Supplemental methods

Lentiviral vectors

Individual single guide RNAs (sgRNAs) were designed using CCTop[1] (https://cctop.cos.uni-heidelberg.de/), and cloned via BsmBI into the SGL40C.EFS.dTomato (Addgene 89395) or SGL40C.EFS.E2Crimson (100894) backbone. Dual sgRNA vectors were generated by inserting a second promoter-sgRNA cassette into an existing sgRNA vector via EcoRI/XhoI. Short hairpin RNAs (shRNAs) were designed using the Adams *et al.* miR-N tool[2] (https://felixfadams.shinyapps.io/miRN/) and cloned via BsmBI into the SIN40C.SFFV.eGFP.miR30n (169278) backbone. Non-targeting sgRNAs and shRNAs were designed against firefly luciferase. *MYNRL15* cDNAs were expressed from the bidirectional LBid.lnc.GFP[3,4] vector. The L40C-CRISPR.EFS.mNeon (170483) all-in-one system was used on primary cells for *in vitro* assays. Stable cell lines were generated using either pLKO5d.SFFV.dCas9-KRAB.P2A.BSD (90332) or pLKO5d.EFS.SpCas9.P2A.BSD (57821). Stable patient-derived xenografts were generated using SIN40C.SFFV.dCas9-KRAB.P2A.mNeon (170482). The sgRNA libraries used in this study were expressed from the following backbones: SGL40C.EFS.dTomato (89395; CRISPRi lncRNA and *MYNRL15* tiling), SGL.EFS.tBFP (173915; gained chromatin interaction region), and SGL.EFS.dTomato.P2A.PAC (173914; CTCF-enriched lncRNA loci).

Lentiviruses

Lentiviral particles were produced by co-transfecting the expression vector and the packaging plasmids pMD2.G and psPAX2 (Addgene 12259 and 12260 respectively) into HEK293T cells using polyethylenimine (PEI). Viral particles were concentrated via ultracentrifugation, and in the case of all-in-one constructs, were further concentrated using Lenti-X™ Concentrator reagent (TaKaRa). Transductions were performed in normal cell culture media, in the presence of Polybrene (Sigma-Aldrich).

LNA-GapmeRs

Custom- antisense LNA-GapmeRs targeting the *MYNRL15* transcript were obtained from Qiagen through their in-house design tool. Negative control B (Qiagen 339515) was used as a non-targeting control. Cells were cultured in media containing 2.5 µM LNA-GapmeR for delivery by unassisted uptake[5]. Fresh LNA-GapmeR was added every 2 days to maintain its concentration in the culture media. LNA-GapmeR sequences can be found in Supplemental Table 7.

CRISPR library design

Guides for the CRISPRi-based targeting of HSPC/AML lncRNAs were designed using the standalone version of CCTop[1] (https://cctop.cos.uni-heidelberg.de/). In brief, the lncRNA genes were annotated using GENCODE v25 (release 03/2016)[6], LNCipedia 4.0 (release 05/2016)[7], and NONCODE v4 (release 01/2014)[8] as previously described[9], and sgRNAs were selected 0 bp to 250 bp downstream of transcription start sites (TSSs)[10]. Three to nine sgRNAs were selected per gene, depending on the number of different TSSs present in the transcript isoforms and the distance between them. Genes with a single TSS, or with multiple TSSs with high transcript-level support (TSL 1 or 2, according to Ensembl annotations) spaced more than 300 bp apart, were targeted using three sgRNAs per TSS in a 0-150 bp window downstream of the respective TSS. Genes with multiple TSSs in close proximity to each other (spaced ≤150 bp apart) were targeted using five sgRNAs in a 0-250 bp window downstream of the first TSS. Guides were prioritized for low off-target binding – a criterion that is built-in to the CCTop tool.

Guides tiling the *MYNRL15* locus were designed by inputting 15 kb of DNA sequence (hg38) symmetrically centered on *MYNRL15* into the CRISPOR[11] (http://crispor.tefor.net/) saturating mutagenesis assistant. To maintain dense tiling of the region (mean coverage: 0.11 sgRNAs per bp), only guides with an MIT specificity score of 0 were excluded.

Guides targeting the 29 protein-coding genes located in the gained distal chromatin interaction region were designed using CCTop[1] (https://cctop.cos.uni-heidelberg.de/). Coding sequences (CDS) from Ensembl v102 (release 11/2020) were used as inputs, and where possible, sgRNAs were selected to target most, if not all, protein-coding isoforms. Guides were prioritized for low off-target binding, and those with low predicted on-target efficacies (CRISPRater[12] score<0.4) were excluded.

Guides targeting CTCF sites in CTCF-enriched lncRNA loci were selected from GuideScan[13] (http://www.guidescan.com/) and CRISPick[14,15] (formerly the Broad GPP sgRNA design tool; https://portals.broadinstitute.org/gppx/crispick/public). CTCF binding sites were determined using ENCODE ChIP-seq peak calling data, and sgRNAs were selected to tile CTCF motifs and/or point-source(s) within the peaks. If both features were within 50 bp of each other, the target region was defined as a 150 bp region centered on the midpoint between the two. If the CTCF motif and point-source were within 100 bp of each other, a 300 bp target region was used. Otherwise, two 80 bp target regions were used for sgRNA selection, centered on the motif and point-source, respectively. Guides located in these target regions were first selected from GuideScan, then topped up from CRISPick in cases where a coverage of 0.15 sgRNAs per bp was not met.

Due to our usage of SGL40C vectors for lentiviral sgRNA delivery, in which sgRNA transcription is driven from a human U6 promoter, guides containing polyT stretches (4 or more) were excluded from all libraries, to avoid premature termination of sgRNA transcription mediated by RNA polymerase III. Guides directed against luciferase and the neomycin resistance cassette were used as non-targeting controls; guides targeting *PPP1R12C* and *SLC22A13* were used as nonessential cutting controls; guides against *MYC*, *MYB*, *ACTB*, *U2AF1*, *RPL9*, and *POL2RA* were used as positive depletion controls. The sgRNA spacer sequences of the four CRISPR libraries used in this study are provided in Supplemental Tables 1-3 and 5.

CRISPR library cloning and screening

Library spacer sequences were purchased from Integrated DNA Technologies, pooled, and cloned via BsmBI into one of the following vectors: SGL40C.EFS.dTomato (Addgene 89395; CRISPRi lncRNA and *MYNRL15* tiling), SGL.EFS.tBFP (173915; gained chromatin interaction region), and SGL.EFS.dTomato.P2A.PAC (173914; CTCF-enriched loci). XL1-Blue supercompetent cells (Agilent 200236) were used for transformation, and subsequently plated on 15 cm LB agar plates containing ampicillin. Colonies were counted from 1 cm^2 areas to ensure sufficient library representation, and then harvested and prepped for plasmid DNA using the QIAGEN Plasmid Maxi Kit. Lentiviral particles were produced as outlined above.

Stable dCas9-KRAB- or Cas9-expressing cell lines were transduced with the sgRNA libraries at an MOI of 0.3, and maintained at 1000-fold representation of the library for 16-18 population doublings. The screens were counted every 2-3 days and split accordingly. Samples were taken at the beginning and end of the screen, to determine differences in sgRNA abundance over time and thereby identify essential genes or regions. Genomic DNA was isolated from these samples via the QIAmp DNA Blood Mini Kit (Qiagen), and the sgRNA cassettes were PCR amplified using the NEBNext® High-Fidelity 2x PCR Master Mix (New England Biolabs) and barcoded primers containing the Illumina P5 and P7 adapter sequences as overhangs. The sgRNA amplicons (~300 bp) were gel purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) and sequenced on an Illumina HiSeq 2000 (50 bp single-end reads).

We applied the MAGeCK (model-based analysis of genome-wide CRISPR-Cas9 knockout)[16] pipeline to process raw reads and call AML dependency genes from the CRISPRi lncRNA, gained chromatin interaction region, and CTCF-enriched loci screens. The *MYNRL15* tiling screens were analysed in R, using DESeq2[17] (Bioconductor) to combine replicates and perform pan-cell line analysis.

Fluorescence-based proliferation assays

Individual proliferation assays were conducted in stable dCas9-KRAB- or Cas9-expressing cell lines for CRISPR/Cas9 experiments, and in wild-type lines for RNAi experiments. Cells were transduced with individual sgRNA or shRNA perturbation constructs at an efficiency of 40-80%, to attain a mixed population allowing for direct competition between transduced and untransduced cells. These cultures were maintained for up to 20 days, during which fluorescence was tracked every 2-3 days via flow cytometry. Depletion curves were generated by normalizing the percentage of fluorescent (i.e. transduced) cells at each time point to both the initial fluorescence (day 0) and the non-targeting control (sgLUC / shLUC) (gating strategy provided in Supplemental Information). For the rescue experiment involving *MYNRL15* cDNAs, sgRNA-expressing cells were transduced a second time with the cDNA expression constructs. The double-positive population was then tracked by flow cytometry.

CRISPR-Cas9 indel and excision validation

PCR-based methods were used to validate CRISPR-Cas9 mediated insertions and deletions (indels; to knock out protein-coding genes) and *MYNRL15* excision via paired sgRNAs. Cells were harvested between days 3 and 5 post-transduction, and genomic DNA was isolated using the Quick-DNA™ Miniprep Kit (Zymo Research). We relied on TIDE[18] (tracking of indels by decomposition) to assess knockout efficiency; thus, we PCR amplified ~700 bp genomic regions centered on the corresponding sgRNA target sites from knockout and control (wild type) samples. The resulting products were subjected to Sanger sequencing, and knockout and wild type sequences were compared in the TIDE online tool (http://shinyapps.datacurators.nl/tide/). To validate *MYNRL15* excision, we performed PCR using primers flanking the region to be excised, thereby allowing us to ascertain deletion based on the size of the PCR product. All PCR primer sequences can be found in Supplemental Table 7.

Flow cytometry and cell sorting

Flow cytometry data were collected on a CytoFLEX B4-R3-V5 or CytoFLEX S V4-B2-Y4-R3 using CytExpert software (Beckman Coulter). Cell sorting was performed on a FACSAria™ II

using FACSDiva™ software, or on a FACSMelody™ using FACSChorus™ software (BD Biosciences). An anti-human CD45 FITC (Beckman Coulter) antibody was used to analyze xenotransplantation experiments. Kaluza 2.1 (Beckman Coulter) or FlowJo™ v10.6 (BD Biosciences) software was used for data analysis.

Quantitative real-time PCR

RNA was isolated from cells using the Quick-RNA™ Microprep or Miniprep Kits (Zymo Research), between days 3 to 5 post-transduction. RNA fractionation was performed as previously described[19], except that we directly lysed the nuclear pellet instead of isolating the nuclear-soluble and chromatin-associated fractions separately. *B2M* and *XIST* were utilized as cytoplasmic and nuclear controls, respectively. The TURBO DNA-free™ Kit (Invitrogen) was used for DNase treatment. Total cDNA was synthesized using the High-Capacity cDNA Reverse Transcription Kit, and gene expression was quantified by real-time PCR using SYBR™ Select Master Mix and gene-specific primers on a StepOnePlus™ Real-Time PCR cycler (all products from Applied Biosystems). *B2M* was used as a housekeeping control. Primers for qRT-PCR can be found in Supplemental Table 7. QuantiTect® primer assays were used to detect *WDR61* and *IMP3* (Qiagen QT00083776 and QT00232330).

Animal experiments

Two-color *in vivo* competition experiments were performed in murine xenograft models of AML as previously described[3,20]. In brief, stable dCas9-KRAB cell lines or *in vivo* expanded patient-derived AML cells (PDXs) were transduced with E2Crimson or dTomato sgRNA vectors, mixed 1:1, and injected via tail vein into irradiated (2.5 Gy), 8-10 week old NOD.Cg-Prkdc^{scid} Il2rgtm^{1Wjl}/SzJ (NSG) recipients. One to two million cells were injected per mouse, and tracked via flow cytometry on peripheral blood samples every 4 weeks. The mice were sacrificed upon leukemia onset, at which point cells were isolated from the bone marrow, spleen, and liver, and analyzed by flow cytometry. An anti-human CD45 antibody was used to track AML cell lines, due to the absence of a fluorescent reporter in the dCas9-KRAB construct

(Addgene 90332). The dCas9-KRAB PDXs were generated using a mNeon-containing construct (170482); thus, fluorescence-based tracking was sufficient. In both cases, the E2Crimson and dTomato cell populations were compared to determine relative proliferation in direct competition. All mice were housed under a 12 hour light / 12 hour dark cycle in a pathogen-free environment at the Martin Luther University Halle-Wittenberg. All animal procedures were approved by the local state authorities (Landesverwaltungsamt Sachsen-Anhalt).

RNA sequencing

RNA was isolated from cells using the Quick-RNA™ Miniprep Kit (Zymo Research) on days 3 and 6 or 7 post-transduction (for ML-2 and K562, respectively; the late time point was selected based on depletion kinetics, see Fig. 2c). PolyA-enriched total cellular RNA sequencing was performed by Novogene Company, Ltd. on an Illumina NovaSeq 6000 using 150 bp paired-end chemistry. The raw sequence data were processed by Novogene using a standard pipeline. In brief, reads were filtered using in-house scripts and mapped to human reference genome hg38 using HISAT2[21], and gene expression was quantified using the featureCounts[22] function in R. Differential expression analysis was conducted in R using DESeq2[17] (Bioconductor). Gene sets from MSigDB v7.2 (H1, C2, C3, C6), custom hematopoietic[9] and chromosome 15 gene sets, and PAF1c-knockout expression signatures[23] were checked for enrichment in the Broad GSEA software[24]. Custom positional gene sets were generated by walking a 1 Mb or 5 Mb window along chromosome 15. Gene ontology analysis was done using the DAVID[25] functional annotation tool (https://david.ncifcrf.gov/summary.jsp).

NG Capture-C

Chromatin conformation capture with selective enrichment for *MYNRL15*-interacting sequences was performed using next generation (NG) Capture-C as previously described[26], with the following modifications: (1) 5-10 million cells were harvested per sample and the DpnII digestion reactions were scaled down accordingly. (2) DNA was sheared to 200 bp fragments using a Branson 450 Digital Sonifier (Marshall Scientific) (time 18 s, amplitude 20%, pulse 0.5 s, pause 1.5 s; repeat 5x). (3) All material from the first capture was used as

input for the second capture. (4) The libraries were sequenced by Novogene Company, Ltd. on an Illumina NovaSeq 6000 (150 bp paired-end reads).

Biotinylated oligonucleotides for library capture were designed using CapSequm2[27] (refer to Supplemental Table 7 for sequences) and purchased from Integrated DNA Technologies. These probes capture a viewpoint corresponding to the candidate *cis*-regulatory region C1. Two biological replicates were prepared per sample and pooled for oligonucleotide capture (multliplexed library capture[26]). The raw sequence data were processed with the capC-MAP package[28] using default settings. Normalized pileups (RPMs; binstep=3000, window=6000) were capped at the 99th percentile and scaled to the highest signal within the sample, such that cross-sample comparisons could be made on a 0 to 1 scale. The tracks were viewed in the UCSC Genome Browser[29] using a smoothing window of 2 pixels, alongside CTCF ChIP-seq data from K562 cells (ENCODE accession no. ENCFF519CXF) and Knight-Ruiz matrix-balanced[30] Micro-C[31] data from H1-hESC cells. Hi-C data from Rao *et al*.[32] were also used to confirm the presence of specific 3D chromatin structures in other cell lines.

ATAC-seq

We performed assay for transposase accessible chromatin sequencing (ATAC-seq) as previously described[33,34]. On day 3 post-transduction, 50,000 cells were sorted and processed using the Illumina Tagment DNA Enzyme and Buffer Kit (20034197). The resulting libraries were sequenced by Novogene Company, Ltd. on an Illumina NovaSeq 6000 (150 bp paired-end reads). The data processing was also performed by Novogene: In brief, raw reads were trimmed and filtered using Skewer[35] and clean reads were aligned to hg19 with BWA[36]. Mitochondrial reads were removed prior to subsequent analysis. Normalized pileups were generated using deepTools[37] and viewed in the Integrated Genomics Viewer (IGV)[38].

CTCF CUT&RUN

We performed CUT&RUN as previously described[39,40]. On day 3 post-transduction, 400,000 cells were sorted and incubated with the following antibodies: rabbit anti-human CTCF (1:50;

Diagenode C15410210), and rabbit IgG (Diagenode C15410206). The pAG/MNase nuclease (Addgene 123461) was produced and purified as previously described[40], after removal of the HA tag. Illumina libraries were constructed from cleaved DNA and sequenced by Novogene Company, Ltd. on an Illumina NovaSeq 6000 (150 bp paired-end reads). For processing the raw data, we used Trimmomatic[41] to remove adapter sequences, followed by Kseq[42] to trim reads containing ≤6 bp of adapter sequence, which are not effectively handled by Trimmomatic. Trimmed reads were aligned to hg38 using bowtie2[43]. The resulting SAM files were converted into BAM format and sorted and indexed using Samtools[44]. Normalized bigWig tracks were generated using bamCoverage from deepTools[37]. The processed data were viewed in the Integrated Genomics Viewer (IGV)[38].

Dual luciferase assays

Dual luciferase assays were performed using the Dual-Luciferase® Reporter Assay System (Promega). The candidate *cis*-regulatory regions C1 and C2 were cloned alone or in combination upstream of the minimal promoter in the pGL4.23 firefly luciferase reporter construct (Promega E8411). A pGL4.7 *Renilla* luciferase reporter construct (Promega E6881) driven from the EF1α promoter was used as a background control. The firefly and *Renilla* vectors were co-transfected into K562 cells at a 20:1 ratio via nucleoporation, using the Lonza 4D-Nucleofector™ and SF Cell Line X Kit S. 24 h post-transfection, cells were harvested and measured on a GloMax® 96 Luminometer (Promega).

CTCF-enriched lncRNA loci

To identify CTCF-bound genic loci, we overlapped ENCODE CTCF ChIP-seq peaks with gene annotations from GENCODE v23 (release 07/2015)[6] using the findOverlaps function from the IRanges package in R (Bioconductor). We confirmed the presence of CTCF motifs in the ChIP-seq peaks using GimmeMotifs[45]. CTCF density was determined by counting the number of CTCF binding sites and normalizing by gene length. Log_{10}-transformed values of this metric followed an approximately normal distribution; thus, we defined elevated CTCF density as greater than 2 standard deviations above the median. Our analysis focused on

loci that produce long coding or noncoding transcripts (>200 nt) and included the following biotypes: protein coding, lncRNA, lincRNA, processed transcript, and pseudogene. Refer to Supplemental Table 6 for a catalog of CTCF-enriched lncRNA loci (C-LNC) across 18 different cell lines and primary cell types.

For the analysis of C-LNCs in the context of AML, gene expression values were obtained from the TCGA[46] and NCI-TARGET[47] AML patient cohorts. C-LNCs were deemed clinically significant if 1) stratifying patients based on their expression yielded a significant difference in event-free or overall survival (P<0.05, log-rank test), or 2) their expression significantly differed in cases harboring any of the following genetic abnormalities compared to cases without: complex karyotype, t(8:21), inv(16), PML-RARA or BCR-ABL translocation, *KMT2A* rearrangement, *FLT3*-ITD, or mutations in *CEBPA*, *NPM1*, *DNMT3A*, *TP53*, *cKIT*, or *WT1* (P<0.05, two-sided t-test).

TCGA/TARGET survival analysis

Event-free survival was defined as the time elapsed between diagnosis and the first event or last follow-up. An event was defined as death from any cause, failure to achieve remission, relapse, and secondary malignancy. Failure to achieve remission was considered an event on day 0. Overall survival was defined as time elapsed between diagnosis and death from any cause or last follow-up. We used the Kaplan-Meier method of estimating survival rates and two-sided log-rank tests to compare differences in survival, as implemented in the survival[48] and survminer[49] packages (base R). DESeq2[17] (Bioconductor) was used to normalize and variance-stabilize read count data[17] from the TCGA[46] and NCI-TARGET[47] AML cohorts. The NCI-TARGET dataset also required batch correction, for which we used sva[50] (Bioconductor). Normalized (and batch corrected) gene expression values were used for all subsequent analyses. For patient stratification, optimal cut-offs were determined via maximally selected log-rank statistics as implemented in the maxstat package[51] (base R). Unsupervised clustering was performed using Rtsne[52] (base R).

Supplemental tables

Supplemental Table 1: CRISPRi lncRNA library sgRNA spacer sequences and screening results.

Supplemental Table 2: *MYNRL15* tiling library sgRNA spacer sequences and screening results.

Supplemental Table 3: Capture-C gained region library sgRNA spacer sequences and screening results.

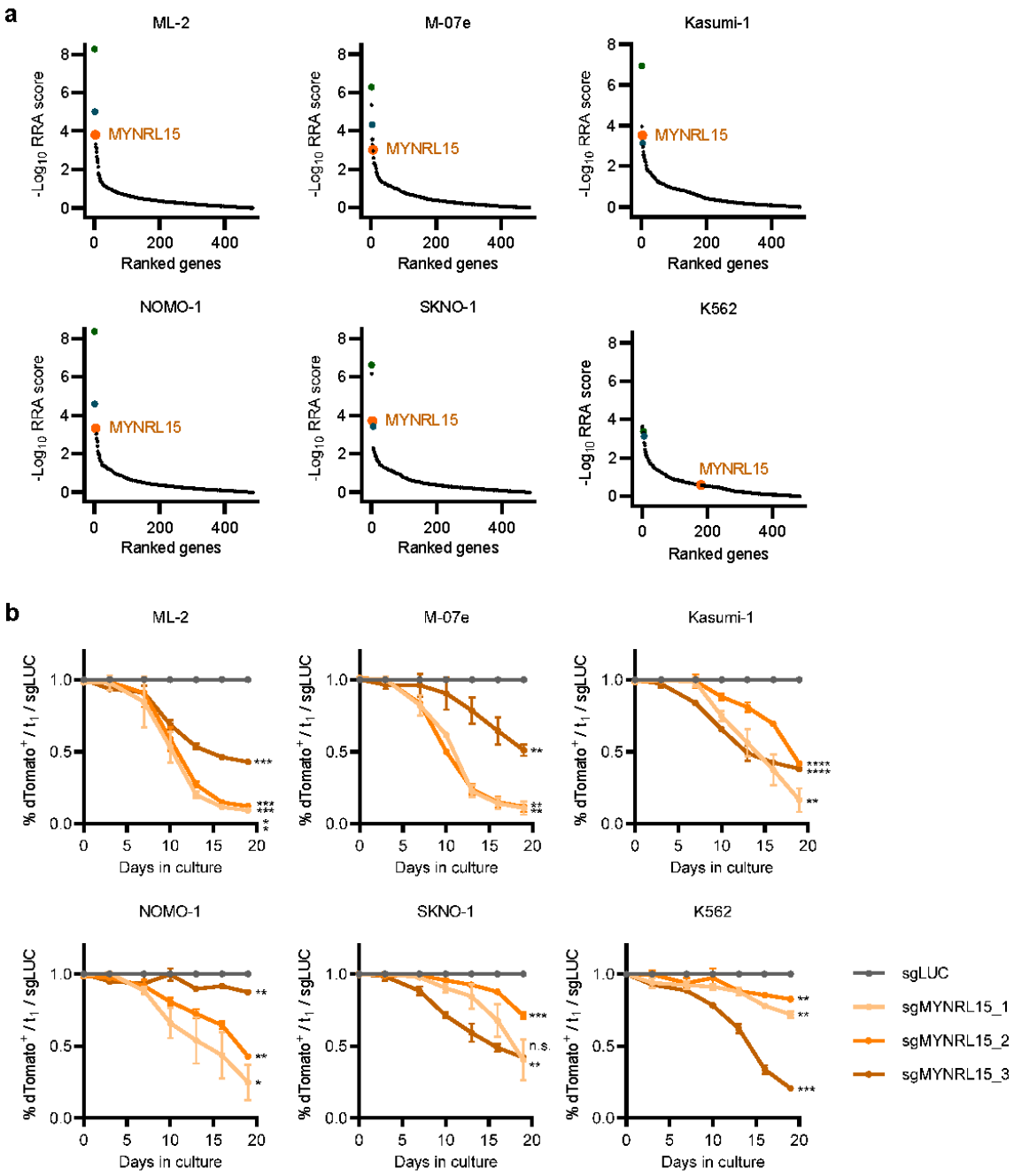
Supplemental Table 4: Patient sample characteristics.

Supplemental Table 5: C-LNC library sgRNA spacer sequences and screening results.

Supplemental Table 6: Catalog of C-LNCs in 18 cell lines and primary cell types.

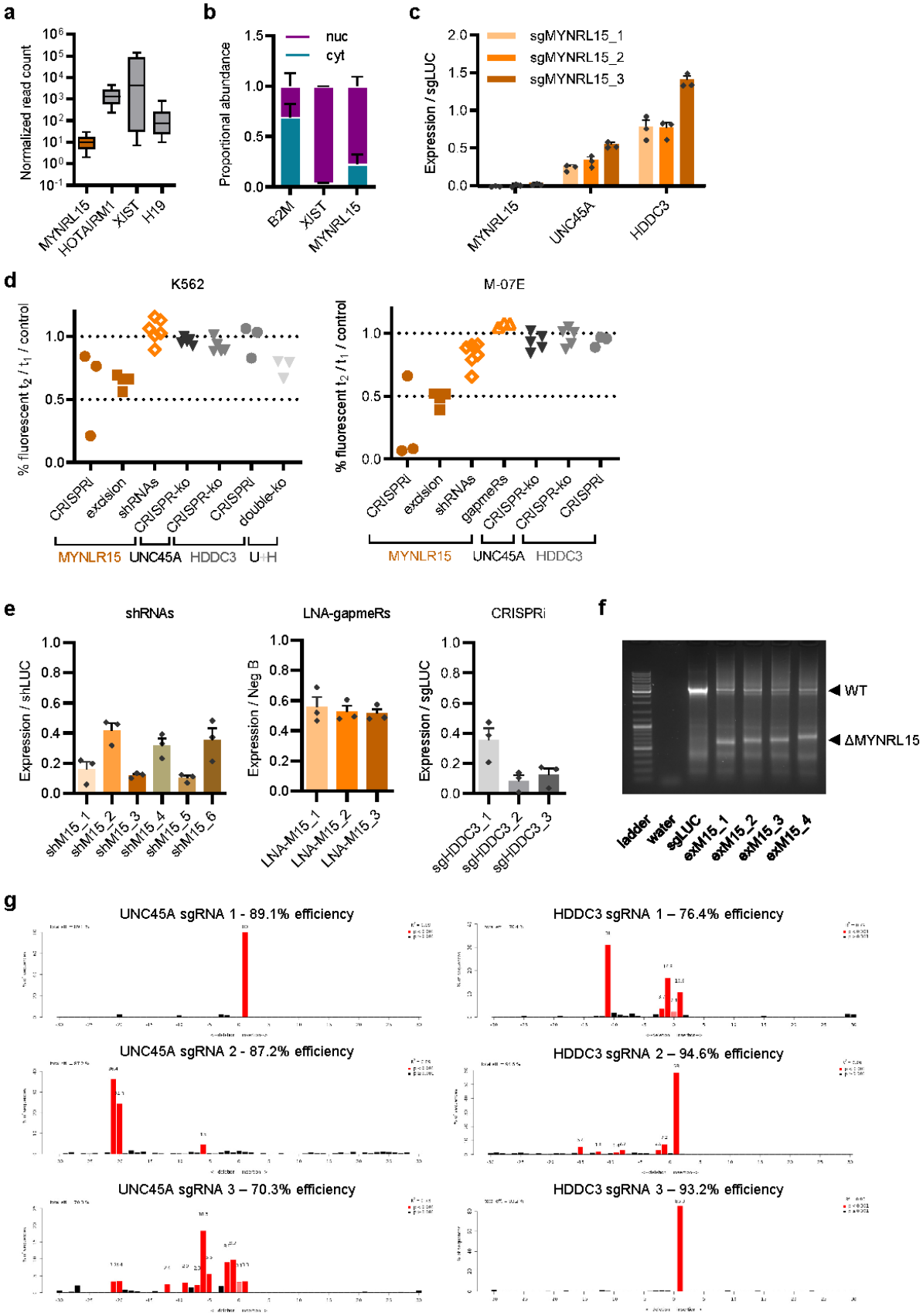
Supplemental Table 7: Sequences of individual sgRNAs, shRNAs, LNA-gapmeRs, Capture-C probes, and (qRT-)PCR primers.

Supplemental figures



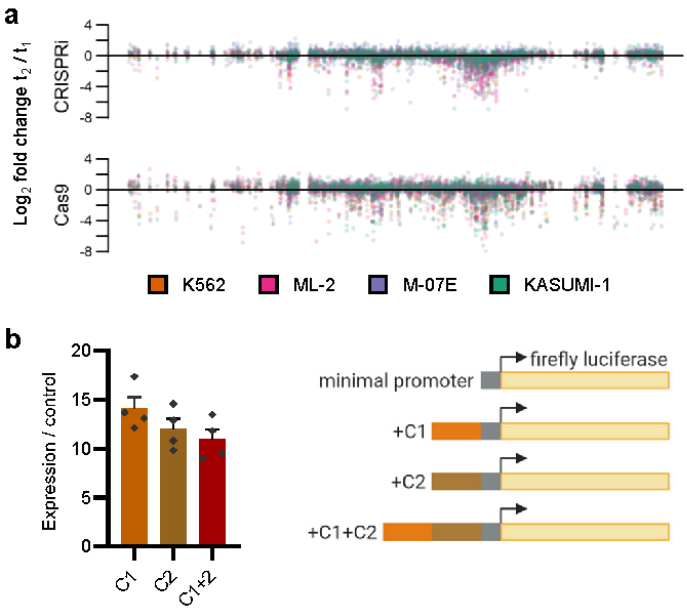
Supplemental Figure 1: Identification and validation of *MYNRL15* dependency in six myeloid leukemia cell lines.

a, MAGeCK essentiality scores from the CRISPRi screen of HSPC/AML lncRNAs. Results from the individual cell lines are shown (n=2 biological replicates each). *MYNRL15* (orange) is highlighted, along with the positive control genes, *MYC* (green) and *MYB* (turquoise). **b**, Fluorescence-based proliferation assays using individual guides targeting *MYNRL15* (n=2 biological replicates, mean ± s.e.m.). The data are normalized to day 0 and to the non-targeting control. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001, n.s. not significant (two-tailed, unpaired t-test).



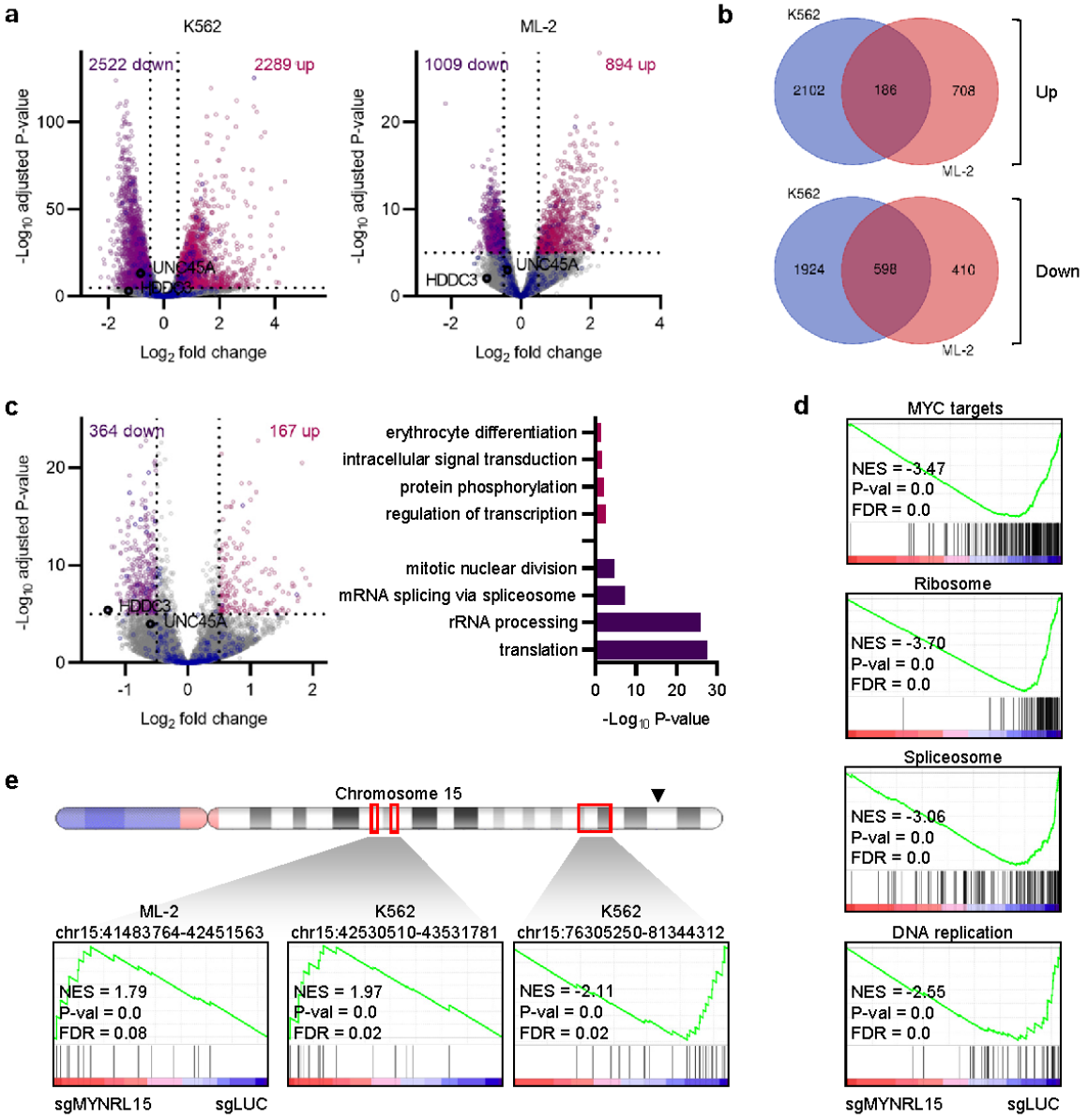
Supplemental Figure 2: Further validation of *MYNRL15* and perturbation approaches.

a, Expression of *MYNRL15* compared to *bona fide* lncRNAs in the NCI-TARGET pediatric AML cohort[47] (n=258; zeros omitted). Midline, median; box limits, lower and upper quartiles; whiskers, 10% and 90% quantiles. **b**, Subcellular localization of *MYNRL15* compared to the control transcripts *XIST* (nuclear) and *B2M* (cytoplasmic), determined via fractionated qRT-PCR (n=2 biological replicates, mean ± s.e.m.). **c**, Expression of *MYNRL15* and its flanking coding genes after targeting the CRISPRi system to the *MYNRL15* TSS, as determined using qRT-PCR (n=3 biological replicates, mean ± s.e.m.; data are normalized to the non-targeting control). **d**, Endpoint depletion values from fluorescence-based proliferation assays using different perturbation strategies in K562 (left) and M-07E (right) cells. Each point represents one vector that was used for perturbation (mean of n=3 biological replicates; data are normalized to day 0 and to the non-targeting control). **e**, qRT-PCR validations of *MYNRL15* knockdown using RNAi (left) and LNA-gapmeRs (center), and of CRISPRi mediated *HDDC3* knockdown (right) (n=3 biological replicates, mean ± s.e.m.; data are normalized to the non-targeting control). **f**, PCR validation of *MYNRL15* excision, using bulk genomic DNA isolated from cells transduced with dual sgRNA vectors, and primers flanking the desired deletion. **g**, Representative TIDE analyses showing the cutting efficiencies of guides targeting *UNC45A* and *HDDC3*.



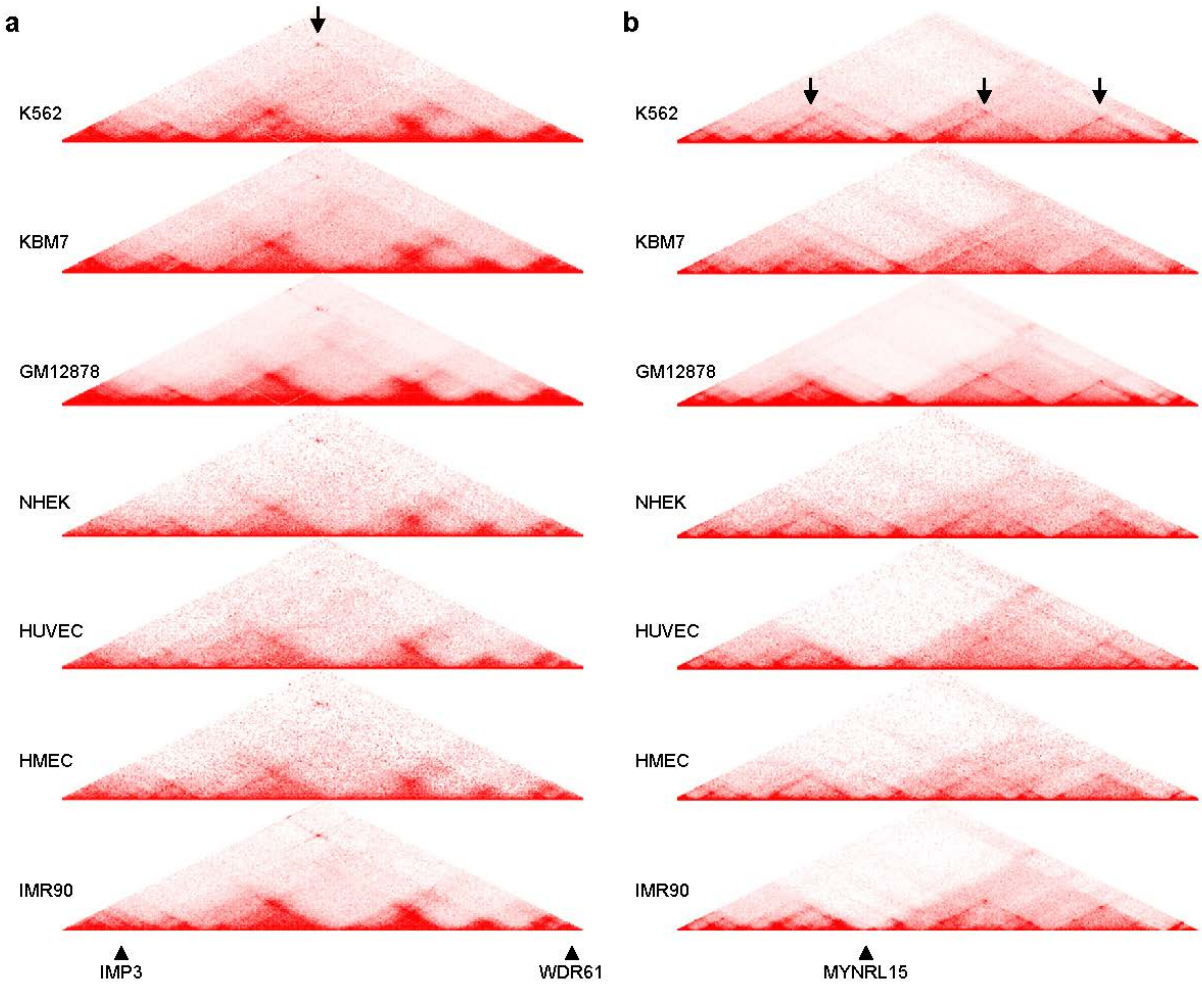
Supplemental Figure 3: *MYNRL15* tiling screen and dual luciferase assay of candidate *cis*-regulatory regions.

a, Results of tiling screens using complementary CRISPRi (top) and CRISPR-Cas9 based (bottom) strategies to interrogate the *MYNRL15* locus. The four tested cell lines are overlaid in different colors (mean of n=2 biological replicates per cell line). **b**, Relative luminescence measured in dual luciferase assays in K562 cells, using reporter constructs containing the C1 and/or C2 sequences cloned upstream of a minimal promoter (n=4 biological replicates, mean ± s.e.m.). The data are normalized to the empty minimal promoter control.



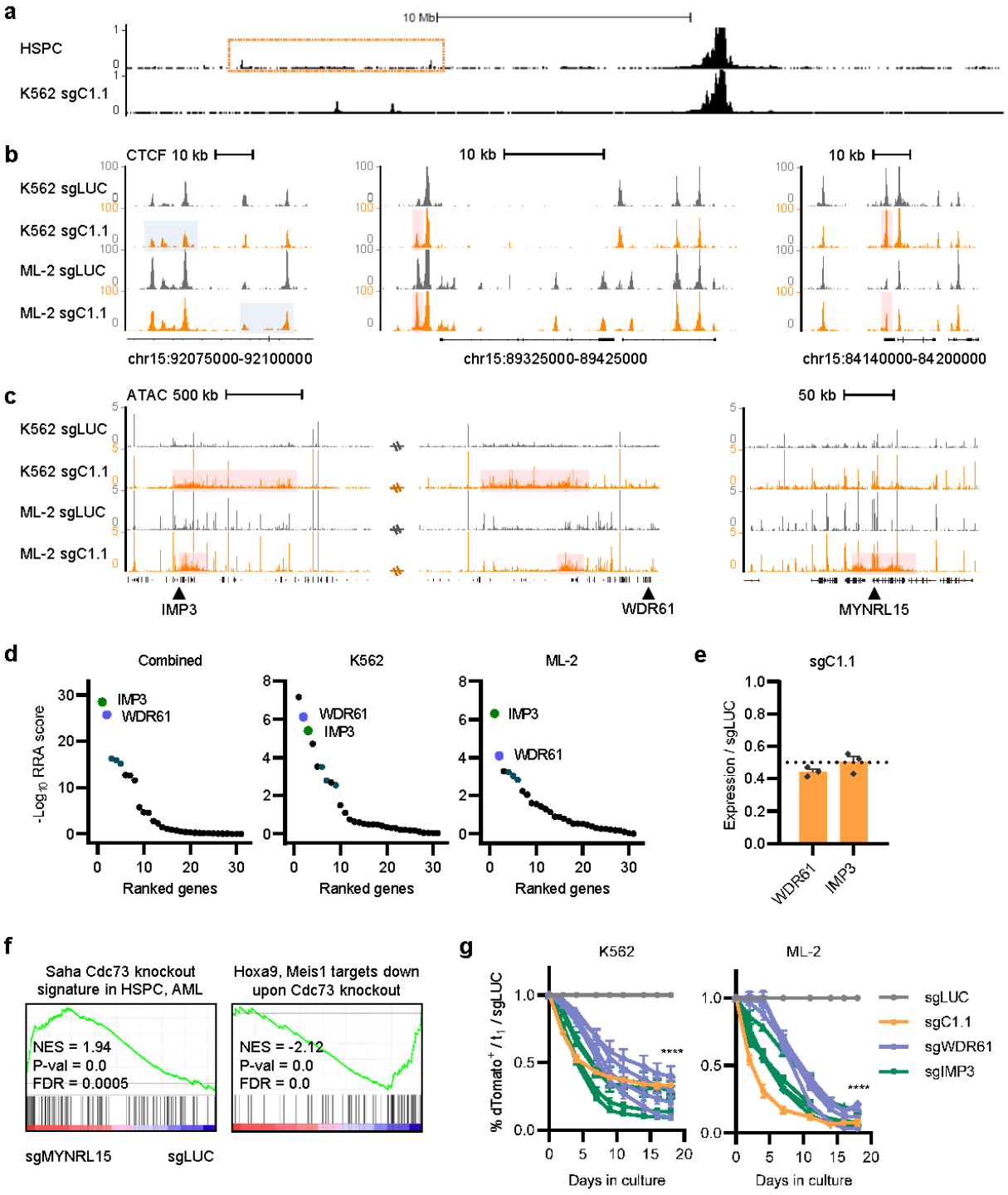
Supplemental Figure 4: Global gene expression profiling following *MYNRL15* perturbation.

a, Volcano plots showing gene expression changes following *MYNRL15* perturbation in K562 (left) and ML-2 (right) cells. Adjusted *P* values were calculated using DESeq2. Up- (pink) and downregulated (purple), and chromosome 15 genes (navy) are highlighted. **b**, Overlap of up- and downregulated genes in K562 and ML-2 cells. **c**, Combined differential expression analysis across K562 and ML-2 cells (left), and enriched gene ontology terms in the differentially expressed gene lists (right; *P* values calculated using DAVID). **d**, Selection of the most strongly downregulated cancer dependency gene sets from GSEA. A combined analysis of K562 and ML-2 cells is shown. **e**, Chromosome 15 gene sets that are not commonly deregulated across K562 and ML-2 cells. The *MYNRL15* locus is indicated with a black arrow. **d-e**, Nominal *P*-values were calculated using GSEA. All analyses compare *MYNRL15* perturbation (via 4 guides, two per cCRE; "sgMYNRL15") to the non-targeting control ("sgLUC"; n=2 biological replicates per guide) across early (day 3) and late (day 6 or 7) time points.



Supplemental Figure 5: Hi-C maps of the *MYNRL15* locus and gained interaction region in multiple cell types.

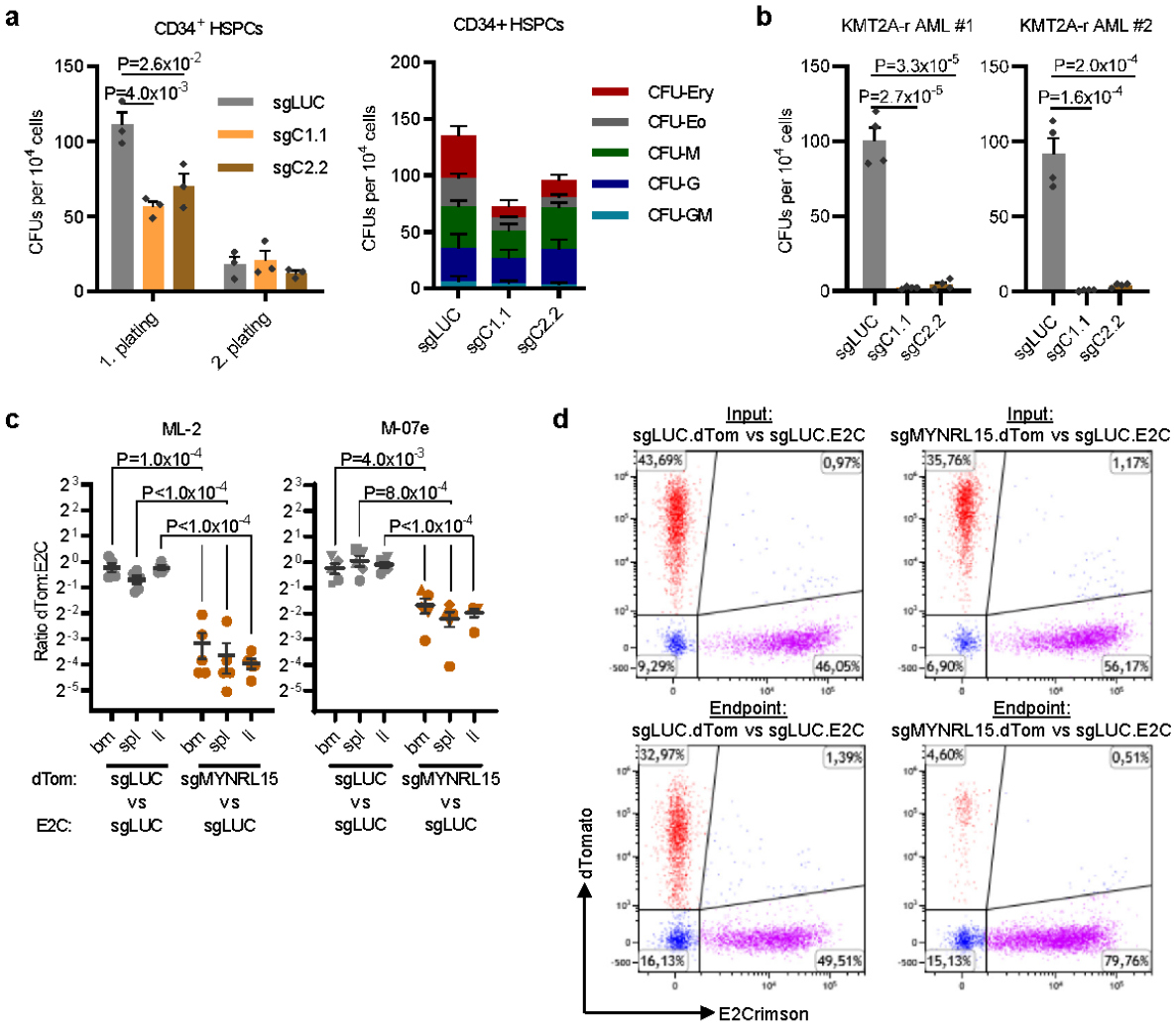
a-b, Hi-C maps from 7 different cell lines[32] including K562s. Knight-Ruiz matrix-balanced values are shown. **a**, Chromatin contacts in the gained distal interaction region. The interaction denoting the hierarchical loop is indicated with a black arrow. **b**, Local chromatin interactions around the *MYNRL15* locus. Contact domains are indicated with black arrows.



Supplemental Figure 6: Additional data from the mechanistic delineation of *MYNRL15* perturbation.

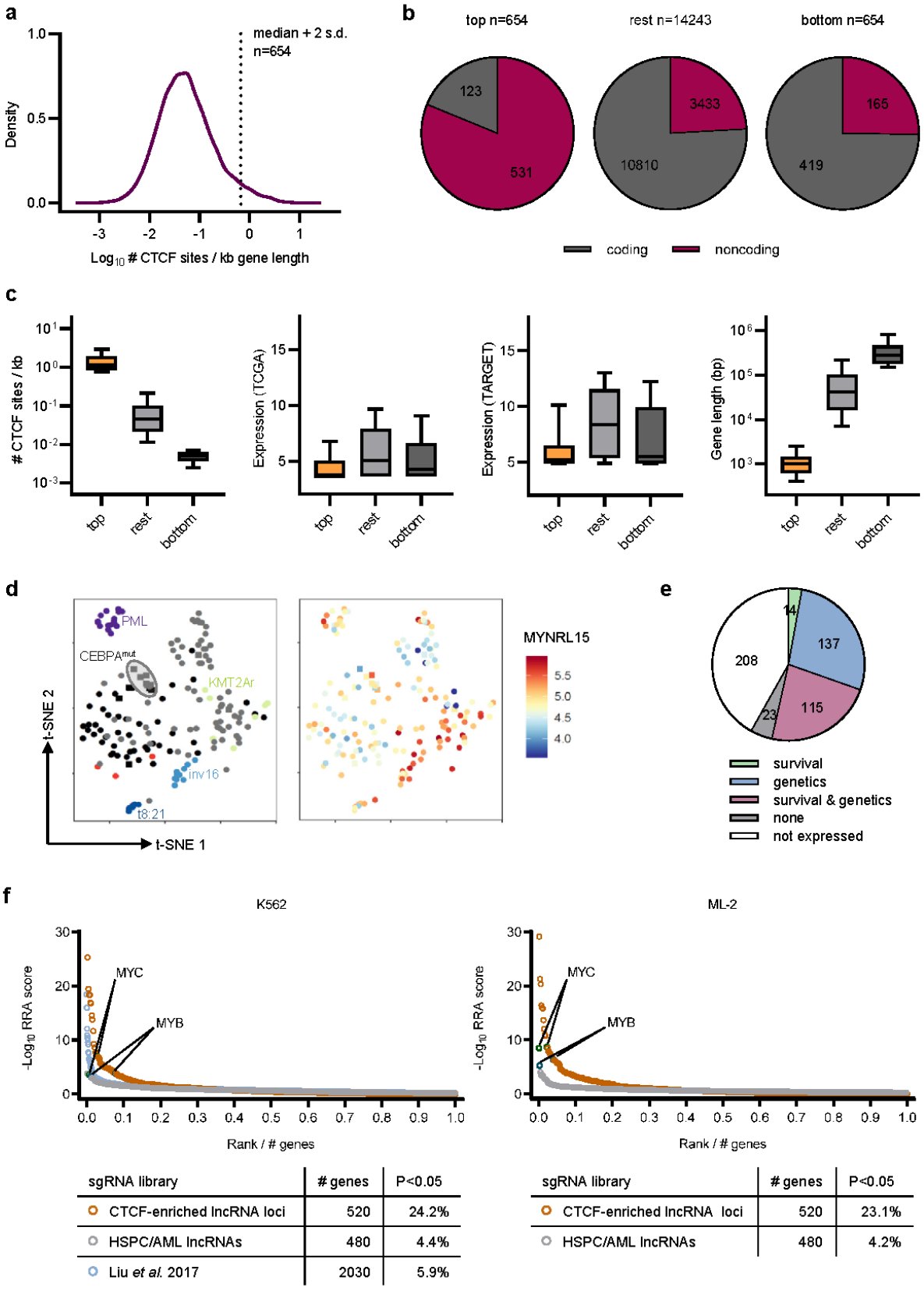
a, Chromosome 15 NG Capture-C interaction profiles from *in vitro* expanded CD34$^+$ HSPCs (n=2 biological replicates; viewpoint in C1; smoothing window 2 pixels). The track from K562 sgC1.1 (*MYNRL15* perturbation) is shown for reference. **b**, CUT&RUN tracks showing other examples of altered CTCF binding downstream of *MYNRL15* (left), and in the area between the gained distal interactions and *MYNRL15* (center and right). Decreases upon *MYNRL15* perturbation are shaded in blue; increases in red. **c**, ATAC-seq tracks showing a wide view

of chromatin accessibility in the gained distal interaction region (left; note the track discontinuity), and in the region around the *MYNRL15* locus (right). Increases are shaded in red. **d**, MAGeCK essentiality scores for the 29 coding genes from the gained chromatin interaction region, which were targeted in CRISPR-Cas9 knockout screens in K562 and ML-2 cells (n=3 biological replicates per cell line). *WDR61* (green) and *IMP3* (lilac) are highlighted, as are the positive controls *U2AF1*, *POL2RA*, and *RPL9* (turquoise). **e**, qRT-PCR validation of *WDR61* and *IMP3* downregulation upon *MYNRL15* perturbation using sgRNA C1.1 (n=3 biological replicates, mean ± s.e.m.; data normalized to the non-targeting control). **f**, PAF1c loss-associated gene sets[23] detected upon *MYNRL15* perturbation. Nominal *P*-values were calculated by GSEA. **g**, Fluorescence-based proliferation assays validating the depletion of K562 and ML-2 cells by WDR61 and IMP3 knockout (4 guides per cell line; n=4 biological replicates, mean ± s.e.m.; data are normalized to day 0 and to the non-targeting control). *MYNRL15* perturbation using sgRNA C1.1 is shown alongside for comparison. ****$P<0.0001$ (two-tailed, unpaired t-test); all conditions shared the same *P*-value.



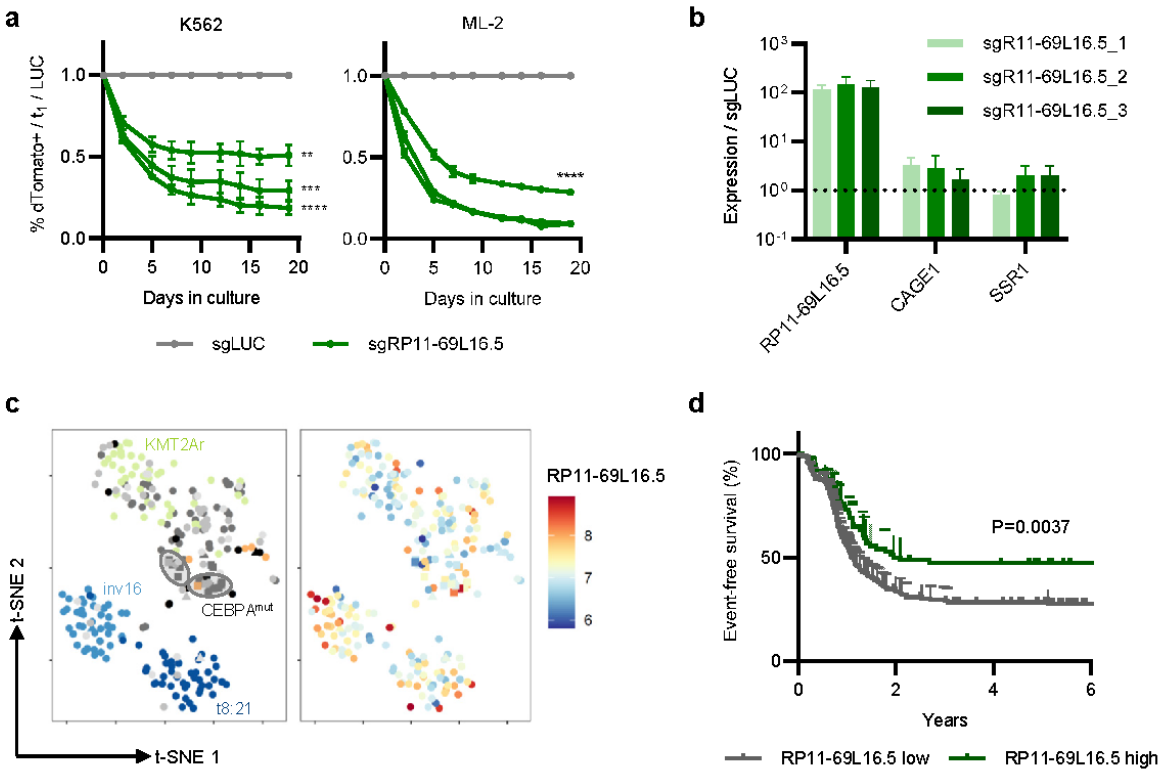
Supplemental Figure 7: *MYNRL15* perturbation in primary CD34⁺ HSPCs and patient-derived AML cells.

a, Colony counts following *MYNRL15* perturbation in CD34⁺ HSPCs from healthy donors (n=3 biological replicates; mean ± s.e.m.). Replating capacity (left) and differentiation (right) were evaluated in parallel. **b**, Colony counts following *MYNRL15* perturbation in two patient-derived AML samples (n=4 biological replicates; mean ± s.e.m.). **c**, Results from two-color *in vivo* competition assays using CRISPRi based *MYNRL15* perturbation in two AML cell lines. The data are displayed as ratios of dTomato⁺ (dTom) to E2Crimson⁺ (E2C) cells in the bone marrow (bm), spleens (spl), and livers (li) of the recipient mice (n=5 mice per group, mean ± s.e.m.). **a-c**, *P* values were calculated using two-tailed, unpaired t-tests. **d**, Representative flow cytometry data from the two-color competitive xenotransplantation assays.



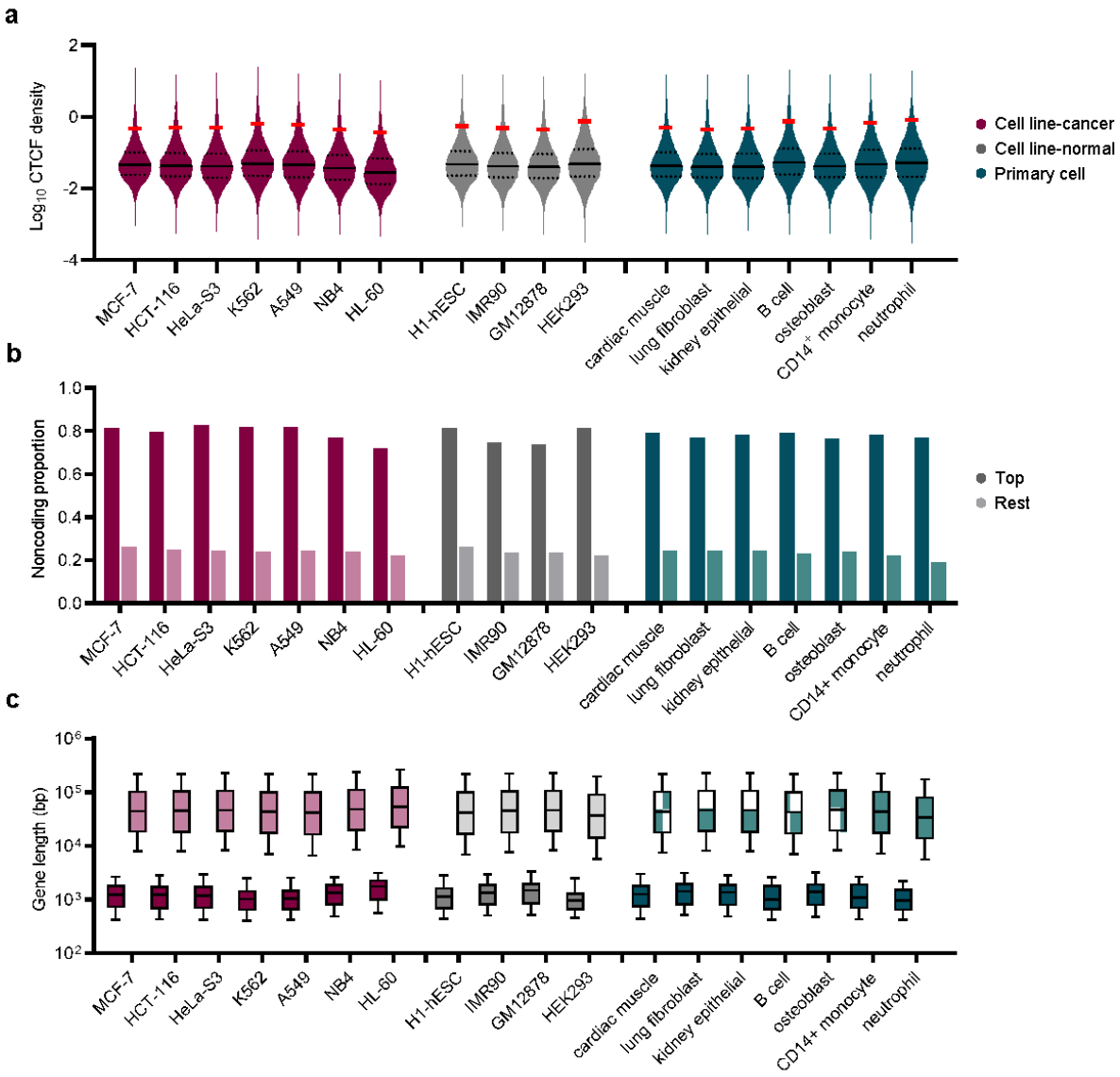
Supplemental Figure 8: Exploration of CTCF density metrics and CTCF-enriched lncRNA loci (C-LNCs).

a, Distribution of log_{10}-transformed values for the CTCF density metric (number of CTCF sites per kb of gene length) in K562 cells. The dashed line indicates the cut-off defined (i.e. median + 2 s.d.) for elevated CTCF binding. **b**, Proportions of coding and noncoding genes in different sections of the ranked list of CTCF-bound loci. **c**, Box plots illustrating predictive features of other loci like *MYNRL15*. From left to right: CTCF density, normalized expression in the TCGA AML[46] and NCI-TARGET AML[47] cohorts, gene length. **b-c**, Graphs compare the top (n=654) CTCF-bound loci to the rest (n=14243) or bottom (< median - 2 s.d., n=654) of the ranked list. Midline, median; box limits, lower and upper quartiles; whiskers, 10% and 90% quantiles. **d**, Unsupervised clustering of the TCGA AML[46] cohort reveals molecular sub-types (left), several of which associate with *MYNRL15* expression (i.e. PML, inv16, *KMT2A*r) (right). **e**, Breakdown of CTCF-enriched loci based on their association with clinical aspects like cytogenetics, mutations, and survival in the NCI-TARGET AML cohort (left). **f**, MAGeCK essentiality scores of CTCF-enriched lncRNA loci (C-LNCs), from CRISPR-Cas9 screens in K562 (left) and ML-2 (right) cells (n=2 biological replicates per cell line; the gene ranks are normalized to library size). The C-LNC screening results (orange) are displayed alongside our initial CRISPRi library (grey) and of Liu *et al.*'s[53] CRISPRi genome-wide lncRNA library (blue). The positions of the positive controls *MYC* and *MYB* are indicated.



Supplemental Figure 9: Example of a myeloid leukemia vulnerability identified from C-LNC screen.

a, Fluorescence-based proliferation assays validating depletion of K562 and ML-2 cells following CRISPR-Cas9 mediated perturbation of CTCF sites in *RP11-69L16.5* (n=3 biological replicates per guide, mean ± s.e.m.; data are normalized to day 0 and the non-targeting control). ***P*<0.01, ****P*<0.001, *****P*<0.0001 (two-tailed, unpaired t-tests); where only one set of asterisks are shown, all conditions share the same *P* value. **b**, Expression of *RP11-69L16.5* and its neighboring coding genes following perturbation, as determined via qRT-PCR (n=3 biological replicates, mean ± s.e.m.; data normalized to the non-targeting control). **c**, Unsupervised clustering of the NCI-TARGET AML[47] cohort with samples colored according to molecular subtype (left) and *RP11-69L16.5* expression (right). **d**, Kaplan-Meier survival curves showing patients with high (n=74) vs low (n=184) expression of *RP11-69L16.5* in the NCI-TARGET AML[47] cohort (survival probability ± 95% C.I.). Five-year event-free survival: 47.2% vs 28.5%. The depicted *P* value was calculated using a two-sided log-rank test.



Supplemental Figure 10: Catalog of C-LNCs in 18 cell lines and primary cell types.

a, Distributions of log_{10}-transformed values for the CTCF density metric (number of CTCF sites per kb of gene length) across 18 cell lines and primary cell types from ENCODE. Mid-line, median; dotted lines, lower and upper quartiles. Red lines indicate the cut-off defined (median + 2 s.d.) for elevated CTCF binding in each cell type. **b**, Proportion of noncoding genes in the top (≥ cut-off) and rest (< cut-off) of CTCF-bound loci. **c**, Box plots illustrating gene length distributions between the top (n=654) and rest (n=14243) of CTCF-bound loci. Midline, median; box limits, lower and upper quartiles; whiskers, 10% and 90% quantiles.

Supplemental references

1 Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* **10**, e0124633, doi:10.1371/journal.pone.0124633 (2015).

2 Adams, F. F. *et al.* An optimized lentiviral vector system for conditional RNAi and efficient cloning of microRNA embedded short hairpin RNA libraries. *Biomaterials* **139**, 102-115, doi:10.1016/j.biomaterials.2017.05.032 (2017).

3 Al-Kershi, S. *et al.* The stem cell-specific long noncoding RNA HOXA10-AS in the pathogenesis of KMT2A-rearranged leukemia. *Blood Adv* **3**, 4252-4263, doi:10.1182/bloodadvances.2019032029 (2019).

4 Emmrich, S. *et al.* miR-99a/100~125b tricistrons regulate hematopoietic stem and progenitor cell homeostasis by shifting the balance between TGFbeta and Wnt signaling. *Genes Dev* **28**, 858-874, doi:10.1101/gad.233791.113 (2014).

5 Stein, C. A. *et al.* Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents. *Nucleic Acids Res* **38**, e3, doi:10.1093/nar/gkp841 (2010).

6 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).

7 Volders, P. J. *et al.* An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* **43**, 4363-4364, doi:10.1093/nar/gkv295 (2015).

8 Zhao, Y. *et al.* NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res* **44**, D203-208, doi:10.1093/nar/gkv1252 (2016).

9 Schwarzer, A. *et al.* The non-coding RNA landscape of human hematopoiesis and leukemia. *Nat Commun* **8**, 218, doi:10.1038/s41467-017-00212-4 (2017).

10 Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661, doi:10.1016/j.cell.2014.09.029 (2014).

11 Concordet, J. P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**, W242-W245, doi:10.1093/nar/gky354 (2018).

12 Labuhn, M. *et al.* Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res* **46**, 1375-1385, doi:10.1093/nar/gkx1268 (2018).

13 Perez, A. R. *et al.* GuideScan software for improved single and paired CRISPR guide RNA design. *Nat Biotechnol* **35**, 347-349, doi:10.1038/nbt.3804 (2017).

14 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).

15 Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416, doi:10.1038/s41467-018-07901-8 (2018).

16 Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554, doi:10.1186/s13059-014-0554-4 (2014).

17 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

18 Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168, doi:10.1093/nar/gku936 (2014).

19 Cabianca, D. S. *et al.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819-831, doi:10.1016/j.cell.2012.03.035 (2012).

20 Bhayadia, R. *et al.* Endogenous Tumor Suppressor microRNA-193b: Therapeutic and Prognostic Value in Acute Myeloid Leukemia. *J Clin Oncol* **36**, 1007-1016, doi:10.1200/JCO.2017.75.2204 (2018).

21 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915, doi:10.1038/s41587-019-0201-4 (2019).

22 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

23 Saha, N. *et al.* The PAF1c Subunit CDC73 Is Required for Mouse Hematopoietic Stem Cell Maintenance but Displays Leukemia-Specific Gene Regulation. *Stem Cell Reports* **12**, 1069-1083, doi:10.1016/j.stemcr.2019.03.010 (2019).

24 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

25 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

26 Davies, J. O. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* **13**, 74-80, doi:10.1038/nmeth.3664 (2016).

27 Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**, 205-212, doi:10.1038/ng.2871 (2014).

28 Buckle, A., Gilbert, N., Marenduzzo, D. & Brackley, C. A. capC-MAP: software for analysis of Capture-C data. *Bioinformatics* **35**, 4773-4775, doi:10.1093/bioinformatics/btz480 (2019).

29 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).

30 Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis* **33**, 1029-1047, doi:10.1093/imanum/drs019 (2012).

31 Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* **78**, 554-565 e557, doi:10.1016/j.molcel.2020.03.003 (2020).

32 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

33 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).

34 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

35 Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182, doi:10.1186/1471-2105-15-182 (2014).

36 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

37 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

38 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

39 Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *Elife* **8**, doi:10.7554/eLife.46314 (2019).

40 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019, doi:10.1038/nprot.2018.015 (2018).

41 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

42 Zhu, Q., Liu, N., Orkin, S. H. & Yuan, G. C. CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. *Genome Biol* **20**, 192, doi:10.1186/s13059-019-1802-4 (2019).

43 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

44 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

45 van Heeringen, S. J. & Veenstra, G. J. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270-271, doi:10.1093/bioinformatics/btq636 (2011).

46 The Cancer Genome Atlas Research Network *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).

47 Bolouri, H. *et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* **24**, 103-112, doi:10.1038/nm.4439 (2018).

48 Aalen, O. O. A linear regression model for the analysis of life times. *Stat Med* **8**, 907-925, doi:10.1002/sim.4780080803 (1989).

49 Scrucca, L., Santucci, A. & Aversa, F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplant* **40**, 381-387, doi:10.1038/sj.bmt.1705727 (2007).

50 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883, doi:10.1093/bioinformatics/bts034 (2012).

51 Lausen, B., Hothorn, T., Bretz, F. & Schumacher, M. Assessment of optimal selected prognostic factors. *Biometrical J* **46**, 364-374, doi:10.1002/bimj.200310030 (2004).

52 van der Maaten, L. J. P. H., G. E. Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).

53 Liu, S. J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, doi:10.1126/science.aah7111 (2017).