allele in the clonal MDS-L cell lines. Mutations are shown in red. Blue nucleotides are in exon 1. Gray nucleotides are in the intron downstream of exon 1.
C. RT-qPCR analysis of *RUNX1a* mRNA normalized to total *RUNX1* mRNA in clonal MDS-L cell lines with normal HNRNPA1 protein levels (WT: n = 4), approximately 50% HNRNPA1 protein (Het: n = 2), and undetectable HNRNPA1 protein (KO: n = 4). Data are mean +/- s.d. ** $p < 0.01$, two-tailed student's t-tests.
D. Western blot showing RUNX1a and tubulin (loading control) protein. The first two lanes contain protein lysates from 293T cells that have been transfected with vector containing untagged RUNX1a cDNA or empty vector. Lane one is a positive control for RUNX1a protein (marked with the blue arrow). The remaining lanes contain protein from the clonal MDS-L cell lines. RUNX1a protein was quantified by normalizing RUNX1a signal intensity to tubulin signal intensity using LI-COR Image Studio software. Normalized protein quantifications are shown below the respective lane.

SUPPLEMENTARY FIGURE 8. *HNRNPA1* and *KHDRBS1* expression during normal hematopoiesis and in AML patients

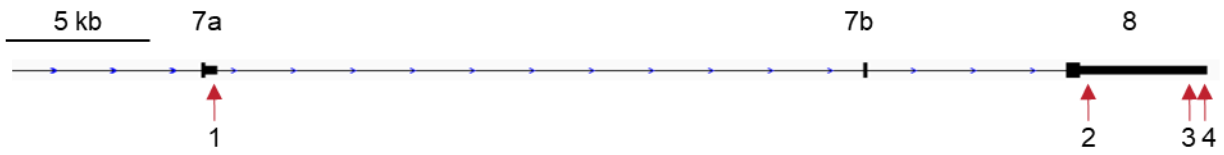
A. *HNRNPA1* mRNA expression in healthy hematopoietic populations. Data are from the HemaExplorer cohort (GSE17054, GSE19599, GSE11864, E-MEXP-1242)[5] obtained from the Bloodspot database[4]. Affymetrix probe 214280_x_at. The blue color labels hematopoietic stem and progenitor populations. Teal indicates myeloid populations. Yellow are lymphoid populations. HSC, hematopoietic stem cells; HPC, hematopoietic progenitor cells; CMP, common myeloid progenitors; GMP, granulocyte monocyte progenitors; MEP, megakaryocyte-erythroid progenitors; PM, promyelocyte; MY, myelocyte; PMN polymorphonuclear cells; NK, natural killer cells; DC, dendritic cells; BM, bone marrow; PB, peripheral blood.
B. *HNRNPA1* mRNA expression in healthy bone marrow and AML patient samples. Data are from the Leukemia MILE study (GSE13159)[6] obtained from the Bloodspot database[4]. Affymetrix probe 214280_x_at. Two-tailed student's t-test; *** $p < 0.001$.
C. *KHDRBS1* mRNA expression in the same dataset as in (A). Affymetrix probe 200040_at.
D. *KHDRBS1* mRNA expression in the same dataset as in (B). Affymetrix probe 20040_at. Two-tailed student's t-test; *** $p < 0.001$.

Supplementary Figure 1. *RUNX1* endogenous poly(A) site usage in common myeloid progenitors (CMPs) via 3'READS



Supplementary Figure 2. Sequence comparison of major *RUNX1* poly(A) sites

A



Poly(A) Site 1

CCUGCAGAAAUCACUUGAUGCACAGCACAUGGGAAGCAUUGUGUGUAUUUAUUAAUCCUUCACAAAGUCUUUGAGAUA
UAUUUUUAUCAAAUAUUUAGCAUGGAUCCCGGUACACUUUCAAUACUUAAUAAAUGGUCAAUGUUAUUCUUUUUCACU
AUU|AUUGCUAUUUAUGUGCUCUUAUUGGCGGCAUUGUUGUAUCACCAGUUAGUUGUACCCAUCGGUCCUCUCCAAGU
CACACAGCUACUCUGCGAUAAUUCCCCCGGUGUCUCCUUAUU

Poly(A) Site 2

GAGGUGUCCGAGGCGACGCACCUCGAGGGUGUCCGCCGGCCCCAGCACCCAGGGGACGCGCUGGAAAGCAAACAGG
AAGAUUCCCGGAGGGAAACUGUGAAUGCUUCUGAUUUAGCAAUGCUGUGAAUAAAAAGAAAGAUUUUAUACCCUUGA|
CUUAACUUUUUAACCAAGUUGUUUAUUCCAAAGAGUGUGGAAUUUUGGUUGGGGUGGGGGGAGAGGAGGGAUGCAAC
UCGCCCUGUUUGGCAUCUAAUUCUUAUUUUUAAUUUUUCCGCACCUUAUCAAUUGCAAAAUGCGUAUUUGCAUUUGGG
UGG

Poly(A) Site 3

ACAGCUAAUAGCAUGGUUCCAAUUUUUUUUAAGUUCACUUUUUUUGUUCUAGGGGAAAUGAAUGUGCAAAAAAAGAAA
AAGAACUGUUGGUUAUUUGUGUUAUUCUGGAUGUAUAAAAAUCAAUGGAAAAAAAUAAACUUUCAAAUU|GAAAUGACG
GUAUAACACAUCUACUGAAAAAGCAACGGGAAAUGUGGUCCUAUUUAAGCCAGCCCCCACCUAGGGUCUAUUUGUGUG
GCAGUUAUUGGGUUUGGUCACAAAACAUCCUGAAAAUUCGUGCGUGGGCUUC

Poly(A) Site 4

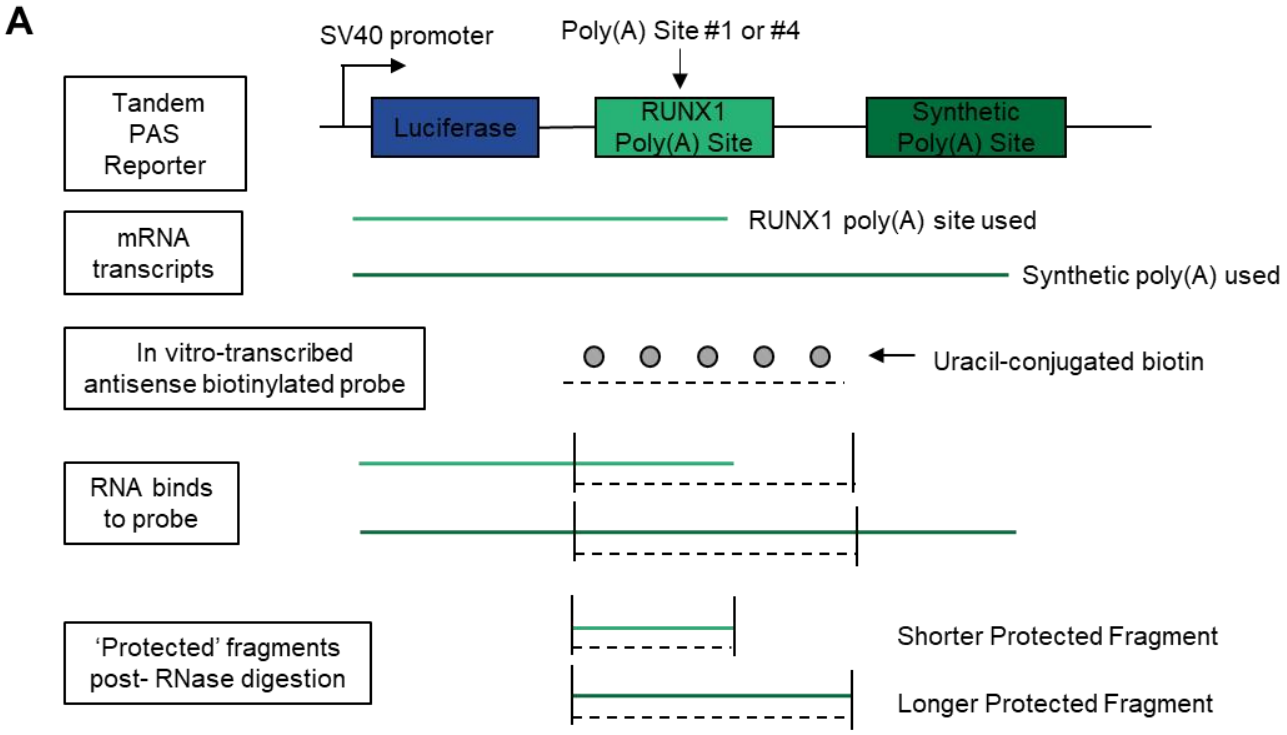
CUCUCCUGGGAGCAUUCGUCGUGCCCAGCCUGAGCAGGGCAGCUGGACUGCUGCUGUUCAGGAGCCACCAGAGCCU
UCCUCUCUUUGUACCACAGUUUCUUCUGUAAAUCCAGUGUUACAAUCAGUGUGAAUGGCAAAUAAACAGUUUGACAA
GUACAUACACCAUA|UCCAAUUGGUUUUGUCUUUCUCUGAGAUGGAGGGUUGGGAAACUGAGUAAGGCCCACUGCACA
CACUAGGAAGAUACCUCCAGUGAACAGCUUCAGGCUGAAGCUAUGUCCUCACCCCCGGUGCUCCUAGCACACUCUGG
CUUCUGAAGGC

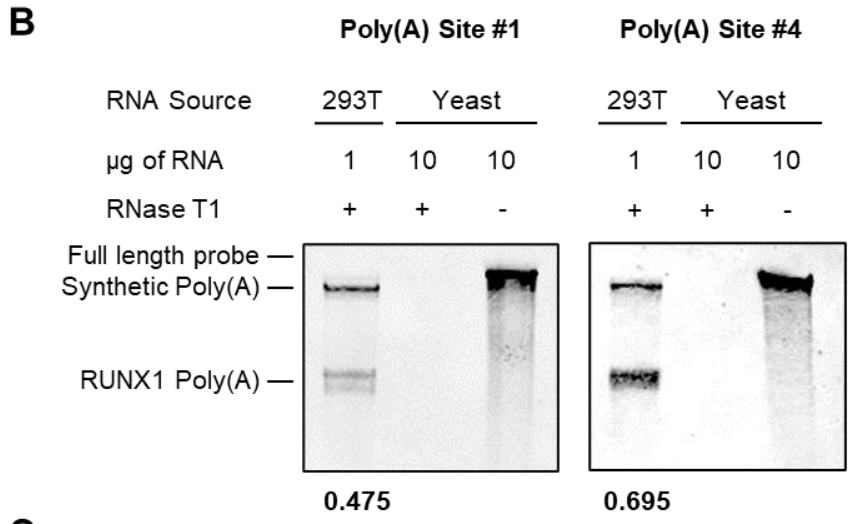
Poly(A) Signal (PAS) Hexamer Downstream G/U-rich elements Upstream UGUA Motif

B

RUNX1 Poly(A) Site ID	Canonical PAS Hexamer (AAUAAA)	Downstream U-rich and G/U-rich Elements (UU dinucleotide)	Upstream UGUA motif
1	✓	✓	✓
2	✓	✓	X
3	✓	X	✓
4	✓	✓	✓✓

Supplementary Figure 3. *RUNX1* poly(A) sites differ in cleavage efficiency



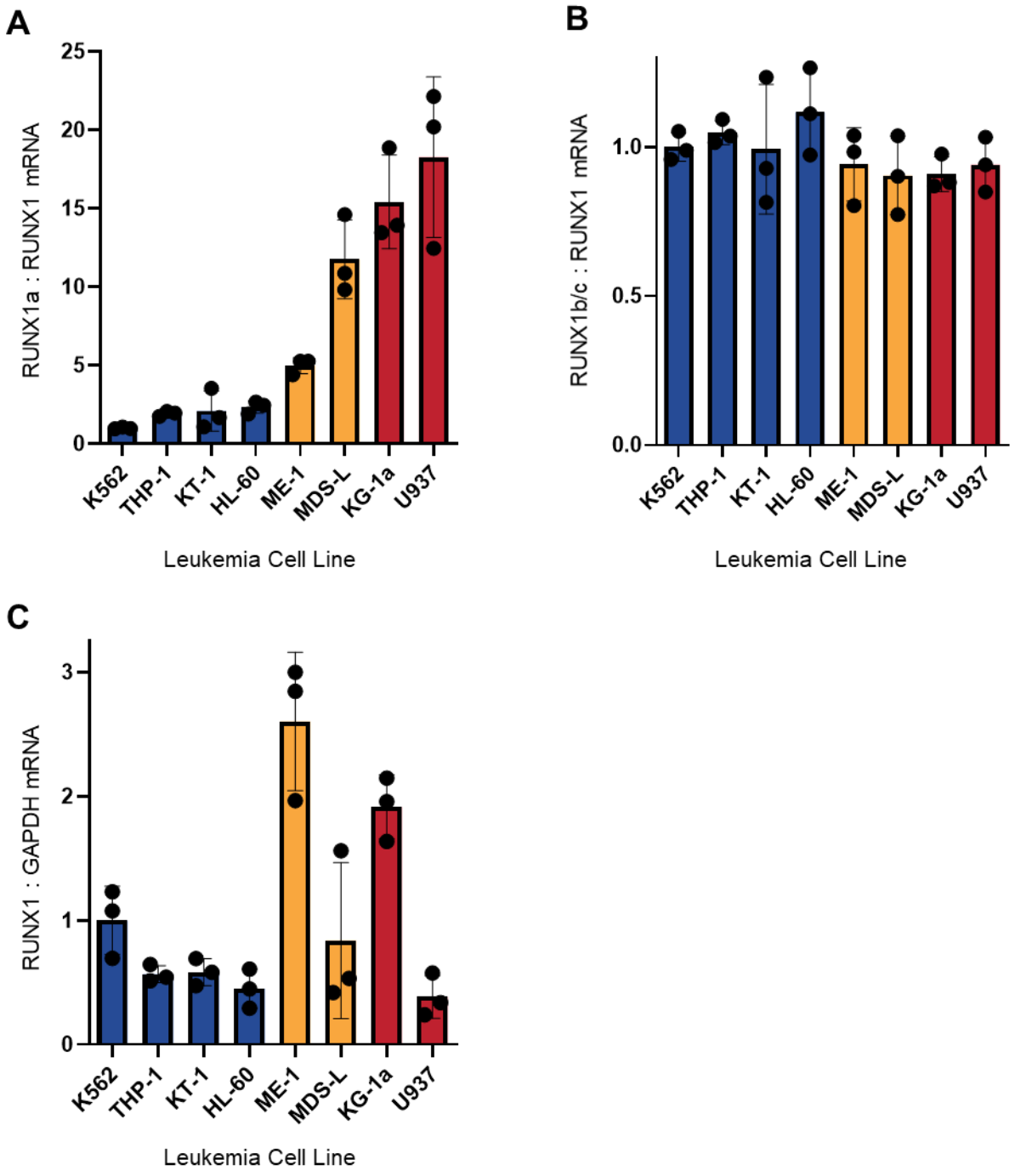


C

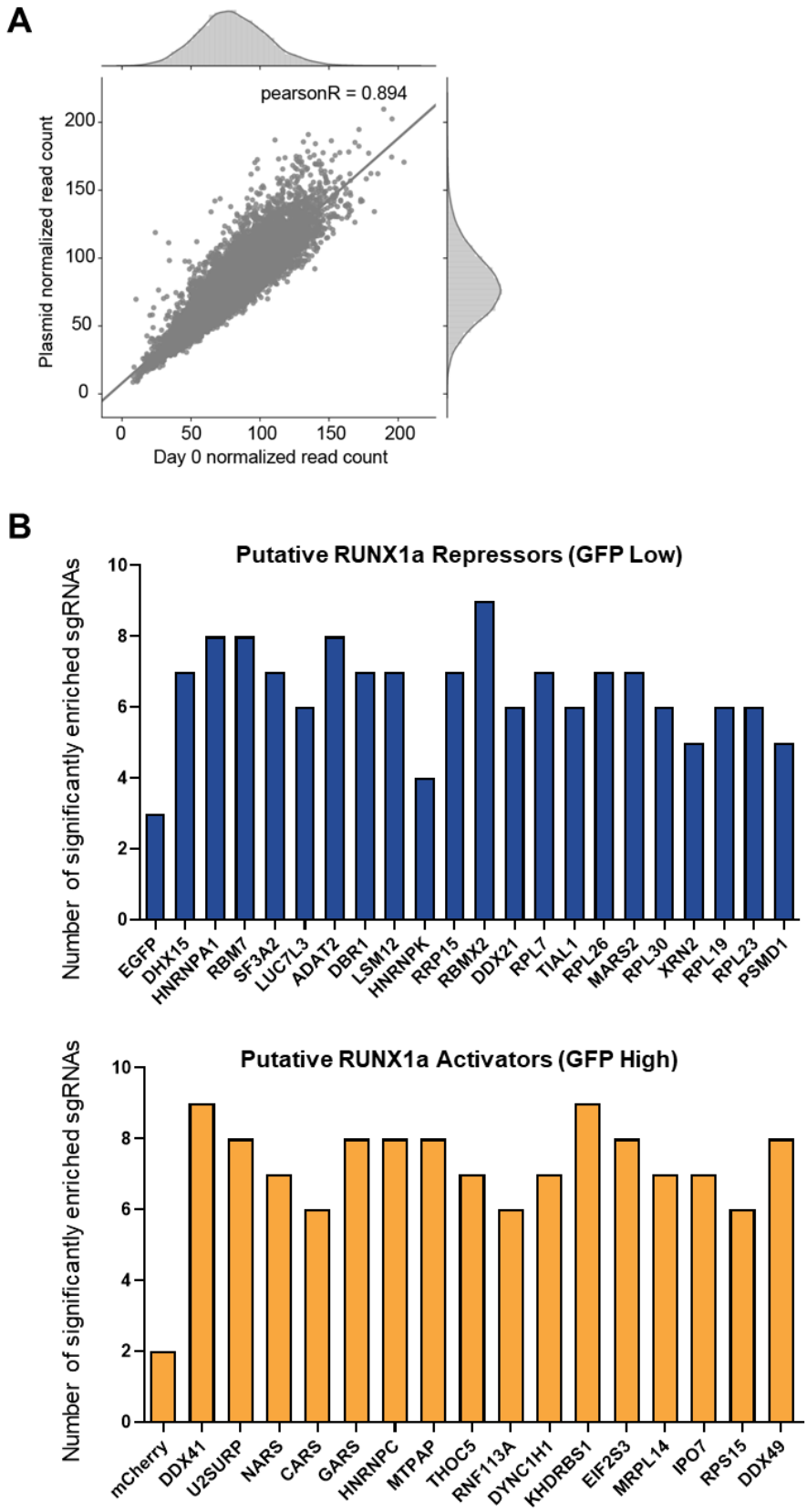
	Synthetic Poly(A)			RUNX1 Poly(A)			
RUNX1 Poly(A) Site	Band Intensity	# Uracil Residues in Probe	Normalized Intensity	Band Intensity	# Uracil Residues in Probe	Normalized Intensity	Cleavage at the RUNX1 Poly(A) Site
#1	4.56	94	0.0485	2.98	68	0.0438	0.4746
#4	4.81	100	0.0481	6.69	61	0.1097	0.6951

Cleavage at the RUNX1 Poly(A) Site = (RUNX1 Normalized Intensity) / (RUNX1 Normalized Intensity + Synthetic Poly(A) Normalized Intensity)

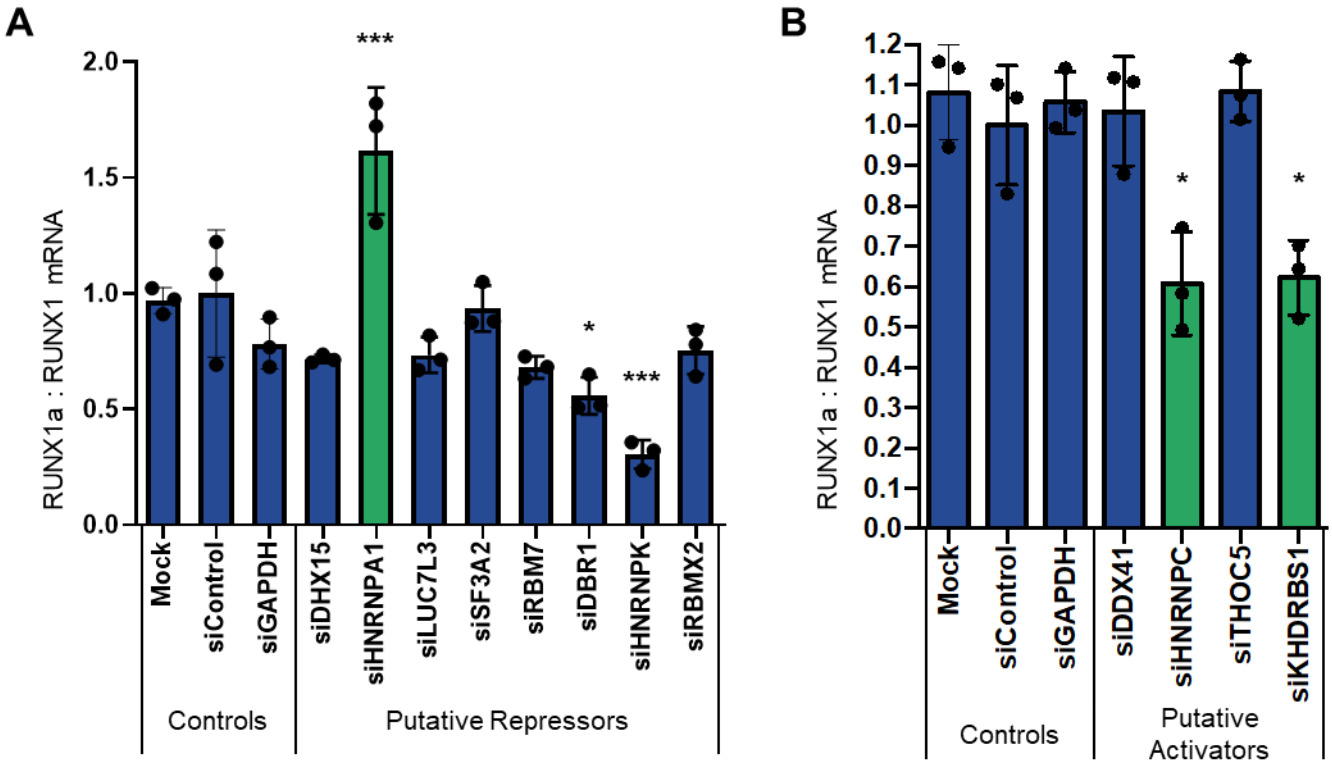
Supplementary Figure 4. Total *RUNX1* and relative *RUNX1* isoform expression in various leukemia cell lines



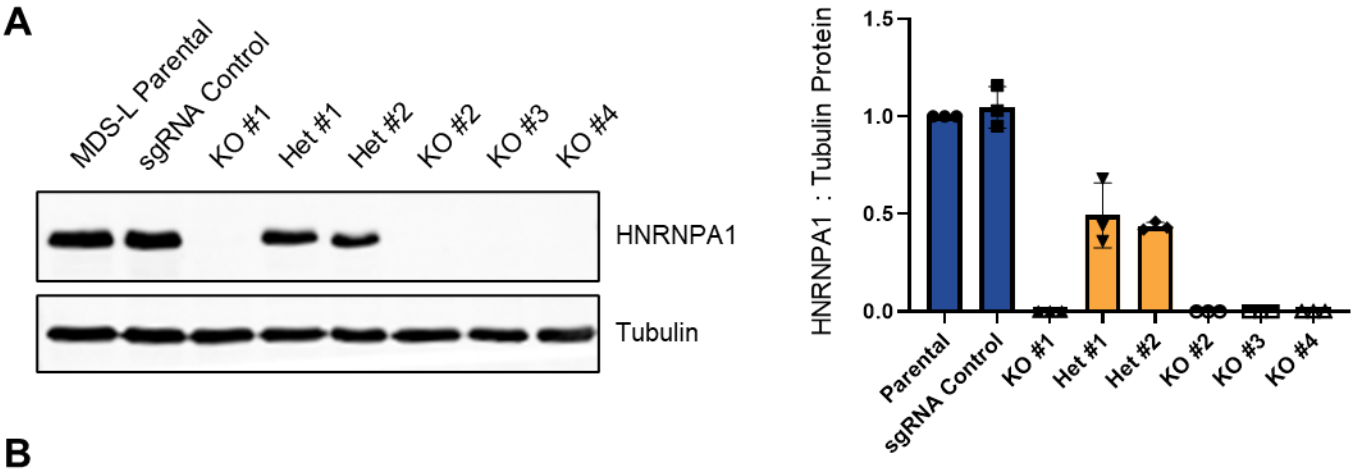
Supplementary Figure 5. CRISPR screen quality control metrics



Supplementary Figure 6. Secondary siRNA screen reveals regulators of *RUNX1* exon 7a usage

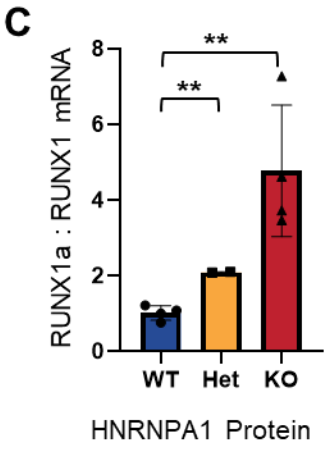


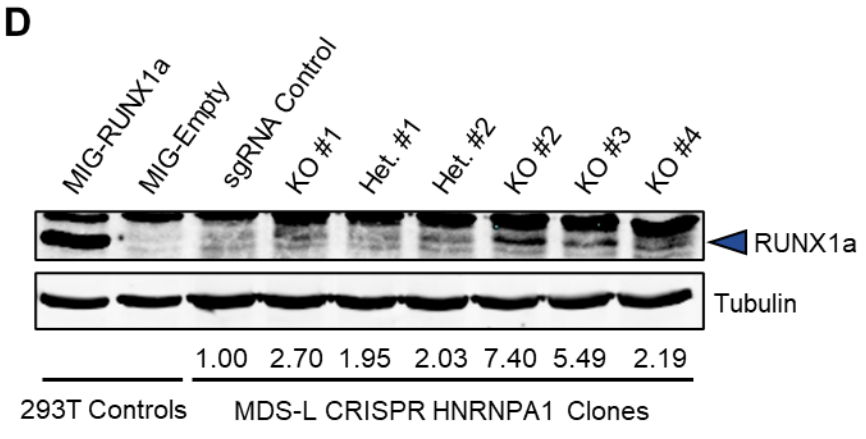
Supplementary Figure 7. CRISPR knockout of HNRNPA1 increases endogenous RUNX1a levels



B

Clone ID	Individual Allele Sequences	Mutation Analysis	Protein Level
sgRNA Control	...TGCCGTCATGTCTAAGTCAGAGGTGAGTTAGGCGCG... ...TGCCGTCATGTCTAAGTCAGAGGTGAGTTAGGCGCG...	No mutations	100%
KO #1	...TGCCGTCATGTCTAAGTGAGTTAGGCGCGTCAGAGGTGA... ...TGCCGTCATGTCTAAGTTCAGAGGTGAGTTAGGCGCG...	13 bp insertion 1 bp insertion	Undetectable
Het #1	...TGCCGTCATGTCTA---CAGAGGTGAGTTAGGCGCG... ...TGCCGTCATGT----------GAGTTAGGCGCG...	3 bp deletion 13 bp deletion	~50%
Het #2	...TGCCGTCATGTC---------GGTGAGTTAGGCGCG... ...----------GGTGAGTTAGGCGCG...	9 bp deletion 344 bp deletion	~50%
KO #2	...TGCCGTCATG-------CAGAGGTGAGTTAGGCGCG... ...TGCCGTCATGT----------GAGTTAGGCGCG...	7 bp deletion 13 bp deletion	Undetectable
KO #3	...TGCCGTCATGTCTAAG----------CG... ...TGCCGTCATGTCTAAGT-AGAGGTGAGTTAGGCGCG...	18 bp deletion 1 bp deletion	Undetectable
KO #4	...TGCCGTCATGTCTAAGT----------TAGGCGCG... ...TGC----------AGAGGTGAGTTAGGCGCG...	11 bp deletion 15 bp deletion	Undetectable





Supplementary Figure 8. *HNRNPA1* and *KHDRBS1* expression during normal hematopoiesis and in AML patients