Myeloid leukemia vulnerabilities at CTCF-enriched long noncoding RNA loci

Michelle Ng
Martin Luther University Halle-Wittenberg
Lonneke Verboon
Martin Luther University Halle-Wittenberg
Hasan Issa
Martin Luther University Halle-Wittenberg
Raj Bhayadia
Martin Luther University Halle-Wittenberg
Oriol Alejo-Valle
Martin Luther University Halle-Wittenberg https://orcid.org/0000-0002-0988-7416
Dorit Borchert
Hannover Medical School
Konstantin Schuschel
Martin Luther University Halle-Wittenberg
Eniko Regenyi
Free University of Berlin
Dirk Reinhardt
University Hospital Essen
Marie-Laure Yaspo
Max Planck Institute for Molecular Genetics
Dirk Heckl
Martin Luther University Halle-Wittenberg
Jan-Henning Klusmann (jan-henning.klusmann@kgu.de)
Goethe University Frankfurt https://orcid.org/0000-0002-1070-0727

Genetics Article

Keywords: long noncoding RNA, C-LNC, MYNRL15

Posted Date: August 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-727909/v1

Myeloid leukemia vulnerabilities at CTCF-enriched long noncoding RNA loci

Abstract word count: 165

Main text word count: 2656

References: 58

Running title: *MYNRL15* in myeloid leukemia

Keywords: long noncoding RNA, myeloid leukemia, CRISPR, chromatin architecture

Abstract

The noncoding genome presents a largely untapped source of biological insights, including thousands of long noncoding RNA (lncRNA) loci. While some produce *bona fide* lncRNAs, others exert transcript-independent *cis*-regulatory effects, and the lack of predictive features renders mechanistic dissection challenging. Here, we describe CTCF-enriched lncRNA loci (C-LNC) as a subclass of functional genetic elements exemplified by *MYNRL15,* a pan-myeloid leukemia dependency identified by an lncRNA-based CRISPRi screen. *MYNRL15* perturbation selectively impairs acute myeloid leukemia (AML) cells over hematopoietic stem / progenitor cells *in vitro*, and depletes AML xenografts *in vivo*. Mechanistically, we show that crucial DNA elements in the locus mediate its phenotype, triggering chromatin reorganization and downregulation of cancer dependency genes upon perturbation. Elevated CTCF density distinguishes *MYNRL15* and 531 other lncRNA loci in K562 cells, of which 43-54% associate with clinical aspects of AML and 18.4% are functionally required for leukemia maintenance. Curated C-LNC catalogs in other cell types will help refine the search for noncoding oncogenic vulnerabilities in AML and other malignancies.

Main

It becomes increasingly clear that the 98% of the human genome that does not encode protein nonetheless contains a wide range of functional elements that are vital for cellular homeostasis[1,2]. These include *cis*-regulatory elements such as enhancers and promoters, insulators and other determinants of genome topology, as well as a large number and variety of non-protein-coding transcripts. Long noncoding RNAs (lncRNAs) in particular comprise a substantial portion of the noncoding transcriptome[3-5] and in recent years, have emerged as important players in diverse cellular processes and contexts[6-8]. The hematopoietic system is no exception, where lncRNAs have been described to regulate cell programming and fate[9], and where their dysregulation has been tied to malignancy[10-16]. LncRNAs present a significant opportunity to extend our understanding of human health and disease; however, the fact remains that the vast majority of lncRNA loci lack functional characterization, and may regulate cellular behaviour in ways yet unknown. Indeed, characterization is often a difficult process complicated by *cis*-regulatory mechanisms unrelated to the transcriptional product[17-23]. Improved functional classification systems are imperative for expediting investigations into lncRNA determinants of pathophysiology, including the search for noncoding oncogenic vulnerabilities.

CRISPRi screens of HSPC/AML lncRNAs identify MYNRL15 as a leukemia dependency

Aiming to identify lncRNAs that contribute to myeloid malignancy, we began by analyzing a noncoding RNA expression atlas of the human blood system encompassing hematopoietic stem cells (HSCs) and their differentiated progeny, as well as pediatric acute myeloid leukemia (AML) samples[16]. In addition to stem cell signatures reminiscent of those previously established for protein-coding genes[24-27], we discovered progenitor- and AML subtype-associated lncRNA profiles that could potentially serve as leukemia-specific targets, given their absence in HSCs (Fig. 1a). To probe this resource for functionality and find novel AML vulnerabilities, we conducted a CRISPRi-based dropout screen of 480 lncRNA genes from 8 distinct signatures in 6 human leukemia cell lines (Fig. 1b). Five were selected to represent relevant

cytogenetic subgroups of AML (ML-2, NOMO-1 [*KMT2A*-rearreanged], SKNO-1, KASUMI-1 [standard risk with t(8:21)], M-07E [high risk with inv(16)(p13.3q24.3)]), and we also included the well-studied erythroleukemia line K562. One candidate emerged as crucial in all six cell lines – *AC068831.3* (ID: ENSG00000224441 in Ensembl v91 [release 12/2017]), which we renamed *MYNRL15* (myeloid leukemia noncoding regulatory locus on chromosome 15; Fig. 1c-d, Extended Data Fig. 1).

MYNRL15 is a low-abundance, nuclear-enriched transcript (Extended Data Fig. 2a-b) originating from chromosome 15, where it is flanked by two protein-coding genes: *UNC45A* and *HDDC3* (Fig. 1c). Given the local effect of the CRISPRi system on nearby genes (Extended Data Fig. 2c), a range of gain- and loss-of-function approaches were necessary in order to delineate the source of the *MYNRL15* knockdown phenotype (Fig. 1e-f, Extended Data Fig. 2d-g). While CRISPR mediated excision of *MYNRL15* recapitulated the effect produced by CRISPRi, repression of the transcript via RNAi and LNA-gapmeRs had little impact on cell viability (Fig. 1e, Extended Data Fig. 2d-f). Both protein-coding neighbors were also dispensable, as determined via CRISPR-Cas9 mediated knockout of *UNC45A* and *HDDC3*, and CRISPRi mediated knockdown of *HDDC3* (Fig. 1e, Extended Data Fig. 2d-g). In addition, overexpression of *MYNRL15* cDNAs failed to rescue the CRISPRi knockdown phenotype (Fig. 1f). Altogether, these results indicate that neither of the flanking protein-coding genes, nor the *MYNRL15* transcript, is responsible for the function of this locus in myeloid leukemia cells, and rather suggest *MYNRL15* as an expressed noncoding regulatory locus.

Functional dissection of the MYNRL15 locus reveals crucial regulatory regions

Given the apparent dispensability of *UNC45A*, *HDDC3*, and the *MYNRL15* transcript itself in leukemia cells, we hypothesized that *MYNRL15* may harbor DNA regulatory elements which drive its leukemia dependency phenotype. To test this hypothesis, we functionally dissected the *MYNRL15* locus via complementary CRISPRi and CRISPR-Cas9 screens tiling a 15 kb area centered on *MYNRL15*. Notably, this area exhibits features characteristic of *cis*-regula-

tory elements (CREs), including H3K4Me1 and H3K27Ac histone marks, DNase hypersensitivity, and transcription factor occupancy (Fig. 2a). The screens uncovered two crucial DNA regions whose accessibility and integrity were required for the maintenance of leukemic cells (Fig. 2b, Extended Data Fig. 3a), and which enhanced reporter gene expression in dual luciferase assays (Extended Data Fig. 3b) – identifying the regions as functional sequences and candidate *cis*-regulatory elements (cCREs C1 and C2). We note that the parallel screening strategy using the dCas9 variant helps alleviate the risk of potential off-target DNA damage-driven phenotypes[28], thus increasing the robustness of the screen. The CRISPR-Cas9 mutagenesis strategy also reiterates that leukemia cells do not appear particularly dependent on the UNC45A and HDDC3 coding sequences, arguing against local enhancer functions on these genes underlying the anti-leukemic effect of *MYNRL15* perturbation.

Aiming to identify target genes and pathways controlled by the *MYNRL15* cCREs, we next performed RNA sequencing (RNA-seq) upon CRISPR-Cas9 mediated perturbation of the locus. We opted for the CRISPR-Cas9 system in order to achieve a more targeted perturbation of *MYNRL15* and attenuate effects on *UNC45A* and *HDDC3* caused by CRISPRi. We selected two guides from each cCRE, all of which robustly depleted K562 and ML-2 leukemia cells (Fig. 2c). This phenotype was underpinned by global changes in gene expression (Fig. 2d, Extended Data Fig. 4a-d), including a dramatic suppression of cancer-essential signatures related to proliferation and metabolism (Fig. 2d, Extended Data Fig. 4d). While these results corroborated *MYNRL15*'s leukemia dependency phenotype – with the downregulated genes being enriched for members of key oncogenic pathways (Fig. 2d, Extended Data Fig. 4c-d) – no clear target genes emerged, leading us to consider the possibility that *MYNRL15* may instead regulate multiple genes in a genomic neighborhood[29] in a more subtle manner. To explore this option, we applied a sliding window approach to gene set enrichment analysis using 1 Mb to 5 Mb sections of chromosome 15. This revealed positional gene sets that were coordinately deregulated upon *MYNRL15* perturbation, including the local region around *MYNRL15* and several distal regions (Fig. 2e, Extended Data Fig. 4e).

Altered chromosome 15 architecture underlies the MYNRL15 perturbation phenotype

Given the deregulation of chromosome 15 neighborhoods upon *MYNRL15* perturbation, we explored whether *MYNRL15* may be involved in chromatin conformation via next generation Capture-C (NG Capture-C)[30], using probes complementary to *MYNRL15* cCRE C1 to enrich for interactions involving the locus. This approach revealed extensive chromatin contacts between *MYNRL15* and sequences within a 500 kb radius, with weaker contacts occurring up to 2 Mb away – a profile that was consistent across K562 and ML-2 cells (Fig. 3a-c). The local interaction peaks demarcate nearby contact domains (Fig. 3c, Extended Data Fig. 5a-b), implicating *MYNRL15* in the 3D organization of this region of chromosome 15. Interestingly, *MYNRL15* perturbation had little impact on this local interaction profile, instead causing cells to gain two long-range interactions 12 Mb and 15 Mb upstream of the locus at the base of a hierarchical loop (Fig. 3b, Extended Data Fig. 5a) – pointing to 3D chromatin reorganization upon *MYNRL15* perturbation that brings the locus into contact with this structure. We further note the presence of distal interactions in this region in $CD34^+$ hematopoietic stem / progenitor cells (HSPCs; Extended Data Fig. 6a), suggesting that *MYNRL15* perturbation in leukemic cells may re-establish the long-range connectivity of normal blood cells. Consistent with *MYNRL15*'s involvement in chromosome 15 conformation, CTCF – a fundamental determinant of genome architecture, which occupies three sites in the *MYNRL15* locus (Fig. 2a) – showed reduced binding at the locus and distal interaction sites following *MYNRL15* perturbation, among other changes (Fig. 3d, Extended Data Fig. 6b). This was accompanied by diffuse gains in chromatin accessibility across the distal interaction sites (Fig. 3e, Extended Data Fig. 6c).

Integrating our findings regarding altered chromatin conformation and gene expression with cancer dependency data, we eventually pinpointed the target genes of *MYNRL15* through a small-scale CRISPR-Cas9 screen of the 29 protein-coding genes located in the gained interaction region. Combining these results with differential expression information, we identified

two downregulated protein-coding genes in the region that also score as leukemia dependencies: *IMP3* and *WDR61* (Fig. 3f, Extended Data Fig. 6d-e). Notably, *WDR61* is a component of the PAF1 complex (PAF1c), which is involved in important transcriptional programs during hematopoiesis and leukemogenesis[31,32]. Gene expression changes induced by Paf1c inactivation[33] were also detected upon *MYNRL15* perturbation (Extended Data Fig. 6f). *IMP3* encodes a component of the 60-80S U3 small nucleolar ribonucleoprotein, which is required for cleavages in pre-18S ribosomal RNA processing[34]. It is a homolog of the yeast Imp3 protein and has not been comprehensively studied in human cells to date. CRISPR-Cas9 mediated knockout of *WDR61* and *IMP3* robustly depleted K562 and ML-2 cells (Extended Data Fig. 6g), recapitulating the *MYNRL15* perturbation phenotype and positioning these genes as targets of *MYNRL15* (Fig. 3g).

AML specificity and potential therapeutic applicability of MYNRL15

To evaluate whether *MYNRL15* dependency is specific to leukemic cells, we leveraged all-in-one lentiviral CRISPR-Cas9 constructs in primary human $CD34^+$ HSPCs and blasts derived from two AML patients (see Supplementary Table 1 for clinical characteristics). The transduced cells were sorted and seeded in methylcellulose-based colony-forming assays. While *MYNRL15* perturbation moderately attenuated colony formation in $CD34^+$ HSPCs, it had little effect on replating capacity and differentiation (Extended Data Fig. 7a). In contrast, AML colony-forming units were virtually eradicated (Extended Data Fig. 7b) – implying that *MYNRL15* perturbation selectively impacts AML cells, and outlining a possible therapeutic window (Fig. 4a).

To assess the therapeutic potential of *MYNRL15* perturbation, we applied CRISPRi-based two-color competitive xenotransplantation assays using AML cell lines and patient-derived xenografts (PDXs) (Fig. 4b). Importantly, *MYNRL15* perturbation significantly impaired the propagation of two AML cell lines and two PDXs (Fig. 4b, Extended Data Fig. 7c-d) in the hematopoietic organs of recipient mice, confirming its capacity to deplete leukemic cells *in*

vivo. Combined with the apparent AML-specific effect of *MYNRL15* perturbation, these results provide a proof-of-principle of how *MYNRL15* perturbation may be leveraged as a therapeutic strategy.

MYNRL15 exemplifies a putative new class of CTCF-bound long noncoding RNA loci

Having implicated *MYNRL15* in 3D genome organization and demonstrated its therapeutic potential, we explored whether *MYNRL15* may belong to a sub-category of biologically relevant lncRNA loci that have thus far been overlooked due to their lack of transcript-specific functions. Given the effect of *MYNRL15* on chromatin architecture and the multiple CTCF binding sites in the locus, we explored CTCF density as a predictive metric for identifying noncoding regulatory loci like *MYNRL15* (Fig. 5a, Extended Data Fig. 8a-c). Interestingly, $\log_{10}$-transformed values of this metric followed a near-normal distribution, and a cut-off of two standard deviations from the median identified 654 genes with elevated CTCF density which were highly enriched for lncRNAs (>80%, n=531 using K562 ChIP-seq data) (Fig. 5a, Extended Data Fig. 8a-b). Inversely, the remainder comprised mainly of protein-coding loci (Fig. 5a, Extended Data Fig. 8b), with *bona fide* lncRNAs such as *PVT1* and *XIST* also featuring in the lower part of the ranked list; others do not appear at all, due to an absence of CTCF binding sites. These observations support CTCF density as a relevant metric for distinguishing a potential subset of lncRNA loci with transcript-independent functions, hereafter referred to as CTCF-enriched lncRNA loci, or C-LNC. Besides elevated CTCF density, these loci also tended to display low gene expression and short genomic length (median 1 kb; Extended Data Fig. 8c) – mirroring *MYNRL15*, which produces a low-abundance transcript and spans 2544 bp on chromosome 15, and providing additional predictive features.

As a first step in determining the relevance of C-LNCs to myeloid leukemia, we tested their association with clinical aspects in two AML patient cohorts[35,36]. This revealed that 43% and 54% of the identified C-LNCs associated with genetically-defined AML subgroups or patient survival in the two AML cohorts, respectively (Fig. 5a-b, Extended Data Fig. 8d-e), suggest-

ing that activity at these loci could underpin aberrant transcription factor programs and/or influence patient outcomes. Furthermore, 18.4% functionally validated as essential for myeloid leukemia maintenance in CRISPR-Cas9 screens tiling the CTCF sites in the loci (Fig. 5c, Extended Data Fig 8f) – a hit identification rate that is substantially higher than what is typically reported for lncRNA essentiality screens[37-39] (ranging from 2-6%), including our own (4.6%; Fig. 5c, Extended Data Fig. 8f). Taken together, these data illustrate the effectiveness of CTCF density metrics in refining functional lncRNA candidate lists, and underline the relevance such loci hold for AML and cancer pathophysiology in general. We provide a catalog of C-LNCs across 18 cell lines and primary cell types (Fig. 5d, Extended Data Fig. 9, Supplementary Table 11) – www.C-LNC.org – as a basis for advancing the discovery of both novel oncogenic vulnerabilities and functional lncRNA loci in other contexts.

Discussion

There is general agreement that the current lncRNA classification system leaves much to be desired, necessitating extensive experimental labor in order to discern between the possible modes of action for any given lncRNA[22]. Through the functional and molecular dissection of *MYNRL15*, we provide evidence for myeloid leukemia vulnerabilities at noncoding regulatory loci involved in chromatin architecture, and demonstrate pathophysiological as well as potential therapeutic relevance. We moreover present predictive metrics based on distinct features of *MYNRL15* – namely, elevated CTCF density, low expression, and short span – and establish their value in distinguishing a functionally and clinically relevant subclass of lncRNA loci (CTCF-enriched lncRNA loci, or C-LNC). A catalog of C-LNCs covering 18 cell lines and primary cell types can be found at www.C-LNC.org or in Supplementary Table 11, providing a basis for extending investigations of C-LNCs into other cellular contexts. These and other efforts aimed at improving the functional classification of lncRNAs[19,22,40,41] will expedite the development of precise and comprehensive annotations, and facilitate the process of discriminating transcript-dependent from -independent functions.

In our study, *MYNRL15* perturbation resulted in the formation of a long-range chromatin interaction, leading to the downregulation of *WDR61* and *IMP3* and tumor suppression. Given the accompanying reduction of CTCF occupancy, we expect this to occur through a mechanism similar to topologically associating domain (TAD) fusion[42-44], although on a larger scale than typically observed for TADs[45]. Alternatively, or perhaps concurrently, the attenuation of CTCF binding upon *MYNRL15* perturbation may strengthen compartmentalization[46] and promote longer-range, higher-order architecture. We note that, while there is substantial overlap between enhancer RNA (eRNA) and lncRNA annotations[47], and while some of our data support a local enhancer-like function for *MYNRL15*, we did not find evidence for locally-driven phenotypes or RNA function, and the long-range architectural changes upon perturbation of the locus especially separate *MYNRL15* from classical eRNAs.

Given the attenuated impact of *MYNRL15* perturbation on normal HSPCs compared to AML cells, we surmise that distal connectivity may be the native conformation of the locus that is lost during leukemic transformation; thus re-introducing it would selectively impair leukemic cells. The oncogenic rewiring of 3D chromatin architecture through mutations and structural variants has been reported in cancer[44,48-51]. However, it is unlikely that genetic alteration underlies *MYNRL15*'s role in leukemia, since the locus is required by cells from varied cytogenetic and mutational backgrounds, and its perturbation drives matching chromatin changes in two divergent cell lines. We speculate instead that *MYNRL15* may be involved in unifying leukemic genome organization signatures – a phenomenon that has long been established for stemness-related expression and epigenetic signatures[24,25,39]. Recent works have begun to implicate aspects of chromatin architecture in cell state transitions during hematopoiesis[52-55] and in the maintenance of leukemic transcription programs[56-58]. We expect future studies will further reveal leukemic 3D genome organization signatures that underpin general oncogenic behaviors, irrespective of mutational drivers.

An important future direction will be to ascertain whether C-LNCs, like *MYNRL15*, contribute to cancer-related chromatin architecture as a class. Based on their elevated CTCF densities

and other shared features, we hypothesize that many may function through similar mechanisms as *MYNRL15*. With our preliminary catalog of C-LNCs spanning various human cell types and tissue contexts (www.C-LNC.org), we provide a resource that will help catalyze future research and lay a foundation for unravelling principles of C-LNC function in healthy and malignant cells. Given the high essentiality rate that we observed in myeloid leukemia, C-LNCs could symbolize a major refinement in the search for both functional lncRNA loci and noncoding oncogenic vulnerabilities across all types of cancer.

Materials and methods

Cells and cell culture

HEK293T cells and the human leukemia cell lines K562, ML-2, NOMO-1, KASUMI-1, SKNO-1, and M-07E were obtained from the German National Resource Center for Biological Material (DSMZ, Braunschweig, Germany) and cultured according to their recommendations. All cell lines were routinely tested for *Mycoplasma* contamination. Human CD34$^+$ hematopoietic stem and progenitor cells (HSPCs) were isolated from mobilized peripheral blood from anonymous healthy donors, and enriched using anti-CD34 immunomagnetic microbeads (Miltenyi Biotech). Acute myeloid leukemia (AML) patient samples were provided by the Berlin-Frankfurt-Münster Study Group (AML-BFM-SG, Essen, Germany), and expanded via serial xenotransplantation in immunocompromised humanized mice. Clinical information for the patient-derived xenografts (PDXs) used in this study can be found in Supplementary Table 1. CD34$^+$ HSPC and PDX maintenance are described below. Informed consent was obtained from all human participants or custodians. All investigations were approved by the local ethics committee of the Martin Luther University Halle-Wittenberg.

Lentiviral vectors

Individual single guide RNAs (sgRNAs) were designed using CCTop[59] (https://cctop.cos.uni-heidelberg.de/), and cloned via BsmBI into the SGL40C.EFS.dTomato (Addgene 89395) or SGL40C.EFS.E2Crimson (100894) backbone. Short hairpin RNAs (shRNAs) were designed

using the Adams *et al.* miR-N tool[60] (https://felixfadams.shinyapps.io/miRN/) and cloned via BsmBI into the SIN40C.SFFV.GFP.miR30n (169278) backbone. Non-targeting control sgRNAs and shRNAs were designed against firefly luciferase. *MYNRL15* cDNAs (transcript IDs: ENST00000448987.1 and ENST00000438890.1; long and short isoforms respectively) were expressed from the bidirectional LBid.lnc.GFP[61,62]. 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 or pLKO5d.EFS.SpCas9.P2A.BSD (90332 and 57821 respectively). Stable PDXs 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 protein-coding), and SGL.EFS.dTomato.P2A.PAC (173914; CTCF-enriched loci).

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).

Individual sgRNA and shRNA sequences are listed in Supplementary Tables 2-3. All plasmids have been deposited in Addgene.

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[63]. Fresh LNA-GapmeR was added every 2 days to maintain its

concentration in the culture media. LNA-GapmeR sequences can be found in Supplementary Table 4.

CRISPR library sgRNA design

Guides for the CRISPRi-based targeting of HSPC/AML lncRNAs were designed using the standalone version of CCTop[59] (https://cctop.cos.uni-heidelberg.de/). In brief, the lncRNA genes were annotated using GENCODE v25 (release 03/2016)[64], LNCipedia 4.0 (release 05/2016)[65], and NONCODE v4 (release 01/2014)[66] as previously described[16], and sgRNAs were selected 0 bp to 250 bp downstream of transcription start sites (TSSs)[67]. 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[68] (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[59] (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[69] score<0.4) were excluded.

Guides targeting CTCF sites in CTCF-enriched gene loci were selected using GuideScan[70] (http://www.guidescan.com/) and CRISPick[71,72] (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/Cas9 libraries used in this study can be found in Supplementary Tables 7-10.

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 protein-coding), or SGL.EFS.dTomato.P2A.PAC (173914; CTCF-enriched loci). XL1-Blue supercompetent cells (Agilent 200236) were used for transformation, and subse-

quently plated on 15 cm LB agar plates containing ampicillin. Colonies were counted 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. Genomic DNA was isolated from cells via the QIAmp DNA Blood Mini Kit (Qiagen) at the beginning and end of the screen, and the sgRNA cassettes were PCR amplified using NEBNext® High-Fidelity 2x PCR Master Mix (New England Biolabs) and primers containing the Illumina P5 and P7 adapter sequences as overhangs. The 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)[73] 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[74] (Bioconductor) to combine replicates and perform pan-cell line analysis.

Hematopoietic assays with primary cells

$CD34^+$ HSPCs were thawed and expanded in StemSpan SFEM (STEMCELL Technologies) containing 1% penicillin/streptomycin (Gibco™), 100 ng/ml SCF, 100 ng/ml FLT3L, 20 ng/ml IL6, 50 ng/ml TPO (cytokines from Peprotech), and 750nM SR1 (STEMCELL Technologies) for 2 days prior to transduction. Cells were transduced in the presence of 4 µg/ml Polybrene (Sigma-Aldrich) on RetroNectin®-coated plates (TaKaRa), using two consecutive rounds of super-concentrated virus ~4 hours apart. Four days post-transduction, HSPCs were sorted and plated in human methylcellulose complete medium HSC003 (R&D Systems) for colony-forming assays. Fifteen thousand cells were initially plated across two 6 mm dishes. The colonies were counted once they had reached a sufficient size (10-14 days).

For assays using patient-derived AML blasts, *in vivo* expanded PDXs were thawed and pre-cultured in StemSpan SFEM (STEMCELL Technologies) containing 1% penicillin/streptomycin (Gibco™), 50 ng/ml SCF, 50 ng/ml FLT3L, 10 ng/ml IL6, 2.5 ng/ml IL3, 10 ng/ml TPO (cytokines from Peprotech), 750 nM SR1 (STEMCELL Technologies), and 35 nM UM171 (STEMCELL Technologies) for 24 to 48 hours. Transductions were carried out in the presence of 2 µg/ml Polybrene (Sigma-Aldrich). The cells were harvested 48 hours post-transduction for xenotransplantation into mice or for colony-forming assays.

Mice and transplantation experiments

Two-color *in vivo* competition experiments were performed in murine xenograft models of AML as previously described[61,75]. 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 peripheral blood every 4 weeks. The mice were sacrificed upon leukemia onset, and cells harvested from the hematopoietic organs (bone marrow, spleen, and liver) were analyzed by flow cytometry. All mice were housed in a specific pathogen-free environment. All animal procedures were approved by the local state authorities (Landesverwaltungsamt Niedersachsen/Sachsen-Anhalt).

Flow cytometry and cell sorting

Flow cytometry analyses were performed on a CytoFLEX B4-R3-V5 or CytoFLEX S V4-B2-Y4-R3 (Beckman Coulter). Cell sorting was performed on a FACSAria™ II or FACSMelody™ (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.

Gene expression analyses

RNA was isolated from cells using the Quick-RNA™ Microprep or Miniprep Kits (Zymo Research). RNA fractionation was performed as previously described[76], except that we lysed the nuclear pellet directly rather than isolating the nuclear-soluble and chromatin-associated fractions separately. The TURBO DNA-free™ Kit (Invitrogen) was used for DNase treatment. cDNA was synthesized using the High-Capacity cDNA Reverse Transcription Kit, and gene expression was quantified via real-time PCR using SYBR™ Select Master Mix and specific primers on a StepOnePlus™ Real-Time PCR cycler (all products from Applied Biosystems). *B2M* was used as a housekeeping control. Primer sequences for qRT-PCR can be found in Supplementary Table 5. To detect *WDR61* and *IMP3*, we used QuantiTect® primer assays (Qiagen QT00083776 and QT00232330 respectively).

RNA sequencing was performed by Novogene Company, Ltd. on an Illumina NovaSeq 6000 (150 bp paired-end reads) using polyA-enriched total cellular RNA. 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[77], and gene expression was quantified using featureCounts[78] in R. The processed count data were subsequently analyzed in R using DESeq2[74] (Bioconductor). Gene sets from MSigDB v7.2 (H1, C2, C3, C6), custom hematopoietic[16] and chromosome 15 gene sets, and PAF1c-knockout gene expression signatures[33] were checked for enrichment via the Broad GSEA software[79]. The custom positional gene sets were generated by walking a 1 Mb or 5 Mb window along chromosome 15 and taking expressed genes within the windows.

NG Capture-C

Chromatin conformation capture with selective enrichment for *MYNRL15*-interacting sequences was performed using next generation (NG) Capture-C as previously described[30], with the following modifications: (1) Five to ten million transduced cells were harvested per sample and the DpnII digestion reactions were scaled down accordingly. Transduced K562 or ML-2 cells were used to assess the effects of *MYNRL15* perturbation. *In vitro* expanded

CD34$^+$ HSPCs (day 3) were used to assess the native conformation of the locus. (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[80] (refer to Supplementary Table 6 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[30]). The raw sequence data were processed with the capC-MAP package[81] 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[82] using a smoothing window of 2 pixels, alongside CTCF ChIP-seq data from K562 cells (ENCODE accession no. ENCFF519CXF) and Knight-Ruiz matrix-balanced[83] Micro-C[84] data from H1-hESC cells. Hi-C data from Rao *et al*.[85] 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[86,87]. 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[88] and clean reads were aligned to hg19 with BWA[89]. Mitochondrial reads were removed prior to subsequent analysis. Normalized pileups were generated using deepTools[90] and viewed in the Integrated Genomics Viewer (IGV)[91].

CTCF CUT&RUN

We performed CUT&RUN as previously described[92,93]. 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[93], 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[94] to remove adapter sequences, followed by Kseq[95] to trim reads containing ≤6 bp of adapter sequence, which are not effectively handled by Trimmomatic. Trimmed reads were aligned to hg38 using bowtie2[96]. The resulting SAM files were converted into BAM format and sorted and indexed using Samtools[97]. Normalized bigWig tracks were generated using bamCoverage from deepTools[90]. The processed data were viewed in the Integrated Genomics Viewer (IGV)[91].

ENCODE datasets

The following K562 ChIP-seq data were used in this study: CTCF (ENCFF519CXF), SMC3 (ENCFF175UEE), H3K27Ac (ENCFF469JMR), H3K4Me1 (ENCFF100FDI), and H3K4Me3 (ENCFF767UON). In addition, the following CTCF ChIP-seq data from other cell lines and primary cells were also used: GM12878 (ENCFF960ZGP), H1-hESC (ENCFF821AQO), A549 (ENCFF535MZG), HeLa-S3 (ENCFF502CZS), IMR90 (ENCFF307XFM), MCF-7 (ENCFF867BUQ), HCT-116 (ENCFF171SNH), HEK293 (ENCFF285QVL), HL-60 (ENCFF432AMS), NB4 (ENCFF456PDQ), CD14+ monocyte (ENCFF300XXC), B cell (ENCFF910TER), neutrophil (ENCFF122IMV), fibroblast lung (ENCFF777ODE), cardiac muscle cell (ENCFF301YXM), kidney epithelial cell (ENCFF674KUN), and osteoblast (ENCFF744PXO).

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)[64] 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[98]. Within each gene, the number of CTCF binding sites was counted and normalized by gene length. Log10-transformed values of this metric followed an approximately normal distribution; thus, we defined elevated CTCF density as over two 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 Supplementary Table 11 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[36] and NCI-TARGET[35] 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 with-

out: 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[99] and survminer[100] packages (base R). DESeq2 (Bioconductor) was used to normalize and variance-stabilize read count data[74] from the TCGA[36] and NCI-TARGET[35] AML cohorts. The NCI-TARGET dataset also required batch correction, for which we used sva[101] (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[102] (base R).

Statistical analyses and definitions

Statistical evaluations of experimental data were carried out in GraphPad Prism 8 using unpaired, two-tailed t tests. Data are presented as mean ± s.d. or s.e.m. as indicated in the figure legends. Statistical analysis of gene expression data (RNA-seq) was performed in R using DESeq2 (Bioconductor). Survival analysis was also done in R using the survival and survminer packages (base R). CRISPR-Cas9 screening data were analyzed with the MAGeCK suite, with the exception of the tiling screens, which were analyzed in R using DESeq2. In all cases, measurements were taken on at least two biological replicates and P<0.05 was considered significant. Sample sizes are indicated in the figure legends. No statistical methods were used to predetermine sample size.

Data availability

All RNA-seq, Capture-C, CUT&RUN, and ATAC-seq data have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE172240. The CRISPR-Cas9 screening data have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under the accession numbers PRJEB44308 and PRJEB44320.

Code availability

All computational tools used in this study are publically available. Please refer to the corresponding Methods sections for links or references to the relevant publications. R scripts are available from the authors upon request.

References

1 Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

2 Encode Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710, doi:10.1038/s41586-020-2493-4 (2020).

3 Fang, S. *et al.* NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res* **46**, D308-D314, doi:10.1093/nar/gkx1107 (2018).

4 Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* **47**, 199-208, doi:10.1038/ng.3192 (2015).

5 Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigo, R. & Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet* **19**, 535-548, doi:10.1038/s41576-018-0017-y (2018).

6 Yao, R. W., Wang, Y. & Chen, L. L. Cellular functions of long noncoding RNAs. *Nat Cell Biol* **21**, 542-551, doi:10.1038/s41556-019-0311-8 (2019).

7 Slack, F. J. & Chinnaiyan, A. M. The Role of Non-coding RNAs in Oncology. *Cell* **179**, 1033-1055, doi:10.1016/j.cell.2019.10.017 (2019).

8 Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22**, 96-118, doi:10.1038/s41580-020-00315-9 (2021).

9 Alvarez-Dominguez, J. R. & Lodish, H. F. Emerging mechanisms of long noncoding RNA function during normal and malignant hematopoiesis. *Blood* **130**, 1965-1975, doi:10.1182/blood-2017-06-788695 (2017).

10 Beck, D. *et al.* A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. *Leukemia* **32**, 263-272, doi:10.1038/leu.2017.210 (2018).

11 Bill, M. *et al.* Expression and functional relevance of long non-coding RNAs in acute myeloid leukemia stem cells. *Leukemia* **33**, 2169-2182, doi:10.1038/s41375-019-0429-5 (2019).

12 De Clara, E. *et al.* Long non-coding RNA expression profile in cytogenetically normal acute myeloid leukemia identifies a distinct signature and a new biomarker in NPM1-mutated patients. *Haematologica* **102**, 1718-1726, doi:10.3324/haematol.2017.171645 (2017).

13 Delas, M. J. *et al.* lncRNA requirements for mouse acute myeloid leukemia and normal differentiation. *Elife* **6**, doi:10.7554/eLife.25607 (2017).

14 Garzon, R. *et al.* Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci U S A* **111**, 18679-18684, doi:10.1073/pnas.1422050112 (2014).

15 Papaioannou, D. *et al.* Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia. *Haematologica* **102**, 1391-1400, doi:10.3324/haematol.2017.166215 (2017).

16 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).

17 Anderson, K. M. *et al.* Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* **539**, 433-436, doi:10.1038/nature20128 (2016).

18 Cho, S. W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**, 1398-1412 e1322, doi:10.1016/j.cell.2018.03.068 (2018).

19 Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452-455, doi:10.1038/nature20149 (2016).

20 Groff, A. F. *et al.* In Vivo Characterization of Linc-p21 Reveals Functional cis-Regulatory DNA Elements. *Cell Rep* **16**, 2178-2186, doi:10.1016/j.celrep.2016.07.050 (2016).

21 Paralkar, V. R. *et al.* Unlinking an lncRNA from Its Associated cis Element. *Mol Cell* **62**, 104-110, doi:10.1016/j.molcel.2016.02.029 (2016).

22 Kopp, F. & Mendell, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**, 393-407, doi:10.1016/j.cell.2018.01.011 (2018).

23 Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet* **21**, 102-117, doi:10.1038/s41576-019-0184-5 (2020).

24 Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med* **17**, 1086-1093, doi:10.1038/nm.2415 (2011).

25 Jung, N., Dai, B., Gentles, A. J., Majeti, R. & Feinberg, A. P. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nat Commun* **6**, 8489, doi:10.1038/ncomms9489 (2015).

26 Krivtsov, A. V. *et al.* Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia* **27**, 852-860, doi:10.1038/leu.2012.363 (2013).

27 Diffner, E. *et al.* Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. *Blood* **121**, 2289-2300, doi:10.1182/blood-2012-07-446120 (2013).

28 Aguirre, A. J. *et al.* Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* **6**, 914-929, doi:10.1158/2159-8290.CD-16-0154 (2016).

29 Joung, J. *et al.* Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* **548**, 343-346, doi:10.1038/nature23451 (2017).

30 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).

31 Milne, T. A. *et al.* Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. *Mol Cell* **38**, 853-863, doi:10.1016/j.molcel.2010.05.011 (2010).

32 Muntean, A. G. *et al.* The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. *Cancer Cell* **17**, 609-621, doi:10.1016/j.ccr.2010.04.012 (2010).

33 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).

34 Granneman, S. *et al.* The human Imp3 and Imp4 proteins form a ternary complex with hMpp10, which only interacts with the U3 snoRNA in 60-80S ribonucleoprotein complexes. *Nucleic Acids Res* **31**, 1877-1887, doi:10.1093/nar/gkg300 (2003).

35 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).

36 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).

37 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).

38 Liu, Y. *et al.* Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat Biotechnol*, doi:10.1038/nbt.4283 (2018).

39 Zhu, S. *et al.* Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat Biotechnol* **34**, 1279-1286, doi:10.1038/nbt.3715 (2016).

40 Kirk, J. M. *et al.* Functional classification of long non-coding RNAs by k-mer content. *Nat Genet* **50**, 1474-1482, doi:10.1038/s41588-018-0207-8 (2018).

41 Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* **17**, 601-614, doi:10.1038/nrg.2016.85 (2016).

42 Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).

43 Despang, A. *et al.* Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet* **51**, 1263-1271, doi:10.1038/s41588-019-0466-z (2019).

44 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).

45 Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nat Rev Genet* **22**, 154-168, doi:10.1038/s41576-020-00303-x (2021).

46 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922, doi:10.1016/j.cell.2017.05.004 (2017).

47 Sartorelli, V. & Lauberth, S. M. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* **27**, 521-528, doi:10.1038/s41594-020-0446-0 (2020).

48 Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458, doi:10.1126/science.aad9024 (2016).

49 Akdemir, K. C. *et al.* Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet* **52**, 294-305, doi:10.1038/s41588-019-0564-y (2020).

50 Katainen, R. *et al.* CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet* **47**, 818-821, doi:10.1038/ng.3335 (2015).

51 Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369-381, doi:10.1016/j.cell.2014.02.019 (2014).

52 Chen, C. *et al.* Spatial Genome Re-organization between Fetal and Adult Hematopoietic Stem Cells. *Cell Rep* **29**, 4200-4211 e4207, doi:10.1016/j.celrep.2019.11.065 (2019).

53 Qi, Q. *et al.* Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. *Blood* **137**, 1327-1339, doi:10.1182/blood.2020005780 (2021).

54 Takayama, N. *et al.* The Transition from Quiescent to Activated States in Human Hematopoietic Stem Cells Is Governed by Dynamic 3D Genome Reorganization. *Cell Stem Cell* **28**, 488-501 e410, doi:10.1016/j.stem.2020.11.001 (2021).

55 Hu, G. *et al.* Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity* **48**, 227-242 e228, doi:10.1016/j.immuni.2018.01.013 (2018).

56 Kloetgen, A. *et al.* Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat Genet* **52**, 388-400, doi:10.1038/s41588-020-0602-9 (2020).

57 Luo, H. *et al.* CTCF boundary remodels chromatin domain and drives aberrant HOX gene transcription in acute myeloid leukemia. *Blood* **132**, 837-848, doi:10.1182/blood-2017-11-814319 (2018).

58 Vilarrasa-Blasi, R. *et al.* Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat Commun* **12**, 651, doi:10.1038/s41467-020-20849-y (2021).

59 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).

60 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).

61 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).

62 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).

63 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).

64 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).

65 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).

66 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).

67 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).

68 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).

69 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).

70 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).

71 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).

72 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).

73 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).

74 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).

75 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).

76 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).

77 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).

78 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).

79 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).

80 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).

81 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).

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

83 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).

84 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).

85 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).

86 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).

87 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).

88 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).

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

90 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).

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

92 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).

93 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).

94 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).

95 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).

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

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

98 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).

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

100 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).

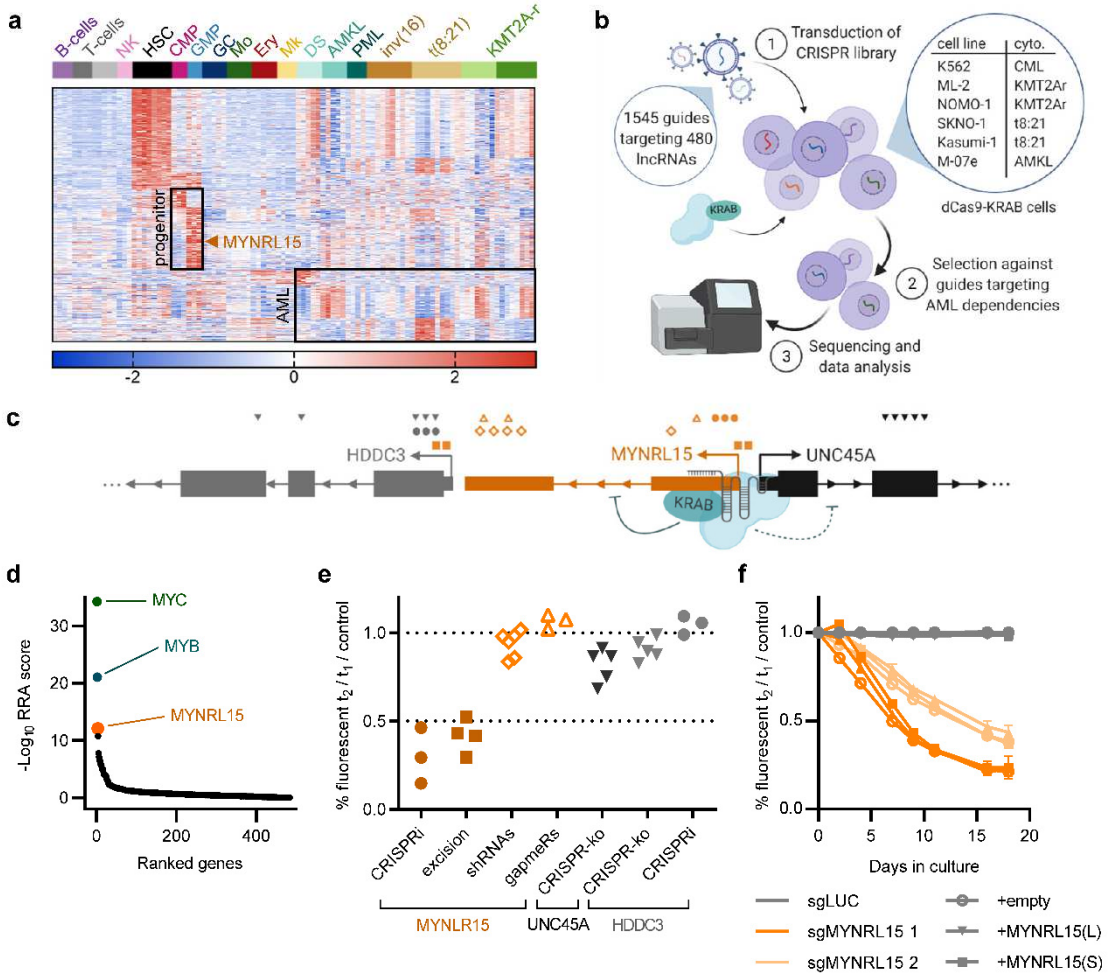
101 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883, doi:10.1093/bioinformatics/bts034 (2012).

102 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).

103 Hsieh, T. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat Methods* **13**, 1009-1011, doi:10.1038/nmeth.4025 (2016).

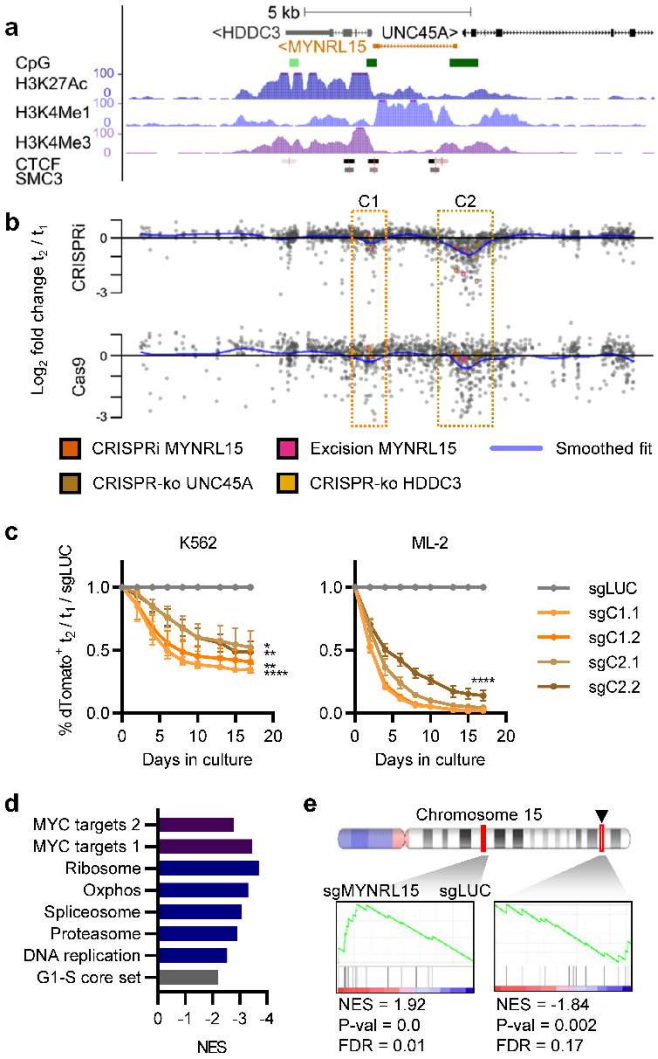
Figures

Fig. 1: CRISPRi screen of HSPC/AML lncRNA signatures identifies *MYNRL15* as a myeloid leukemia dependency.



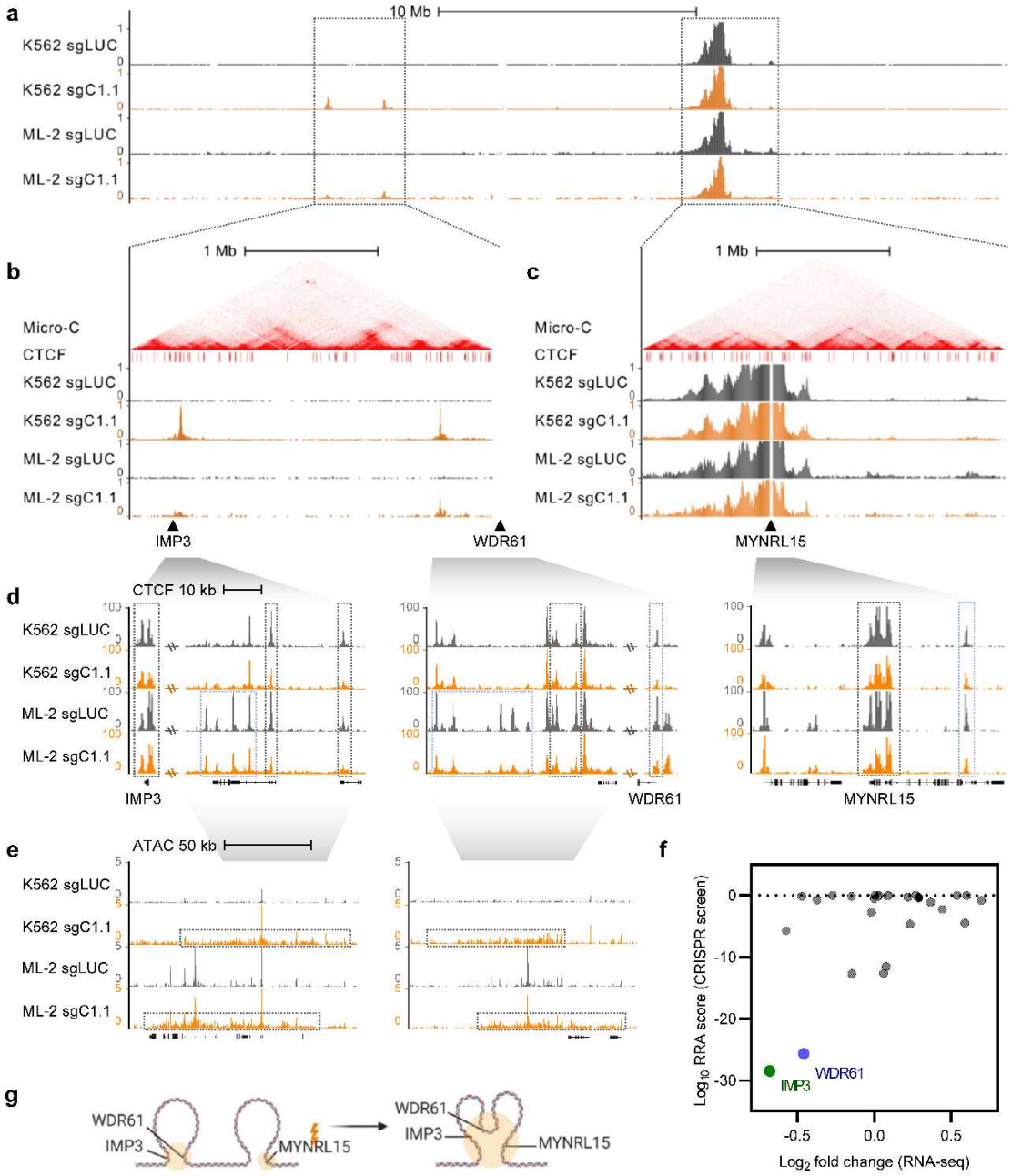
a, Expression of HSPC/AML lncRNAs across 12 normal blood cell populations and 46 pediatric AML samples[16]. Signatures of particular therapeutic interest are outlined. Natural killer cell (NK), hematopoietic stem cell (HSC), common myeloid progenitor (CMP), granulocyte-monocyte progenitor (GMP), granulocyte (GC), monocyte (Mo), erythroid precursor (Ery), megakaryocyte (Mk), myeloid leukemia of Down syndrome (DS), non-DS megakaryoblastic leukemia (AMKL), promyelocytic leukemia (PML), *KMT2A*-rearranged leukemia (*KMT2A*-r). **b**, Workflow for screening HSPC/AML lncRNAs. **c**, Schematic of the *MYNRL15* locus (not to scale), including target sites of the different perturbation approaches used. Target gene: *MYNRL15* (orange), *UNC45A* (black), *HDDC3* (grey). Perturbation strategy: CRISPRi (filled circle), dual sgRNA mediated excision (filled square), RNAi (empty diamond), LNA-gapmeR (empty triangle), CRISPR-Cas9 knockout (filled triangle). **d**, Pan-cell line analysis of the CRISPRi screen identifies *MYNRL15* as the top hit behind the positive controls *MYC* and *MYB*. **e**, Endpoint depletion values from proliferation assays using different perturbation strategies in ML-2 cells. Each point corresponds to a vector used for perturbation (mean of n=3 shown). **f**, Proliferation assays using cDNA overexpression to rescue the *MYNRL15* CRISPRi depletion phenotype (n=2; mean ± s.e.m.). Long isoform (L), short isoform (S).

Figure 2: Functional dissection of the *MYNRL15* locus reveals crucial regulatory regions.



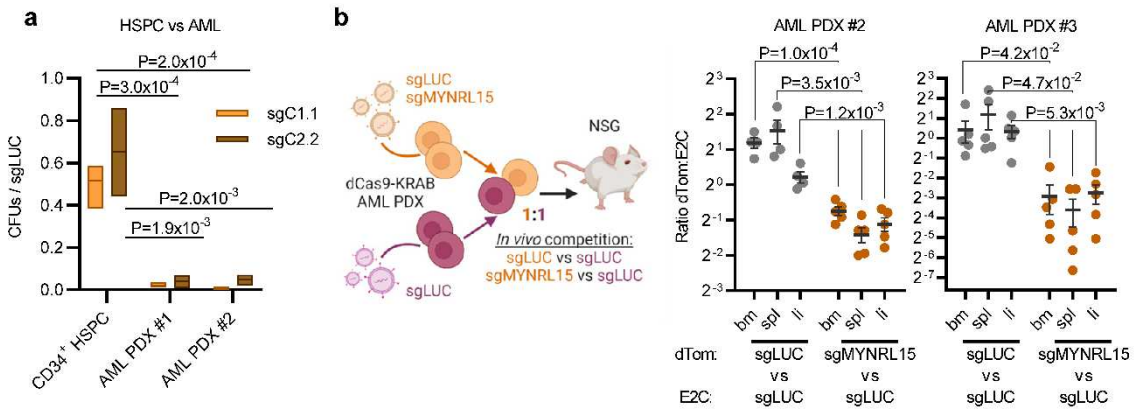
a, Tracks from the UCSC Genome Browser showing from top to bottom: gene annotations, CpG islands, histone marks, and CTCF and cohesin occupancy (ChIP-seq from ENCODE). **b**, Tiling screens of the *MYNRL15* locus using complementary CRISPRi (top) and CRISPR-Cas9 based (bottom) strategies (mean of 4 cell lines; n=2 per cell line). Pre-tested sgRNAs are depicted in color. A smoothed fit curve is shown in blue. The two identified leukemia-essential cCREs are outlined. **c**, Individual proliferation assays using sgRNAs from C1 and C2 for CRISPR-Cas9 based perturbation of the *MYNRL15* locus (n=3, mean ± s.e.m.; 2 guides per cCRE). * P<0.05; ** P<0.01; **** P<0.0001; where only one set of asterisks is shown, all conditions shared the same P-value. **d-e**, GSEA comparing *MYNRL15* perturbation (using 4 guides, 2 per cCRE; “sgMYNRL15”) to the non-targeting control (n=2 per guide; “sgLUC”) in combined analyses of K562 and ML-2 cells. **d**, Normalized enrichment scores (NES) of cancer dependency gene sets upon *MYNRL15* perturbation. **e**, Two chromosome 15 gene sets that are downregulated upon *MYNRL15* perturbation.

Figure 3: *MYNRL15* perturbation alters genome architecture on chromosome 15, leading to downregulation of *WDR61* and *IMP3*.



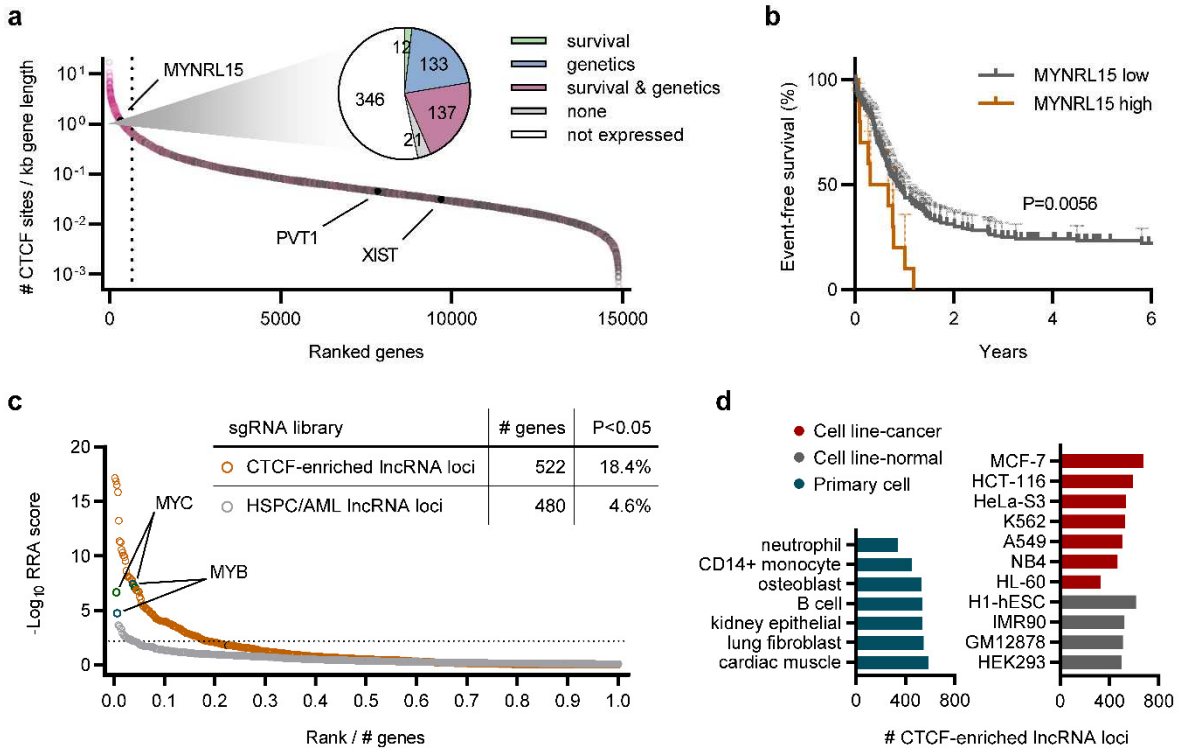
a, NG Capture-C interaction profiles on chromosome 15 in K562 and ML-2 cells, using one guide targeting *MYNRL15* (sgC1.1) and a non-targeting control (sgLUC) (n=2; viewpoint in C1; smoothing window 2 pixels). **b-c**, Close-ups of the gained interaction region and the region surrounding *MYNRL15*, with ENCODE CTCF ChIP-Seq and Micro-C[103] tracks from the UCSC Genome Browser. **d**, CTCF occupancy at the *MYNRL15* locus and distal interaction sites, as determined via CUT&RUN. The *MYNRL15*, *IMP3*, and *WDR61* loci, and gained interaction sites are outlined with black boxes. Note the track discontinuity in the left and center views. **e**, Chromatin accessibility at the gained interaction sites, as determined via ATAC-seq. **f**, Integrative analysis depicting CRISPR-Cas9 screening scores and differential expression upon *MYNRL15* perturbation for the 29 coding genes in the gained interaction region. **g**, Model of chromosome 15 reorganization following *MYNRL15* perturbation.

Figure 4: AML specificity and therapeutic potential of *MYNRL15*.



a, Direct comparison of the impact of *MYNRL15* perturbation on colony-forming capacity in CD34⁺ HSPCs and two AML PDXs (HSPC, n=3; PDXs, n=4 each; data presented as mean ± range). **b**, Setup (left) and results (right) of direct two-color *in vivo* competition assays testing CRISPRi mediated perturbation of *MYNRL15* in AML PDXs. The data are presented as ratios of dTomato⁺ (dTom) to E2Crimson⁺ (E2C) cells in the bone marrow (bm), spleen (spl), and liver (li) of transplanted mice at the experimental endpoint (n=4 in the AML PDX #2 control group; otherwise n=5).

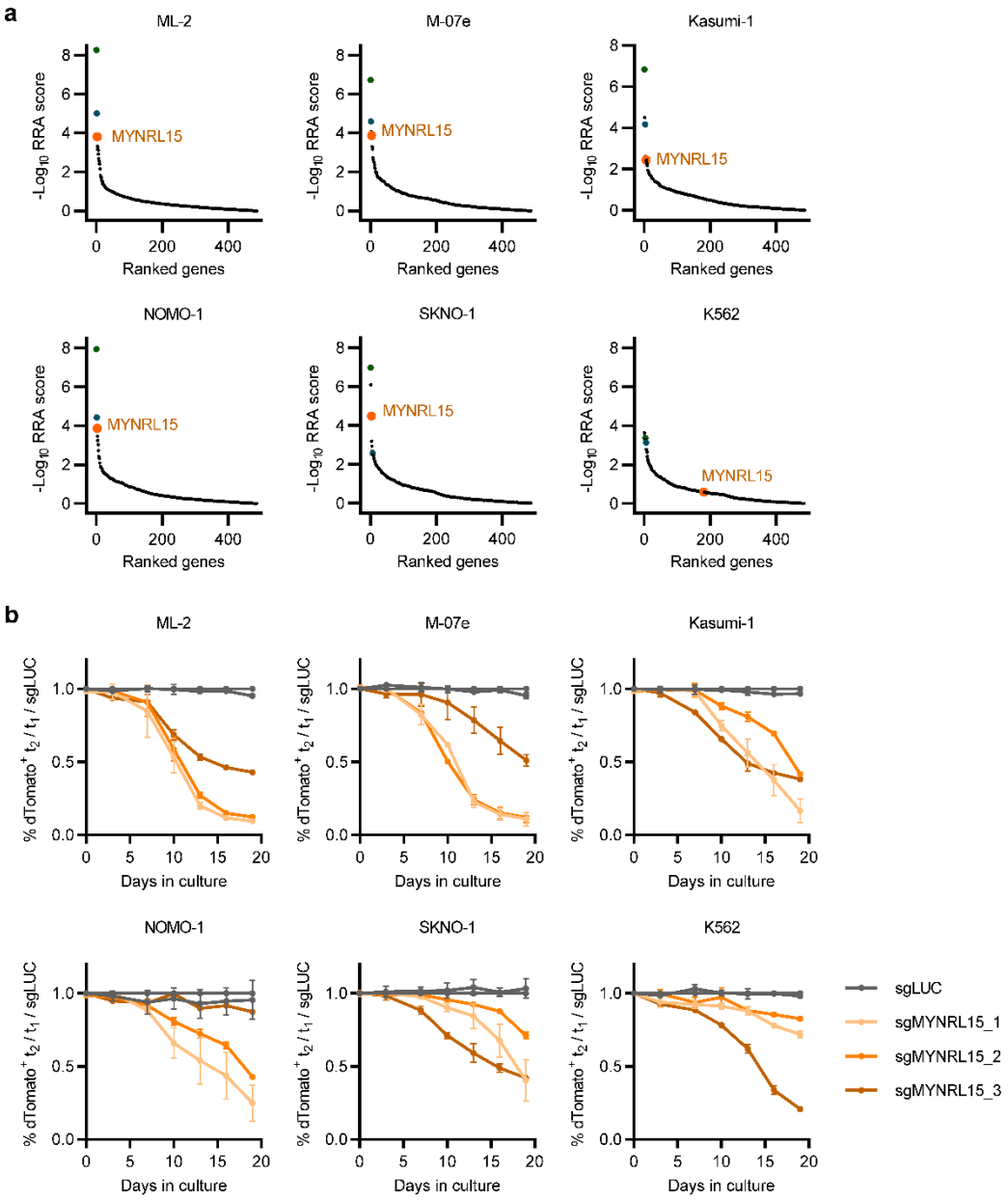
Figure 5: A subclass of clinically relevant, functionally validated CTCF-enriched lncRNA loci.



a, Ranked list of protein-coding and lncRNA loci ordered by CTCF density. The positions of *MYNRL15*, and the *bona fide* lncRNAs *PVT1* and *XIST* are marked. Non-CTCF-bound loci are not shown. The dashed line indicates the cut-off defined for elevated CTCF binding (i.e. median + 2 s.d. of $\log_{10}$-transformed values). Inset: breakdown of CTCF-enriched loci based on their association with clinical characteristics like cytogenetics, mutations, and survival in the TCGA AML cohort. **b**, Kaplan-Meier survival curves of patients with high (n=10) versus low (n=171) expression of *MYNRL15* in the TCGA AML cohort[36] (survival probability ± 95% C.I.). Five-year event-free survival: 23.2% vs 0.0%. **c**, Ranked list of CTCF-enriched lncRNA loci ordered by essentiality, as determined via CRISPR-Cas9 screening and MAGeCK analysis (orange). Results from our initial CRISPRi lncRNA library are displayed alongside (grey) for comparison. Gene ranks are normalized to library size. The positive controls *MYC* and *MYB* are indicated. **d**, Numbers of CTCF-enriched lncRNA loci (C-LNC) in 18 different cell lines and primary cell types.

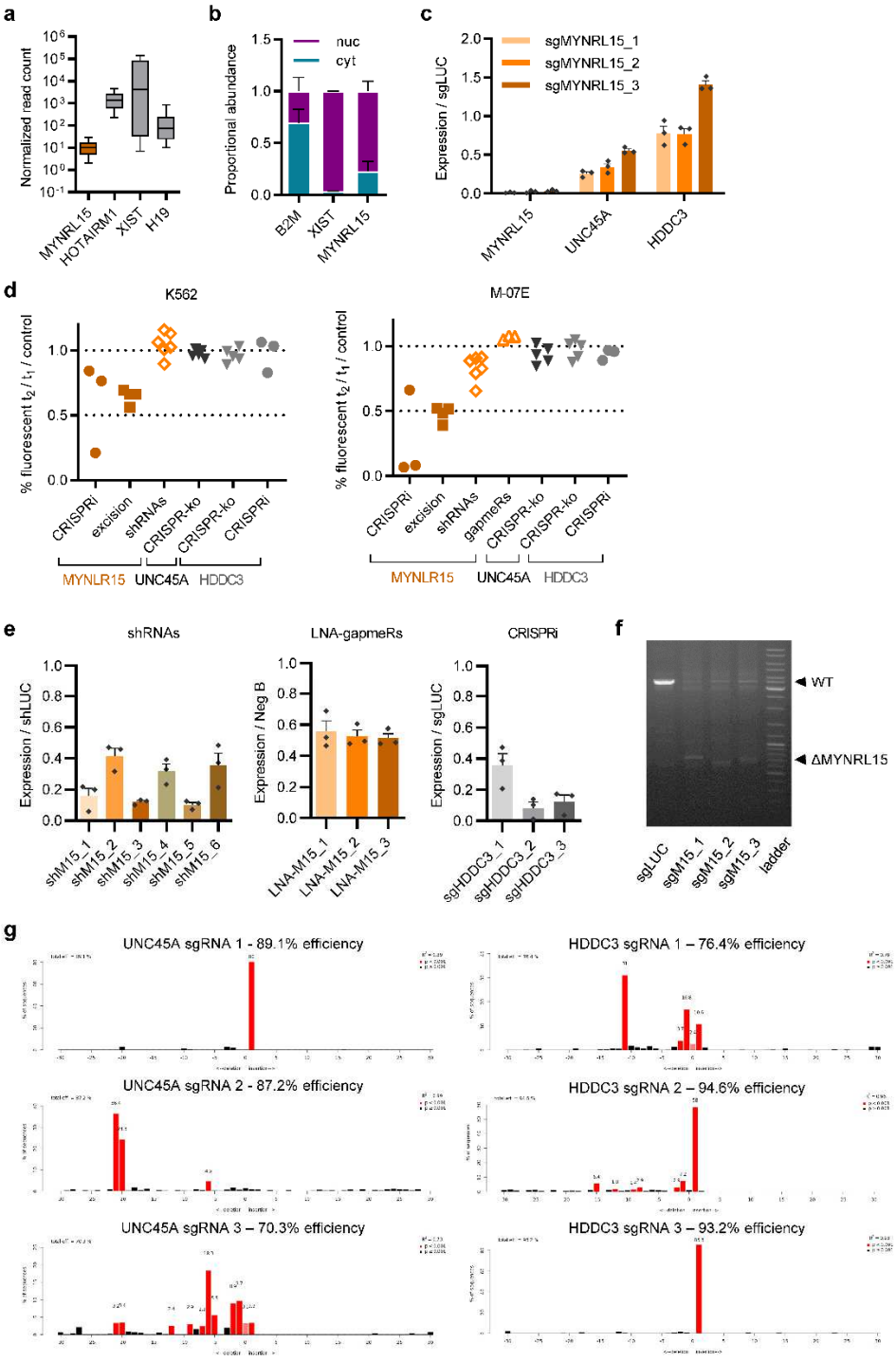
Extended data figures

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



a, Ranked lists from MAGeCK analysis identifying *MYNRL15* (orange) among the top hits from the CRISPRi screen in all cell lines. The positive control genes, *MYC* (green) and *MYB* (turquoise), are also colored. **b**, Individual validation of guides targeting *MYNRL15* via proliferation assays (n=2; mean ± s.e.m.).

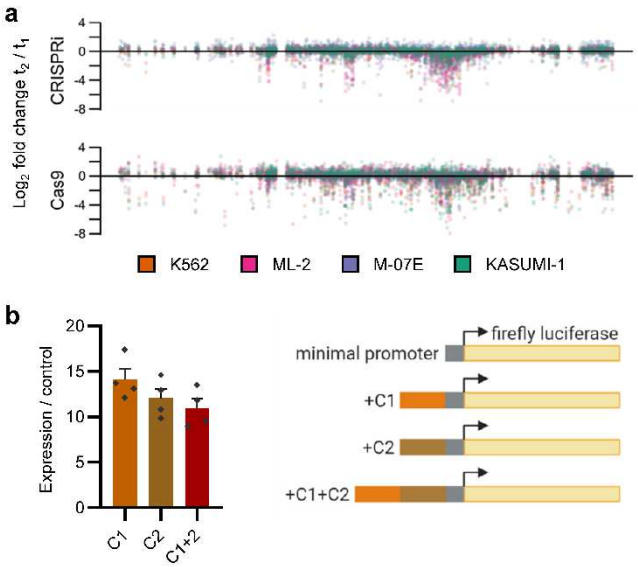
Extended Data Figure 2: Further validation of *MYNRL15* and perturbation approaches.



a, Expression of *MYNRL15* compared to *bona fide* lncRNAs in the NCI-TARGET pediatric AML cohort[35] (n=258; zeros omitted). Midline, median; box limits, lower and upper quartiles; whiskers, 10% and 90% quantiles. **b**, Subcellular localization of *MYNRL15* compared to control transcripts *XIST* (nuclear) and *B2M* (cytoplasmic), determined via fractionated qRT-PCR (n=2). **c**, Targeting the CRISPRi system to the *MYNRL15* TSS causes concurrent *UNC45A* knockdown, as determined by qRT-PCR (n=3). **d**, Endpoint depletion values from proliferation assays using different perturbation strategies in K562 (left) or M-07E (right) cells. Each point corresponds to one vector that was used for perturbation (mean of n=3). **e**, qRT-PCR validations of *MYNRL15* knockdown using RNAi (left; n=3) and LNA-gapmeRs (center; n=3),

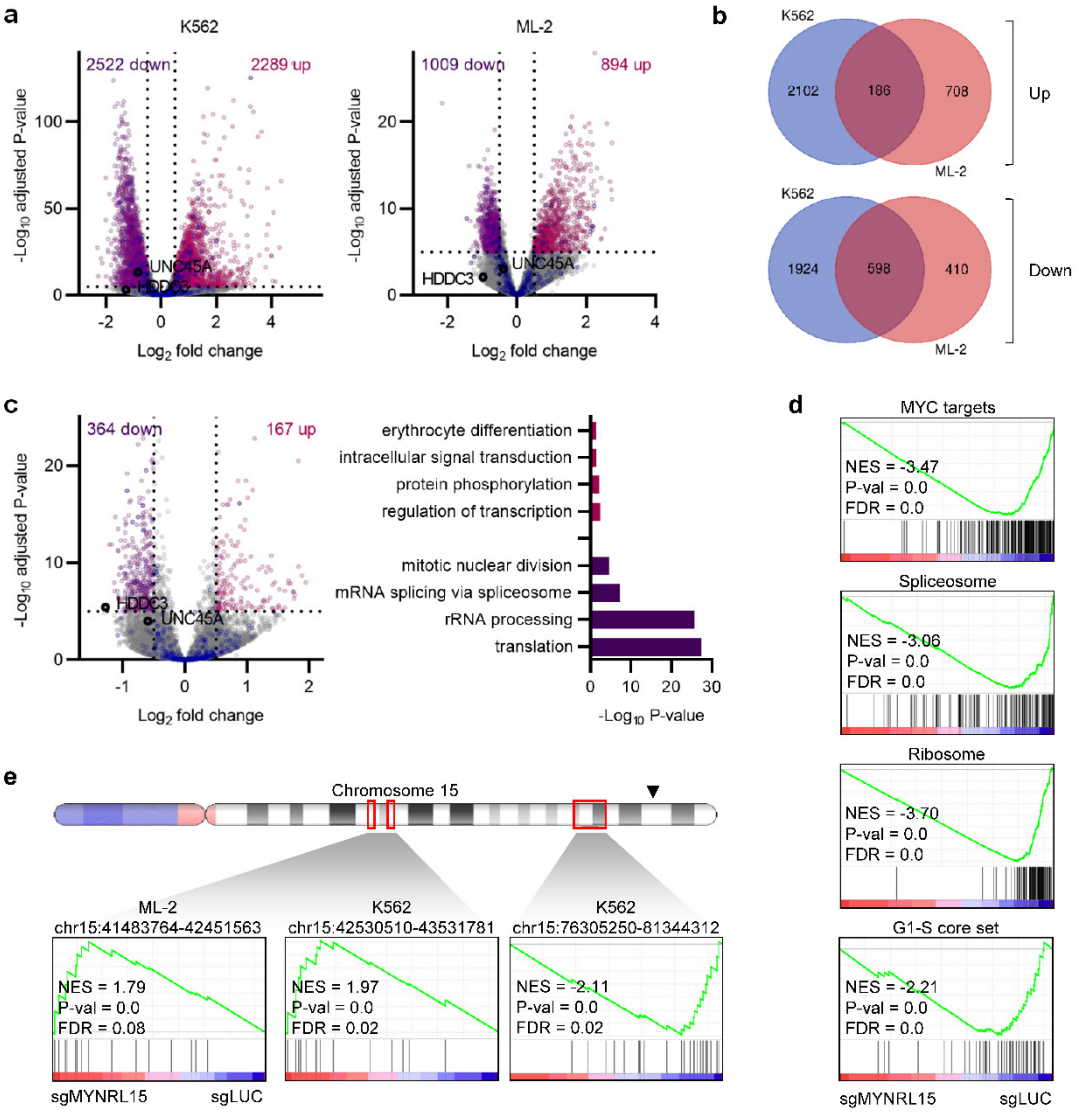
and of *HDDC3* knockdown via CRISPRi (right; n=3) **f**, PCR validation of *MYNRL15* excision, using bulk genomic DNA isolated from cells transduced with dual sgRNA vectors. **g**, Representative TIDE analyses showing the cutting efficiencies of guides targeting *UNC45A* and *HDDC3*. Where error bars are shown, data are presented as mean ± s.e.m.

Extended Data Figure 3: Extended data from the *MYNRL15* tiling screen.



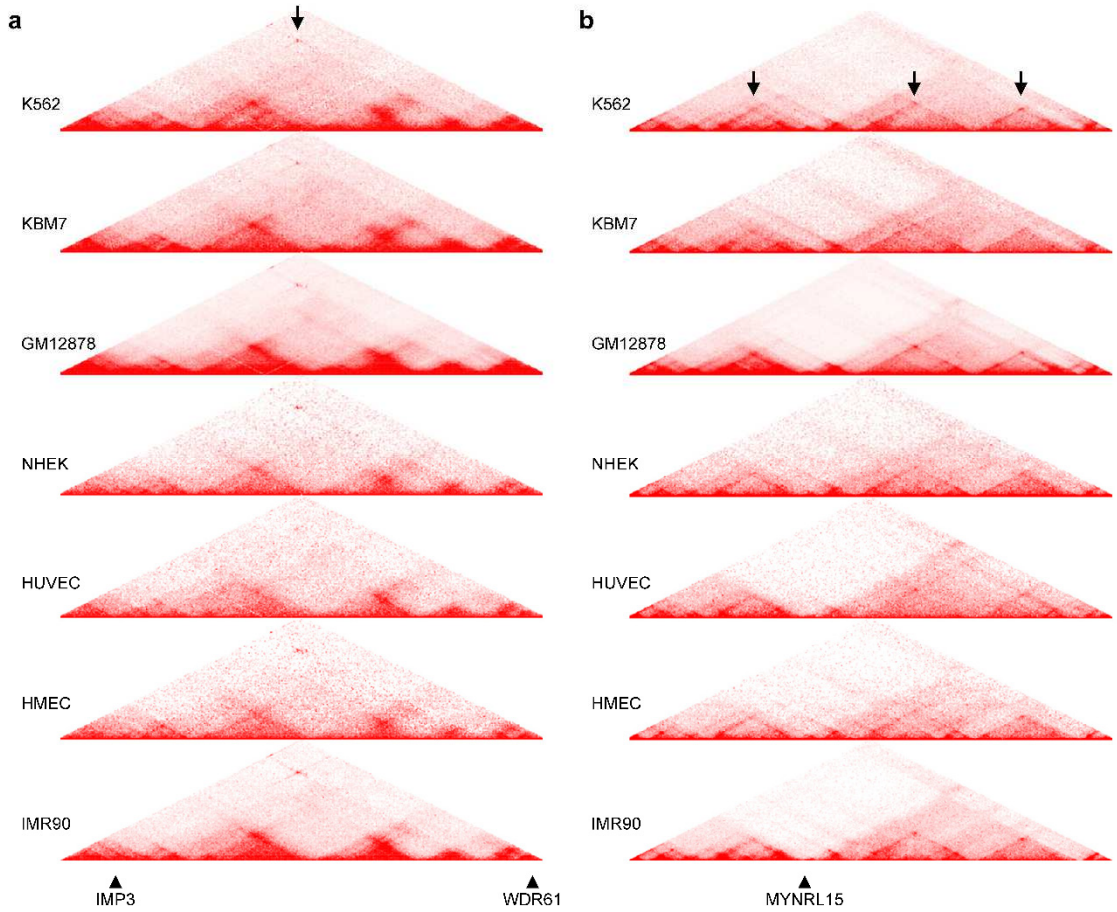
a, Tiling screens of the *MYNRL15* locus using CRISPRi (top) and CRISPR-Cas9 (bottom). The four tested cell lines are shown in different colors (mean of n=2 per cell line). **b**, Dual luciferase assays in K562 cells, using reporter constructs containing C1 and/or C2 upstream of a minimal promoter (n=4; mean ± s.e.m.).

Extended Data Figure 4: Global gene expression profiling upon *MYNRL15* perturbation.



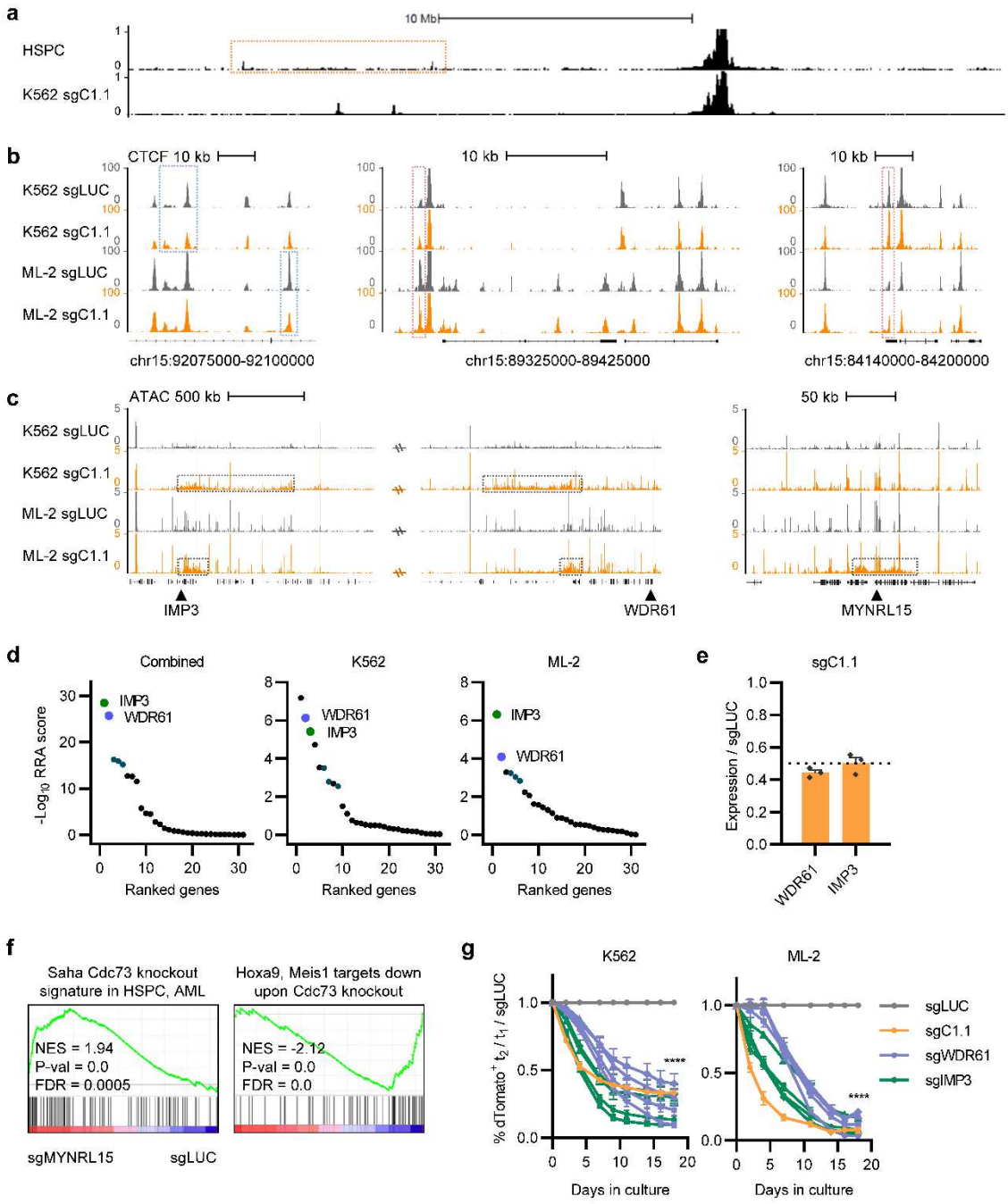
a, Differentially expressed genes following *MYNRL15* perturbation in K562 (left) and ML-2 (right) cells. Up- (pink) and downregulated (purple), and chromosome 15 genes (navy) are shown in color. **b**, Overlap of up- and downregulated genes in K562 and ML-2 cells. **c**, Combined differential expression analysis of K562 and ML-2 cells (left), and enriched gene ontology terms in the differentially expressed gene lists (right). **d**, A selection of the most significantly dysregulated 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 locus is indicated by a black arrow. All analyses compare *MYNRL15* perturbation (4 guides, two per cCRE; “sgMYNRL15”) to the non-targeting control (n=2 per guide; “sgLUC”).

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



a-b, Hi-C maps from 7 different cell lines[85] 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.

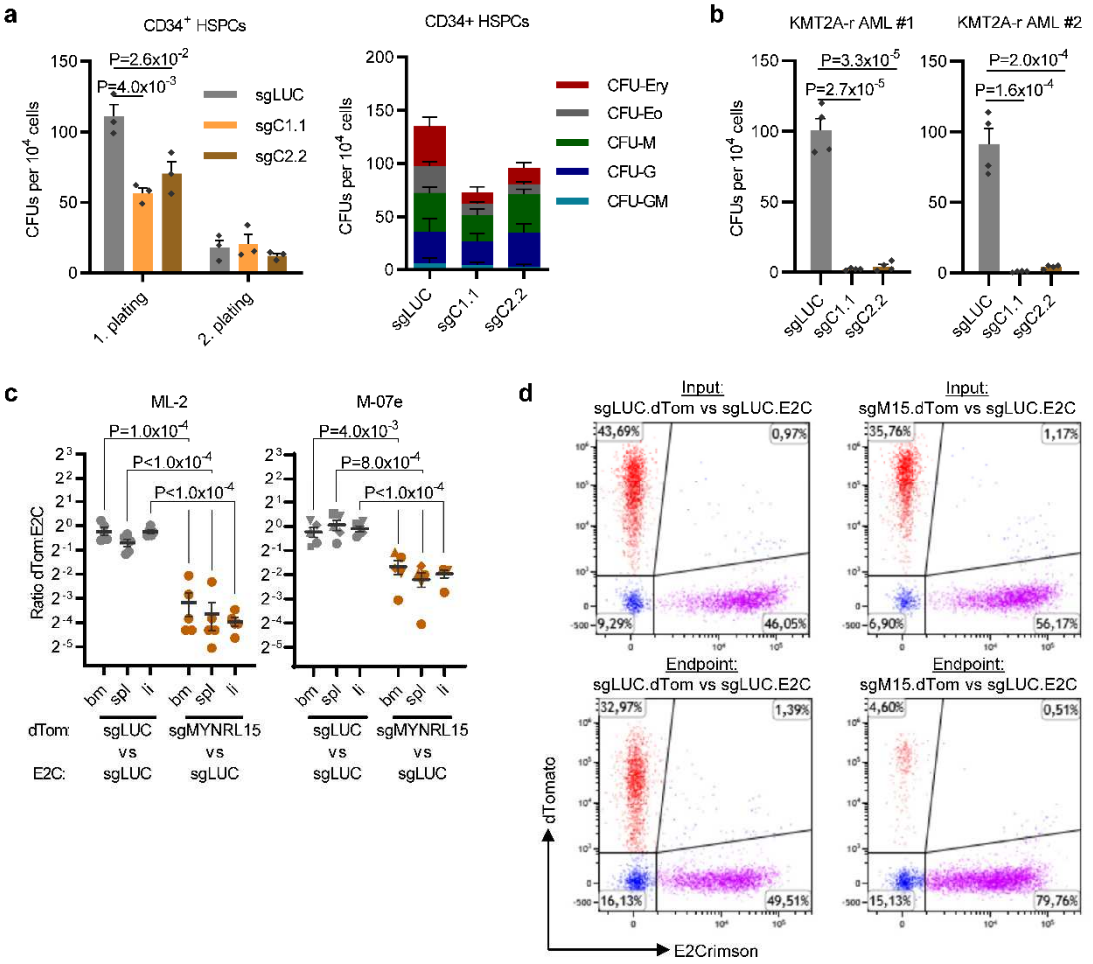
Extended Data Figure 6: Extended data from the mechanistic delineation of *MYNRL15* perturbation.



a, Chromosome 15 NG Capture-C interaction profiles from HSPCs (n=2; viewpoint in C1; smoothing window 2 pixels). The K562 sgC1.1 track is shown for reference. **b**, CUT&RUN tracks showing other examples of altered CTCF occupancy downstream of *MYNRL15* (left), and in the intervening space between the gained distal interactions and the *MYNRL15* locus (center and right). Decreases upon *MYNRL15* perturbation are outlined in blue; increases in pink. **c**, ATAC-seq tracks showing a wide view of chromatin accessibility in the gained distal interaction region (left; note the track discontinuity), and around the *MYNRL15* locus (right). **d**, MAGeCK analysis of CRISPR-Cas9 screens of the 29 coding genes from the gained distal interaction region (n=3). *WDR61* and *IMP3* are depicted in color, as are the positive controls *U2AF1*, *POL2RA*, and *RPL9* (turquoise). **e**, qRT-PCR validation of *WDR61* and *IMP3*

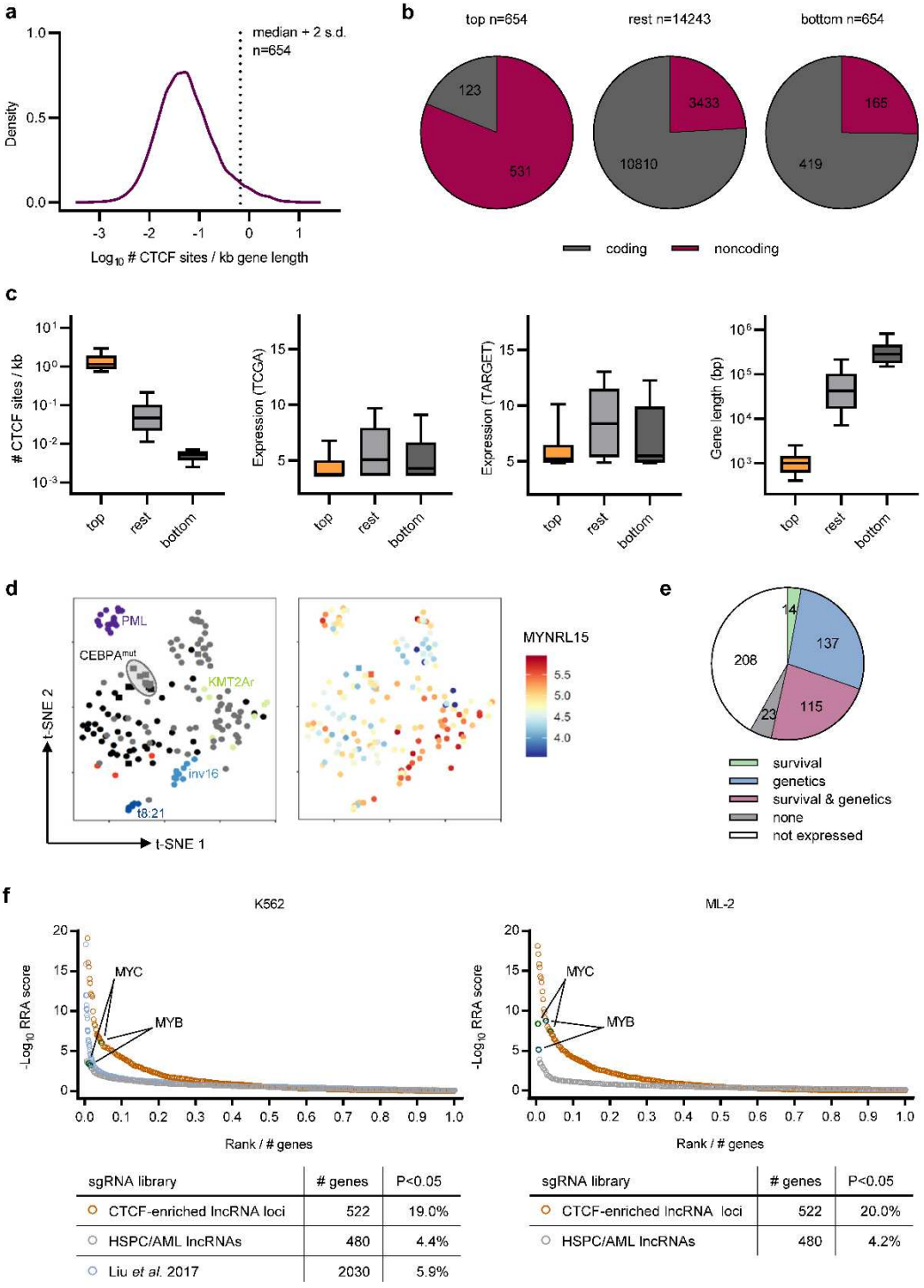
downregulation upon *MYNRL15* perturbation using sgRNA C1.1. **f**, Retrieval of PAF1c loss-associated gene sets upon *MYNRL15* perturbation in our RNA-seq data. **g**, Individual proliferation assays validating depletion of K562 and ML-2 cells upon *WDR61* and *IMP3* knockout, compared to *MYNRL15* perturbation using sgRNA C1.1 (n=4, mean ± s.e.m.; 4 guides each targeting *WDR61* and *IMP3*). ****P<0.0001; all conditions shared the same P-value.

Extended Data Figure 7: Extended data from *MYNRL15* perturbation experiments in primary and patient-derived cells.



a, Colony counts upon *MYNRL15* perturbation in CD34⁺ HSPCs from healthy donors (n=3; mean ± s.e.m.). Replating (left) and differentiation (right) assays were performed in parallel. **b**, Colony counts following *MYNRL15* perturbation in two patient-derived AML samples (n=4; mean ± s.e.m.). **c**, Representative flow cytometry data from two-color competitive xenotransplantation assays.

Extended Data Figure 8: Defining a new subclass of CTCF-enriched lncRNA loci.



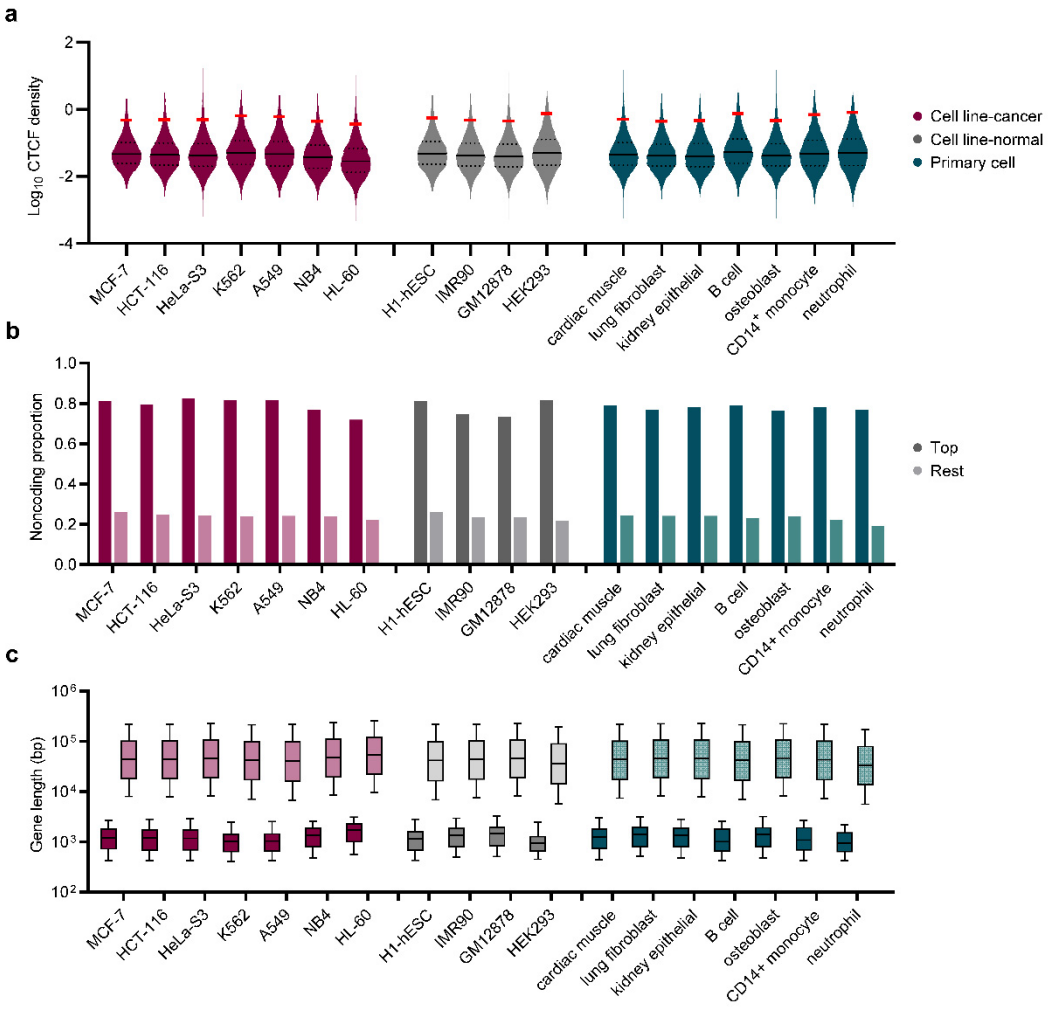
sgRNA library	# genes	P<0.05
CTCF-enriched lncRNA loci	522	19.0%
HSPC/AML lncRNAs	480	4.4%
Liu *et al.* 2017	2030	5.9%

sgRNA library	# genes	P<0.05
CTCF-enriched lncRNA loci	522	20.0%
HSPC/AML lncRNAs	480	4.2%

a, Distribution of $\log_{10}$-transformed values for the CTCF density metric (number of CTCF sites per kb of gene length). The dashed line indicates the cut-off defined (median + 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 for other loci like *MYNRL15*. From left to right: CTCF density, normalized expression in the TCGA AML cohort[36], normalized expression in the NCI-TARGET AML cohort[35], gene length. The top CTCF-bound loci are compared to the rest, or to the bottom of the ranked list in all plots. Midline, median; box limits, lower and upper quartiles; whiskers, 10% and 90% quantiles. **d**,

Unsupervised clustering of the TCGA cohort groups patients based on molecular subtypes (left), several of which are associated with *MYNRL15* expression (i.e. PML, inv16, *KMT2A*r) (right). **e**, Breakdown of CTCF-enriched loci based on their association with clinical aspects such as cytogenetics, mutations, and survival in the NCI-TARGET cohort (left). **f**, Ranked lists of CTCF-enriched lncRNA loci ordered by essentiality, as determined using MAGeCK on the results of the CRISPR-Cas9 screens in K562 (left) and ML2 (right) cells. The C-LNC screening results (orange) are displayed alongside those from the CRISPRi lncRNA library (grey) and Liu *et al.* 2017 (blue). Gene ranks are normalized to library size. The positive controls *MYC* and *MYB* are indicated.

Extended Data Figure 9: 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. Midline, median; dotted lines, lower and upper quartiles. The red lines indicate the cut-off defined (median + 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 differences in gene length distributions between the top and rest of CTCF-bound loci. Midline, median; box limits, lower and upper quartiles; whiskers, 10% and 90% quantiles.

Supplementary tables

Supplementary Table 1: Patient sample information

	AML PDX #1	AML PDX #2	AML PDX #3
Used in CFU assays	Yes	Yes	No
Used in transplants	No	Yes	Yes
Gender	Female	Male	Female
Age at diagnosis (years)	7	16	15
WBC ($\times 10^9$/L)	585.0	69.7	2.5
Hemoglobin (g/dl)	8.3	10.7	11
BM blasts (%)	84	93	90
CNS	No	No	No
SCT	Yes	No	No
Molecular genetics	Not determined	$NRAS^{mut}$	Not determined
Cytogenetics	46,XX,t(9;11)(p22;q23)[8]/50,XX,i-dem,+3,+8,+18,+19[15]	42~44,XY,t(6;11)(q27;q23)[cp2]/51,i-dem,+X,+der(6)t(6;11)(q27;q23),+8,+19,+21[5]	47,XX,+19[12]/46,XX[5]
Response	CCR	NR	CCR
Relapse	No	No	Yes
Death	No	Yes	Yes

Supplementary Table 2: sgRNA sequences

Target	System	Spacer
LUC	All systems	CCGCTGGAAGATGGAACCGC
LUC	All systems	GGGCATTTCGCAGCCTACCG
MYNRL15 TSS	CRISPRi	GTGCACTTCTGCTGCGGTCG
MYNRL15 TSS	CRISPRi	GCACGAGGTCTACGGTCATC
MYNRL15 TSS	CRISPRi	GGAGCGCGCCCGGGCAGGGG
MYNRL15 front	Excision	AGGTTGTCTCGTGCCCGCGC
MYNRL15 front	Excision	TCCGACGCAAGAGTGGGGCG
MYNRL15 back	Excision	GTCGGCCCCATCCGCGCGAT
MYNRL15 back	Excision	TGATGTAGGGGGTCCCCTCG
MYNRL15 C1	CRISPR-Cas9	TGCGGCCGCAGGGGCAGGGA
MYNRL15 C1	CRISPR-Cas9	CCCTGGCGCCGGGGAGGCCC
MYNRL15 C2	CRISPR-Cas9	GGAGCGCGCCCGGGCAGGGG
MYNRL15 C2	CRISPR-Cas9	GCCCCGGCCGGCCACCCCCG
UNC45A	CRISPR-Cas9	GTTCAAATGTGGAGACTACG
UNC45A	CRISPR-Cas9	AGGCCGTTCTGCACCGGAAC
UNC45A	CRISPR-Cas9	GGGGCGTCGCGTCCAGACCC
UNC45A	CRISPR-Cas9	CGCCCTGGCGGCCTACACTC
UNC45A	CRISPR-Cas9	GGTGGCAGGCGGCCCGGTTC
HDDC3	CRISPR-Cas9, CRISPRi	TGATGTAGGGGGTCCCCTCG
HDDC3	CRISPR-Cas9, CRISPRi	GTGCTTGCGAGCCGCGAAGT

HDDC3	CRISPR-Cas9, CRISPRi	GCTGGAGGCTGCCGACTTCG
HDDC3	CRISPR-Cas9	TGATTCCCGCCTCGTGGGTC
HDDC3	CRISPR-Cas9	GGCCCTGCTCCATGACACGG
IMP3	CRISPR-Cas9	ACGCGGAACTGGTCGCGTTC
IMP3	CRISPR-Cas9	GGAGCAAGGGCACGTACGCG
IMP3	CRISPR-Cas9	CGGGGTCGGTAACCACGTCA
IMP3	CRISPR-Cas9	GGAGCTCTGCGACTTCGTCA
WDR61	CRISPR-Cas9	GACCTTCACCAGGTCATCTA
WDR61	CRISPR-Cas9	CTGGCTAGGTATTTCCCATC
WDR61	CRISPR-Cas9	GCTCAGCGTGCCAGCCAAAT
WDR61	CRISPR-Cas9	GAATGCAACGTTCAGCACCC

Supplementary Table 3: shRNA sequences

Target	Sense strand
LUC	TGGCTACATTCTGGAGACATA
LUC	CCGCCTGAAGTCTCTGATTAA
MYNRL15	CAGGCTTATGTTCTTCTTGCA
MYNRL15	AGCAGAAGTGCACGAGGTCTA
MYNRL15	CCGGTCAGCTCCAGAGGAATT
MYNRL15	CCGCTGAATTAGCCTCCACGA
MYNRL15	AAGAGCCAGGCTTATGTTCTT
MYNRL15	CGCACCTAAGCTGTCCCCGCA

Supplementary Table 4: LNA-gapmeR sequences

Target	Sequence
Negative control B	GCTCCCTTCAATCCAA
MYNRL15	TCGTGGAGGCTAATTC
MYNRL15	GGTGCAAGAAGAACAT
MYNRL15	GACAGCTTAGGTGCGC

Supplementary Table 5: qRT-PCR primer sequences

Target	Direction	Sequence
MYNRL15	Forward	CAGGCTTATGTTCTTCTTGCA
MYNRL15	Reverse	AGCAGAAGTGCACGAGGTCTA
UNC45A	Forward	CCGGTCAGCTCCAGAGGAATT
UNC45A	Reverse	CCGCTGAATTAGCCTCCACGA
HDDC3	Forward	AAGAGCCAGGCTTATGTTCTT
HDDC3	Reverse	CGCACCTAAGCTGTCCCCGCA
B2M	Forward	TCTCTCTTTCTGGCCTGGAG
B2M	Reverse	AATGTCGGATGGATGAAACC

Supplementary Table 6: Oligonucleotides for NG Capture-C

Target	Sequence
C1 front	GATCATGCCTACCGTTCCAATTGTTATGAGGCTTAAACGGTATATCG

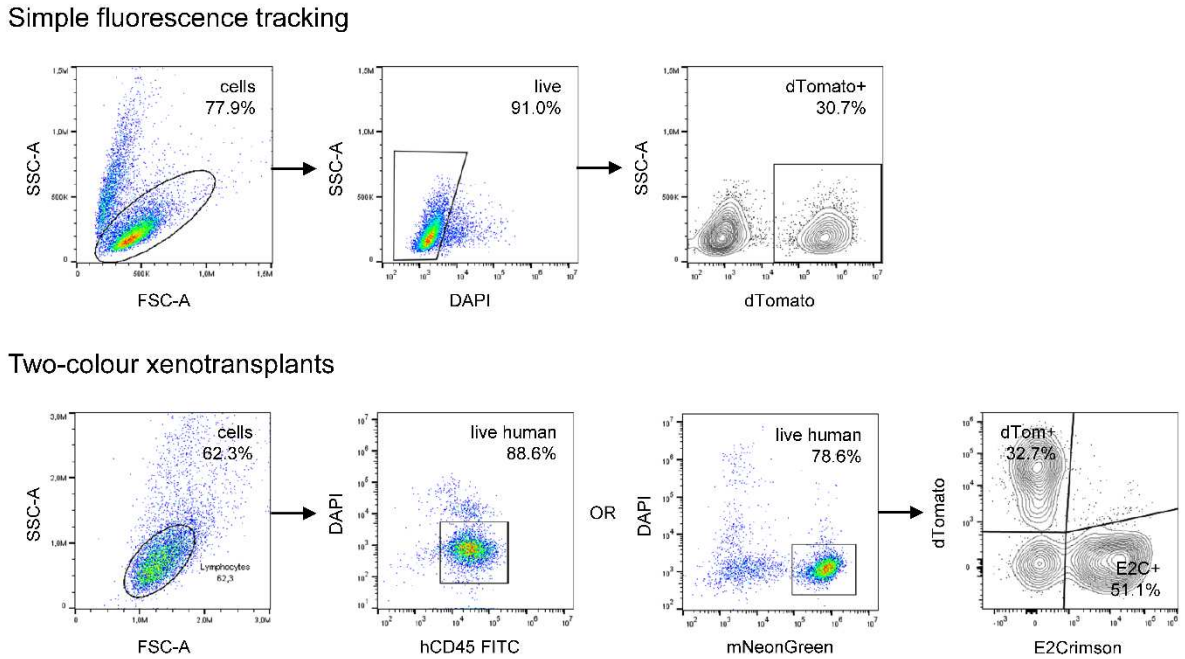
	CACTAAGCTTGGGACAAAACAGGTGCTCAATGAGGTGCACGCCCCCC ACGTTTCCTTCCCAGTCCCCACCACC
C1 back	CAGCCTCCAGCAGCTGCGCCGCCTCAGAGCCCATCGCGCGGATGGGG CCGACGACTGCGGCCGCAGGGGCAGGGACGGAACGTTTACAGCGCCC CCTGGCGCCGGGGAGGCCCGGGGATC
C2 front	GATCCAGAGCCAAGCGCCCCGCCCCTGCCCGGGCGCGCTCCCTCCTT AGCCCTGCCCCTCTCTGACCCCACCTCCGACGCAAGAGTGGGGCGGG GCAGCTGCCGGTGGCGTCCCGAACCC
C2 back	CCCCCGAGCCCCGGCCGGCCACCCCCGGGGTGCGTACCCAACCCCCG CGCCATCACCCCTTCGCACCCGCCCTGACCATCCCTGGCCTCCTTCT CCCCATCCATGAGGCTCGCCCCGATC

Supplementary Table 7-11:

Supplementary Tables 7-10 contain spacer sequences for the CRISPR libraries used in this study. Supplementary Table 11 contains a catalog of CTCF-enriched lncRNA loci (C-LNC) in 18 different cell lines and primary cell types. All four are large tables and are thus attached as separate files.

Supplementary information

Gating strategies for flow cytometry:



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable7CRISPRilncRNAlibrary.xlsx
- SupplementaryTable8MYNRL15tilinglibrary.xlsx
- SupplementaryTable9CaptureCgainedlibrary.xlsx
- SupplementaryTable10CLNClibrary.xlsx
- SupplementaryTable11CLNCcatalog.xlsx