SUPPLEMENTAL METHODOLOGY

Lentiviral generation

Lentiviral particles were generated following our optimized protocol.[1] In brief, HEK293T cells were plated overnight to reach 80-85% confluency on the next day. Cells were then co-transfected with the viral expression vector plus packaging plasmids (pMD2.G and psPAX2, Addgene) using Lipofectamine 2000 (Life Technologies). At 48 h and 72 h thereafter, culture supernatants were collected and filtered through a 0.45-mm PVDF filter (Millipore). Viruses were further concentrated using PEG-it® Virus Precipitation Solution (System Biosciences).

Plasmid generation

LOUP cDNA in pCMV-SPORT6 plasmid (Dharmacon) was sub-cloned into the lentiviral pCDH-MSCV-MCS-EF1-copGFP expression vector that carries the copGFP marker (System Biosciences). Short hairpin RNAs (shRNA) targeting Renilla (shControl) and RUNX1-ETO (shRUNX1-ETO) were cloned into lentiviral vector containing GFP in an optimized 'miRE' context.[2] shRNA sequences are provided in Table S3.

Generation of *LOUP*-depleted U937 cells (CRISPR/Cas9)

In order to deplete *LOUP*, we employed CRISPR/Cas9 genome-editing platform which introduces small insertion and deletion (indel) mutations in the *LOUP* gene via the non-homologous end-joining (NHEJ) DNA repair mechanism.[3,4] FUCas9Cherry[5] (Addgene) was used as expression vector to generate mCherry-Cas9 lentiviral particles as described above. U937 cells were transduced with these particles using the TransDux® reagent (System Biosciences). Cas9-stable cells were then selected by several rounds of Fluorescence-activated cell sorting (FACS) sorting for mCherry positivity. *LOUP*-targeting sgRNAs were designed using Cas-Designer[6] and cloned into the pLVx U6se EF1a sfPac vector which carries eGFP (kind gift from Dr. Iannis Aifantis). To avoid disruption of the URE, known to

be critical for *PU.1* induction,[7] we designed single-guide RNAs (sgRNA) targeting two distinct regions of the *LOUP* gene: (1) the *LOUP* intronic area downstream of the URE, and (2) the intronic area immediately upstream of the second exon of the *LOUP* gene (~ 15 kb downstream from the URE). Cas9-stable cells were then transduced with eGFP-sgRNA lentiviruses. Cells expressing high levels of both eGFP and mCherry were FACS sorted, one cell per well, into 96-well plates. Genomic DNA from cell clones were isolated using DNeasy Blood & Tissue Kit kit (QIAGEN) and used for PCR amplifying CRISPR/Cas9 target sites. PCR products were sequenced and indel profile were analyzed by Inference of CRISPR edits (ICE) software.[8] Cell clones having homozygous indels were verified by Sanger sequencing. Primer and sgRNA sequences are provided in Table S3.

Cord blood CD34$^+$ cell transduction and myeloid differentiation culture

Isolation and lentiviral transduction of human cord blood CD34$^+$ cells were performed following described protocols.[1,9] Briefly, purified CD34$^+$ cells were cultured in expansion culture (IMDM supplemented with 20% BIT 9500 (Stem Cell Technologies, Vancouver), 100 ng/ml FLT3L, 100 ng/ml SCF, 100 ng/ml TPO, 20 ng/ml IL-6 - all from Peprotech, Cranbury)) and transduced with lentiviral particles using the TransDux® reagent (System Biosciences). Transduced cells were cultured in myeloid differentiation culture (IMDM with 20% BIT 9500 (Stem Cell Technologies, Vancouver), 100 ng/ml SCF, 10 ng/ml FLT3L, 20 ng/ml IL-6, 20 ng/ml IL-3, 20 ng/ml GM-CSF, and 20 ng/ml G-CSF)) and selected with 2 ug/ml puromycin. Cytospin slides of cultured cells were stained with Camco Stain Pak (Cambridge Diagnostic). Images were acquired with a Nikon Eclipse microscope, 60/0.80 magnification and the SPOT Insight2 camera.

Generation of CRISPR activation cells (CRISPRa)

sgRNAs targeting the region 500 bp upstream of the *LOUP* transcriptional start site were designed using Cas-Designer.[6] The sgRNAs were then cloned into the pXR502 plasmid as previously described.[10] K562 cells stably expressing dCas9-VP64 were generated via lentiviral delivery of dCas9-VP64-Blast[11] and Blasticidin selection. dCas9-VP64 stable cells were transduced with lentiviruses that

packaged the sgRNA-cloned pXR502 plasmids as previously described.[10] One-day post-transduction, cells were selected with puromycin for 2-3 days before collection for analysis. sgRNA sequences are provided in Table S3.

Plasmid transfections

K562 cells, in exponential growth, were electroporated with expression plasmids using program T16, kit V (Lonza). Electroporated cells were incubated at 37°C overnight in a 5% CO2 incubator. The next day, cells were changed to fresh medium. Cells were harvested at 48 h after electroporation.

RUNX1-ETO inducible Tet-Off U937 cell culture

U937 cells with conditional RUNX1-ETO expression was previously established.[12,13] Cells were stably transfected with the tetracycline transactivator (tTA) under the control of a tetracycline responsive element and pUHD RUNX1-ETO (also called pUHD-AML1/ETO or pUHD-CBF2T1). Cells were maintained in RPMI 1640 medium supplemented with 10% FBS, 1mM L-glutamine, 0.5 µg puromycin, 1 ug/ml G418, and 1 µg/ml of tetracycline. To induce RUNX1-ETO expression, cells were washed in with RPMI 1640 blank medium to remove tetracycline before being cultured for 48 h in the above-mentioned medium without tetracycline. Expression of RUNX1-ETO was confirmed by western blot using AML1 antibody (#4334, Cell Signaling Technology) that is capable of detecting both RUNX1 and RUNX1-ETO.

Cellular fractionation, RNA extraction, RT-PCR and qPCR analysis

Cultured cells were washed with phosphate-buffered saline (PBS). Total RNA was extracted with Trizol reagent (Invitrogen) or PureLink™ RNA Mini Kit (Ambion) and treated with RNase-free DNase I (Roche) to remove contaminated genomic DNA. polyA$^-$ and polyA$^+$ RNAs were isolated from total RNA

using the Poly(A)Purist™ MAG Kit (Ambion) following manufacturer's protocol. Isolation of RNA from subcellular fractions was performed as previously described[14] with modifications. In brief, cells were lysed in cytosolic lysis solution (10 mM HEPES pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 % NP40, 1 mM DTT plus protease and RNase inhibitors) for 10 minutes on ice. After centrifugation, the supernatant was collected as the cytoplasmic fraction for cytosolic RNA isolation. After washing in cytosolic lysis solution, the nuclear pellet was used for nuclear RNA isolation. To collect nucleoplasm and chromatin fractions, the nuclear pellet was further lysed with nuclear lysis solution (20 mM HEPES pH 7.9, 1.5 mM MgCl2, 450 nM NaCl, 0.2 mM EDTA, 25% glycerol, 1 mM DTT, plus protease and RNase inhibitors). After centrifugation, the nuclear-soluble fraction (nucleoplasm) was collected as supernatant and the chromatin-associated fraction was collected as the pellet. RNAs from collected fractions were extracted with Trizol reagent and treated with RNase-free DNase I (Roche).

For RT-PCR, RNA was reverse-transcribed by using the SuperScript® III Reverse Transcriptase (Invitrogen). Red Taq Pro Complete (Denville Scientific) was used to amplify designated amplicons. For qPCR assays, cDNA was generated by the QuantiTect Rev. Transcription Kit (Qiagen) which also includes additional DNA contamination removal. iQ SYBR Green Supermix (Biorad) was used for PCR quantitation in a RotorGene cycler (Corbett). Relative quantification was performed using the ddCt method.[15] To calculate *LOUP* transcript numbers per cell, *LOUP* DNA fragments amplified by RT-PCR from HL-60 cDNA were cloned into the pSCAmpKan plasmid (Agilent). *LOUP* RNA fragments were *in vitro*-transcribed by using a MAXIscript™ Transcription Kit (Ambion). The RNA fragments were used to generate a standard curve for absolute quantification in qRT-PCR assays. Primers and probes used for all PCR assays are provided in Table S3.

Fluorescence-activated cell sorting and analysis

Cell populations were isolated for RNA extraction as previously described.[16] Briefly, mononuclear cells were isolated from bone marrow, spleen and peripheral blood after lysing red blood cells with ACK lysis buffer.[17] Single cell suspensions were stained with fluorochrome-conjugated antibodies (Biolegend and eBioscience) and FACS-sorted based on the following markers. LT-HSC: $Lin^-c\text{-}Kit^+Sca\text{-}1^+CD150^+CD48^-$; ST-HSC: $Lin^-c\text{-}Kit^+Sca\text{-}1^+CD150^-CD48^+$; LMPP: $Lin^-c\text{-}Kit^+Sca\text{-}1^+CD34^+Flt3^+$; MEP: $Lin^-c\text{-}Kit^+Sca\text{-}1^-CD34^-CD16/32^-$; CMP: $Lin^-c\text{-}Kit^+Sca\text{-}1^-CD34^+CD16/32^-$; GMP: $Lin^-c\text{-}Kit^+Sca\text{-}1^-CD34^+CD16/32^+$; Myeloid: $Mac1^+Gr1^+$. Myeloid surface marker staining and FACS analysis were performed following previously described procedures.[18] Cells were stained with PACBLUE-CD11b (BioLegend). Stained cells were analyzed using LSRII flow cytometer (BD Biosciences) and FlowJo software (Tree Star)

Transcript mapping by P5-linker ligation and 3' RACE

The 5' end of the *LOUP* transcript was identified using the P5-linker ligation method as described previously.[19] Briefly, single-stranded cDNAs were generated from HL-60 $polyA^+$ RNA by using SuperScript III reverse transcriptase (Life Technologies) with *LOUP*-specific nested primer #1. Double-strand cDNAs were then synthesized from single-stranded cDNA using a SuperScript™ Double-Stranded cDNA Synthesis Kit (Life Technologies) and blunt-ended by NEBNext End Repair Enzym Module (New England Biolabs). After purification, these cDNAs were ligated with the P5-splinkerette adapter and purified. All purification steps were done by using QIAquick PCR Purification Kit (QIAGEN). Ligated products were then purified and used as templates for PCR using a P5 primer and *LOUP*-specific nested primers #1 and #2 with Phusion Hot Start DNA polymerase (Finnzymes). P5-linker ligation products were gel purified using a QIAgen Gel Extraction Kit (QIAGEN), sub-cloned into the pSCAmpKan vector, and transformed into competent bacteria using a StrataClone Blunt PCR Cloning Kit (Agilent). The 3'RACE assay was performed using a 2^{nd} Generation 5'/3' RACE Kit (Roche) according to manufacturer's instructions. In brief, cDNA was generated from HL-60 $polyA^+$ RNA using

oligo(dT)-anchor primer mix. Overlapping RACE products were then amplified from cDNA using an anchor primer and *LOUP*-specific primers. RACE products were sub-cloned into the pSCAmpKan vector and transformed into competent bacteria using a StrataClone Cloning Kit (Agilent). Plasmids containing P5-linker and RACE products were purified from bacteria, sequenced, and assembled. Primer information is in Table S3.

Northern blotting

10 ug polyA$^-$ and polyA$^+$ RNAs were dissolved and heat denatured in sample buffer containing formamide, MOPS and formaldehyde. Denatured RNAs were separated on a 1% denaturing agarose gel containing formaldehyde, MOPS and EtBr, before being transferred to Brightstar-plus positively charged nylon membrane (Life Technologies). The *LOUP* probe was PCR amplified with primers described in Table S3. The PCR product was sub-cloned into cloned into the pSCAmpKan vector using a StrataClone PCR Cloning Kit (Agilent). The probe sequence was verified by Sanger sequencing. The probe was released from the vector by restriction enzyme digestion and gene purification. The *LOUP* probe was radiolabeled using the Random Primed DNA Labeling Kit (Roche). Northern blot analysis was performed with ExpressHyb™ Hybridization Solution (Clontech) following the manufacturer's protocol.

Quantitative Chromosome Conformation Capture (3C-qPCR)

3C-qPCR experiments were performed by adapting described methods.[20-22] Briefly, 1x10^6 cells were crosslinked using 1% formaldehyde in PBS at room temperature for 10 minutes. The crosslinking reaction was stopped by adding 0.125 M Glycine and incubated for 5 minutes at room temperature followed by 15 minutes on ice. Crosslinked cells were then washed with ice-cold PBS and lysed in 3C lysis buffer (10 mM Tris-HCl, pH 8.0; 10 mM NaCl; Igepal CA-630 0.2% (vol/vol); 1X protease inhibitor cocktail (Sigma)) with 15 Dounce homogenizer strokes. After centrifugation, nuclear pellets were

washed in 1x restriction enzyme buffer before being lysed with 0.1% SDS in 1x restriction enzyme buffer at 65^0C for 10 minutes. After incubation, the chromatin solution was supplemented with 1% Triton X-100 and digested by ApoI restriction enzyme (New England Biolabs) at 37^0C overnight with rotation. The following day, 1.5% SDS was added to the reaction and enzyme activity was inhibited by incubating at 65^0C for 30 minutes. Nearby DNA ends of digested chromatin were joined by T4-ligase (New England Biolabs) at 16^0C for 2 h. Bound proteins, including histones, were removed by proteinase K digestion at 65^0C overnight. The DNA library was extracted by phenol/chloroform using phase-lock gel tubes (5PRIME) and ethanol precipitation. RNA was removed by incubating 3C libraries with RNase A (Lucigen) at 37^0C for 15 minutes. TaqMan real-time PCR quantifications of ligation products were performed, using primers and probes as documented in Table S3.

Chromatin Isolation by RNA Purification (ChIRP)

ChIRP assays were performed as previously described[23,24] with additional modifications. Briefly, to preserve RNA-chromatin interactions, cells were first crosslinked with 2 mM EGS at room temperature for 45 minutes. After washing cells with ice-cold PBS, cells were further crosslinked with 3% paraformaldehyde for 15 minutes at room temperature. The crosslinking reaction was quenched with 0.125 M glycine for 5 minutes at room temperature. Crosslinked cells were washed in ice-cold PBS and lysed in sonication buffer (20 mM Tris pH 8, 150 mM NaCl, 0.1% SDS, 1% Triton-X, 2 mM EDTA, 1 mM PMSF) supplemented with cOmplete™, Mini Protease Inhibitor Cocktail (Sigma-Aldrich) and SUPERase In RNase Inhibitor (Invitrogen). After sonication and centrifugation, the supernatant containing sheared chromatin was collected and incubated with biotinylated anti-sense DNA tiling probes in hybridization buffer (750 mM NaCl, 1% Triton, 0.1% SDS, 50 mM Tris-Cl pH 7.0, 1 mM EDTA, 15% formamide, 1 mM PMSF) supplemented with cOmplete™, Mini Protease Inhibitor Cocktail and SUPERase In RNase Inhibitor. Hybridized chromatin fragments were captured using Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen). Captured chromatin fragments was either used for extracting chromatin-bound RNA by Trizol reagent or for DNA isolation. Chromatin-bound *LOUP* was quantitated

from chromatin-bound RNA by qRT-PCR. Enrichment of the URE and the PrPr were evaluated by qPCR. Probes used in the ChIRP assay were designed using the online probe designer at singlemoleculefish.com and are listed in Table S3.

DNA pull-down assay (DNAP)

DNAP was performed as described previously with minor modifications.[25] Briefly, the nuclear extract was pre-cleared with Dynabeads™ MyOne™ Streptavidin C1 for 30 minutes at 4°C then incubated overnight with biotinylated oligonucleotides in binding buffer (10 mM HEPES pH 7.9; 100 mM KCl, 5 mM MgCl2, 1 mM EDTA, 10% glycerol, 1 mM DTT, 0.5% NP-40, 1 mM DTT) supplemented with 1x protease inhibitor cocktail (Sigma-Aldrich). Beads were washed with binding buffer then added to the binding reaction. After 1-hour incubation, beads were washed five times with binding buffer. DNA-bound proteins were eluted from beads and subjected to SDS-PAGE and immunoblotting.

RNA pull-down assay (RNAP) and RNA-Protein interaction prediction

RNAP were performed essentially as described previously[26] with few modifications. Briefly, biotinylated RNA was *in vitro*-transcribed using the MAXIscript™ Transcription Kit (Ambion). The DNA template was removed by DNAseI treatment. Transcribed RNA was purified using a RNeasy Mini Kit (QIAGEN). Purified RNA was denatured by heating to 90°C for 2 minutes following incubation on ice for 2 minutes in RNA structure buffer (10 mM Tris pH 7, 0.1 M KCl, 10 mM MgCl2). Denatured RNA was then shifted to room temperature for 20 minutes to form proper secondary structure. Nuclear extracts were treated with RNase-free DNase I (Roche) to remove genomic DNA and pre-cleared with Dynabeads™ MyOne™ Streptavidin C1 or Streptavidin agarose beads (Invitrogen) in binding buffer I (150 mM KCl, 25 mM Tris pH 7.4, 0.5 mM DTT, 0.5% NP40, 1 mM PMSF) supplemented with cOmplete™, Mini Protease Inhibitor Cocktail, and SUPERase In RNase Inhibitor. Pre-cleared extracts were then incubated with biotinylated RNAs in binding buffer I for 1 hour. Beads were washed with

binding buffer I then added to the binding reaction. After 1-hour incubation, beads were washed five times with binding buffer I. RNA-bound proteins were eluted from beads and subjected to SDS-PAGE and immunoblotting. For recombinant proteins (full-length RUNX1 (OriGene Technologies), and Runt domain (MyBiosource)), binding buffer II (50 mM Tris-Cl 7.9, 10% Glycerol, 100 mM KCl, 5 mM MgCl2, 10 mM β-ME, and 0.1% NP-40) was used.

In silico prediction of RNA-Protein interactions were performed using the catRAPID Fragments algorithm in which protein-RNA interaction propensities were predicted based on calculation of secondary structure, hydrogen bonding, and van der Waals contributions.[27] To identify RIP-seq peaks containing sequences similar to R1 or R2, a .bed file with coordinates and peak IDs was prepared. FASTA nucleotide sequences corresponding to the peaks were extracted by getfasta algorithm within The BEDtools suite.[28] Output files were as input for blast2sequences (http://blast.ncbi.nlm.nih.gov/). Top blast hits were analyzed using catRAPID against RUNX1.

RNA Immunoprecipitation sequencing and qPCR (RIP-seq and RIP-qPCR)

RIP was performed following a protocol reported by Hendrickson et al[29] with modifications. Briefly, cells were crosslinked in 0.1% formaldehyde at room temperature for 10 minutes. The crosslinking reaction was quenched for 5 min at room temperature with 0.125 M glycine. Crosslinked cells were washed with ice-cold PBS. Cell pellet was lysed in RIPA lysis buffer (50 mM Tris (pH 8), 150 mM KCl, 0.1 % SDS, 1 % Triton-X, 5 mM EDTA, 0.5 % sodium deoxycholate, 0.5 mM DTT) supplemented with protease inhibitor cocktail (Thermo Scientific) and 100 U/ml RNaseOUT™ (Invitrogen). After sonication, cell lysate was pre-cleared by incubating with Dynabeads® Protein G (Invitrogen). Beads were then captured and removed using a magnet. Pre-cleared lysate was incubated with anti-RUNX1 antibody or IgG (Abcam) at 4°C for 2 hours before adding 50 µl of Dynabeads® Protein G to capture antibodies. After washing, beads were kept at -20^{0}C or proceeded to incubation with reverse-crosslinking buffer (3× PBS (without Mg or Ca), 6% N-lauroyl sarcosine, 30 mM EDTA, 15 mM DTT) supplemented with Proteinase K (Ambion) and RNaseOUT together with the input sample. Captured RNAs were extracted

with Trizol reagent. Contaminated DNA was removed from extracted RNA by DNAseI from RNase-Free DNase Set (QIAGEN) then ribosomal RNA was removed using the Ribo-Zero™ Magnetic Gold Kit (Epicentre). RNA was further purified using RNeasy MinElute Cleanup Kit (QIAGEN). RNA quality was determined using the RNA 6000 Pico Kit on a Bioanalyzer (Agilent). Purified RNA was used for qRT-PCR as described elsewhere and cDNA library construction using the Truseq stranded total RNA library prep kit (Illumina) according to the manufacturer's protocol. The libraries were pooled together and subjected to pair-end sequencing on a Nextseq500 (Illumina) to achieve 2×40 bp reads.

Chromatin Immunoprecipitation and qPCR (ChIP-qPCR)

ChIP was performed as previously described.[30] Briefly, $2x10^6$ U937 cells were crosslinked with 1% formaldehyde (formaldehyde solution, freshly made: 50 mM HEPES-KOH; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 11% formaldehyde) for 10 minutes at room temperature. The crosslinking reaction was stopped by incubating with 0.125 M glycine for 5 minutes at room temperature. Crosslinked cells were washed twice with ice-cold PBS (freshly supplemented with 1 mM PMSF). The cell pellet was then lysed for 10 minutes on ice and chromatin was fragmented by sonication (25 cycles, 30 seconds on, 60 seconds off, high power, Bioruptor). The chromatin solution was incubated with 10 µg antibody overnight at 4°C. Protein A magnetic beads (NEB) were used to capture antibody-bound chromatin. After washing, chromatin was reverse-crosslinked and treated with proteinase K overnight at 65°C. Beads were then removed using a magnet and the chromatin solution was treated with RNase treatment (Epicentre) for 30 minutes at 37°C. ChIP DNA was extracted with Phenol:Chloroform:Isoamyl Alcohol 25:24:1, pH:8 (Sigma-Aldrich) and then precipitated with an equal volume of isopropanol in the presence of glycogen. The DNA pellet was dissolved in 30 µl of TE buffer for qPCR analyses. Fold enrichment was calculated using the formula $2^{(-\Delta\Delta Ct(ChIP/IgG))}$. Primer sets used for ChIP-qPCR are listed in Table S3.

RIP-seq and ChIP-seq data analyses

RIP-seq samples were demutliplexed. Reads were deduplicated by Clumpify from the BBtools suite, sourceforge.net/projects/bbmap/) with the parameters “dedupe spany addcount”. Adaptor quality trimming and filtering was performed by BBDuck from the BBtools suite with the parameters “ktrim=l hdist=2”. Low quality reads/bases were removed by Trimmomatic[31] with the parameters: “LEADING:28 SLIDINGWINDOW:4:26 TRAILING:28 MINLEN:20”. The processed reads were then aligned to Human genome build 38 (hg38) by the STAR aligner [32] with the parameters “--outFilterScoreMinOverLread 0.05 --outFilterMatchNminOverLread 0.05 --outFilterMultimapNmax 30 --outSAMprimaryFlag AllBestScore”. Coverage maps were generated using bamCoverage (part of the deepTools suite[33] with default parameters. Peak calling was performed using HOMER (v4.10).[34] RUNX1 peaks with at least two-fold enrichment over IgG control were selected for annotation using HOMER. Peaks were assigned to a gene locus by satisfying at least one of the following location criteria: a nearest transcription start site, on a promoter, or on a transcript body. Ensemble 97 human gene CRCh38.p12 was used to retrieve gene annotation information through Biomart in Ensembl.[35] For ChIP-seq and Dnase-seq data, raw reads were downloaded from GEO (RUNX1 ChIP-seq in THP-1 cells: GSM2108052; RUNX1-ETO and H3K9Ac ChIP-seq, and Dnase-seq in Kasumi-1: GSE29222). Read quality were evaluated by FastQC.[36] Where necessary, reads with low-quality were trimmed by trim_galore.[37] Genome alignment, coverage maps, and peak calling were performed using software packages as above. ChIP-seq peaks with at least ten-fold enrichment over surrounding 10 kb region were selected for annotation using HOMER. BigWig files were uploaded and viewed via the UCSC genome browser.

The following gene tracks were from published data deposited in GEO and were processed via the Cistrome pipeline.[38] H3K27ac overlay track includes monocytes (GSM2679933), THP-1 (GSM2544236), and HL-60 (GSM2836486). The H3K4me1 overlay track includes monocyte (GSM1435532), HL-60 (GSM2836484) and THP-1 (GSM3514951). H3K4me3 overlay track includes monocytes (GSM1435535), HL-60 (GSM945222), and THP-1 (GSM2108047). The DNAse-seq overlay

281 track includes monocytes (GSM701541) and HL-60 (GSM736595). RUNX1 ChIP-seq tracks include
282 $CD34^+$ cells from healthy donors (GSM1097884), and an AML patient with FLT3-ITD and no other
283 defined mutations (GSM1581788). The CAGE track (reverse strand and max counts) was imported
284 from the FANTOM5 project.[39]

285

286 **RNA sequencing data analysis (RNA-seq)**

287 Raw sequencing reads (FASTQ files) of the Human Body Map data set were downloaded from
288 AEArrayExpress (E-MTAB-513). Read quality was assessed by FastQC.[36] Reads with low-quality were
289 trimmed by trim_galore.[37] The *LOUP* transcript was integrated into the Ensembl human cDNA catalog
290 GRCh38 and transcript levels were quantified against this catalog using the Salmon software.[40] AML
291 RNA-seq were downloaded from TCGA and transcript counts were determined. For RNA-seq track
292 visualization, the following RNA-seq raw data were downloaded from GEO: THP-1 (GSM1843218), HL-
293 60 (GSM1843216), $CD34^+$ HSPC (GSM1843222), Monocyte (GSM1843224), and Jurkat
294 (GSM2260195). BigWig files were generated using packages as described in ChIP-seq and RIP-seq
295 analyses and viewed via the UCSC genome browser.

296

297 **Single-cell RNA-seq (scRNA-seq) data analyses**

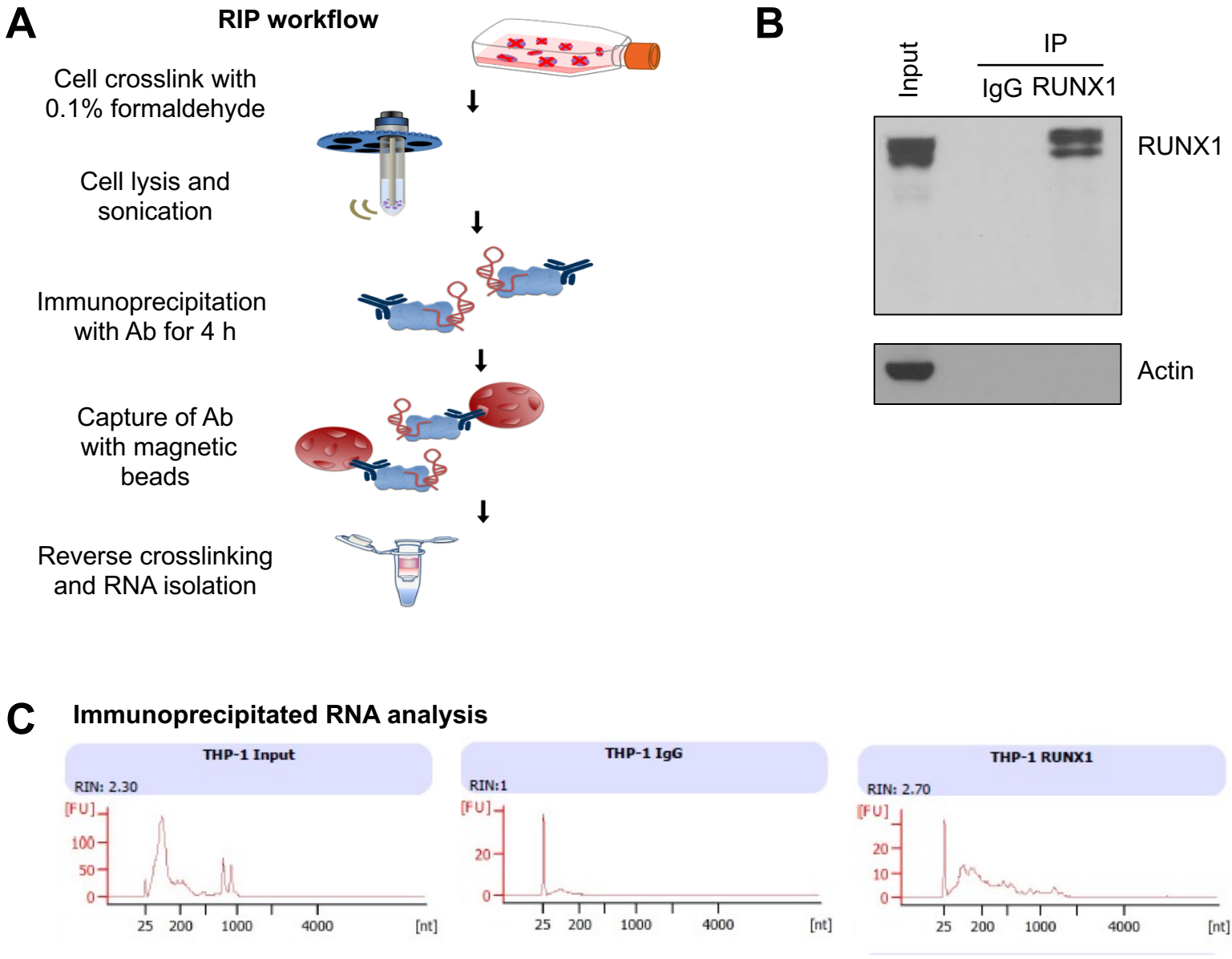
298 Raw fastq files data of mononuclear cells isolated from peripheral blood and bone marrow were
299 obtained from the 10x Genomics public datasets repository
300 (https://www.10xgenomics.com/resources/datasets/) and pooled together to maximize coverage of
301 hematopoietic cell lineages. Transcripts were mapped to the human transcriptome using Cell Ranger
302 (10x Genomics) with a custom hg38 gtf containing the *LOUP* transcript details. Subsequent analyses
303 were performed in R (v3.6.2) using the previously published Bioconductor workflow with minor
304 modifications.[41] Filtering criteria were as below. First, cells with library sizes more than three median
305 absolute deviations (MADs) below the median library or four MAD's above the median library size were

filtered out. Second, cells with a total number of expressed genes (>= 1 read) more than three MADs below the median total number of expressed genes or four MAD's above the median total number of expressed genes were filtered out. Third, cells with a total percentage of expressed genes originating from mitochondrial DNA more than eight MADs above the median were filtered out. A doublet score was then computed to estimate the percentage of barcodes for two or more cells as previously described.[42] Cells with a doublet score of 0.99 were excluded. Expression of each cell was normalized by a size factor approach as previously described[43] resulting in $\log_2$(normalize_expression) values. Principle component and t-Distributed Stochastic Neighbor Embedding (tSNE) analyses revealed no significant batch effects to be regressed out for the samples. To account for dropouts which are found more frequently for genes with lower expression magnitude in scRNA-seq,[44] cells with undetectable *LOUP* and *PU.1* transcripts were referred as *LOUP*$^-$/*PU.1*$^-$ and cells with detectable *LOUP* and *PU.1* transcripts were referred as *LOUP*$^+$/*PU.1*$^+$. Expression data visualization was performed using SPRING software.[45] Briefly, a graph of cells connected to their nearest neighbors in gene expression space was determined. This was then projected into two dimensions using a force-directed graph layout. Identity of each cell was inferred using the Blueprint-Encode annotation which includes normalized expression values of 259 bulk RNA-seq samples generated from pure and defined cell populations.[46,47] This annotation was integrated in the SingleR R package.[48] Annotated cells were grouped into major definitive cell lineages as described in the text. Gene Ontology (GO) analysis was performed using the Database for Annotation, Visualization and Integrated Discovery functional annotation tool (http://david.abcc.ncifcrf.gov). Significance of over-represented Gene Ontology biological processes was examined based on $-\log_{10}$ of corrected *p*-values from Bonferroni-corrected modified Fisher's exact test.[49] A list of enriched genes in *LOUP*$^+$/*PU.1*$^+$ group *vs*. *LOUP*$^-$/*PU.1*$^-$ group was generated using SPRING software.[45] Upregulated genes (Z-score >1) was used for GO analysis.

Prediction of coding potential with PhyloCSF

The cross-species multiple sequence comparisons result of 46 species (i.e., multiz100way) was downloaded from the UCSC genome browser (https://genome.ucsc.edu). Guided by the GENCODE gene annotation (ver. 28), the alignment of the longest isoform of each gene was extracted from alignments of cross-species multiple sequence comparisons. The alignment was analyzed by PhyloCSF[50] with 58mammals mode. All possible coding reading frames on the same strand were scanned. The maximal score was used.

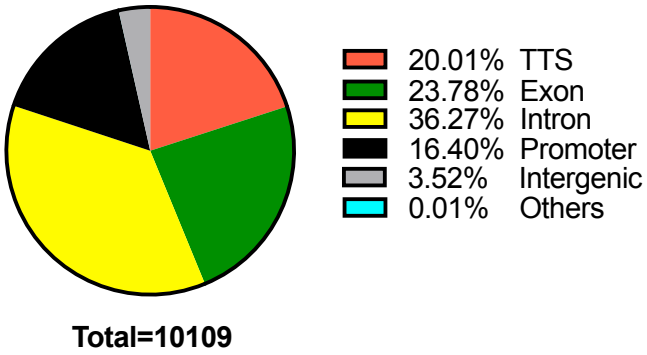
SUPPLEMENTAL FIGURES


A
RIP workflow
Cell crosslink with 0.1% formaldehyde
Cell lysis and sonication
Immunoprecipitation with Ab for 4 h
Capture of Ab with magnetic beads
Reverse crosslinking and RNA isolation
B
Input
IP
IgG
RUNX1
RUNX1
Actin
C
Immunoprecipitated RNA analysis
THP-1 Input
RIN: 2.30
THP-1 IgG
RIN:1
THP-1 RUNX1
RIN: 2.70
[FU]
[nt]
25
200
1000
4000
D
RUNX1 RIP-seq and ChIP-seq flowchart (THP-1 cells)
RUNX1 RIP-seq
hg38 GRCh38.p12
(226,950 transcripts)
RUNX1 ChIP-seq
(GSE79899)
10,109 peaks
Gene annotation
59,598 unique genes
Gene annotation
24,185 peaks
6,035 genes
13,272 genes

FigS1

E

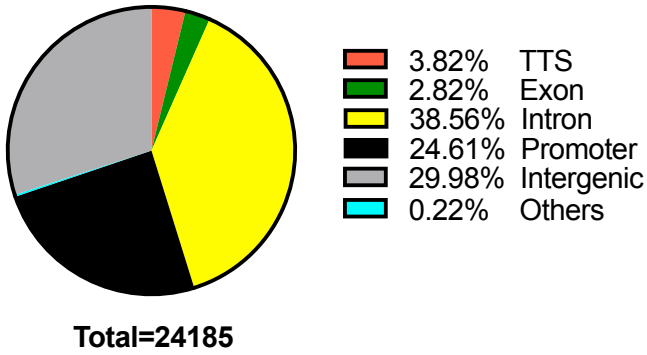
RIP-seq, THP-1 cells



Total=10109

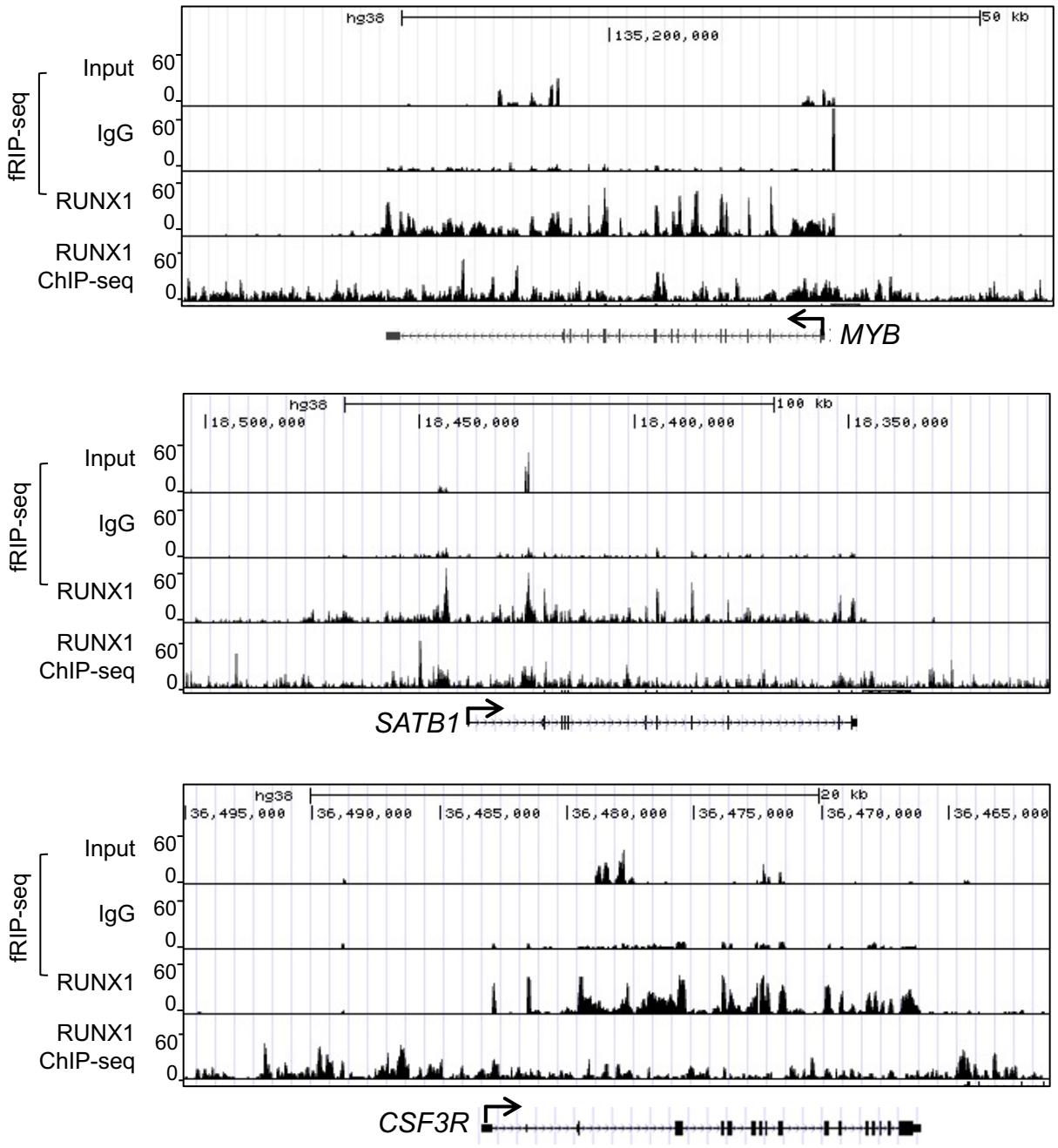
F

ChIP-seq, THP-1 cells

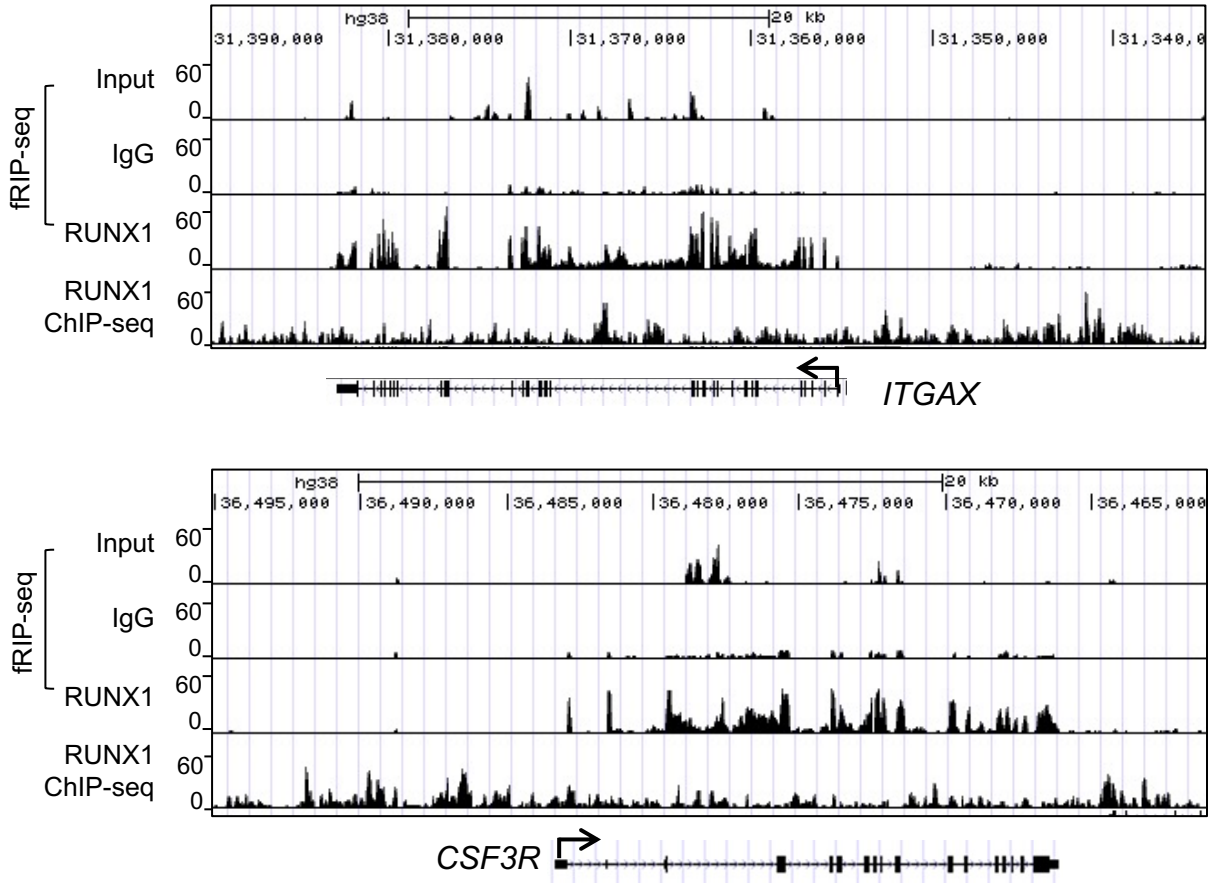


Total=24185

G



FigS1


hg38
20 kb
31,390,000
31,380,000
31,370,000
31,360,000
31,350,000
31,340,0
fRIP-seq
Input
IgG
RUNX1
RUNX1
ChIP-seq
60
0
ITGAX
hg38
20 kb
36,495,000
36,490,000
36,485,000
36,480,000
36,475,000
36,470,000
36,465,000
fRIP-seq
Input
IgG
RUNX1
RUNX1
ChIP-seq
60
0
CSF3R

FigS1

Figure S1. Identification of gene loci exhibiting concurrent RUNX1-RNA and -DNA interactions, Related to Figure 1

(A) Workflow of RUNX1-RIP procedure. Ab: antibody.

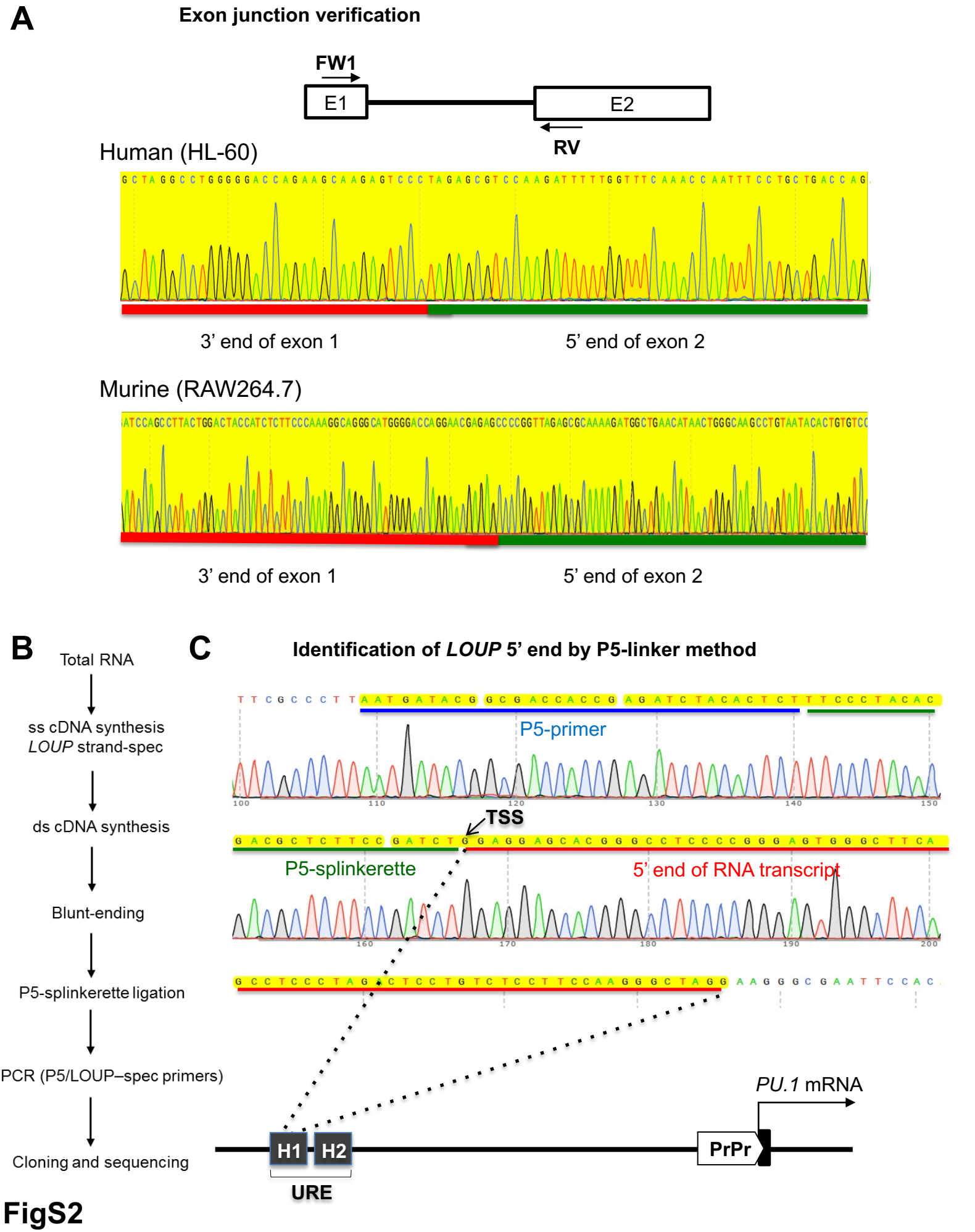
(B) Immunoblot detection of RUNX1 and actin proteins immunoprecipitated from THP-1 cell lysate using anti-RUNX1 antibody and Rabbit IgG isotype control.

(C) Bioanalyzer analysis of RNAs captured by anti-RUNX1 antibody and IgG control plus input RNAs.

(D) Analysis flowchart of RUNX1 RIP-seq and ChIP-seq analyses.

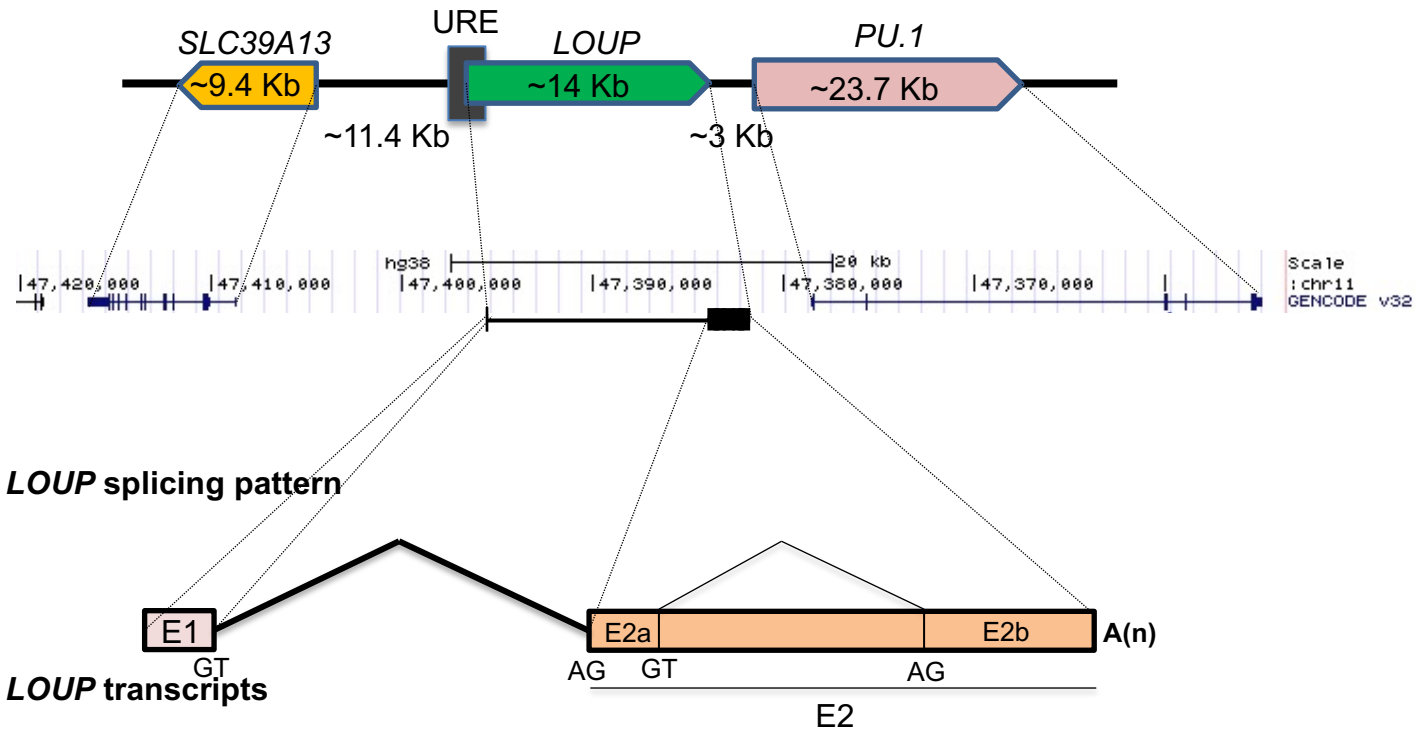
(E and F) Pie charts showing distribution of RUNX1 RIP-seq peaks and RUNX1 ChIP-seq peaks at different genomic locations.

(G) Examples of the myeloid gene loci having both RUNX1 RIP peaks and RUNX1 ChIP-seq peaks from THP-1 cells.


A
Exon junction verification
FW1
E1
E2
RV
Human (HL-60)
3' end of exon 1
5' end of exon 2
Murine (RAW264.7)
3' end of exon 1
5' end of exon 2
B
Total RNA
ss cDNA synthesis
LOUP strand-spec
ds cDNA synthesis
Blunt-ending
P5-splinkerette ligation
PCR (P5/LOUP–spec primers)
Cloning and sequencing
C
Identification of LOUP 5' end by P5-linker method
P5-primer
TSS
P5-splinkerette
5' end of RNA transcript
PU.1 mRNA
H1
H2
URE
PrPr

FigS2

D **Genomic location**



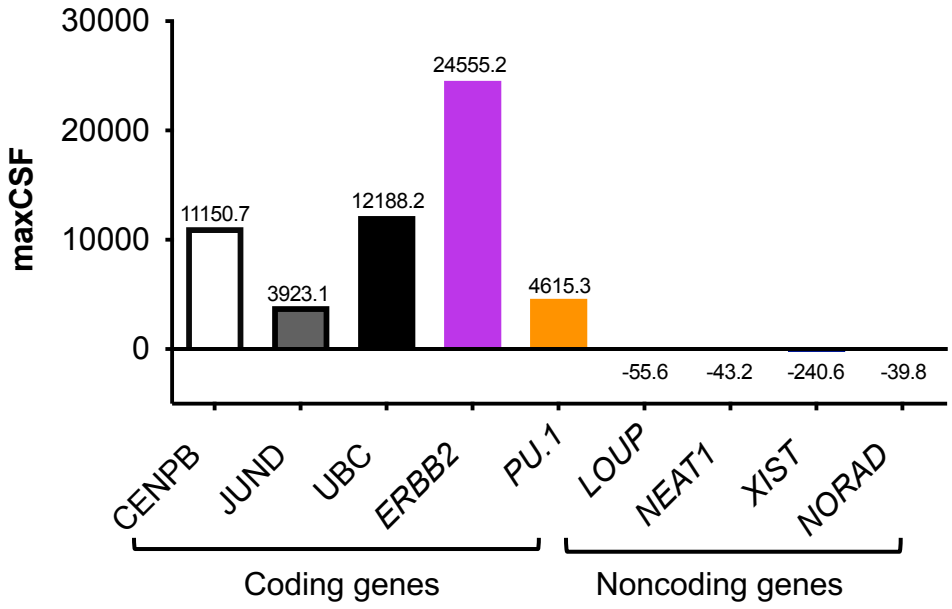
***LOUP* splicing pattern**

***LOUP* transcripts**

~2.3 Kb + A(n)

~1.0 Kb + A(n)

E **Coding potential analysis**



FigS2

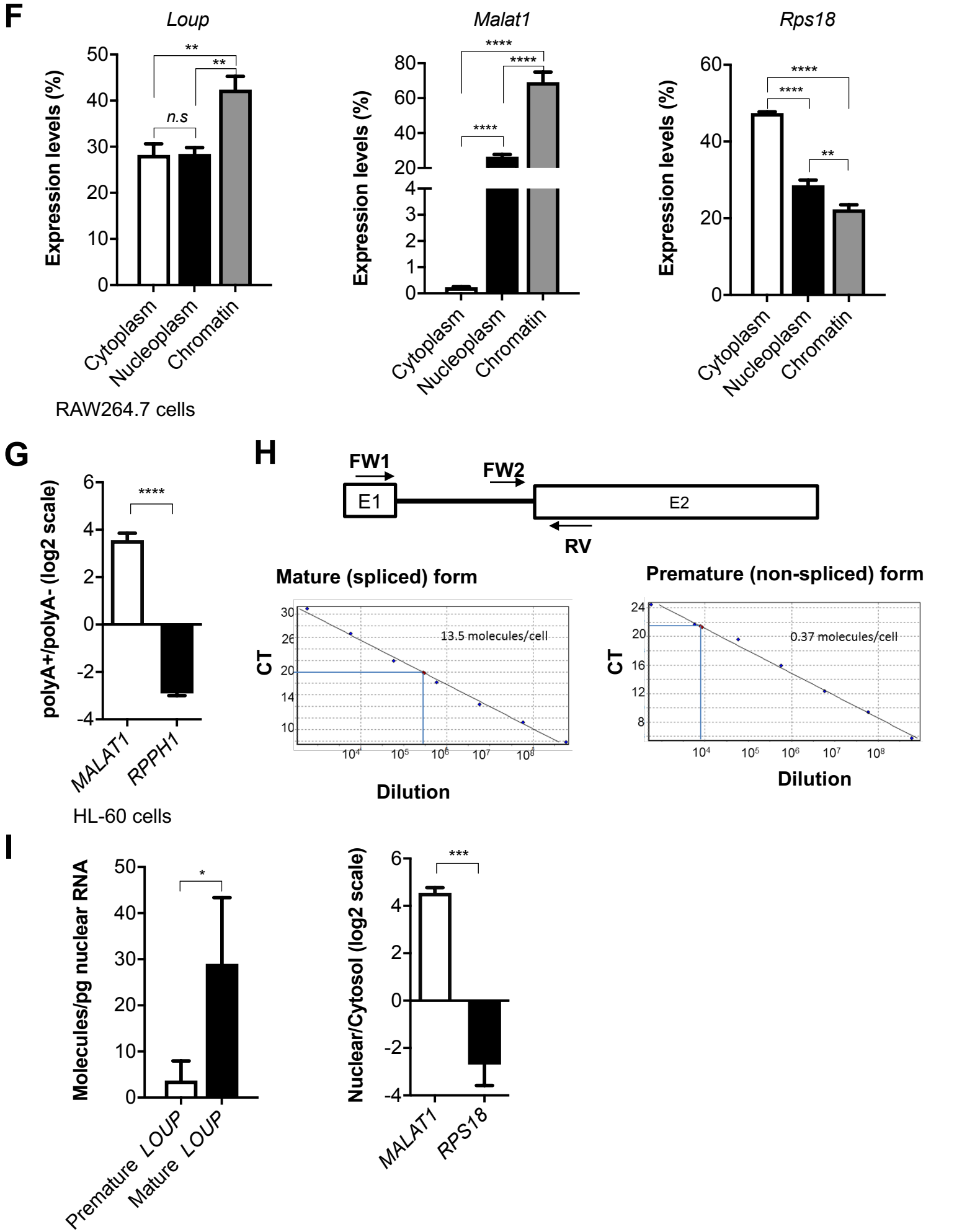

F
Loup
Malat1
Rps18
Expression levels (%)
n.s
Cytoplasm
Nucleoplasm
Chromatin
RAW264.7 cells
G
polyA+/polyA- (log2 scale)
MALAT1
RPPH1
HL-60 cells
H
FW1
FW2
E1
E2
RV
Mature (spliced) form
Premature (non-spliced) form
CT
13.5 molecules/cell
0.37 molecules/cell
Dilution
I
Molecules/pg nuclear RNA
Premature LOUP
Mature LOUP
Nuclear/Cytosol (log2 scale)
MALAT1
RPS18
HL-60 cells
FigS2

Figure S2. Transcript map and molecular features of *LOUP*, Related to Figure 2

(A) RT-PCR confirmation of exon-exon junction. Upper panel: Schematics of the PCR amplicon and primer locations. Lower panels: DNA sequencing of PCR products from human (HL-60) and murine (RAW264.7) cells.

(B) Workflow of 5' end mapping by the P5-linker ligation method.

(C) P5-linker ligation assay determining the 5' end of *LOUP* transcript. Upper panel: DNA sequencing analysis showing locations of the P5-primer, P5-splinkerette, and transcription start site (TSS). Lower panel: Schematic diagram of the *PU.1* locus. Shown are the URE element with the two homology regions H1 and H2.

(D) Schematic diagram showing the relative genomic location of *LOUP* and two neighbor genes: *PU.1* and *SLC39A13* as well as exact locations on the UCSC genome browser track (top), the splicing pattern of *LOUP* (middle), and resultant transcripts (bottom). E1: Exon 1, E2: Exon 2. E2a and E2b are exons derived from an additional splicing event within Exon 2. Exon boundaries were mapped by 3'RACE and RT-PCR.

(E) PhyloCSF analysis of *LOUP* and other known coding and noncoding genes. Shown are coding potential scores.

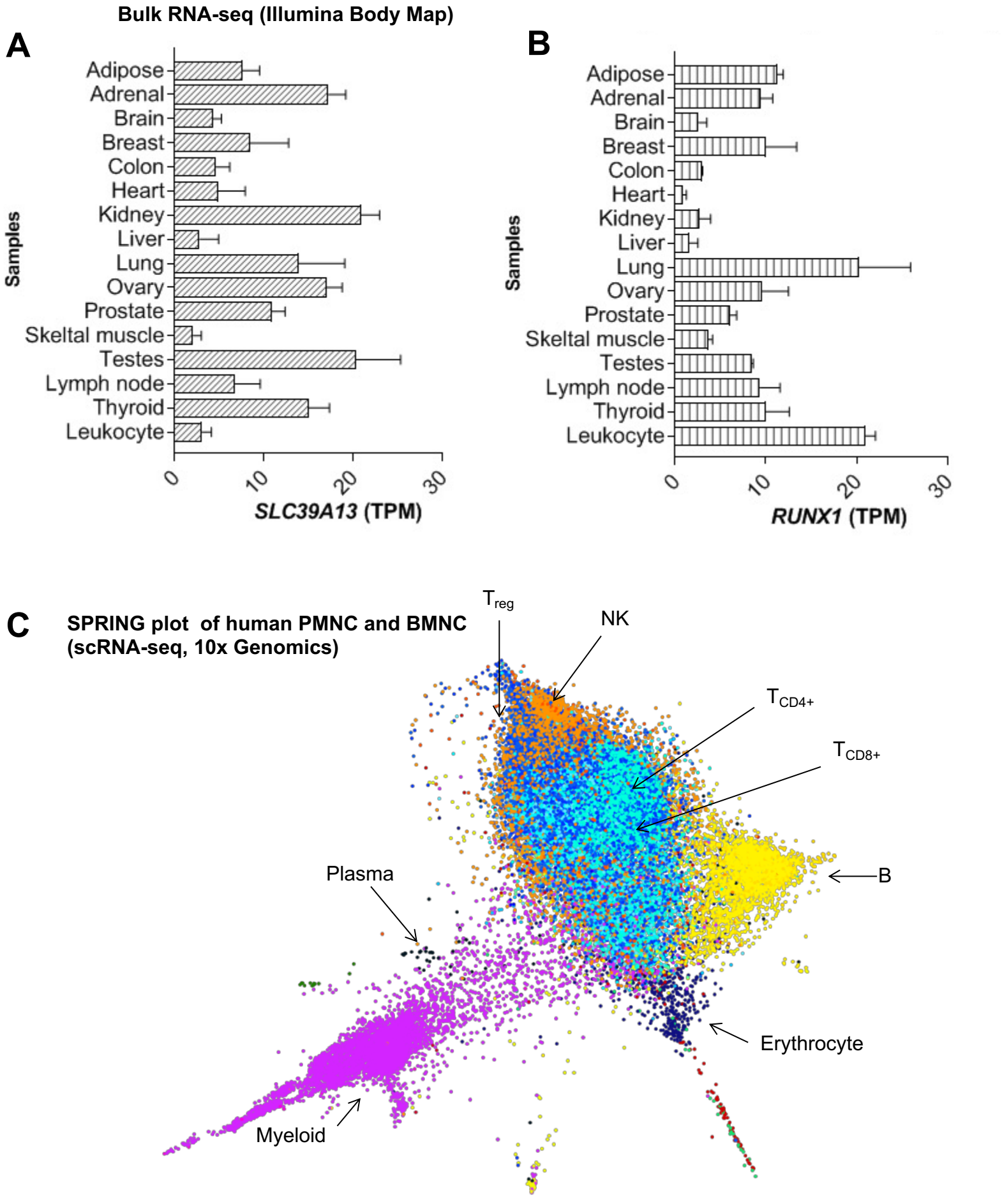
(F) qRT-PCR analysis of *Loup* RNA in subcellular fractions isolated from RAW264.7 cells. Fraction enrichment controls include *Malat1* (chromatin) and *Rps18* (cytoplasm) [51].

(G) qRT-PCR analysis of fraction enrichment controls including *MALAT1* (polyA$^+$) and *RPPH1* (polyA$^-$) (right panel).

(H) Measurement of transcript numbers per HL-60 cell. Upper panel: Schematic diagram of amplified amplicons showing primer locations for non-spliced *LOUP* (FW2-RV) and spliced *LOUP* (FW1-RV). Lower panels: qRT-PCR with RNA standard curve for spliced and non-spliced forms.

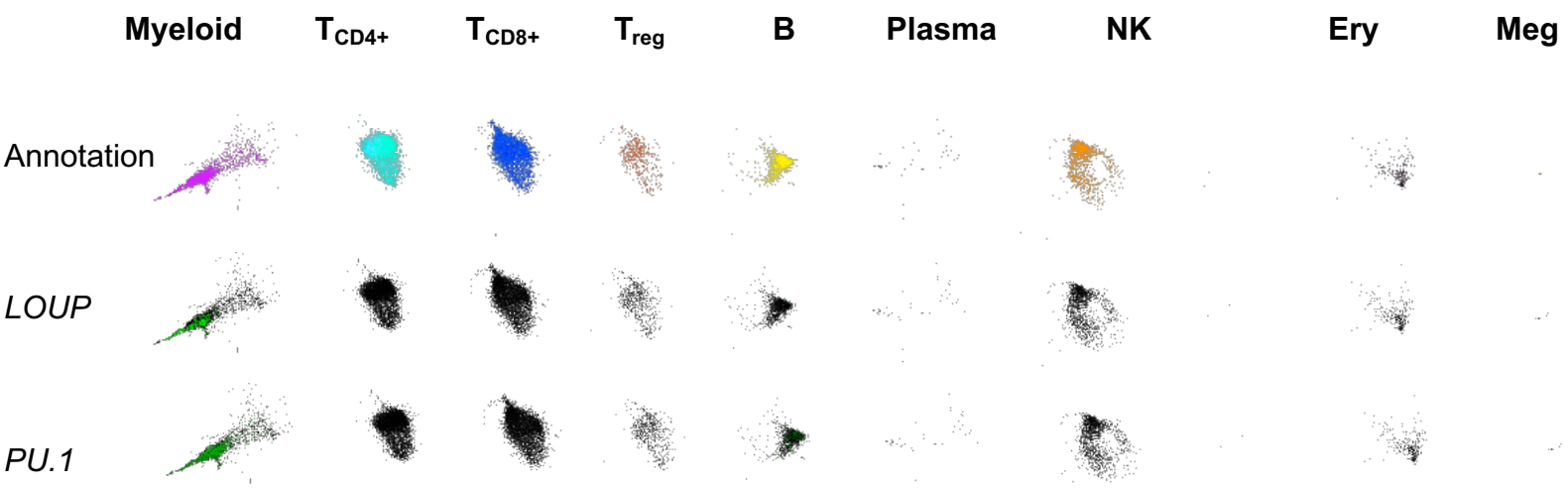
(I) Left panel: qRT-PCR analysis of *LOUP* forms in the nucleus. Right panel: Fraction enrichment controls include *MALAT1* (nucleoplasm) and *RPS18* (cytoplasm).

Error bars indicate SD (n=3). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$, ****$p < 0.0001$, *n.s*: not significant.

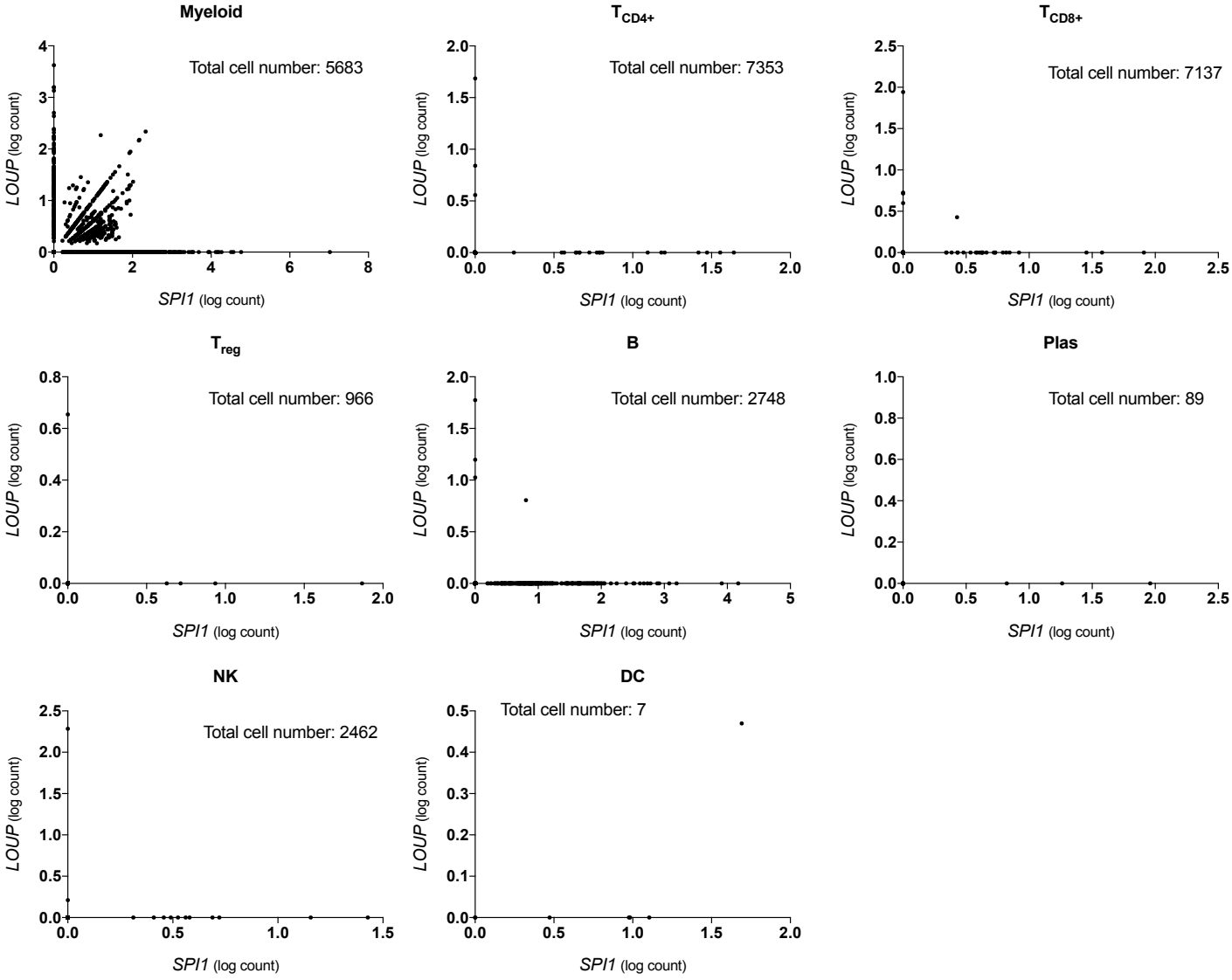

Bulk RNA-seq (Illumina Body Map)
A
Adipose
Adrenal
Brain
Breast
Colon
Heart
Kidney
Liver
Lung
Ovary
Prostate
Skeltal muscle
Testes
Lymph node
Thyroid
Leukocyte
Samples
0
10
20
30
SLC39A13 (TPM)
B
Adipose
Adrenal
Brain
Breast
Colon
Heart
Kidney
Liver
Lung
Ovary
Prostate
Skeltal muscle
Testes
Lymph node
Thyroid
Leukocyte
Samples
0
10
20
30
RUNX1 (TPM)
C
SPRING plot of human PMNC and BMNC
(scRNA-seq, 10x Genomics)
Treg
NK
TCD4+
TCD8+
Plasma
B
Erythrocyte
Myeloid

FigS3

D scRNA-seq (10x Genomic)



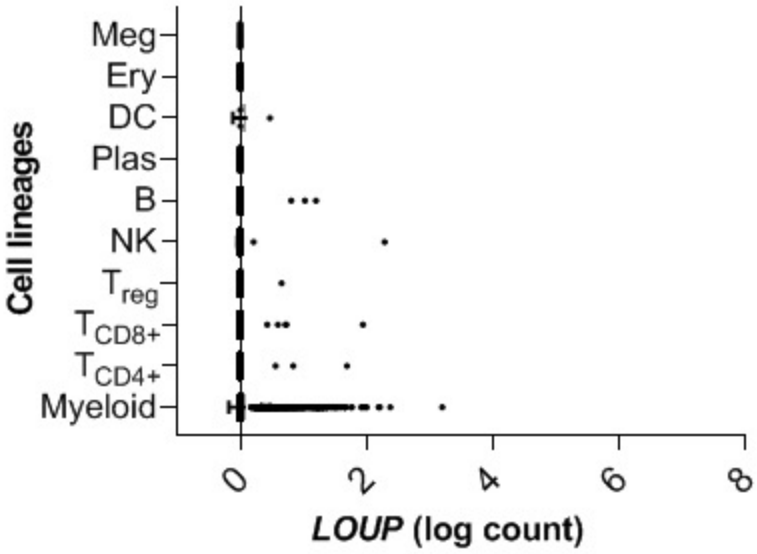
E



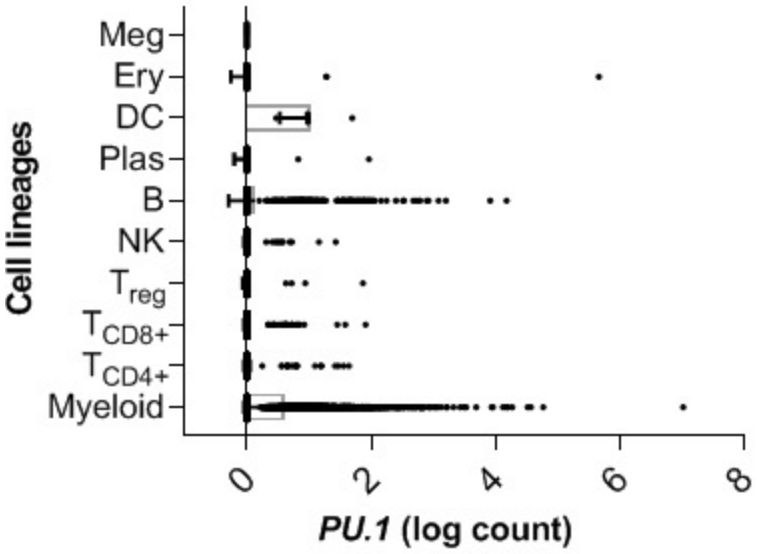
FigS3

scRNA-seq (10x Genomic)

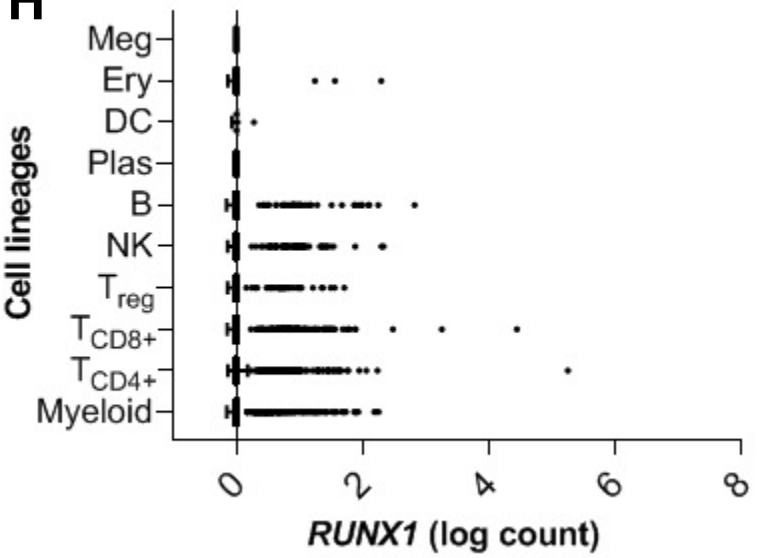
F



G

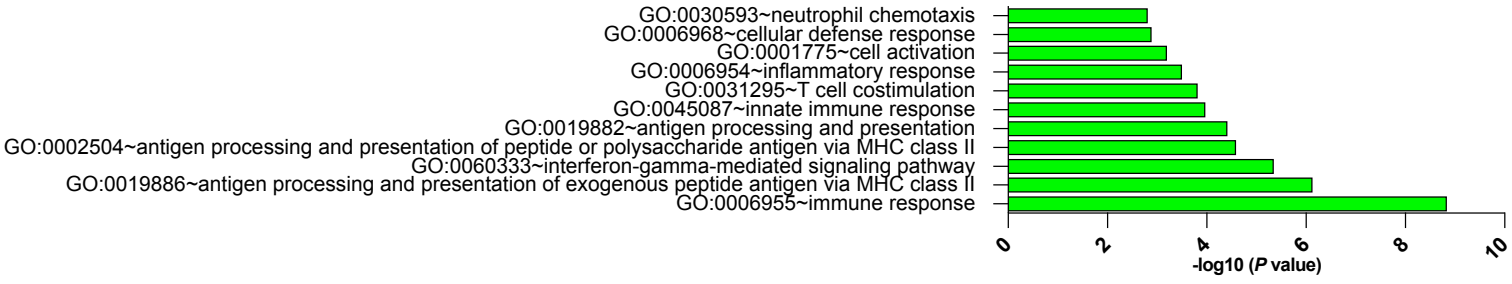


H



I

GO analysis of genes enriched in *LOUP*$^{+}$/*PU.1*$^{+}$ cells (10x Genomic)



FigS3

Figure S3. Gene expression profiles in normal tissues and cell lineages, Related to Figure 3

(A and B) Transcript profiles of *SLC39A13* and *RUNX1* in human tissues. Shown are transcript counts from the Illumina Body Map dataset. Error bars indicate SD (n=2).

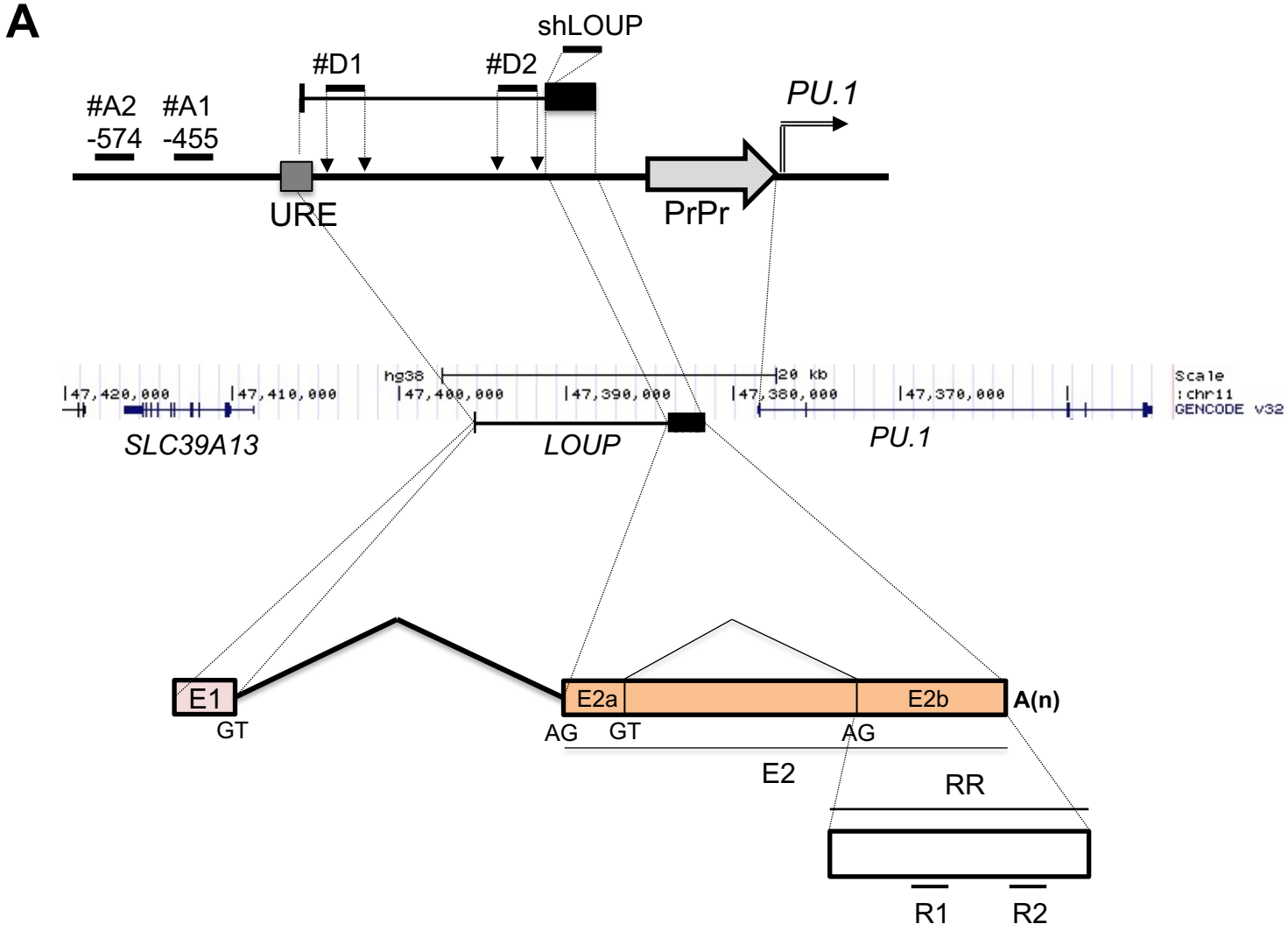
(C) SRING plot analysis of the 10x Genomic scRNA-seq dataset showing color-coded definitive blood lineages using the Blueprint-Encode annotation.[48] Annotated cells in sub-populations were grouped into major cell populations (see methods for details).

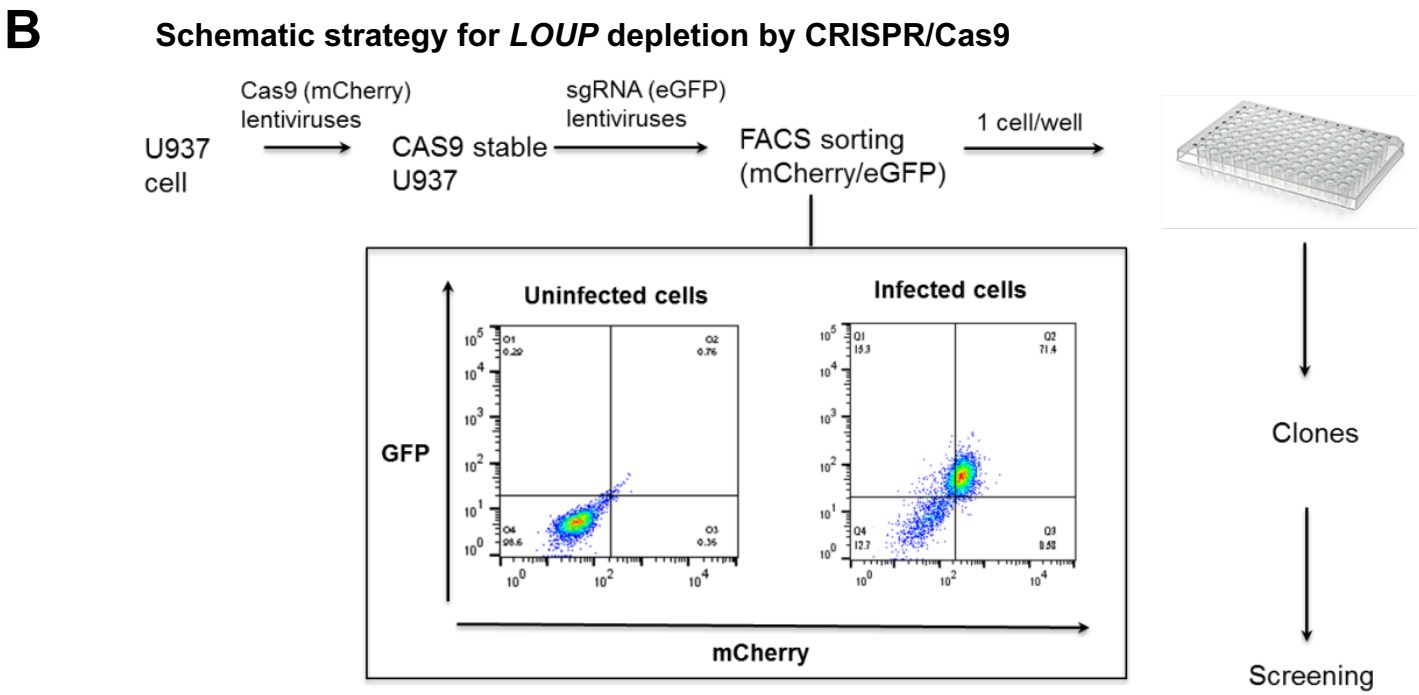
(D) SRING plot analysis showing *LOUP* and *PU* RNA levels in each cell population. Upper panel: Color-coded cell populations extracted from the SPRING plot (C). Middle and bottom panels: locations of individual cells expressing *LOUP* and *PU.1*, respectively (green dots).

(E) Scatter plots for each blood cell lineage of the 10x Genomic scRNA-seq dataset showing the level of expression of *PU.1* mRNA versus the level of expression of *LOUP*. Each dot on the graph represents an individual cell.

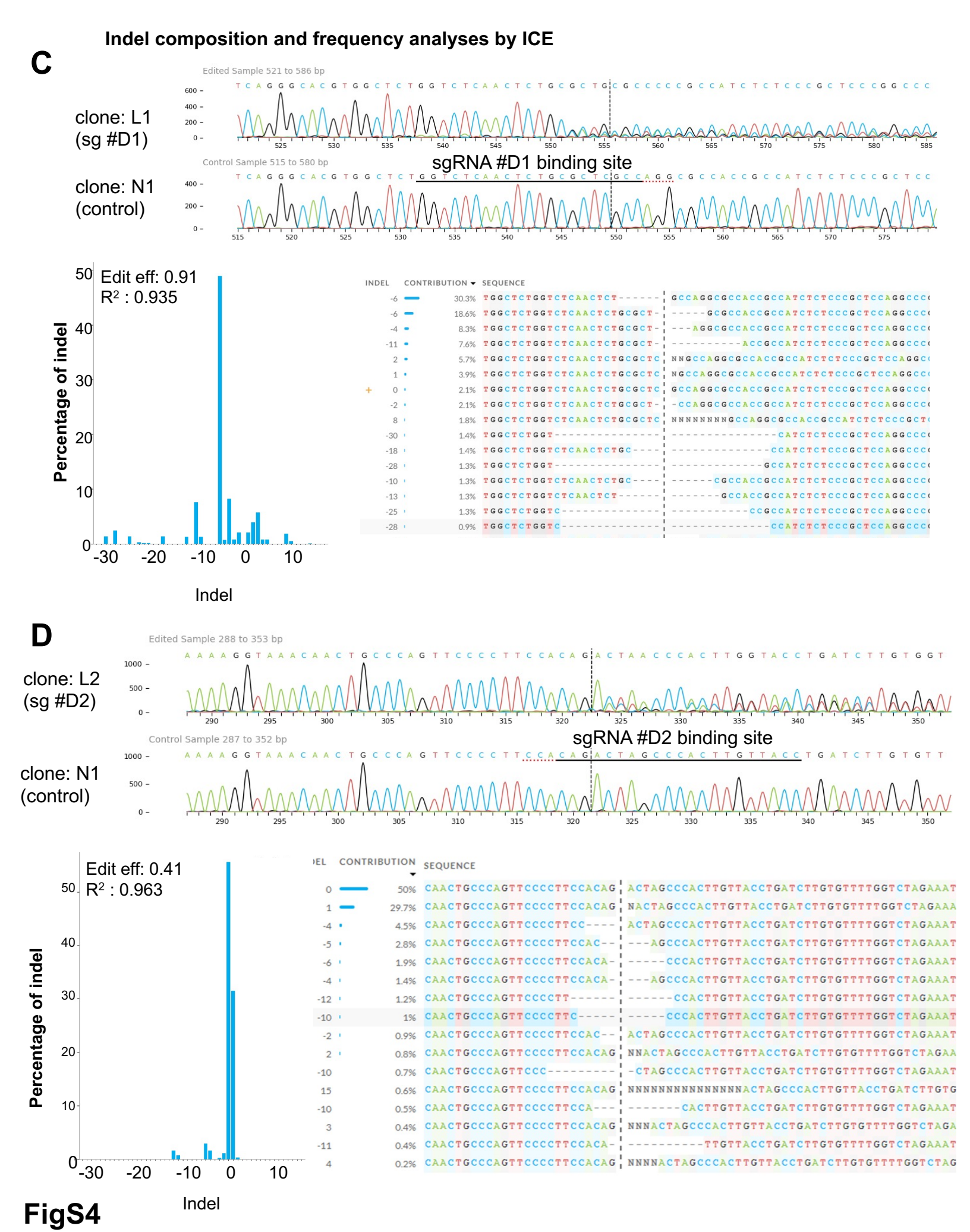
(F, G, and H) Transcript profiles of *LOUP* RNA, and *PU.1* and *RUNX1* mRNAs in blood cell lineages of the 10x Genomic scRNA-seq dataset. Each dot on the graph represents an individual cell.

(I) GO analysis for enrichment of biological processes using a list of genes upregulated in $LOUP^{+}/PU.1^{+}$ cells as compared to $LOUP^{l\text{-}}/PU.1^{+}$ cells. Top enrichment GO terms are shown.


A
shLOUP
#D1
#D2
#A2
#A1
-574
-455
PU.1
URE
PrPr
hg38
20 kb
47,420,000
47,410,000
47,400,000
47,390,000
47,380,000
47,370,000
Scale
:chr11
GENCODE v32
SLC39A13
LOUP
PU.1
E1
GT
E2a
E2b
A(n)
AG
GT
AG
E2
RR
R1
R2


B
Schematic strategy for LOUP depletion by CRISPR/Cas9
U937
cell
Cas9 (mCherry)
lentiviruses
CAS9 stable
U937
sgRNA (eGFP)
lentiviruses
FACS sorting
(mCherry/eGFP)
1 cell/well
Uninfected cells
Infected cells
GFP
mCherry
Clones
Screening

FigS4


Indel composition and frequency analyses by ICE
C
clone: L1
(sg #D1)
clone: N1
(control)
Edited Sample 521 to 586 bp
Control Sample 515 to 580 bp
sgRNA #D1 binding site
Edit eff: 0.91
R² : 0.935
Percentage of indel
Indel
INDEL
CONTRIBUTION
SEQUENCE
D
clone: L2
(sg #D2)
clone: N1
(control)
Edited Sample 288 to 353 bp
Control Sample 287 to 352 bp
sgRNA #D2 binding site
Edit eff: 0.41
R² : 0.963
Percentage of indel
Indel
CONTRIBUTION
SEQUENCE
FigS4

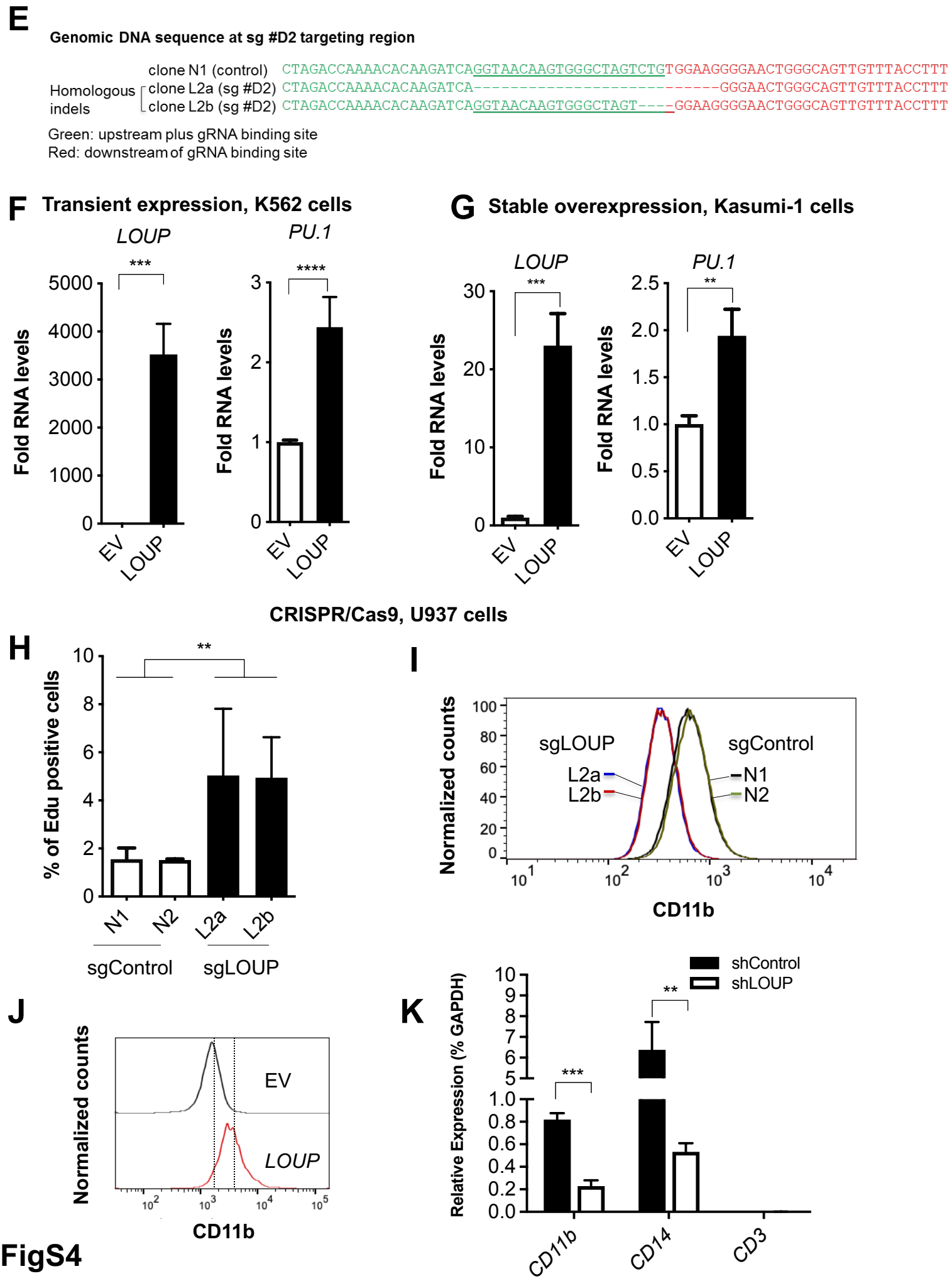

E
Genomic DNA sequence at sg #D2 targeting region
clone N1 (control) CTAGACCAAAACACAAGATCAGGTAACAAGTGGGCTAGTCTGTGGAAGGGGAACTGGGCAGTTGTTTACCTTT
Homologous indels
clone L2a (sg #D2) CTAGACCAAAACACAAGATCA----------GGGAACTGGGCAGTTGTTTACCTTT
clone L2b (sg #D2) CTAGACCAAAACACAAGATCAGGTAACAAGTGGGCTAGT----GGAAGGGGAACTGGGCAGTTGTTTACCTTT
Green: upstream plus gRNA binding site
Red: downstream of gRNA binding site
F
Transient expression, K562 cells
LOUP
PU.1
Fold RNA levels
EV
LOUP
G
Stable overexpression, Kasumi-1 cells
CRISPR/Cas9, U937 cells
H
% of Edu positive cells
N1
N2
L2a
L2b
sgControl
sgLOUP
I
Normalized counts
sgLOUP
sgControl
L2a
L2b
N1
N2
CD11b
J
EV
LOUP
K
Relative Expression (% GAPDH)
shControl
shLOUP
CD11b
CD14
CD3
FigS4

Figure S4. Effects of gain and loss of *LOUP* expression, Related to Figure 4

(A) Schematic diagram showing the genomic location and transcript pattern of *LOUP*. Top diagram: Location of *LOUP* gene relative to the URE and the PrPr. Shown are sgRNA-binding sites for *LOUP* depletion using CRISPR/Cas9 technology (#D1 and #D2) and for *LOUP* induction by CRISPR/dCas9-VP64 technology (#A1 and #A2), as well as shRNA binding site for *LOUP* knock-down. Distance from the TSS of *LOUP* is indicated in bp. Middle diagram: Genomic locations of *LOUP* and its neighboring genes, *SLC39A13* and *PU.1*, on the USCS genome browser track. Bottom diagram: The splicing pattern of *LOUP*. E1: Exon 1, E2: Exon 2. E2a and E2b are exons derived from an additional splicing event within Exon 2 (see also Figure S2D). White boxed area depicts a repetitive region (RR) of 670 bp. The schematic diagram underneath illustrates mature *LOUP* lncRNA containing the RR region. R1 and R2: two regions with high catRAPID *in silico* prediction interaction scores (see also Figure 6E).

(B) Schematic of strategy for *LOUP* depletion. Included is a FACS sorting scheme for isolation of cells expressing both mCherry (Cas9) and eGFP (sgRNAs).

(C and D) Inference of CRISPR edits (ICE) analyses for indel composition and frequency of CRISPR/Cas9 cell clones. Top panels: Trace file segments of amplified genomic regions surrounding sgRNA-binding sites (#D1 and #D2 *LOUP* sgRNAs) in the edited (upper panel) and the control (lower panel) samples. Dotted red underline: Protospacer adjacent motif (PAM) sequence. Solid black underline: guide sequences. Expected cut sites are denoted as vertical dotted lines. Bottom-left panel: Indel efficiency analysis. Bottom-right panel: Indel distribution analysis. Dashed lines indicate deletion length.

(E) Genomic PCR and Sanger sequencing confirmation of U937 cell clones with *LOUP* homozygous indels (L2a and L2b) and control (N1).

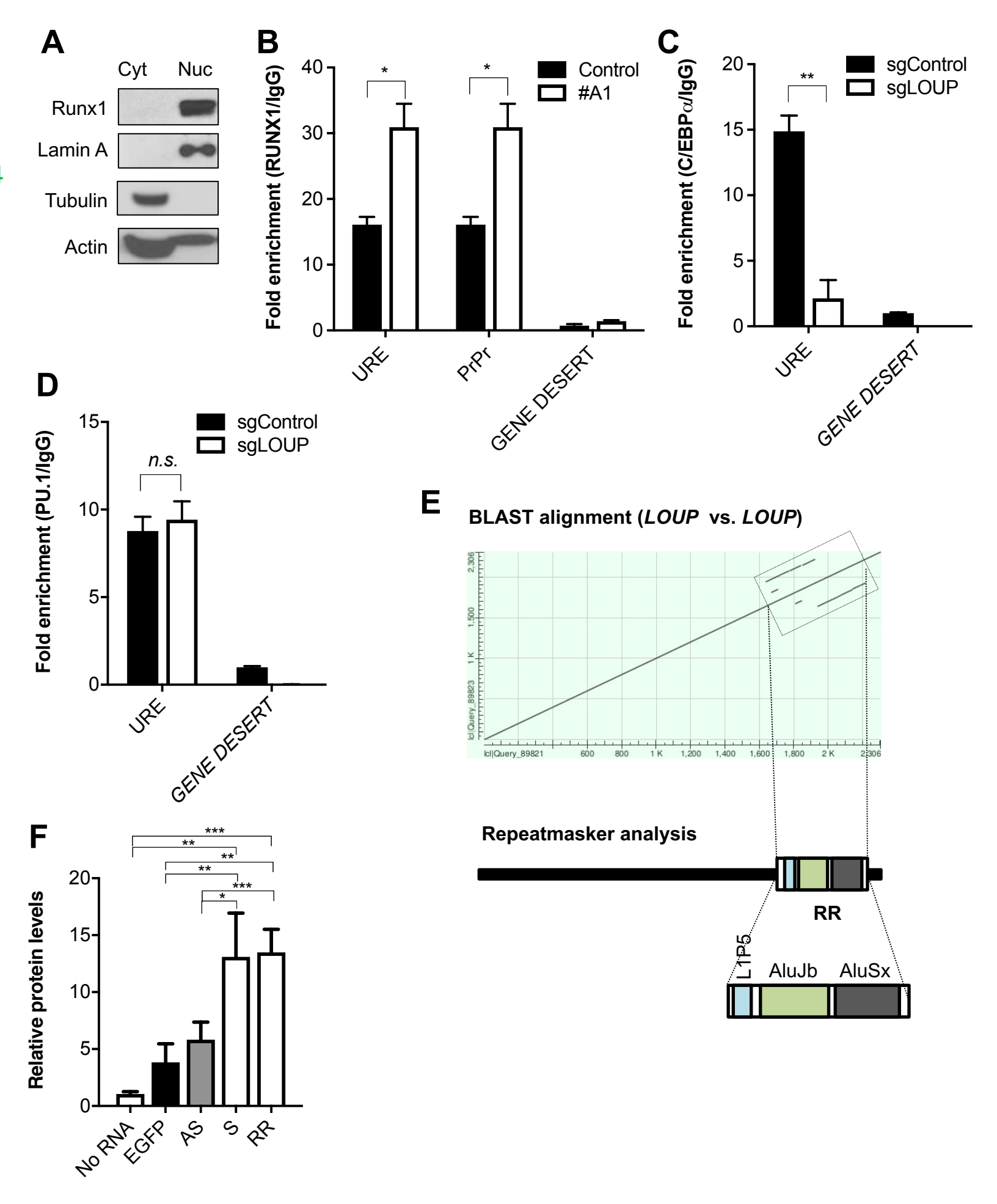
(F and G) qRT-PCR expression analysis for *LOUP* (left panels) and *PU.1* (right panels). F) K562 cells transiently transfected with *LOUP* cDNA or empty vector (EV) by electroporation. G) Kasumi-1 cells stably transfected with *LOUP* cDNA or empty vector (EV) by lentiviral transduction.

(H) Edu incorporation was measured by flow cytometry for cell proliferation. U937 sgRNA cell clones with *LOUP* indels (L2a and L2b) and controls (N1 and N2).

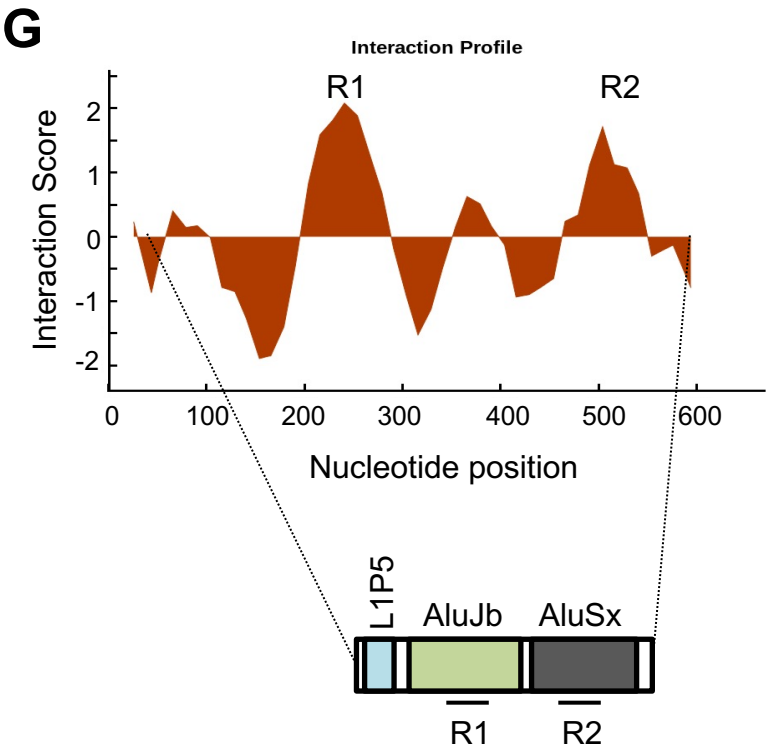
(I and J) Representative flow cytometry results of CD11b myeloid marker expression. I) U937 sgRNA cell clones with *LOUP* indels (L2a and L2b) and controls (N1 and N2). J) Kasumi-1 cells stably transfected with *LOUP* cDNA or empty vector (EV) by lentiviral transduction.

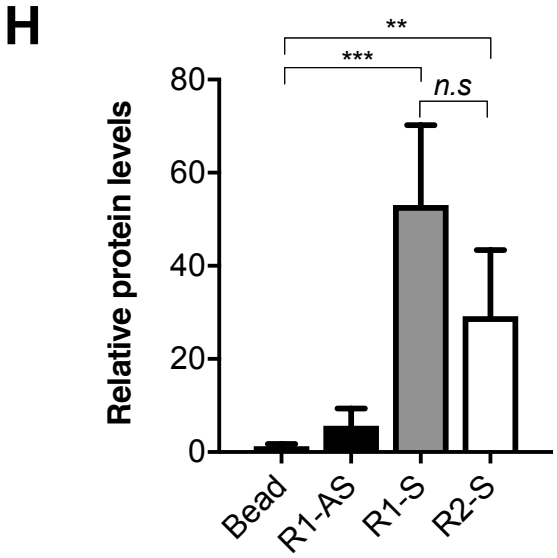
(K) qRT-PCR expression analysis for myeloid marker *CD11b*, macrophage marker *CD14*, and T-cell marker *CD3* after knocking down *LOUP* in cord blood $CD34^+$ HSPC cells following myeloid induction with cytokines.

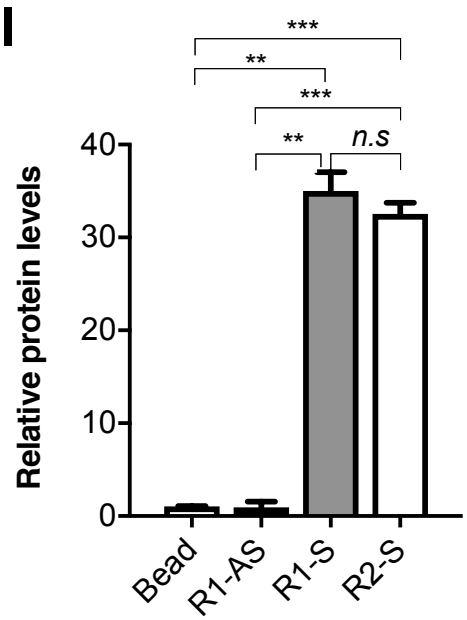
Error bars indicated SD (n=3), $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$.


A
Cyt
Nuc
Runx1
Lamin A
Tubulin
Actin
B
Fold enrichment (RUNX1/IgG)
Control
#A1
URE
PrPr
GENE DESERT
C
Fold enrichment (C/EBPα/IgG)
sgControl
sgLOUP
URE
GENE DESERT
D
Fold enrichment (PU.1/IgG)
sgControl
sgLOUP
n.s.
URE
GENE DESERT
E
BLAST alignment (LOUP vs. LOUP)
Repeatmasker analysis
RR
L1P5
AluJb
AluSx
F
Relative protein levels
No RNA
EGFP
AS
S
RR

FigS5


G
Interaction Profile
R1
R2
Interaction Score
2
1
0
-1
-2
0
100
200
300
400
500
600
Nucleotide position
L1P5
AluJb
AluSx
R1
R2


H
**
***
n.s
Relative protein levels
80
60
40
20
0
Bead
R1-AS
R1-S
R2-S


I
***
**
***
**
n.s
Relative protein levels
40
30
20
10
0
Bead
R1-AS
R1-S
R2-S

FigS5

Figure S5. *LOUP* interacts with RUNX1, Related to Figure 6

(A) Immunoblot of RUNX1 and control proteins in nuclear and cytosol fractions from U937 cells.

(B) ChIP-qPCR analysis of RUNX1 occupancy at the URE and the PrPr. K562 dCas9-VP64-stable cells infected with *LOUP*-targeting (#A1) or non-targeting (control) sgRNAs. PCR amplicons include the URE (contains known RUNX1-binding motif at the URE), PrPr (contains putative RUNX1-binding motif in the PrPr) and GENE DESERT (a genome region that is devoid of protein-coding genes).

(C and D) ChIP-qPCR analysis for occupancy of C/EBPα and PU.1 at the URE. *LOUP*-depleted U937 (sgLOUP, L2a) and control (sgControl, N1) clones were used. PCR amplicons include the URE (contains known RUNX1-binding motif at the URE) and GENE DESERT (a genome region that is devoid of protein-coding genes).

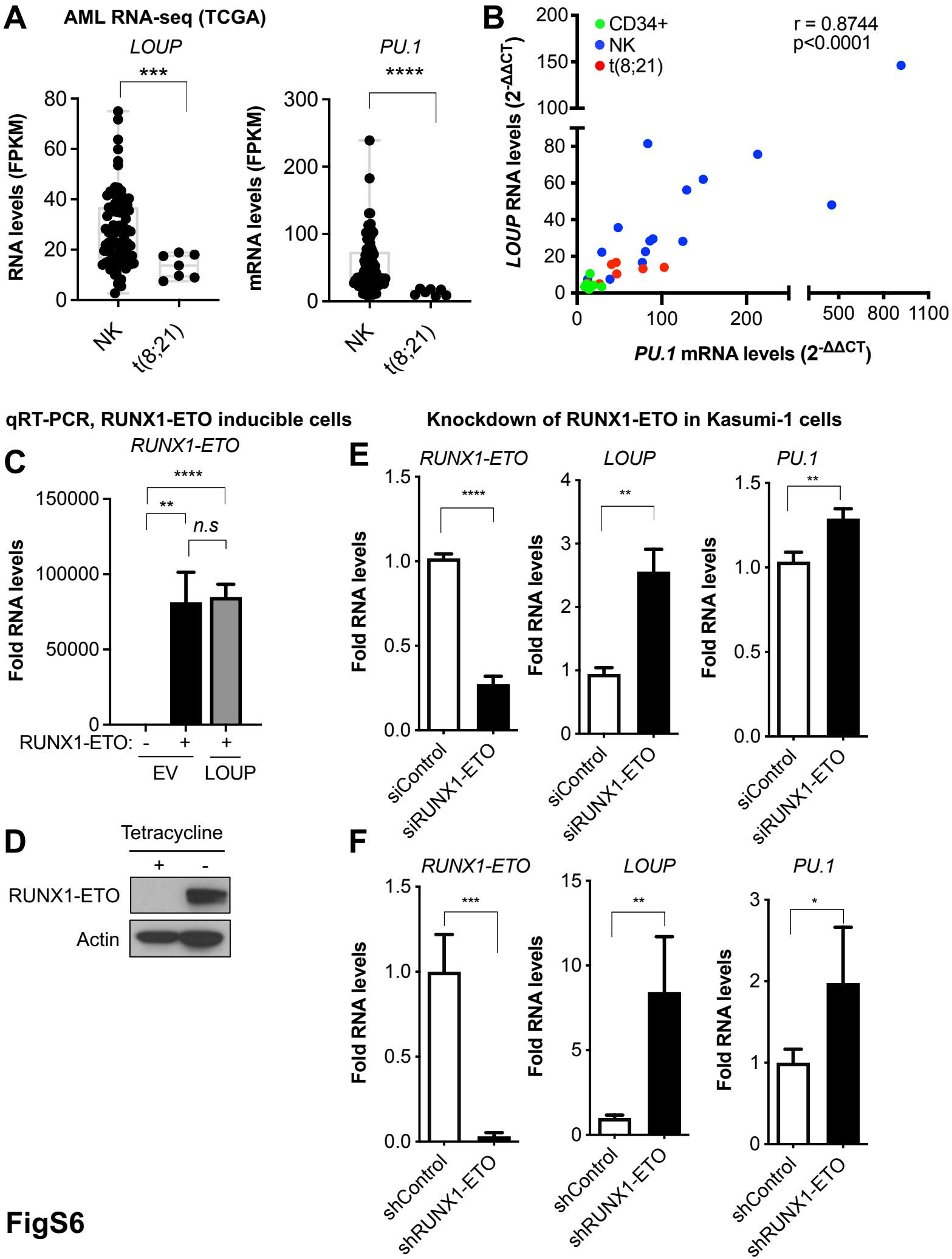
(E) Nucleotide identity plot generated from alignment of *LOUP* to itself using discontinuous megablast algorithm from BLAST (http://blast.ncbi.nlm.nih.gov/). Boxed area depicts a repetitive region (RR) of 670 bp. The schematic diagram underneath illustrates mature *LOUP* lncRNA containing the RR region

(F) Quantitation of RUNX1-*LOUP* interaction. Intensities of immunoblot bands from RNA pull-down analysis (Bead: no RNA control; EGFP: *EGFP* mRNA control; AS: full-length antisense control; S: full-length sense, and RR: repetitive region) were analyzed by ImageJ software. Data are shown relative to Bead (no RNA control samples). Error bars indicated SD (n=3).

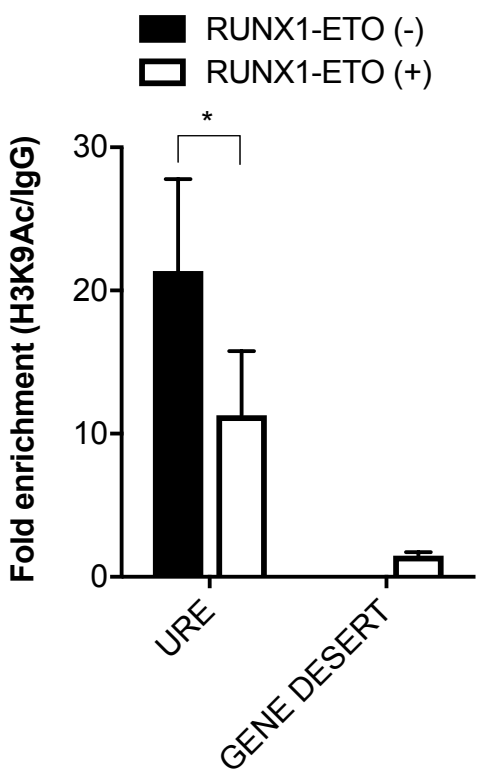
(G) *In silico* prediction of RR-RUNX1 interaction by catRAPID Fragments algorithm. R1 and R2: two regions with high interaction scores.

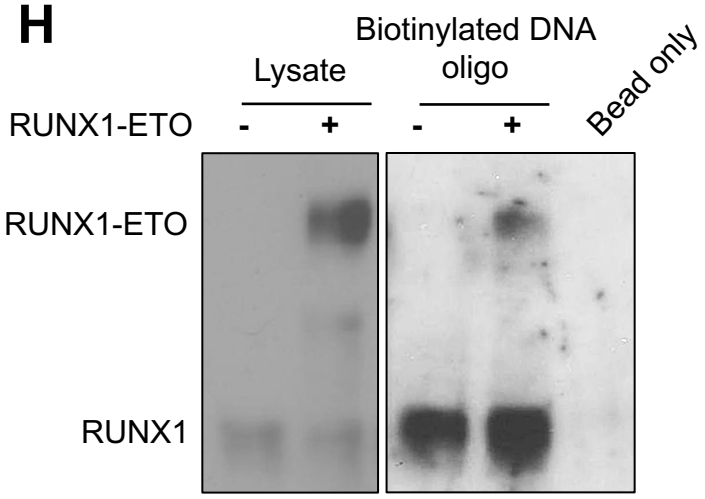
(H and I) Quantitation for RNAP binding analysis of R1 and R2 with recombinant full-length and Runt domain of RUNX1. Intensities of immunoblot bands from RNA pull-down analysis (Bead: no RNA control; R1-AS (R1 antisense control); R1-S (R1 sense); and R2-S (R2 sense)) were analyzed by ImageJ software. Data are shown relative to Bead (no RNA control samples).

Error bars indicated SD (n=3), *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$, *n.s*: not significant.


A
AML RNA-seq (TCGA)
LOUP
PU.1
***
****
RNA levels (FPKM)
mRNA levels (FPKM)
NK
t(8;21)
B
CD34+
NK
t(8;21)
r = 0.8744
p<0.0001
LOUP RNA levels (2-ΔΔCT)
PU.1 mRNA levels (2-ΔΔCT)
qRT-PCR, RUNX1-ETO inducible cells
C
RUNX1-ETO
n.s
Fold RNA levels
RUNX1-ETO: - + +
EV
LOUP
D
Tetracycline
RUNX1-ETO
Actin
Knockdown of RUNX1-ETO in Kasumi-1 cells
E
RUNX1-ETO
LOUP
PU.1
siControl
siRUNX1-ETO
F
shControl
shRUNX1-ETO

FigS6


G
RUNX1-ETO (-)
RUNX1-ETO (+)
*
Fold enrichment (H3K9Ac/IgG)
30
20
10
0
URE
GENE DESERT


H
Biotinylated DNA
Lysate
oligo
Bead only
RUNX1-ETO
-
+
-
+
RUNX1-ETO
RUNX1

FigS6

Figure S6. RUNX1-ETO inhibits *LOUP* expression, Related to Figure 7

(A) Transcript count for *LOUP* levels in AML patient samples (RNA-seq data was retrieved from TCGA portal. NK: normal karyotype (n=87), t(8;21): t(8;21) karyotype (n=7); Mann-Whitney U test: ***p < 0.001, ****p < 0.0001.

(B) Scatter plot showing the correlation between *LOUP* and *PU.1* RNA levels measured by qRT-PCR. Each dot represents a human sample. $CD34^+$: healthy subjects (green dots, n=7), NK: AML patients with normal karyotype (blue dots, n=14), t(8;21): AML patients with t(8;21) karyotype (red dots, n=7). Spearman's rank correlation coefficient (r) was computed.

(C) qRT-PCR expression analysis of *RUNX1-ETO* mRNA in tetracycline-inducible RUNX1-ETO-expressing (Tet-Off) U937 cells (RUNX1-ETO +/-: with/without induction). Cells were transfected with empty vector (EV) and *LOUP* cDNA. Error bars indicate SD (n=3), **p < 0.01; ****p < 0. 0001; *n.s*: not significant.

(D) Immunoblot of RUNX1-ETO and actin control in tetracycline-inducible RUNX1-ETO-expressing (Tet-Off) U937 cells 48 h following tetracycline withdrawal.

(E and F) qRT-PCR expression analysis of *RUNX1-ETO* (left panel), *LOUP* RNA (middle panel) and *PU.1* mRNA (right panel). E) Kasumi-1 cells transfected with mismatch siRNA (siControl) and RUNX1-ETO targeting siRNA (siRUNX1-ETO). Error bars indicate SD (n=3), **p < 0.01, ****p < 0.0001. F) Kasumi-1 cells transfected with Renilla-targeting shRNA (shControl) and RUNX1-ETO targeting shRNA (shRUNX1-ETO). Error bars indicate SD (n=4), *p < 0.05, **p < 0.01, ***p < 0.001.

(G) ChIP-qPCR analysis of H3K9Ac occupancy at the URE in U937 cells with inducible expression of RUNX1-ETO (+/-: with/without induction). PCR amplicons include the URE (contains known RUNX1-binding motif at the URE) and GENE DESERT (a genome region that is devoid of protein-coding genes). Error bars indicate SD (n=3), *p < 0.05.

(H) DNA pull-down assay showing binding of RUNX1-ETO to the RUNX1-binding motifs at the URE. Proteins captured by biotinylated DNA oligo containing RUNX1-binding motif in U937 nuclear lysate with inducible expression of RUNX1-ETO (+/-: with/without induction) were detected by immunoblot.

SUPPLEMENTAL TABLE LEGENDS

Table S1. List of myeloid genes, Related to Figure 1. 78 myeloid genes defined by their known roles in myeloid development or myeloid molecular markers.

Table S2. List of myeloid genes displaying both RUNX1-RIP and RUNX1-ChIP peaks, Related to Figure 1. 15 myeloid genes displaying both RUNX1-RIP and RUNX1-ChIP peaks defined by their known roles in myeloid development or myeloid molecular markers.

Table S3. Table of oligonucleotide information. Primers, probes, and sgRNAs for various assays used in this article.

Table S4. List of enriched genes in *LOUP*$^{+}$/*PU.1*$^{+}$ cells, Related to Figure 3.

Table S5. List of R1 with catRapid scores in RIP-seq data, Related to Figure 3.

Table S6. List of R2 with catRapid scores in RIP-seq data, Related to Figure 3.

QUANTITATION AND STATISTICAL ANALYSIS

In general, quantitation and statistical tests were performed using GraphPad Prism 8.0 software. Data are shown as mean ± SD. The paired two-tailed Student's t-test (otherwise specified in respective figure legends) was used to calculate statistical significance of differences between two experimental groups. $p \leq 0.05$ was considered statistically significant.

DATA SETS

Data are available on the Gene Expression Omnibus database under GEO Series accession number GEO: GSE140459.

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and request for resources and reagents should be directed to and will be fulfilled by the Lead Contacts, Daniel G. Tenen (daniel.tenen@nus.edu.sg) and Bon Q. Trinh (btrinh@bidmc.harvard.edu)

SUPPLEMENTAL REFERENCES

1. Trinh BQ, Barengo N, Kim SB, Lee JS, Zweidler-McKay PA, Naora H. The homeobox gene DLX4 regulates erythro-megakaryocytic differentiation by stimulating IL-1beta and NF-kappaB signaling. *J Cell Sci*. 2015;128(16):3055-3067.
2. Fellmann C, Hoffmann T, Sridhar V, et al. An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep*. 2013;5(6):1704-1713.
3. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816-821.
4. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31(3):233-239.
5. Aubrey BJ, Kelly GL, Kueh AJ, et al. An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. *Cell Rep*. 2015;10(8):1422-1432.
6. Park J, Bae S, Kim JS. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics*. 2015;31(24):4014-4016.
7. Li Y, Okuno Y, Zhang P, et al. Regulation of the PU.1 gene by distal elements. *Blood*. 2001;98(10):2958-2965.
8. Hsiau TTM, Kelsey Waite, Joyce Yang, Reed Kelso, Kevin Holden, Rich Stoner. Inference of CRISPR Edits from Sanger Trace Data
. *BioRxiv*. 2018.
9. Wang L, Gural A, Sun XJ, et al. The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science*. 2011;333(6043):765-769.
10. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308.
11. Konermann S, Brigham MD, Trevino AE, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015;517(7536):583-588.
12. Burel SA, Harakawa N, Zhou L, Pabst T, Tenen DG, Zhang DE. Dichotomy of AML1-ETO functions: growth arrest versus block of differentiation. *Mol Cell Biol*. 2001;21(16):5577-5590.
13. Pabst T, Mueller BU, Harakawa N, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat Med*. 2001;7(4):444-451.
14. Lee S, Kopp F, Chang TC, et al. Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell*. 2016;164(1-2):69-80.
15. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402-408.
16. Zhang H, Alberich-Jorda M, Amabile G, et al. Sox4 is a key oncogenic target in C/EBPalpha mutant acute myeloid leukemia. *Cancer Cell*. 2013;24(5):575-588.
17. Zhang P, Iwasaki-Arai J, Iwasaki H, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity*. 2004;21(6):853-863.
18. Mueller BU, Pabst T, Fos J, et al. ATRA resolves the differentiation block in t(15;17) acute myeloid leukemia by restoring PU.1 expression. *Blood*. 2006;107(8):3330-3338.
19. Melo CA, Drost J, Wijchers PJ, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell*. 2013;49(3):524-535.
20. Hagege H, Klous P, Braem C, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*. 2007;2(7):1722-1733.
21. Staber PB, Zhang P, Ye M, et al. Sustained PU.1 levels balance cell-cycle regulators to prevent exhaustion of adult hematopoietic stem cells. *Mol Cell*. 2013;49(5):934-946.
22. Deng W, Blobel GA. Detecting Long-Range Enhancer-Promoter Interactions by Quantitative Chromosome Conformation Capture. *Methods Mol Biol*. 2017;1468:51-62.
23. Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). *J Vis Exp*. 2012(61).
24. Trimarchi T, Bilal E, Ntziachristos P, et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell*. 2014;158(3):593-606.

558 25. Trinh BQ, Barengo N, Naora H. Homeodomain protein DLX4 counteracts key transcriptional control
559 mechanisms of the TGF-beta cytostatic program and blocks the antiproliferative effect of TGF-beta.
560 *Oncogene*. 2011;30(24):2718-2729.
561 26. Tsai MC, Manor O, Wan Y, et al. Long noncoding RNA as modular scaffold of histone modification
562 complexes. *Science*. 2010;329(5992):689-693.
563 27. Bellucci M, Agostini F, Masin M, Tartaglia GG. Predicting protein associations with long noncoding
564 RNAs. *Nat Methods*. 2011;8(6):444-445.
565 28. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
566 *Bioinformatics*. 2010;26(6):841-842.
567 29. Hendrickson GD, Kelley DR, Tenen D, Bernstein B, Rinn JL. Widespread RNA binding by
568 chromatin-associated proteins. *Genome Biol*. 2016;17:28.
569 30. Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and
570 lineage-committed cells. *Nature*. 2007;448(7153):553-560.
571 31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
572 *Bioinformatics*. 2014;30(15):2114-2120.
573 32. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*.
574 2013;29(1):15-21.
575 33. Ramirez F, Ryan DP, Gruning B, et al. deepTools2: a next generation web server for deep-
576 sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-165.
577 34. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors
578 prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-
579 589.
580 35. Hunt SE, McLaren W, Gil L, et al. Ensembl variation resources. *Database (Oxford)*. 2018;2018.
581 36. Andrews S. FastQC: a Quality Control Tool for High Throughput Sequence Data. Babraham
582 Bioinformatics version 0115; 2016.
583 37. Krueger F. Trim Galore: A wrapper script to automate quality and adapter trimming as well as
584 quality control. 045. Babraham Bioinformatics
585 ; 2017.
586 38. Zheng R, Wan C, Mei S, et al. Cistrome Data Browser: expanded datasets and new tools for gene
587 regulatory analysis. *Nucleic Acids Res*. 2019;47(D1):D729-D735.
588 39. de Rie D, Abugessaisa I, Alam T, et al. An integrated expression atlas of miRNAs and their
589 promoters in human and mouse. *Nat Biotechnol*. 2017;35(9):872-878.
590 40. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware
591 quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419.
592 41. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-
593 seq data with Bioconductor. *F1000Res*. 2016;5:2122.
594 42. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-
595 Cell Transcriptomic Data. *Cell Syst*. 2019;8(4):281-291 e289.
596 43. Lun AT, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data
597 with many zero counts. *Genome Biol*. 2016;17:75.
598 44. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression
599 analysis. *Nat Methods*. 2014;11(7):740-742.
600 45. Weinreb C, Wolock S, Klein AM. SPRING: a kinetic interface for visualizing high dimensional single-
601 cell expression data. *Bioinformatics*. 2018;34(7):1246-1248.
602 46. Martens JH, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes.
603 *Haematologica*. 2013;98(10):1487-1489.
604 47. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*.
605 2012;489(7414):57-74.
606 48. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a
607 transitional profibrotic macrophage. *Nat Immunol*. 2019;20(2):163-172.
608 49. Dennis G, Jr., Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and
609 Integrated Discovery. *Genome Biol*. 2003;4(5):P3.

610 50. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein
611 coding and non-coding regions. *Bioinformatics*. 2011;27(13):i275-282.
612 51. West JA, Davis CP, Sunwoo H, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active
613 chromatin sites. *Mol Cell*. 2014;55(5):791-802.

614