Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Myeloid leukemia vulnerabilities at CTCF-enriched long noncoding RNA loci

Michelle Ng
Martin Luther University Halle-Wittenberg
Lonneke Verboon
Goethe University Frankfurt
Hasan Issa
Goethe University Frankfurt
Raj Bhayadia
Goethe University Frankfurt
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: December 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-727909/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Myeloid leukemia vulnerabilities at CTCF-enriched long noncoding RNA loci

Michelle Ng[1], Lonneke Verboon[2], Hasan Issa[2], Raj Bhayadia[2], Oriol Alejo Valle[1], Dorit Borchert[3], Konstantin Schuschel[2], Eniko Regenyi[1,4], Dirk Reinhardt[5], Marie-Laure Yaspo[4], Dirk Heckl[1,*], Jan-Henning Klusmann[2,*]

[1] Department of Pediatric Hematology and Oncology, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
[2] Department of Pediatrics, Goethe University Frankfurt, Frankfurt (Main), Germany
[3] Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany
[4] Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany
[5] Clinic for Pediatrics III, University Hospital Essen, Essen, Germany

* Co-corresponding authors; these authors contributed equally

Contact information for correspondence:

Prof. Dr. Jan-Henning Klusmann
jan-henning.klusmann@kgu.de
+49 151 4072 7844

or

Prof. Dirk Heckl
dirk.heckl@uk-halle.de
+49 152 5352 0105

Abstract word count: 235 (max 250)

Text word count: 3960 (max 4000)

Figure count: 5 (max 7)

References: 69 (max 100)

Running title: *MYNRL15* in myeloid leukemia

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

Key points

- CTCF-enriched lncRNA locus *MYNRL15* represents a novel myeloid leukemia dependency involved in chromatin architecture
- Elevated CTCF density characterizes a set of lncRNA loci enriched in leukemia vulnerabilities

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 a lack of predictive features renders mechanistic dissection challenging. Here, we describe *MYNRL15,* a CTCF-enriched lncRNA locus and pan-myeloid leukemia dependency initially identified by expression-guided CRISPR interference screens. We show that accessibility and integrity of the *MYNRL15* locus is required for myeloid leukemia maintenance; its perturbation selectively impairs acute myeloid leukemia (AML) cells compared to hematopoietic stem and progenitor cells *in vitro*, and depletes AML xenografts *in vivo*. While the *MYNRL15* transcript and neighboring protein-coding genes appear dispensable, dense CRISPR tiling of the locus revealed two crucial candidate *cis*-regulatory DNA elements which drive the perturbation phenotype. Disruption of these elements triggers the formation of a tumor-suppressive, long-range chromatin interaction. By integrating transcriptome profiling with a CRISPR-Cas9 knockout screen of genes from the gained interaction region, we pinpointed two downregulated, potent cancer dependency genes as effectors of *MYNRL15* disruption: *WDR61* and *IMP3*. Finally, guided by distinctive features of the *MYNRL15* locus, we find that elevated CTCF density characterizes a set of lncRNA loci enriched in leukemia vulnerabilities (22.6-24.2% essentiality rate). A catalog of CTCF-enriched lncRNA loci (C-LNCs) in 18 cell types representing different cancer entities and tissues is provided with this study, towards refining the search for noncoding oncogenic vulnerabilities in leukemia and other malignancies.

Introduction

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.

Methods

Additional method details can be found in the Supplemental Material.

Lentiviral vectors

Individual sgRNA and shRNA sequences are listed in Supplemental Table 7. The sgRNA sequences of the four CRISPR libraries in this study are provided in Supplemental Tables 1-3 and 5. Please refer to the Supplemental Material for details on CRISPR library design, cloning, and screening, as well as for a complete list of expression plasmids and Addgene IDs.

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. 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) samples were provided by the Berlin-Frankfurt-Münster Study Group (AML-BFM-SG, Essen, Germany), and expanded via serial xenotransplantation in mice. 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.

Hematopoietic assays

$CD34^+$ 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. Four days post-transduction, the cells were sorted and plated in human methylcellulose complete medium HSC003 (R&D Systems) for colony-forming assays. Fifteen thousand cells were initially plated over 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 samples 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), and 750 nM SR1 and 35 nM UM171 (both from STEMCELL Technologies) for 24-48 hours. Transductions were conducted 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.

Animal experiments

Two-color *in vivo* competition experiments were performed in murine xenograft models of AML as previously described[24,25]. 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. All mice were housed 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^{TM} 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. Differential expression analysis was conducted in R using DESeq2[26] (Bioconductor). Gene sets from MSigDB v7.2 (H1, C2, C3, C6), custom hematopoietic[16] and chromosome 15 gene sets, and PAF1c-knockout expression signatures[27] were checked for enrichment in the Broad GSEA software[28]. Custom positional gene sets were generated by walking a 1 Mb or 5 Mb window along chromosome 15. Gene ontology analysis was performed using the DAVID[29] online 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[30], with minor modifications. Transduced K562 or ML-2 cells (day 3 post-transduction) and *in vitro* expanded $CD34^+$ HSPCs (day 3) were used to evaluate the effect of *MYNRL15* perturbation and the native conformation of the locus, respectively. We used biotinylated oligonucleotides (sequences in Supplemental Table 7) corresponding to a viewpoint in the candidate *cis*-regulatory region C1 to enrich for interactions involving the locus. Two biological replicates were prepared per sample and pooled for oligonucleotide capture. The raw sequencing data were processed with the capC-MAP package[31].

C-LNC discovery

To identify CTCF-bound genic loci, we overlapped ENCODE CTCF ChIP-seq peaks with gene annotations from GENCODE v23 (release 07/2015)[32]. 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 > 2 s.d. above the median. Our analysis focused on loci that produce long (>200 nt) coding or noncoding transcripts and included the following biotypes: protein coding, lncRNA, lincRNA, processed transcript, and pseudogene.

Statistical analyses

Statistical evaluations of experimental data were carried out in GraphPad Prism 9 using two-tailed, unpaired t tests. Data are presented as mean ± s.d. or s.e.m. as indicated in the figure legends. Statistical analyses of gene expression data (RNA-seq) were carried out in R using DESeq2. Survival analyses were also performed in R using the Kaplan-Meier method and two-sided log-rank tests as implemented in the survival and survminer packages. CRISPR-Cas9 screening data were analyzed using MAGeCK to call essential genes, with the exception of the tiling screens, which were analyzed in R using DESeq2. $P<0.05$ 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 (reviewer access token: alstwikmtdyjvst). Raw sequencing data from the CRISPR-Cas9 screens have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under the accession numbers PRJEB44308 and PRJEB44320. K562 ChIP-seq data from ENCODE were utilized in this study: CTCF (ENCFF519CXF), SMC3 (ENCFF175UEE), H3K27Ac (ENCFF469JMR), H3K4Me1 (ENCFF100FDI), H3K4Me3 (ENCFF767UON). In addition, CTCF ChIP-seq data from 18 cell lines and primary cell types were also used (ENCODE accession numbers listed in Supplemental Table 6). Micro-C[33] and Hi-C[34] data were accessed via the UCSC Genome Browser. Gene expression data from adult and pediatric AML patients were obtained from TCGA (https://gdc.cancer.gov/access-data) and NCI-TARGET (https://ocg.cancer.gov/programs/target/data-matrix), respectively.

Results

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

We previously developed 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[35-38], 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). The cell lines 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]) – and we also included the well-studied erythroleukemia line K562. Cells stably expressing dCas9-KRAB were transduced with a sgRNA library targeted to lncRNA transcription start sites (Supplemental Table 1), and the sgRNA abundances were quantified via next generation sequencing at the start and end of 18 population doublings in order to identify essential lncRNAs. One candidate emerged as crucial in all six tested cell lines – *AC068831.3* (ID: ENSG00000224441 in Ensembl v85 [release 07/2016]), which we renamed *MYNRL15* (myeloid leukemia noncoding regulatory locus on chromosome 15; Fig. 1c-d, Supplemental Fig. 1, Supplemental Table 1).

MYNRL15 is a low-abundance, nuclear-enriched transcript (Supplemental 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 neighboring genes (Supplemental 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, Supplemental Fig. 2d-g). While CRISPR mediated excision of *MYNRL15* using paired sgRNA vectors recapitulated the effect produced by CRISPRi, repression of the transcript via shRNAs and LNA-gapmeRs had little impact on cell viability (Fig. 1e, Supplemental Fig. 2d-f). Both protein-coding neighbors were also dispensable, as determined by individual as well as combined CRISPR-Cas9 mediated knockout of UNC45A and HDDC3, and by CRISPRi mediated knockdown of *HDDC3* (Fig. 1e, Supplemental Fig. 2d-g). In addition, overexpression of *MYNRL15* cDNAs failed to rescue the CRISPRi knockdown phenotype (Fig. 1f). Altogether, these data 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*. Cell lines (K562, ML-2, M-07E, KASUMI-1) stably expressing either dCas9-KRAB or Cas9 were transduced with a sgRNA library covering the region at a mean density of 0.11 sgRNAs per bp (Supplemental Table 2), with the expectation that crucial areas would be demarcated by hubs of depleting sgRNAs. Notably, the *MYNRL15* locus contains several regions that exhibit features characteristic of *cis*-regulatory elements, such as H3K4Me1 and H3K27Ac histone marks, DNase hypersensitivity, and transcription factor occupancy including multiple CTCF and cohesin binding sites (Fig. 2a). The screens uncovered two crucial DNA segments whose accessibility and integrity were required by leukemia cells (Fig. 2b, Supplemental Fig. 3a). Both enhanced reporter gene expression in dual luciferase assays (Supplemental Fig. 3b) – nominating these as functional sequences and candidate *cis*-regulatory elements (cCREs C1-2). We note that inclusion of the dCas9 variant in the screening strategy alleviated concerns over potential off-target DNA damage-driven phenotypes[39] and increased our confidence in the results of the screen. The CRISPR-Cas9 mutagenesis strategy also reiterated that leukemia cells do not seem particularly dependent on the UNC45A and HDDC3 coding sequences, arguing that local enhancer functions on these genes are unlikely to underlie the anti-leukemic effect of *MYNRL15* perturbation.

Aiming to identify the target genes and pathways controlled by the *MYNRL15* locus, we next performed RNA sequencing following the disruption of cCREs C1 and C2 via CRISPR-Cas9 mediated induction of DNA double-strand breaks (hereafter referred to simply as *MYNRL15* perturbation). We opted for the CRISPR-Cas9 system in an effort to obtain a more targeted perturbation of *MYNRL15* and mitigate 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 depletion phenotype was underpinned by global changes in gene expression (Fig. 2d, Supplemental Fig. 4a-d), including the dramatic suppression of cancer dependency signatures related to proliferation and metabolism across both cell lines and two

time points (Fig. 2d, Supplemental Fig. 4d). While these results corroborate *MYNRL15*'s leukemia dependency phenotype – with the downregulated genes being enriched for members of crucial oncogenic pathways (Fig. 2d, Supplemental Fig. 4c-d) – no obvious target genes emerged, leading us to consider the alternative that *MYNRL15* may regulate multiple genes in a genomic neighborhood[40] in a more subtle manner. To explore this possibility, we applied a sliding window approach to gene set enrichment analysis using 1 Mb and 5 Mb sections of chromosome 15. This approach revealed positional gene sets that were coordinately deregulated upon *MYNRL15* perturbation, including the local area around *MYNRL15* and a distal upstream region (Fig. 2e), as well as others that were restricted to K562 or ML-2 cells (Supplemental 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. We elected to focus on cCRE C1 because of the strong divergent H3K27Ac / H3K4me1 signal and CTCF binding site overlapping the element, as well as the potency of its perturbation phenotype (Fig. 2b-c). Capture-C 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 remained consistent between K562 and ML-2 cells (Fig. 3a-c). The local interaction peaks demarcate nearby contact domains (Fig. 3c, Supplemental 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 anchor of a hierarchical loop (Fig. 3b, Supplemental Fig. 5a), indicating 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 and progenitor cells (HSPCs; Supplemental Fig. 6a), suggesting

that *MYNRL15* perturbation in leukemic cells may restore the long-range connectivity of normal blood cells. Consistent with *MYNRL15*'s involvement in chromosome 15 conformation, CTCF – a fundamental determinant of genome topology, which occupies three sites in the *MYNRL15* locus (Fig. 2a) – showed diminished binding at the locus and gained interaction sites following *MYNRL15* perturbation, among other subtle changes (Fig. 3d, Supplemental Fig. 6b). This was accompanied by diffuse gains in chromatin accessibility across the gained interaction sites (Fig. 3e, Supplemental Fig. 6c).

Having narrowed down the source of the *MYNRL15* perturbation effect to the gained chromatin interaction regions, we lastly conducted a small CRISPR-Cas9 knockout screen of the 29 protein-coding genes located in this region (Supplemental Table 3). Thus, by integrating the chromatin conformation and transcriptomic changes with leukemia dependency data, we finally pinpointed two putative downstream effector genes of *MYNRL15* perturbation: *IMP3* and *WDR61* (Fig. 3f, Supplemental Fig. 6d-e). Both are located in the gained chromatin interaction region and are downregulated upon *MYNRL15* perturbation, in addition to scoring as leukemia dependencies. *WDR61* is a component of the PAF1 complex (PAF1c), which is involved in key transcriptional programs during hematopoiesis and leukemogenesis[41,42]. Expression profiles induced by Paf1c inactivation[27] were also detected upon *MYNRL15* perturbation (Supplemental Fig. 6f). *IMP3* encodes a component of the 60-80S U3 small nucleolar ribonucleoprotein, which is required for early cleavages during pre-18S ribosomal RNA processing[43]. It is a homolog of the yeast Imp3 protein and has yet to be comprehensively studied in human cells to date. CRISPR-Cas9 mediated knockout of *WDR61* and *IMP3* robustly depleted K562 and ML-2 cells (Supplemental Fig. 6g), recapitulating the *MYNRL15* perturbation phenotype and positioning these genes as its effectors (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 Supplemental Table 4 for patient 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 (Supplemental Fig. 7a). In contrast, AML colony-forming units were virtually eradicated (Supplemental 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, Supplemental Table 4 for patient characteristics). Importantly, *MYNRL15* perturbation impaired the propagation of two AML cell lines and two PDXs in recipient mice (Fig. 4b, Supplemental Fig. 7c-d), confirming its capacity to deplete leukemic cells *in vivo*. Combined with the selective impairment of AML cells via *MYNRL15* perturbation, these results provide a proof-of-principle of how *MYNRL15* perturbation may be leveraged as a therapeutic strategy.

CTCF density metrics predict other myeloid leukemia dependency lncRNA loci

Having implicated *MYNRL15* in 3D genome organization and demonstrated its therapeutic potential, we explored whether features of *MYNRL15* could be used to identify other 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 harbored in the locus, we explored CTCF density as a predictive metric for uncovering loci like *MYNRL15* (Fig. 5a, Supplemental Fig. 8a-c). 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, Supplemental Fig. 8a-b). The remaining loci comprised mainly of protein-coding genes (Fig. 5a, Supplemental Fig. 8b), with *bona fide* lncRNAs such as *PVT1* and *XIST* also featuring in the lower part of the ranked list; others do not feature at all, due to an absence of CTCF binding sites. These

observations support CTCF density as a relevant metric for distinguishing a putative subset of lncRNA loci with transcript-independent functions, hereafter referred to as CTCF-enriched lncRNA loci, or C-LNCs. In addition to elevated CTCF density, these loci also tend to display low gene expression and short genomic length (median 1 kb; Supplemental 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[44,45]. 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, Supplemental Fig. 8d-e), suggesting that activity at these loci could underpin aberrant transcription factor programs and/or influence patient outcomes. Further, 22.6-24.2% functionally validated as essential for leukemia maintenance in CRISPR-Cas9 screens tiling CTCF sites in the loci (Fig. 5c, Supplemental Fig 8f, Supplemental Table 5) – a hit identification rate that is 4-10-fold higher than what is typically reported for lncRNA essentiality screens[46-48] (range: 2-6%), including our own initial screen of HSPC/AML lncRNA signatures (4.6%; Fig. 5c, Supplemental Fig. 8f). We validated one top hit from the C-LNC screen: *RP11-69L16.5* (ID: ENSG00000220472 in Ensembl v85 [release 07/2016]). CRISPR-Cas9 mediated perturbation produced a robust anti-leukemic phenotype accompanied by strong upregulation of the transcript (Supplemental Fig. 9a-b). High *RP11-69L16.5* expression accordingly associated with improved event-free survival in the NCI-TARGET[44] AML cohort (Supplemental Fig. 9c-d). These data illustrate the effectiveness of CTCF density metrics in refining functional lncRNA candidate lists, and underline the relevance such loci could hold for AML and cancer pathophysiology in general. We provide a catalog of C-LNCs spanning 18 cell lines and primary cell types (Fig. 5d, Supplemental Fig. 10, Supplemental Table 6) – 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 utility in distinguishing a set of lncRNA loci (CTCF-enriched lncRNA loci, or C-LNC) that is 4-10-fold enriched for myeloid leukemia dependencies compared to other lncRNA essentiality screens[46-48]. Extending these findings, we provide a catalog of C-LNCs across 18 cell lines and primary cell types at www.C-LNC.org or in Supplemental Table 6, as a basis for investigating C-LNCs in other cellular contexts. These and other efforts aimed at improving the functional classification of lncRNAs[19,22,49,50] will expedite the development of accurate and comprehensive annotations, and facilitate the process of delineating 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[51-53], although on a larger scale than typically observed for TADs[54]. Alternatively, or perhaps concurrently, the attenuation of CTCF binding upon *MYNRL15* perturbation may strengthen compartmentalization[55] and promote longer-range, higher-order architecture. We note that, while there is substantial overlap between enhancer RNA (eRNA) and lncRNA annotations[56], 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 instead demonstrate long-range architectural changes fol-

lowing perturbation of the *MYNRL15* locus, thereby separating *MYNRL15* from classical eRNAs. Unlike *MYNRL15*, eRNAs also tend to be unspliced and are predicted to be less than 150 nt in length; however, these definitions are far from binding[56].

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[53,57-60]. 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[35,36,48]. Recent works have begun to implicate aspects of chromatin architecture in cell state transitions during hematopoiesis[61-64] and in the maintenance of leukemic transcription programs[65-67]. We expect future studies will further reveal leukemic 3D genome organization signatures that underpin general oncogenic behaviors, irrespective of mutational drivers.

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

Acknowledgements

We thank D. Downes and J. R. Hughes (MRC Weatherall Institute, Oxford) for providing the NG Capture-C protocol and consulting on experimental design. We also thank A. Schwarzer (Hannover Medical School), V. Amstislavskiy (Centogene), F. F. Adams (Roche), and J. L. Mateo (University of Oviedo) for bioinformatics guidance, and A. Kovacsovics (MPI for Molecular Genetics, Berlin), A. Navarrete Santos, L. Gack, K. Huke, K. Menge, C. Beyer, and R. Winkler (Martin Luther University Halle-Wittenberg) for technical assistance. This work was supported by the following funding: J.H.K. – ERC Horizon 2020 program (714226), St. Baldrick's Robert J. Arceci Innovation Award; D.H. – German Cancer Aid (111743); M.N. – Hannover Biomedical Research School.

Author contributions

M.N., L.V., H.I., R.B., O.A.V., and D.B. performed experiments and analyzed the results. M.N., K.S., E.R., and M.L.Y. conducted bioinformatic analyses. D.R. provided patient samples. D.H. and J.H.K. designed and supervised the study. M.N., D.H., and J.H.K. crafted the manuscript and figures. All authors revised and approved the manuscript.

Conflict of interest

D.R. has advisory roles for Celgene Corporation, Novartis, Bluebird Bio, Janssen, and receives research funding from CLS Behring and Roche. J.H.K. has advisory roles for Bluebird Bio, Novartis, Roche and Jazz Pharmaceuticals.

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 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).

25 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).

26 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).

27 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).

28 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).

29 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).

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 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).

32 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).

33 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).

34 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).

35 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).

36 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).

37 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).

38 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).

39 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).

40 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).

41 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).

42 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).

43 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).

44 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).

45 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).

46 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).

47 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).

48 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).

49 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).

50 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).

51 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).

52 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).

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

54 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).

55 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).

56 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).

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

58 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).

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

60 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).

61 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).

62 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).

63 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).

64 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).

65 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).

66 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).

67 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).

Figure legends

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), Down syndrome myeloid leukemia (DS), non-DS megakaryoblastic leukemia (AMKL), promyelocytic leukemia (PML), *KMT2A*-rearranged leukemia (*KMT2A*-r). **b**, Workflow for screening HSPC/AML lncRNAs. Stable dCas9-RKAB myeloid leukemia cell lines were transduced with a lentiviral sgRNA library and maintained for 18 population doublings, before and after which samples were sequenced to determine sgRNA abundances and call essential genes. **c**, Schematic of the *MYNRL15* locus, including target sites of the different perturbation constructs (not to scale). Target gene: *MYNRL15* (orange), *UNC45A* (black), *HDDC3* (grey), *UNC45A* and *HDDC3* (light grey). Perturbation strategy: CRISPRi (filled circle), dual sgRNA mediated excision (filled square), RNAi (empty diamond), LNA-gapmeRs (empty triangle), CRISPR-Cas9 mediated knockout (filled triangle). **d**, Gene essentiality scores from pan-cell line MAGeCK analysis of the CRISPRi screen (6 cell lines, n=2 biological replicates per cell line). *MYNRL15* is highlighted as the top hit behind *MYC* and *MYB*, the positive controls. **e**, Endpoint depletion values from fluorescence-based proliferation assays using different perturbation strategies. Each point represents one vector used for perturbation (mean of n=3 biological replicates shown). **f**, Fluorescence-based proliferation assays using *MYNRL15* cDNAs to rescue the CRISPRi depletion phenotype (n=2 biological replicates, mean ± s.e.m.; double positive sgRNA+cDNA cells are shown). $^{**}P<0.01$ (two-tailed, unpaired t-test); all conditions shared the same *P*-value. Colors denote sgRNA vectors, shapes denote cDNA vectors. Long isoform (L), short isoform (S). **e-f**, The data are normalized to day 0 and to the non-targeting control (sgLUC). These experiments were performed in ML-2 cells.

Figure 2: Functional dissection of *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 (K562 ChIP-seq data from ENCODE). **b**, Tiling screen results using complementary CRISPRi (top) and CRISPR-Cas9 based (bottom) strategies to interrogate the *MYNRL15* locus (mean of 4 cell lines, n=2 biological replicates per cell line). Previously tested sgRNAs are depicted in color. A smoothed fit curve is shown in blue. The two identified leukemia-essential cCREs, C1 and C2, are outlined. **c**, Fluorescence-based proliferation assays using classical CRISPR-Cas9 and individual sgRNAs from C1 and C2 to achieve perturbation of the *MYNRL15* locus (n=3 biological replicates, mean ± s.e.m.; 2 guides per cCRE, 4 in total). The data are normalized to day 0 and to the non-targeting control (sgLUC). $^{*}P<0.05$, $^{**}P<0.01$, $^{****}P<0.0001$ (two-tailed, unpaired t-tests); where only one set of asterisks is shown, all conditions share the same *P*-value. **d-e**, GSEA comparing the *MYNRL15* perturbation group (using all 4 guides from **c**; “sgMYNRL15”) to the non-targeting control (“sgLUC”) in a combined analysis across early (day 3) and late (day 6 7, respectively) time points of ML-2 and K562 cells (n=2 biological replicates per guide). **d**, Normalized enrichment scores (NES) of cancer dependency gene

sets downregulated upon *MYNRL15* perturbation. Colors correspond to MSigDB collections: H1 hallmark (purple), C2 KEGG pathways (blue). ****P=0 (nominal P values from GSEA). **e**, Two chromosome 15 gene sets deregulated by *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 biological replicates; viewpoint in C1; smoothing window 2 pixels). **b-c**, Close-ups of the gained interaction region and the region around *MYNRL15*, alongside K562 CTCF ChIP-Seq and H1-hESC Micro-C[33] tracks from the UCSC Genome Browser. **d**, CTCF occupancy around the *MYNRL15* locus and gained interaction sites, as determined via CUT&RUN. Regions of reduced occupancy are shaded in blue. Note the track discontinuity in the left and center views. **e**, Chromatin accessibility around the gained interaction sites, as determined via ATAC-seq. Regions of increased accessibility are shaded in red. **d-e**, Tracks show K562 and ML-2 cells with *MYNRL15* perturbation (sgC1.1) compared to a non-targeting control (sgLUC) (n=1). **f**, Integrative analysis of CRISPR-Cas9 screening scores (MAGeCK; n=2 per cell line) and differential expression following *MYNRL15* perturbation (DESeq2; comparison as in Fig. 2d-e) for the 29 coding genes in the gained interaction region. A combined analysis of K562 and ML-2 cells is shown. **g**, Model of chromosome 15 reorganization upon *MYNRL15* perturbation.

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

a, Comparison of the impact of *MYNRL15* perturbation on colony-forming capacity in $CD34^+$ HSPCs (n=3 biological replicates) and two AML PDXs (n=4 biological replicates each). The data are normalized to the non-targeting control (mean ± range). **b**, Setup (left) and results (right) of direct two-color *in vivo* competition assays involving 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), spleens (spl), and livers (li) of recipient mice (n=4 in the AML PDX #2 control group, otherwise n=5 mice per group; mean ± s.e.m.). **a-b**, P values were calculated using two-tailed, unpaired t-tests.

Figure 5: Elevated CTCF density characterizes a set of clinically relevant, functionally validated lncRNA loci.

a, Ranked list of protein-coding and lncRNA loci ordered by CTCF density in K562 cells. 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[45]. **b**, Kaplan-Meier survival curves of patients with high (n=10) vs low (n=171) expression of *MYNRL15* in the TCGA AML cohort[45] (survival probability ± 95% C.I.). Five-year event-free survival: 0.0% vs 23.2%. The depicted P value

was calculated using a two-sided log-rank test. **c**, Gene essentiality scores for CTCF-enriched lncRNA loci (C-LNCs), as determined via CRISPR-Cas9 screening and MAGeCK analysis of K562 and ML-2 cells (orange). Our initial CRISPRi lncRNA screen is shown alongside (grey; combined analysis of K562 and ML-2) for comparison. In both cases, the scores are derived from n=2 biological replicates per cell line. Gene ranks are normalized to library size. The positive control genes *MYC* and *MYB* score similarly across CRISPR-Cas9 and CRISPRi based screens. **d**, Numbers of C-LNCs in 18 other cell lines and primary cell types.

Supplementary Files

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

- SupplementaryTable1CRISPRilncRNAlibraryscreen.xlsx
- SupplementaryTable2MYNRL15tilinglibraryscreen.xlsx
- SupplementaryTable3CaptureCgainedlibraryscreen.xlsx
- SupplementaryTable4patientsampleinformation.xlsx
- SupplementaryTable5CLNClibraryscreen.xlsx
- SupplementaryTable6CLNCcatalog.xlsx
- SupplementaryTable7individualoligos.xlsx
- LNC666supplementBloodV1.pdf