LETTER

Corrected: Publisher Correction

https://doi.org/10.1038/s41586-018-0453-z

The *NORAD* lncRNA assembles a topoisomerase complex critical for genome stability

Mathias Munschauer[1]*, Celina T. Nguyen[1], Klara Sirokman[1], Christina R. Hartigan[1], Larson Hogstrom[1], Jesse M. Engreitz[1], Jacob C. Ulirsch[1,2,3], Charles P. Fulco[1], Vidya Subramanian[1], Jenny Chen[1,4], Monica Schenone[1], Mitchell Guttman[5], Steven A. Carr[1] & Eric S. Lander[1,6,7]*

The human genome contains thousands of long non-coding RNAs[1], but specific biological functions and biochemical mechanisms have been discovered for only about a dozen[2–7]. A specific long non-coding RNA—non-coding RNA activated by DNA damage (*NORAD*)—has recently been shown to be required for maintaining genomic stability[8], but its molecular mechanism is unknown. Here we combine RNA antisense purification and quantitative mass spectrometry to identify proteins that directly interact with *NORAD* in living cells. We show that *NORAD* interacts with proteins involved in DNA replication and repair in steady-state cells and localizes to the nucleus upon stimulation with replication stress or DNA damage. In particular, *NORAD* interacts with RBMX, a component of the DNA-damage response, and contains the strongest RBMX-binding site in the transcriptome. We demonstrate that *NORAD* controls the ability of RBMX to assemble a ribonucleoprotein complex—which we term *NORAD*-activated ribonucleoprotein complex 1 (NARC1)—that contains the known suppressors of genomic instability topoisomerase I (TOP1), ALYREF and the PRPF19–CDC5L complex. Cells depleted for *NORAD* or RBMX display an increased frequency of chromosome segregation defects, reduced replication-fork velocity and altered cell-cycle progression—which represent phenotypes that are mechanistically linked to TOP1 and PRPF19–CDC5L function. Expression of *NORAD* in *trans* can rescue defects caused by *NORAD* depletion, but rescue is significantly impaired when the RBMX-binding site in *NORAD* is deleted. Our results demonstrate that the interaction between *NORAD* and RBMX is important for *NORAD* function, and that *NORAD* is required for the assembly of the previously unknown topoisomerase complex NARC1, which contributes to maintaining genomic stability. In addition, we uncover a previously unknown function for long non-coding RNAs in modulating the ability of an RNA-binding protein to assemble a higher-order ribonucleoprotein complex.

NORAD stands out among long non-coding RNAs (lncRNAs) because it (1) is highly conserved relative to other lncRNAs, (2) is abundantly expressed in many cell types, (3) is upregulated upon DNA damage and (4) induces chromosomal instability and aneuploidy when deleted. This phenotype is intriguing as little is known about the roles of lncRNAs in maintaining a stable genome. A model for lncRNA function suggests that lncRNAs can serve as assembly scaffolds for ribonucleoprotein complexes[6,7], yet this model has been explored in only a few cases. The mechanism that connects the *NORAD* lncRNA to chromosomal instability remains unknown.

Two recent studies have reported PUMILIO, a highly abundant cytoplasmic RNA-binding protein with no known role in genomic stability, as the sole *NORAD*-interacting protein[8,9]. However, these results were obtained from in vitro mixing of exogenous *NORAD* fragments with cytoplasmic extracts, which may not accurately represent the protein contacts of *NORAD* in living cells (Supplementary Note 1).

To reveal the direct interactions of *NORAD* with proteins in live cells, we captured and identified *NORAD*-interacting proteins by combining RNA antisense purification (RAP) with quantitative liquid chromatography–mass spectrometry using isobaric mass tag quantification (RAP MS) (Fig. 1a). HCT116 colon carcinoma cells were treated with 365-nm light after 4-thiouridine labelling[10], which covalently crosslinks proteins to RNA but not to other proteins. lncRNA–protein complexes were purified by RNA hybrid selection with antisense oligonucleotides that target *NORAD*, under denaturing and reducing conditions at high temperature to minimize the co-purification of indirectly bound proteins[2] (Fig. 1a). To identify specific interactors with *NORAD*, we quantitatively compared the resulting proteins to those captured in purifications with antisense oligonucleotides that target the well-characterized 'RNA component of mitochondrial RNA processing endoribonuclease' (*RMRP*), which is not expected to interact with the same proteins as *NORAD* [11]. We analysed biological replicate purifications in a single 4-plex iTRAQ cassette, quantifying 1,361 proteins that each had more than two unique peptides (Fig. 1b). The control purification captured about 85% of *RMRP* transcripts (Extended Data Fig. 1) and enriched the target RNA approximately 550-fold versus input RNA. We found 12 strongly enriched proteins (mean $\log_2$(iTRAQ ratio (NORAD/RMRP)) < -1.6, $P < 0.05$, moderated *t*-test) (Fig. 1c), including 8 of the 10 known core components of the *RMRP* complex[11] and one previously identified candidate rRNA- and/or tRNA-processing factor[12].

We then analysed *NORAD* antisense purifications. Experiments captured 82% of endogenous *NORAD* (Extended Data Fig. 1) (about 80-fold enrichment versus input RNA). We reproducibly identified 45 proteins that met our enrichment criteria (mean $\log_2$(iTRAQ ratio (NORAD/RMRP)) > 1.6, $P < 0.05$, moderated *t*-test) (Fig. 1b). This set of proteins is highly specific to *NORAD*, in that 41 out of the 45 proteins (Fig. 1c) were not among 219 promiscuous binders (Supplementary Note 3). The RNA-binding protein PUMILIO2 (PUM2) was indeed present in our *NORAD* interactome, but it ranked 185th out of the 265 proteins we detected (mean $\log_2$(iTRAQ ratio (NORAD/RMRP)) > 0.5) and did not meet our cut-off for strongly enriched proteins.

Notably, many of the 41 *NORAD*-interacting proteins have key roles in nuclear processes such as DNA unwinding, replication and repair (including PURA, PURB, TAF15, ALYREF, SFPQ, SRSF1, RBM14, DDX17, RBMX and its retrogene RBMXL1). Twenty-nine (71%) of the forty-one proteins localize to the nucleus, nucleoplasm or chromatin, whereas only two (5%) localize exclusively to the cytoplasm (Fig. 1d). The interactome thus points towards an important nuclear function of *NORAD*.

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [3]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA, USA. [4]Division of Health Sciences and Technology, MIT, Cambridge, MA, USA. [5]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [6]Department of Biology, MIT, Cambridge, MA, USA. [7]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. *e-mail: mathias@broadinstitute.org; eric@broadinstitute.org

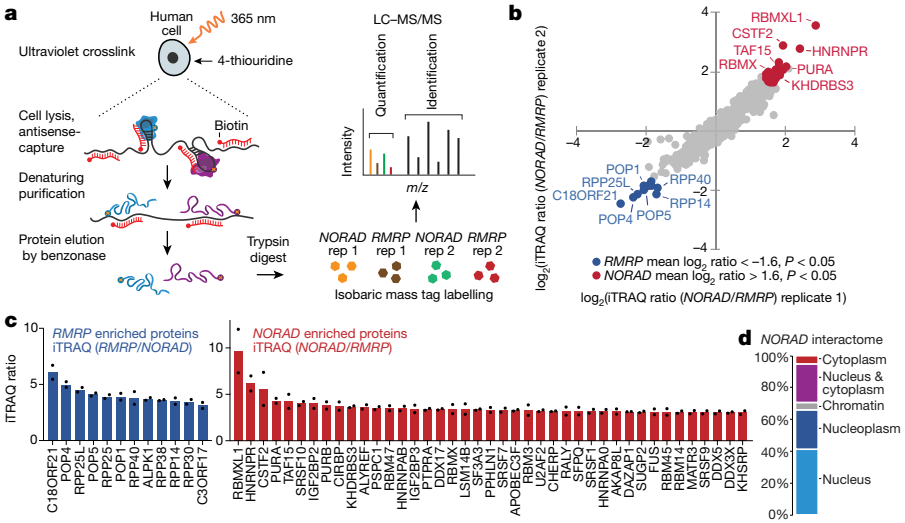


Fig. 1 | *NORAD* directly binds many nuclear proteins in living cells. **a**, Schematic overview of RAP MS. **b**, Quantification of *NORAD*- and *RMRP*-interacting proteins. Scatter plot of $\log_2$-transformed iTRAQ ratios from two biological replicates is shown. Adjusted *P* value, two-tailed moderated *t*-test (Supplementary Note 2). Rep, replicate. **c**, iTRAQ ratios of *RMRP*- (blue) and *NORAD*- (red) enriched proteins. Columns represent the mean of two biological replicate experiments, individual data points are shown (Supplementary Tables 1, 2). **d**, Subcellular localization of *NORAD*-interacting proteins.

Given the overrepresentation of nuclear proteins, we used single-molecule RNA fluorescent in situ hybridization (smRNA FISH) to assess the subcellular localization of *NORAD* in intact cells. In contrast to previous reports that characterized *NORAD* as being located exclusively in the cytoplasm[8], we found that on average 40–50% of *NORAD* transcripts in HCT116 cells reside in the nucleus (Fig. 2a and Extended Data Fig. 2a, b). We confirmed the nuclear localization by subcellular fractionation and quantitative PCR with reverse transcription (RT–qPCR) (Extended Data Fig. 2c, d). Notably, when cells were challenged with DNA damage and replication stress, *NORAD* was upregulated (Extended Data Fig. 2e) and its nuclear localization increased markedly (to about 85%), whereas the localization patterns of control RNAs were unaffected (Fig. 2a, b). Given this shift in localization, we performed RAP experiments with and without DNA damage to confirm that the interactions of *NORAD* with several candidate binders also occur under conditions of DNA damage (Extended Data Fig. 2f, g).

Among the *NORAD*-interacting proteins, we focused on RBMX, the knockdown phenotype of which (impaired DNA damage repair[13] and premature sister-chromatid separation[14]) is closely related to the previously reported *NORAD* knockout phenotype[8]. To explore this connection, we quantified the frequency of chromosome-segregation defects in response to depletion of *NORAD* or RBMX by imaging mitotic cells. We achieved >90% reduction in *NORAD* expression (estimated by RT–qPCR, RNA-sequencing and smRNA FISH) by CRISPR interference (KRAB–dCas9) targeted to the *NORAD* promoter (Fig. 2a and Extended Data Fig. 3a, b). For both wild-type and knockdown cells, we imaged 100 anaphase nuclei and calculated the frequency of DAPI-positive anaphase bridges. Consistent with previous reports[8], *NORAD* depletion caused a significant increase (2.2-fold) in segregation defects (Fig. 2c, d). Importantly, these defects were rescued by expression of full-length *NORAD* in *trans* (Fig. 2d), indicating that the defects are dependent on the *NORAD* RNA. Depletion of RBMX (Extended Data Fig. 3a) caused a comparable increase (2.6-fold) in the frequency of anaphase bridges (Fig. 2d). By contrast, depletion of the cytoplasmic protein PUM2 (Extended Data Fig. 3a) caused no substantial increase in segregation defects (Extended Data Fig. 3c). We reasoned that the interaction between *NORAD* and RBMX may hold important mechanistic insights into *NORAD* function.

To explore this interaction, we mapped RBMX-binding sites on *NORAD* by crosslinking and immunoprecipitation (CLIP). We covalently coupled proteins to RNA using ultraviolet crosslinking[10] and immunopurified RBMX with a specific antibody. We isolated and sequenced RNA crosslinked to RBMX. RBMX displays unusually strong and specific binding to the 5′ end of *NORAD* (Fig. 2e). The RBMX-binding site in *NORAD* extends over more than 800 nucleotides and covers about 15% of *NORAD*—making it eight times larger than the majority of RBMX-binding sites (Extended Data Fig. 3d) and the strongest RBMX-binding region in the transcriptome (Fig. 2f). This unusual binding pattern suggests that *NORAD* serves as a high-affinity binding target for RBMX and contains many RBMX-binding sites. A multiple sequence alignment of *NORAD* transcripts, which was assembled de novo from RNA-sequencing data from 11 mammalian species (Extended Data Fig. 3e), suggests that the RBMX-binding region in *NORAD* is transcribed and conserved throughout mammalian evolution. Next, we performed CLIP for three additional RNA-binding proteins and showed that the RBMX-binding region does not bind PUMILIO, FUBP1 or FUBP3 (Extended Data Fig. 3f).

To confirm that the *NORAD*–RBMX interaction occurs in the nucleus, we performed RBMX RNA immunoprecipitation (RIP) in nuclear and cytoplasmic extracts and showed that over 99% of the total RBMX RIP signal is indeed nuclear (Extended Data Fig. 3g). Consistent with this result, immunofluorescence microscopy suggests that RBMX localizes exclusively to the nucleus (Extended Data Fig. 3h). Finally, depletion of RBMX did not affect subcellular localization of *NORAD* (Extended Data Fig. 3i).

We speculated that *NORAD* might use its large RBMX-binding site to assemble a ribonucleoprotein complex. To examine the role of *NORAD* in such a complex, we sought to identify proteins that bind RBMX and determine whether their interaction with RBMX was dependent on *NORAD*. We performed co-immunoprecipitation and mass spectrometry (co-IP MS) experiments and compared the quantitative enrichment of RBMX-interacting proteins in cells with and without *NORAD* knockdown (Fig. 3a). Importantly, we used a nonspecific RNA and DNA nuclease (benzonase) to ensure that RBMX-binding is direct, rather than being mediated by RNA.

Among the top 11 proteins that bound to RBMX only in the presence of *NORAD*, 7 are linked to DNA replication or repair (Fig. 3b).

Six of these proteins (TOP1, TOP1MT, PRPF19, CDC5L, BCAS2 and MEPCE) were not detected in *NORAD* RAP MS data or were not among the top 200 enriched proteins, which suggests that they bind directly to RBMX and do not interact strongly with *NORAD* (Extended Data Fig. 4a). We further confirmed by western blot the absence of TOP1 in *NORAD* antisense purifications (Extended Data Fig. 4b) and showed that levels of TOP1, RBMX, PRPF19 and CDC5L proteins

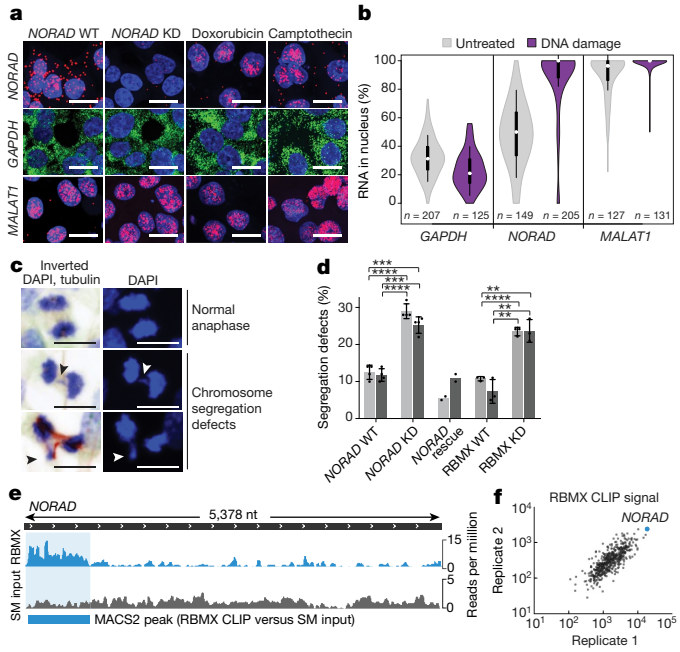


Fig. 2 | *NORAD* and RBMX are present in the nucleus, affect chromosome segregation and interact via a distinct sequence element. **a**, smRNA FISH of *NORAD*, *GAPDH* and *MALAT1* in wild-type (WT), knockdown (KD) and DNA-damaged cells. *GAPDH*, cytoplasmic reference; *MALAT1*, nuclear reference. Scale bar, 20 μm. Images are representative of one experiment; three independent experiments were performed. **b**, Quantification of smRNA FISH experiments. Circles, medians; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range; polygons, extreme values. **c**, Normal anaphase nuclei and chromosome segregation defects in *NORAD* knockdown cells. Scale bar, 5 μm. Images are representative of four independent experiments. **d**, Quantification of chromosome segregation defects. One hundred anaphase nuclei are scored for each condition. Grey bars represent different single-guide RNAs. Values are mean ± standard deviation (*NORAD*, $n = 4$, *NORAD* rescue, $n = 2$, RBMX, $n = 3$). $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, two-tailed Student's *t*-test. **e**, RBMX CLIP data plotted across *NORAD* RNA. SM input, size-matched input for proteins migrating at the same molecular weight range as RBMX. MACS2-identified RBMX peak is shown. **f**, CLIP signal for all RBMX peaks enriched over SM input in two biological replicates.

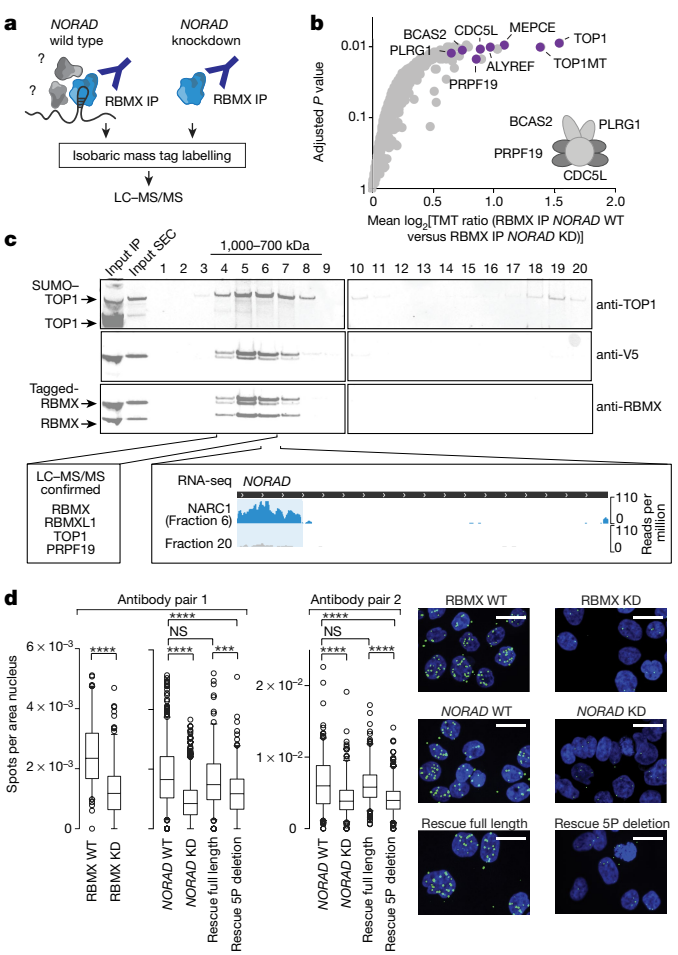


Fig. 3 | *NORAD* modulates RBMX and is essential for assembly of a topoisomerase complex. **a**, Illustration of RBMX co-IP MS experiments. **b**, Quantification of RBMX-interacting proteins in wild-type versus *NORAD* knockdown cells. Mean $\log_2$ TMT ratios from two biological replicates are shown. Adjusted *P* value, two-tailed moderated *t*-test (Supplementary Note 2). Purple, *NORAD*-dependent RBMX interactors functionally linked to DNA replication and repair. Inset, illustration of PRPF19–CDC5L complex. **c**, Western blot of Flag–RBMX–V5 co-IP followed by size-exclusion chromatography. Fractions 4–6 were pooled for mass spectrometry (Supplementary Table 5). Fraction 6 and fraction 20 were used for RNA sequencing. Western blots are representative of one experiment; three independent experiments were performed. SEC, size-exclusion chromatography. **d**, Proximity ligation assay for RBMX and TOP1. Two different antibody pairs were used. Centre lines, medians; box limits, 25th and 75th percentiles; whiskers, 5th and 95th percentiles; dots, outliers. $^{****}P < 0.0001$, $^{***}P < 0.001$, NS, not significant, two-tailed Mann–Whitney *U* test. Sample size antibody pair 1: RBMX WT, $n = 102$; RBMX KD, $n = 193$; *NORAD* WT, $n = 521$; *NORAD* KD, $n = 559$; rescue full length, $n = 176$; rescue 5P(5′ RBMX-binding site) deletion, $n = 171$. Sample size antibody pair 2: *NORAD* WT, $n = 209$; *NORAD* KD, $n = 232$; rescue full length $n = 211$; rescue 5P deletion, $n = 290$. Representative images are shown. Scale bar, 20 μm.

were not changed upon *NORAD* depletion (Extended Data Fig. 4c). PRPF19, CDC5L and BCAS2, together with PLRG1, make up the core of the human PRPF19–CDC5L complex, and both PRPF19–CDC5L and TOP1 have important roles in DNA replication and genomic stability, as previously reviewed[15,16]. TOP1 suppresses genome instability by preventing interference between replication and transcription[17]. This involves relieving torsional stress in DNA (that is, supercoiling) and suppressing the accumulation of RNA–DNA hybrids (R-loops)[18]. Both R-loops and supercoiled DNA impair replication-fork progression and can lead to genomic instability[17,19]. Stalled replication forks activate the DNA-damage response though ATR signalling. CDC5L binds and activates ATR[20], and the E3 ligase PRPF19 ubiquitylates RPA, enhancing ATRIP–ATR recruitment to stalled replications forks[21].

The precise roles of the remaining two *NORAD*-dependent RBMX interactors in maintaining genomic stability are less well understood. MEPCE binds to the 5′ cap of 7SK[22] and was reported in several studies that aimed to identify proteins involved in the DNA damage response[13,23]; however, its exact function in this process remains unknown. Unlike the six proteins above that were not found by RAP to interact strongly with *NORAD*, a seventh protein—ALYREF—was identified as a strong *NORAD* binder. ALYREF is part of the human TREX complex and interacts with the 5′ end of many RNAs, including *NORAD* (Extended Data Fig. 4d), to facilitate RNA export from the nucleus[24]. ALYREF contributes to genomic stability by suppressing R-loops[24], as does TOP1.

We performed reciprocal co-IP and western blots to confirm that TOP1, ALYREF and CDC5L interact with RBMX and also contact each other (Extended Data Fig. 4e), suggesting that these proteins may constitute a complex. To test whether such a complex exists, we generated cell lines that express epitope-tagged RBMX and performed co-IP experiments (using benzonase to digest unprotected RNA and DNA) followed by native elution and size-exclusion chromatography. Western blot analysis of size-fractionated co-IP samples showed that RBMX,

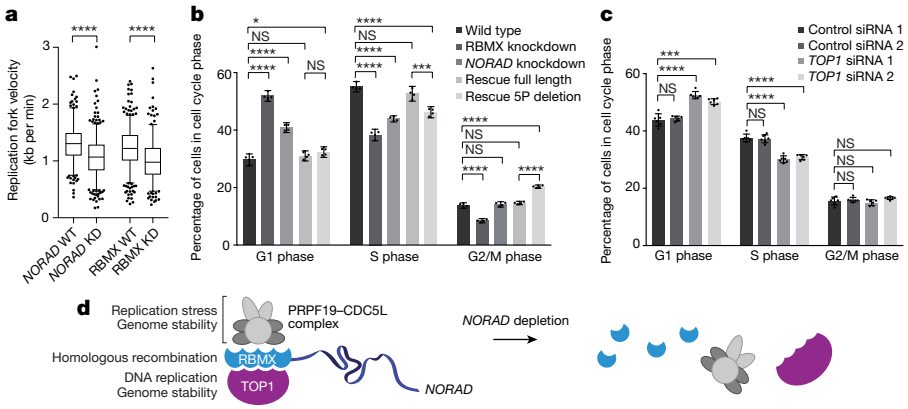


Fig. 4 | Depletion of *NORAD* and NARC1 components affects replication-fork velocity and cell-cycle progression. a, Replication-fork velocity measured by DNA combing (Supplementary Tables 8–10). Centre lines, medians; box limits, 25th and 75th percentiles; whiskers, 5th and 95th percentiles; dots, outliers. ****$P < 0.0001$, two-tailed Mann–Whitney *U* test. Sample size: *NORAD* WT, $n = 280$; *NORAD* KD, $n = 428$; RBMX WT, $n = 438$; RBMX KD, $n = 296$. **b**, Cell-cycle analysis by flow cytometry. CRISPR interference was used to deplete RBMX and *NORAD* ($n = 4$; Supplementary Table 11). Fluorescence-activated cell sorting histograms are shown in Extended Data Fig. 5e. **c**, As in **b**, but for TOP1. RNA interference was used to deplete TOP1 ($n = 6$; Supplementary Table 12). Fluorescence-activated cell sorting histograms are shown in Extended Data Fig. 5f. Values are mean ± 95% confidence interval. ****$P < 0.0001$, ***$P < 0.001$, *$P < 0.05$, NS, not significant, two-tailed Welch's *t*-test. **d**, Model illustrating *NORAD* function.

TOP1 and PRPF19 are part of a 700–1,000-kDa complex (Fig. 3c and Extended Data Fig. 4f). The majority of TOP1 in this complex displays an approximately 50-kDa size shift, consistent with a known SUMO-1 modification of TOP1 proteins that are associated with transcriptionally active or replicating chromatin[25]. Mass spectrometry confirmed that—in addition to RBMX, TOP1 and PRPF19—RBMXL1, which is encoded by an *RBMX* retrogene, is a component of this complex (Fig. 3c). Finally, we speculated that this complex protects *NORAD* from benzonase digestion. We constructed sequencing libraries using RNA extracted from various size-exclusion chromatography fractions. Notably, RNA footprints that matched the previously identified RBMX-binding site in *NORAD* were present only in fractions that contained the complex (Fig. 3c). These data demonstrate that *NORAD* is a physical part of the captured complex.

We next used proximity ligation assays to show that the RBMX–TOP1 interaction occurs in the nucleus, is disrupted by *NORAD* depletion and is rescued by re-introducing full-length *NORAD* into *NORAD*-depleted cells (Fig. 3d and Extended Data Fig. 4g). Importantly, rescue is strongly impaired if the rescue construct lacks the RBMX-binding region (Fig. 3d and Extended Data Fig. 4g).

Our results indicate that *NORAD* modulates the ability of RBMX to interact with other proteins that appear not to bind *NORAD* directly—namely, TOP1 and the core PRPF19–CDC5L complex. Given the distinct molecular composition of this *NORAD*-dependent RBMX complex and the functional importance of its components, we name it *NORAD*-activated ribonucleoprotein complex 1 (NARC1).

Many NARC1 components have prominent roles in maintaining genomic stability. Although individual components such as RBMX or PRPF19 have been reported to contribute to mRNA splicing[13,16], we did not observe global changes in mRNA splicing upon *NORAD* depletion in HCT116 cells (Extended Data Fig. 5a). In other cell types, RBMX and CDC5L can influence the expression of BRCA2 and BRCA1[13,26]. However, *BRCA1* and *BRCA2* were not among differentially expressed genes (Supplementary Table 7) and their proteins levels were not noticeably different in *NORAD*-depleted and normal HCT116 cells (Extended Data Fig. 4c).

Given the essential role of *NORAD* in assembling NARC1, we speculated that *NORAD* depletion may cause a TOP1-related phenotype and directly or indirectly alter DNA replication, which can lead to chromosome segregation defects and genomic instability[27,28]. To assess the functional consequence of *NORAD* depletion on replication, we used the DNA combing technique and measured replication-fork velocity at the single-molecule level. Analysis of over 250 replication forks in wild-type and knockdown cells confirmed that *NORAD* and RBMX depletion significantly reduced replication-fork velocity (Fig. 4a); the observed effect-size was comparable to previously published TOP1 knockdown data[18]. Thus, *NORAD* may directly or indirectly affect DNA replication even in the absence of additional DNA damage stimulus.

We tested whether *NORAD* depletion also affects cell-cycle progression. We labelled newly synthesized DNA with 5-ethynyl-2′-deoxyuridine (EdU) and measured EdU incorporation and total DNA content by fluorescence-activated cell sorting. We observed a clear decrease in S phase accompanied by increased G1 phase in *NORAD*-, RBMX- and TOP1-depleted cells (Fig. 4b, c and Extended Data Fig. 5b–f). Consistent with these findings, impaired replication-fork progression has been linked to chromosome mis-segregation[28] (as observed in *NORAD* and RBMX knockdown cells), which in turn can trigger a cell-cycle arrest in the subsequent G1 phase[29]. Importantly, a G1 arrest alone cannot explain the reduction in replication-fork velocity observed above. We next examined whether the *NORAD*–RBMX interaction is important for this effect on cell-cycle progression. Expression of full-length *NORAD* in *trans* was sufficient to rescue cell-cycle defects in *NORAD*-knockdown cells (Fig. 4b and Extended Data Fig. 5b, e). By contrast, a *NORAD* construct that lacks the RBMX-binding site decreased S phase and increased G2/M phase; this contrasts with *NORAD* knockdown and may point towards an altered molecular function of truncated *NORAD* (Fig. 4b and Extended Data Fig. 5b, e). Deletion of the RBMX-binding site may therefore act as a dominant negative alteration, which indicates that the RBMX-binding region is required for correct *NORAD* function.

Our results link the known function of members of the NARC1 complex (particularly TOP1) in preventing replication stress and genome instability[16,18,19,27] to the role of *NORAD* in suppressing aneuploidy. Importantly, we demonstrate that the RBMX-binding region in *NORAD* contributes to *NORAD* function, presumably by promoting NARC1 assembly.

It has widely been suggested that lncRNAs participate in assembling groups of proteins, but lncRNA–protein complexes have been fully characterized for only a few lncRNAs; these include *XIST*[2], *TERC*[3], *NEAT1*[4], *MALAT1*[5] and *HOTAIR*[6]. Our results demonstrate that *NORAD* is essential for the assembly of the ribonucleoprotein complex NARC1, which physically links proteins known to be involved in DNA replication or repair but not known to act together. We suggest a model in which depletion of *NORAD* or deletion of its RBMX-binding site disrupts NARC1, which alters replication-fork velocity and impairs

cell-cycle progression. It is tempting to speculate that altered DNA replication causes cells to accumulate the observed chromosome segregation defects, a known cause of genomic instability and aneuploidy[28,29] (Fig. 4d). While our data demonstrates a central role of NARC1 in the *NORAD* phenotype, other proteins or complexes may contribute to additional aspects of *NORAD* function.

The precise mechanism or mechanisms by which *NORAD* promotes NARC1 assembly remain to be elucidated but might include (1) inducing a conformational change in RBMX, (2) recruiting a large number of RBMX molecules to its 5′ end to create a protein interaction scaffold or (3) using other direct interactions to bring NARC1 members into close proximity. RBMX encodes a large low-complexity domain that can self-assemble and undergo phase separation in vitro[30]. Binding of RBMX to *NORAD* may nucleate the formation of higher-order RBMX assemblies that facilitate binding of other proteins that contain a low-complexity domain.

In addition to these structural features, *NORAD* has unusual functional features in that *NORAD* localization to the nucleus can be triggered by DNA damage, which may allow cells to rapidly assemble NARC1 or to re-localize pre-assembled complexes without the need for additional protein synthesis.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0453-z.

Received: 10 January 2018; Accepted: 17 July 2018;
Published online 27 August 2018.

1. Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
2. McHugh, C. A. et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
3. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl Acad. Sci. USA* **101**, 10024–10029 (2004).
4. Clemson, C. M. et al. An architectural role for a nuclear noncoding RNA: *NEAT1* RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
5. Tripathi, V. et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
6. Tsai, M.-C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
7. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018).
8. Lee, S. et al. Noncoding RNA *NORAD* regulates genomic stability by sequestering PUMILIO proteins. *Cell* **164**, 69–80 (2016).
9. Tichon, A. et al. A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**, 12209 (2016).
10. Baltz, A. G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
11. Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. *RNA* **16**, 1725–1747 (2010).
12. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
13. Adamson, B., Smogorzewska, A., Sigoillot, F. D., King, R. W. & Elledge, S. J. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat. Cell Biol.* **14**, 318–328 (2012).
14. Matsunaga, S. et al. RBMX: a regulator for maintenance and centromeric protection of sister chromatid cohesion. *Cell Reports* **1**, 299–308 (2012).
15. Pommier, Y., Sun, Y., Huang, S. N. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* **17**, 703–721 (2016).
16. Mahajan, K. hPso4/hPrp19: a critical component of DNA repair and DNA damage checkpoint complexes. *Oncogene* **35**, 2279–2286 (2016).
17. García-Muse, T. & Aguilera, A. Transcription-replication conflicts: how they occur and how they are resolved. *Nat. Rev. Mol. Cell Biol.* **17**, 553–563 (2016).
18. Tuduri, S. et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324 (2009).
19. Gan, W. et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
20. Zhang, N., Kaur, R., Akhter, S. & Legerski, R. J. Cdc5L interacts with ATR and is required for the S-phase cell-cycle checkpoint. *EMBO Rep.* **10**, 1029–1035 (2009).
21. Maréchal, A. et al. PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. *Mol. Cell* **53**, 235–246 (2014).
22. Jeronimo, C. et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell* **27**, 262–274 (2007).
23. Boeing, S. et al. Multiomic analysis of the UV-induced DNA damage response. *Cell Reports* **15**, 1597–1610 (2016).
24. Domínguez-Sánchez, M. S., Barroso, S., Gómez-González, B., Luna, R. & Aguilera, A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet.* **7**, e1002386 (2011).
25. Li, M., Pokharel, S., Wang, J.-T., Xu, X. & Liu, Y. RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nat. Commun.* **6**, 6720 (2015).
26. Abbas, M., Shanmugam, I., Bsaili, M., Hromas, R. & Shaheen, M. The role of the human psoralen 4 (hPso4) protein complex in replication stress and homologous recombination. *J. Biol. Chem.* **289**, 14009–14019 (2014).
27. Miao, Z.-H. et al. Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses. *Cancer Res.* **67**, 8752–8761 (2007).
28. Burrell, R. A. et al. Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494**, 492–496 (2013).
29. Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J. P. L. & Medema, R. H. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* **333**, 1895–1898 (2011).
30. Liu, N. et al. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* **45**, 6051–6063 (2017).

Acknowledgements We thank C. A. McHugh for help setting up RAP MS; K. A. Knouse for help imaging mitotic cells; C. W. Garvie for help with size-exclusion chromatography; L. Ludwig for help with cell-cycle data visualization; L. Gaffney for artwork; T. Wang, B. Cleary, S. R. Grossman, M. Yassour, C. M. Vockley, B. Cimini, K. W. Karhohs, M. Doan, S. A. Myers, D. R. Mani and V. G. Sankaran for discussions. M.M. is supported by a Deutsche Forschungsgemeinschaft (DFG) Research Fellowship. J.M.E. is supported by the Harvard Society of Fellows. M.G. is supported by an NIH Director's Early Independence Award (DP5OD012190), the NIH 4DN program Nucleome Project (U01 DA040612 and U01 HL130007), the New York Stem Cell Foundation, the Edward Mallinckrodt Foundation, Sontag Foundation, Searle Scholars Program, Pew-Steward Scholars program and funds from the California Institute of Technology. M.G. is a NYSCF-Robertson Investigator. Work in the Lander Laboratory is supported by the Broad Institute.

Reviewer information *Nature* thanks R. Bonasio, S. Diederichs and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions M.M. and E.S.L. conceived and designed the study. M.M. performed and analysed all experiments; C.T.N. and K.S. assisted with several experiments, including CLIP. V.S. helped with RNA FISH experiments. C.T.N. analysed RNA FISH and PLA images. C.R.H. collected mass spectrometry data under supervision of M.S. and S.A.C. and helped with design of mass spectrometry experiments. M.G. helped setting up RAP MS, L.H. and J.M.E. developed computational tools and analysed CLIP data. J.M.E. and C.P.F. helped with CRISPR interference and contributed analytical ideas. J.C. performed evolutionary analysis. J.C.U. performed differential gene expression and alternative splicing analysis. M.M. and E.S.L. wrote the manuscript with input from all authors. E.S.L. supervised the work and obtained funding.

Competing interests The Broad Institute holds patents and has filed patent applications on technologies related to other aspects of CRISPR.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0453-z.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0453-z.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.M. or E.S.L.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Tissue culture. We maintained HCT116 cells (ATCC) in McCoy's 5A (Thermo Fisher Scientific) with 10% heat-inactivated FBS (HIFBS, Thermo Fisher Scientific), 1 mM sodium pyruvate, 2mM L-glutamine, and 100 units/ml streptomycin and 100 mg/ml penicillin. Cells were grown at 37 °C and 5% CO_2 atmosphere.

Lentivirus production. We plated 700,000 HEK293T cells in 6-well tissue culture plates and grew them for 24 h before transfecting with 1 μg dVPR, 300 ng VSVG, and 1.2 μg transfer plasmid using FuGene HD (Promega). Sixteen hours after transfection we changed the medium to DMEM with 20% HIFBS. At 48 h post-transfection, we collected viral supernatants and filtered them through a 0.45 μM syringe filter before use.

Generation of CRISPR interference cell lines. We generated inducible CRISPR interference (CRISPRi) cell lines by transducing HCT116 cells with a construct expressing rtTA linked by IRES to a neomycin resistance cassette expressed from an EF1α promoter (ClonTech) and selecting with 200 μg/ml G418 (Thermo Fisher Scientific). Next, rtTA-expressing HCT116 cells were transduced with a previously described KRAB–dCas9 construct linked by IRES to BFP[31]. We selected for cells expressing BFP by fluorescence-activated cell sorting. Inducible *NORAD*, RBMX and PUM2 knockdown cell lines were generated by transducing stable CRISPRi lines with sgRNAs (expressed from a previously described sgOpti backbone[31]) and selecting with 1 μg/ml puromycin.

RAP MS. To capture endogenous *NORAD* transcripts, we designed and synthesized 5′ biotinylated 90-mer DNA oligonucleotides (Integrated DNA Technologies) antisense to the target RNA sequence. We used 26 probes that covered the entire *NORAD* sequence, with the exception of regions that matched to other transcripts or genomic regions as previously described[32]. For *NORAD* and *RMRP* antisense purifications we grew 500 million HCT116 cells per RNA. We supplemented cell culture medium with a final concentration of 200 μM 4-thiouridine and grew cells for 8 h before crosslinking. Cells were washed once with PBS and then crosslinked on ice using 0.8 J/cm^2 of 365-nm ultraviolet light in a Stratalinker (Stratagene). Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 500*g* for 5 min and flash-frozen in liquid nitrogen for storage at −80 °C. Preparation of total cell lysates was performed as previously described[2]. For antisense purification of crosslinked protein–RNA complexes we included the following modifications to the previously described procedure: all buffers were pre-heated to 55 °C. We used 50 μg pooled antisense probes for 500 million lysed cells. For pre-clear of lysates and capture of RNA/DNA hybrids we used 5 ml streptavidin Dynabeads MyOne C1 magnetic beads (Thermo Fisher Scientific) for 500 million cells. Elution of captured proteins from streptavidin beads was achieved by digesting nucleic acids using 250 U of benzonase (Millipore), 25 U RNase A and 1000 U RNase T1 (Thermo Fisher Scientific) for 8 h at 37 °C. Trichloroacetic acid-precipitated proteins were reconstituted in 8 M urea and 50 mM Tris-HCl pH 7.8 and stored at −20 °C until processing for mass spectrometry.

Protein digestion for RAP MS. RAP-captured proteins were resuspended in 40 μl of digestion buffer (8 M urea, 50 mM Tris-HCl pH 7.8), reduced (1 μl of 160 mM DTT, 30 min, room temperature) and alkylated (1.6 μl of 250 mM IAA, 45 min, dark, room temperature), followed by a 2 h Lys-c digestion (0.1 μg per sample) at room temperature. Next, the samples were diluted with 120 μl of 100 mM Tris-HCl pH 7.8 to a final concentration of <2 M urea, and 0.5 μg of trypsin was added for overnight digestion at room temperature with agitation. Samples were quenched with 8.5 μl of formic acid and desalted on 4-punch STAGE-Tips as previously described[33].

iTRAQ labelling of peptides and BRP fractionation for RAP MS. Desalted peptides were labelled with iTRAQ4[34] reagent according to the manufacturer's instructions (AB Sciex). Peptides were dissolved in 30 μl of 50 mM triethylamonium bicarbonate (TEAB) pH 8.5 and labelling reagent was added in 70 μl of ethanol. Samples were incubated with labelling reagent for 1 h with agitation, and the reaction was quenched with 5 μl of 1 M Tris-HCl pH 7.8. Differentially labelled peptides were subsequently mixed and prepared for BRP fractionation on 50 mg SepPak columns according to the following protocol: cartridges were prepared for desalting by equilibrating with methanol, 50% acetonitrile (ACN), 1% formic acid and 3 washes with 0.1% TFA. Samples were loaded on to the cartridge and washed 3 times with 1% formic acid. A pH switch was performed before the collection of BRP fractions with 5 mM ammonium formate at pH 10. BRP fractions were collected at the following ACN concentrations: 10% ACN in 5 mM ammonium formate; 15% ACN in 5 mM ammonium formate; 20% ACN in 5 mM ammonium formate; 30% ACN in 5 mM ammonium formate; 40% ACN in 5 mM ammonium formate; and 50% ACN in 5 mM ammonium formate.

Co-immunoprecipitation and MS. To capture RBMX-interacting proteins, we grew 15 million inducible CRISPRi cells with stably integrated *NORAD* sgRNAs for each immunoprecipitation experiment. For *NORAD* depletion samples, we induced knockdown by supplementing cell culture medium with 0.5 μg/ml doxycycline for 72 h, while *NORAD* wild-type samples were grown without doxycycline. Cells were washed in PBS, trypsinized and collected by centrifugation. Cell pellets were washed twice with ice-cold PBS and cell numbers were counted and normalized between knockdown and wild-type samples. Fresh cell pellets containing 15 million cells were lysed in 375 μl co-IP lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP40, 0.1% sodium deoxycholate, and Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific)). Lysates were incubated on ice for 30 min and mixed by pipetting every 5–10 min to enhance nuclear lysis. Lysates were cleared by centrifugation at 14,000*g* for 10 min at 4 °C and insoluble material was removed. We pre-cleared lysates by incubating with 50 μl protein A magnetic beads (Thermo Fisher Scientific) for 30 min at 4 °C. Meanwhile, 900 ng RBMX antibody (Cell Signaling #14794) was pre-coupled to 50 μl protein A beads for 45 min at room temperature. We determined the total protein concentration in pre-cleared lysates by BCA assay in triplicates and normalized all samples to contain exactly 2.5 mg total protein. To non-specifically digest all DNA and RNA, we added 50 U benzonase and 1 mM $MgCl_2$ to all lysates. Free RBMX antibody was removed from magnetic beads and benzonase-treated lysates were added to beads and incubated overnight at 4 °C. The next day, supernatant was removed and beads were washed twice in 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.05% NP40, followed by two washes in 50 mM Tris-HCl pH 7.5 and 150 mM NaCl. After the last wash, beads were overlaid with 10 μl PBS and immediately subjected to sample preparation for mass spectrometry and TMT labelling.

On-bead protein digestion for co-IP MS. Following immunoprecipitation, washed beads were resuspended in 90 μl of digestion buffer (2 M urea, 50 mM Tris-HCl pH 7.8, 2 mM DTT, 0.005 μg/ml sequencing-grade trypsin) and incubated for 1 h with agitation at 700 rpm. The supernatant was removed and placed in a fresh tube. Beads were washed two times with 60 μl of 2 M urea in 150 mM Tris-HCl pH 7.8, and washes were combined with the supernatant. This procedure was repeated twice to ensure complete removal of proteins from the beads. Supernatants were combined and proteins were reduced (3.5 μl of 500 mM DTT, 30 min, room temperature) and alkylated (9 μl of IAA, 45 min, room temperature, dark), before digestion with 4 μg of trypsin overnight at room temperature with agitation. Samples were acidified (1% formic acid) and desalted on Waters 10 mg Oasis HLB cartridges.

TMT labelling of peptides and BRP Fractionation for co-IP MS. Desalted peptides were labelled with TMT6 reagent according to the manufacturer's instructions (Thermo Fisher Scientific). Peptides were dissolved in 25 μl of HEPES pH 8.5 and 0.2 mg of TMT labelling reagent was added to each sample in 10 μl of ACN. Samples were incubated with labelling reagent for 1 h with agitation. Next, the reaction was quenched with 2 μl of 5% hydroxylamine. Differentially labelled peptides were subsequently mixed and prepared for BRP fractionation on 50 mg SepPak columns according to the following protocol: cartridges were prepared for desalting by equilibrating with methanol, 50% ACN, 1% formic acid and 3 washes with 0.1% TFA. Samples were loaded on to the cartridge and washed 3 times with 1% formic acid. A pH switch was performed with 5 mM ammonium formate at pH 10, collected and run as fraction 1. Subsequent fractions were collected at the following ACN concentrations: 10% ACN in 5 mM ammonium formate; 15% ACN in 5 mM ammonium formate; 20% ACN in 5 mM ammonium formate; 30% ACN in 5 mM ammonium formate; 40% ACN in 5 mM ammonium formate; and 50% ACN in 5 mM ammonium formate.

LC–MS/MS Analysis (RAP MS and co-IP MS). Reconstituted peptides were injected onto a capillary column (Picofrit with 10-μm tip opening, 75-μm diameter, New Objective) packed in-house with 20 cm C18 silica material (1.9 μm ReproSil-Pur C18-AQ medium, Dr Maisch GmbH), and separated on an online nanoflow EASY-nLC 1000 UHPLC system (Thermo Fisher Scientific). Columns were heated to 50 °C in column heater sleeves (Phoenix-ST) to reduce back-pressure during the gradient.

RAP MS experiments. Peptides were separated at a flow rate of 200 nl/min with a linear 120-min gradient from 100% solvent A (3% ACN, 0.1% formic acid) to 35% solvent B (90% ACN, 0.1% formic acid) for 82 min, followed by a 3-min linear increase from 35 to 90% B with a 5-min hold at 60% B before increasing to 90% B for 3 min and holding for 20 min, and equilibrating back at 50% B for 10 min to end the gradient.

Co-IP MS experiments. Peptides in each BRP fraction were separated at a flow rate of 200 nl/min over a linear gradient of 100% A to 20% B for 28 min, with a linear increase from 20% B to 60% B for 16 min, and a hold at 90% B for 5 min before returning to 50% B.

Peptides were analysed on an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) operated in data-dependent mode. Higher-energy collision dissociation tandem mass spectrometry (HCD MS/MS) scans (resolution = 17,500 for iTRAQ and TMT methods) were taken after each MS1 scan (resolution = 70,000) on the top 12 most abundant ions using an AGC target of

3×10^6 ions for MS1 and 5×10^4 ions for MS2. The isolation widths for MS/MS ions were 1.6 for iTRAQ and TMT methods. The maximum ion fill-time for MS/MS scans was 120 ms, the HCD-normalized collision energy was 29; dynamic exclusion time was set to 20 s, and peptide match and isotope exclusion functions were enabled.

Quantification and identification of peptides and proteins (RAP MS and co-IP MS). All mass spectra were processed using the Spectrum Mill software package v.6.01 pre-release (Agilent Technologies), which includes modules developed for iTRAQ and TMT6-based quantification. Precursor ion quantification was done using extracted ion chromatograms for each precursor ion. The peak area for the extracted ion chromatogram of each precursor ion subjected to MS/MS was calculated in the intervening high-resolution MS1 scans of the LC–MS/MS runs using narrow windows around each individual member of the isotope cluster. Peak widths in both time and *m*/*z* domains were dynamically determined on the basis of mass spectrometry scan resolution, precursor charge and *m*/*z*, subject to quality metrics on the relative distribution of the peaks in the isotope cluster versus theoretical. Similar MS/MS spectra acquired on the same precursor *m*/*z* in the same dissociation mode with ± 60 s were merged. MS/MS spectra with precursor charge >7 and poor quality MS/MS spectra, which failed the quality filter by having a sequence tag length less than 1, were excluded from searching.

For peptide identification, MS/MS spectra were searched against the human Uniprot database to which a set of common laboratory contaminant proteins was appended. Search parameters included: ESI-QEXACTIVE-HCD scoring parameters, trypsin or Lys-c/trypsin enzyme specificity with a maximum of 2 missed cleavage, 40% minimum matched peak intensity, ± 20 ppm precursor mass tolerance, ± 20 ppm product mass tolerance, and carbamidomethylation of cysteins and isobaric labelling of lysines and N-termini as fixed modifications in the RAP MS (iTRAQ) and the immunoprecipitation mass spectrometry (TMT6) experiments with no fixed modification on lysines or N-termini for the size-exclusion chromatography experiment. Oxidation of methionine, N-terminal acetylation and deamidated (N) were allowed as variable modifications, with a precursor MH^+ shift range from −18 to 64 Da. Identities interpreted for individual spectra were automatically designated as valid by optimizing score and delta rank1–rank2 score thresholds separately for each precursor charge state in each LC–MS/MS run, while allowing a maximum target-decoy-based false-discovery rate (FDR) of 1.0% at the spectrum level.

In calculating scores at the protein level and reporting the identified proteins, redundancy is addressed in the following manner: the protein score is the sum of the scores of distinct peptides. A distinct peptide is the single highest scoring instance of a peptide detected through an MS/MS spectrum. MS/MS spectra for a particular peptide may have been recorded multiple times (that is, from different precursor charge states, isolated from adjacent BRP fractions or modified by oxidation of Met), but are still counted as a single distinct peptide. When a peptide sequence over eight residues long is contained in multiple protein entries in the sequence database, the proteins are grouped together and the highest scoring one and its accession number are reported. In some cases in which the protein sequences are grouped in this manner, there are distinct peptides that uniquely represent a lower scoring member of the group (isoforms or family members). Each of these instances spawns a subgroup, and multiple subgroups are reported and counted towards the total number of proteins identified. iTRAQ and TMT ratios were obtained from the protein comparisons export table in Spectrum Mill. To obtain iTRAQ or TMT protein ratios, the median was calculated over all of the distinct peptides assigned to a protein subgroup in each replicate. For RAP MS we required each protein to be detected with more than two unique peptides. To enable precise quantification, we limited our analysis to peptides that are uniquely assigned to a specific protein isoform or family member. For co-IP MS we required more than four unique peptides. For statistical analysis, we used the Limma package[35] in R (https://www.r-project.org/) to calculate multiple comparison adjusted *P* values using a moderated *t*-test.

Native RBMX co-IP and size-excusion chromatograpy. To capture native RBMX complexes, we generated stable HCT116 cell lines that express Flag–RBMX–V5. For co-IP and size-exclusion chromatography experiments, we grew 200 million cells. Cells were collected by scraping culture dishes, washed once with PBS and pelleted by centrifugation at 500*g* for 5 min. Fresh cell pellets were lysed in 8 ml co-IP lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP40, 0.1% sodium deoxycholate, and Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific)). Lysates were incubated on ice for 30 min and mixed by pipetting every 5–10 min to enhance nuclear lysis. Lysates were cleared by centrifugation at 14,000*g* for 10 min at 4 °C and insoluble material was removed. We pre-cleared lysates by incubating with 2.5 ml protein G magnetic beads (Thermo Fisher Scientific) for 30 min at 4 °C. Meanwhile, 600 μg Flag M2 antibody (Sigma # F1804) was pre-coupled to 2.5 ml protein G beads for 45 min at room temperature. To non-specifically digest DNA and RNA, we added 500 U benzonase to the cell lysate. Free Flag M2 antibody was removed from magnetic beads and benzonase-treated lysates were added to beads and incubated overnight at 4 °C. The next day, supernatant was removed and beads were washed twice in 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.05% NP40, followed by two washes in 50 mM Tris-HCl pH 7.5 and 150 mM NaCl. After the last wash, protein complexes were eluted using 250 μg Flag-peptide in 500 μl of 150 mM NaCl, 25 mM Tris pH 7.5, 0.05% IGEPAL. Elutions were incubated 1 h at 4 °C with agitation. Eluates were separated from beads and filtered using a 0.2-μm membrane filter. Size-exclusion chromatography of the RBMX complex was performed using a Superose 6 Increase 10/300 column (GE Healthcare) equilibrated in 150 mM NaCl, 25 mM Tris pH 7.5, 0.05% IGEPAL. We injected 400 μl of the eluate onto the column at a flow rate of 0.4 ml/min and collected 0.5-ml fractions. Two hundred and fifty microliters of each fraction was subjected to trichloroacetic acid-precipitation to concentrate proteins. Protein content was analysed by western blotting and mass spectrometry.

For mass spectrometry analysis, proteins were reduced, alkylated and denatured at 90 °C for 5 min, spun down and loaded separated by SDS–PAGE. The gel was run in 1× MES SDS–PAGE running buffer at 175 V for 40 min, after which it was stained for 2 h in SimplyBlue Safe Stain (Thermo Fisher Scientific) and destained in water overnight. The gel lane was cut into 4 fractions, diced and destained with 50% ACN, 50% 100mM ammonium bicarbonate. Destaining buffer was removed and gel pieces were dehydrated with 300 μl of ACN. ACN was aspirated once the gel pieces were white. One-and-a-half micrograms of trypsin was added to each of the 4 fractions in 100 μl of 100 mM ammonium bicarbonate (pH 8) and incubated overnight at 37 °C. The supernatant from each fraction was collected into in a fresh tube, and the peptides were extracted from the gel pieces by washing twice with 60% ACN, 0.1% formic acid and collecting the extract in the tube with the initial supernatant. Finally, gel pieces were dehydrated with ACN, which was collected with the rest of the extract. Fractions were then dried using a Speedvac concentrator, reconstituted in 3% ACN and 0.1% formic acid, and desalted on C18 Stage Tips[33]. Eluate from each fraction was transferred to HPLC vials, dried down and reconstituted in 5 μl of 3% ACN, 5% formic acid and run on an EasyNLC 1200 coupled to an Orbitrap Q Exactive Plus mass spectrometer. The previously described method for co-IP MS experiments (see 'Co-IP MS experiments' above) was used for analysis, with the only difference being a normalized collision energy of 25, which is routinely used for label-free peptide analysis.

To extract RNA from size-exclusion chromatography fractions containing the protein complex as well as control fractions we Trizol-extracted the remaining 250 μl of sample, and isolated RNA using Direct-zol columns (Zymo Research). We removed rRNA with the NEBNext rRNA Depletion Kit (New England Biolabs) by following the manufacturer's instructions. Finally, we constructed RNA-sequencing libraries using the SMARTer smRNA-Seq Kit (Clontech) by following the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 2500 instrument to an average read depth of 15–20 million reads with 50-bp read 1 and 60-bp read 2. We trimmed 5 bp from the beginning of read 1 and 15 bp from the beginning of read 2 before mapping. Reads aligning to rRNA were removed from downstream analysis[36]. Reads were then mapped to hg19 using Bowtie2. Mapping results were restricted to the single best alignment found for any given read. Discordant alignments of paired-end reads were excluded from analysis. Data normalization was performed by scaling coverage values by (1,000,000/total mapped read count).

CLIP. The CLIP protocol below is extensively based on three previously published CLIP methods: irCLIP[37], PAR-CLIP[38] and eCLIP[39].

We constructed the pre-adenylated irCLIP adaptor as previously described[37]. All other oligonucleotides were synthesized as described in the irCLIP protocol, with the exception of reverse transcription primers. We replaced ethyleneglycol spacers with three deoxyuridines and modified the 5′ end of reverse transcription primers to reflect the nucleotide preference of CircLigase II (general structure: /5phos/RNNNNN-6nt-barcode-NNNNNTACCCTTCGCTTCACACACAAG/ideoxyU//ideoxyU//ideoxyU/TACTGAACCGC).

For each CLIP experiment we grew 20 million HCT116 cells in medium supplemented with 200 μM 4-thiouridine for 8 h. Cells were washed once with PBS and then crosslinked on ice using 0.2 J/cm^2 of 365-nm ultraviolet light in a Stratalinker. Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 500*g* for 5 min and flash-frozen in liquid nitrogen for storage at −80 °C. To prepare cell lysates, pellets were thawed on ice and resuspended in NP40 lysis buffer (50 mM HEPES pH 7.5, 150 mM KCl, 2 mM EDTA, 1% (v/v) NP40, 0.25 mM DTT, complete EDTA-free protease inhibitor cocktail) and incubated on ice for 10 min. We sonicated cell lysates using a Branson Digital Sonifier with a microtip set at 5 W power for a total of 1 min 30 s in intermittent pulses (0.7-s on, 2.3-s off), followed by RNase I (Thermo Fisher Scientific) digestion (0.5 U/μl, 10 min at 23 °C). Subsequently, we added 15 μl/ml Murine RNase Inhibitor (New England Biolabs), followed by DNA digestion (20 U TURBO DNase (2 U/μl; Thermo Fisher Scientific), 2.5 mM $MgCl_2$ and 0.5 mM CaCl) for 20 min at 37 °C. We incubated samples on ice for 10 min before clearing lysates by centrifugation at

15,000*g* for 15 min. Insoluble material was removed and total protein concentration was determined by BCA assay. Cell lysates were flash-frozen and stored in batches of 10 mg total protein at −80 °C.

For each immunoprecipitation experiment, lysates (10 mg total protein) were thawed on ice and pre-cleared by incubating with protein A/G magnetic beads (using 30 µl/mg total protein) for 30 min at 4 °C. In the meantime, antibodies (6 µg/mg total protein) were coupled to protein A/G magnetic beads (using 30 µl/mg total protein) at room temperature for 45 min (antibodies used: RBMX, Cell Signaling #14794; ALYREF, Bethyl # A302-892A; PUM1, Bethyl # A302-577A; V5, Abcam # ab27671). We removed unbound antibody and added the pre-cleared lysates to antibody-coupled beads and incubated overnight at 4 °C. The following day, we washed the beads 3 times in IP wash buffer (50 mM HEPES pH 7.5, 300 mM KCl, 0.5% (v/v) NP40, 0.25 mM DTT, complete EDTA-free protease inhibitor cocktail), followed by one wash in FastAP buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 100 mM KCl, 0.02% Triton X-100). Immunopurified protein–RNA complexes were dephosphorylated by resuspending beads in 25 µl FastAP mix (18.5 µl H_2O, 2.5 µl 10× FastAP buffer (Thermo Fisher Scientific), 2.5 U FastAP enzyme (1 U/µl; Thermo Fisher Scientific), 0.5 µl Murine RNase Inhibitor (New England Biolabs)) and incubating for 20 min at 37 °C. In the meantime, we prepared polynucleotide kinase mix (56 µl H_2O, 10 µl 10× PNK buffer (New England Biolabs), 1 µl Murine RNase Inhibitor, 7 µl T4 PNK (10 U/µl; New England Biolabs), 1 µl TURBO DNase) and added 75 µl to each 25 µl sample and incubated 20 min at 37 °C. Beads were separated on a magnet and dephosphorylation reaction was removed before washing beads once in RNA ligation buffer without DTT (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$). Next, 3′ ligation was performed by resuspending beads in 20 µl ligation mix (3 µl H_2O, 2 µl 10× T4 RNA ligation buffer (New England Biolabs), 1 µl DMSO, 1 µl RNase inhibitor, 15 pmoles pre-adenylated 3′ adaptor, 10 µl 50% PEG 8000, 2 µl T4 RNA Ligase 1 High Concentration (New England Biolabs)) using low-retention pipette tips and incubated overnight at 16 °C with agitation. The next day we added 7 µl 4× NuPAGE LDS Sample Buffer (Thermo Fisher Scientific) to ligation reactions and incubated samples for 10 min at 75 °C. Protein–RNA complexes were resolved by SDS–PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 h, followed by transfer to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific). Protein–RNA complexes were visualized using the Odyssey Clx infrared imager (LI-COR) and desired complexes were excised from membrane using a clean scalpel. Membrane pieces were immediately subjected to proteinase K treatment by adding 250 µl proteinase K solution (4 mg/ml Proteinase K (New England Biolabs), 100 mM Tris-HCl pH 7.5, 150 mM NaCl, 12.5 mM EDTA, 1% (w/v) SDS) and incubating 1 h at 55 °C. Following proteinase K treatment, RNA was phenol-chloroform extracted using Heavy Phase Lock Gel tubes (5Prime) and purified with the Zymo RNA Clean & Concentrator-5 kit by following the manufacturer's instructions for small and large RNAs. We eluted RNA in 7 µl H_2O and combined it with 10 pmoles of reverse transcription primer. Samples were heated to 72 °C for 2 min and snap-cooled on ice. Reverse transcription was performed with the SuperScript III kit (Thermo Fisher Scientific) by combining RNA samples with 4 µl 5× First Strand Buffer, 2 µl 0.1M DTT and 6 µl 2mM dNTPs. Samples were incubated at 50 °C for 3 min before adding 1 µl SuperScript III reverse transcription enzyme (200 U/µl) and incubating 1 h at 42 °C. For each reverse transcription reaction, we washed 5 µl MyOne streptavidin C1 beads twice in NT2 buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$ and 0.0005% NP40) and then resuspended the beads in 100 µl NT2 buffer. The resuspended beads were then added to the reverse transcription reaction and the mixture was subsequently incubated for 15 min at room temperature. Beads were washed twice in streptavidin wash buffer (20 mM Tris-HCl pH 7.5, 120 mM NaCl, 2 5mM KCl, 5 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate) and twice in PBS to remove unbound reverse transcription primer. Finally, we resuspended beads in 10 µl of freshly prepared elution buffer (8.25 µl H2O, 1 µl 1 µM elution oligonucleotide (CTGAACCGCTCTTCCGATCT), 0.75 µl of 50 mM $MnCl_2$) and heated the reaction for 5 min at 95 °C, 1 min at 75 °C, followed by a ramp of 0.1 °C per s to 60 °C and holding at 60 °C for 15 min. Once the 60-°C incubation temperature was reached, we prepared CircLigase mix (2 µl H_2O, 2 µl 10× CircLigase-II buffer (Epicentre), 0.25 µl 50mM $MnCl_2$, 4 µl 5M betaine, 1 µl CircLigase-II (Epicentre)) and added 10 µl to each elution without removing beads. We incubated the reaction 2 h at 60 °C. Following incubation, we added 2 reaction volumes Agencourt AMPure XP beads (Beckman Coulter) and 5 reaction volumes isopropanol and incubated 15 min at room temperature. Supernatant was removed and beads were washed once with 80% ethanol (v/v), air-dried and eluted by resuspending dry beads in 25 µl H_2O, heating for 2 min at 95 °C and isolating supernatants. We used 12 µl cDNA for PCR amplification (program: 2 min at 98 °C followed by 12–15 cycles of 98 °C for 15 s, 65 °C for 30 s and 72 °C for 30 s, followed by a final 20-min extension at 72 °C) in a 50 µl reaction using 25 µl 2× NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs), 12 µl H_2O and 1 µl of 25µM PCR1 primer mix containing P3_PCR1 (GCATTCCTGCTGAACCGCTCTTCCGATCT) and P6_PCR1 (TTTCCCCTTGTGTGTGAAGCGAAGGGTA) primers. PCR reactions were subjected to two consecutive rounds of purification using 1.5 volumes of Agencourt AMPure XP beads and two 70% ethanol washes. DNA was eluted in 14 µl H_2O and subjected to a second PCR amplification (program: 2 min at 98 °C followed by 3–6 cycles of 98 °C for 15 s and 72 °C for 45 s, followed by a final 2-min extension at 72 °C) in a 50 µl reaction using 12 µl purified PCR1, 25 µl 2× NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs), 12 µl H_2O and 1 µl of 25 µM PCR2 primer mix containing P3_PCR2 (CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT) and P6_PCR2 (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCCT TGTGTGTGAAGC GAAGGGTA) primers. Upon completion of PCR amplification, we added 10 µl ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific) and incubated reactions for 15 min at 37 °C. Reactions were purified using 1.1 volumes of Agencourt AMPure XP beads and two 70% ethanol washes, followed by elution of air-dried beads in 10 µl H_2O. The concentration of final libraries was determined with the Qubit dsDNA HS Assay (Thermo Fisher Scientific) and library sizes were analysed on a High Sensitivity DNA Bioanalyzer Chip (Agilent).

Size-matched input libraries (SM input)[39] were prepared by resolving 1–2% of input lysates by SDS–PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 h. SDS–PAGE gels were transferred to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific) and proteins migrating at the molecular weight range of the target protein were excised using a clean scalpel. RNA was released by proteinase K treatment and purified as described in the previous section. We performed end-repair of input RNA by adjusting RNA volume to 19.5 µl with H_2O and adding 2.5 µl 10× FastAP buffer (Thermo Fisher Scientific), 2.5 U FastAP enzyme (1 U/µl; Thermo Fisher Scientific), 0.5 µl Murine RNase Inhibitor (New England Biolabs)) and incubating for 20 min at 37 °C. In the meantime, we prepared Polynucleotide kinase mix (56 µl H_2O, 10 µl 10× PNK buffer (New England Biolabs), 1 µl Murine RNase Inhibitor, 7 µl T4 PNK (10 U/µl; New England Biolabs), 1 µl TURBO DNase) and added 75 µl to each 25 µl sample and incubated samples for 20 min at 37 °C. RNA was purified with the Zymo RNA Clean & Concentrator-5 kit using the manufacturer's instructions for small and large RNAs. RNA was eluted in 5 µl H_2O and combined with 25 µl ligation mix (3 µl 10× T4 RNA ligation buffer (New England Biolabs), 1.5 µl DMSO, 1.5 µl RNase inhibitor, 15 pmoles pre-adenylated 3′ adaptor, 15 µl 50% PEG 8000, 3 µl T4 RNA Ligase 1 High Concentration (New England Biolabs)) using low-retention pipette tips and incubated for 2 h at 23 °C with agitation. Ligation reactions were purified to remove free 3′ adaptor using two consecutive Silane bead purifications. For each reaction, we washed 15 µl Silane beads (Thermo Fisher Scientific) twice in 1 ml RLT buffer (Qiagen), resuspended beads in 90 µl RLT and combined 90 µl beads in RLT with 30 µl ligation reaction. We added 0.7 volumes 100% ethanol and incubated mixtures 10 min at room temperature. Supernatant was removed and beads were washed twice with 70% ethanol before eluting air-dried beads in 9 µl H_2O. We used 7 µl of the eluted RNA for reverse transcription and proceeded with the library preparation as described in the above section.

Computational analysis of CLIP data. We sequenced CLIP and corresponding SM input libraries on an Illumina HiSeq 2500 to an average read depth of 30–50 million reads with 52-bp read 1 and 35-bp read 2. The first read includes a 6-nt barcode added during reverse transcription (see 'CLIP' above). After processing to separate samples based on inline barcodes, sequencing reads collected from all CLIP experiments were first mapped to hg19 using TopHat (v.2.0.8)[40]. Reads aligning to rRNA were removed from downstream analysis, as previously described[36]. Duplicate reads were identified and removed using Picard's MarkDuplicates program. Peak calling was performed with the MACS2[41] algorithm to identify genomic coordinates where experimental conditions (protein IP) were significantly enriched for reads relative to size-matched controls (SM input). Peak calling was performed without a shifting model and the band width to compute fragment size was set to 100 bp. Significant peaks are reported with FDR correction of $q = 0.05$. Significant peaks were further filtered to include only regions with an average minimum depth of two reads in the size-matched control condition. To compile significant results across replicate experiments, we intersected the intervals from the peak calling output of each replicate. Normalized coverage in the intersection peaks was first calculated separately for each replicate as the average depth at a given peak divided by the total number of reads after correcting for the observed duplication rate. The mean of the relative fold change between the two replicates was calculated for each peak and peaks that did not show a twofold or greater change in both replicates were excluded. We report a CLIP signal score for a given peak as the product of enrichment (average fold change) and the peak length (see Supplementary Table 3).

RNA-sequencing and analysis. We performed RNA-sequencing on cells that stably expressed individual sgRNAs targeting the *NORAD* promoter, with or without doxycycline-induced KRAB–dCas9 expression. We performed at least 2 biological replicate experiments for knockdown and control conditions after 24 h, 48 h and 96 h of KRAB–dCas9 induction. RNA-sequencing libraries were constructed

as previously described[36]. Reads were pseudo-aligned to hg19 (ENSEMBL transcripts) using kallisto[42] with an index of either 31 or 21 k-mers. Estimated counts were collapsed across transcripts into genes and differential expression analysis was performed using DESeq2[43]. Genes with an absolute $\log_2$(fold change) > 1 and FDR < 0.05 were considered as differentially expressed. *P* values for differential gene expression were corrected using the Benjamini–Hochberg procedure to derive an FDR.

Alternative splicing analysis. Percentage spliced in (PSI) for different exons or introns was calculated using SUPPA2[44] based on isoform transcripts per million (TPM) estimates from kallisto for skipping exon, alternative 5′ or 3′ splice sites, mutually exclusive exons, retained intron and alternative first or last exons. Differential PSI was calculated using diffSplice[45] with the parameters '–area 1000–tpm-threshold 5–lower-bound 0.00 -gc'. Events with a change in PSI $> 20\%$ and FDR < 0.05 were considered as differentially used. *P* values across putative splicing events were corrected using the Benjamini–Hochberg procedure to derive an FDR.

RNA extraction and RT–qPCR. We extracted RNA from 20,000–50,000 cells per experiment in RLT buffer (Qiagen) using Dynabeads MyOne Silane beads (Thermo Fisher Scientific), treated samples with TURBO DNase (Thermo Fisher Scientific) and cleaned again with Silane beads. We used AffinityScript reverse transcriptase (Agilent Technologies) and random nonamer primers to convert RNA to cDNA. We performed qPCR using SYBR Green I Master Mix (Roche) and calculated differences using the $\Delta\Delta C_t$ method versus *GAPDH*. To achieve power to detect small effects in gene expression, we performed three technical qPCR replicates (from the same cDNA) and took the median value for further analysis. We used the following RT–qPCR primers in this study. *RBMX* forward primer: CAGTTCGCAGTAGCAGTGGA, *RBMX* reverse primer: TCGAGGTGGACCTCCATAAC; *NORAD* forward primer: CTCTGCTGT GGCTGCCC, *NORAD* reverse primer: GGGTGGGAAAGAGAGGTTCG; *PUM2* forward primer: GGGAGCTTCTCACCATTCAATG, *PUM2* reverse primer: CCA TGAAAACCCTGTCCAGATC; *GAPDH* forward primer: AGCCACATCGC TCAGAC AC, *GAPDH* reverse primer: GCCCAATACGACCAAATCC; *MALAT1* forward primer: AGTTCAGTGTTGGGGCAATC, *MALAT1* reverse primer: GTTCTTCCGCTCAAATCCTG; *TOP1* forward primer: TCGAAGCGG ATTTCCGATTGA, *TOP1* reverse primer: CTTTGTGCCGGTGTTCTCGAT.

Co-IP western blot. Co-IP experiments were carried out as described above (see 'Co-immunoprecipitation and MS'). The following antibodies were used for immunoprecipitation reactions: RBMX, Cell Signaling #14794, or Santa Cruz Biotechnology # sc-14581; ALYREF, Bethyl # A302-892A; TOP1, Bethyl # A302-589A; CDC5L, Bethyl #A301-681A. Following the last washing step, we resuspended beads in 20 μl Pierce IgG Elution Buffer (Thermo Fisher Scientific) and incubated them for 20 min at room temperature with agitation. We collected supernatants and added 7 μl 4× NuPAGE LDS Sample Buffer (Thermo Fisher Scientific), followed by a 3-min incubation at 95 °C. Proteins were resolved by SDS–PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 h, followed by transfer to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific). Proteins larger than 150 kDa in size were resolved on NuPAGE 3–8% Tris-Acetate Gels (Thermo Fisher Scientific). Western blots were performed using the iBind Western System (Thermo Fisher Scientific). For protein detection, we used the following primary antibodies: RBMX, Cell Signaling #14794, or Santa Cruz Biotechnology #sc-14581; ALYREF, Santa Cruz Biotechnology #sc-32311; TOP1, Santa Cruz Biotechnology # sc-32736; CDC5L, Santa Cruz Biotechnology #sc-81220. We used the following secondary antibodies: IRDye 680RD Goat anti-Mouse IgG (H + L) (LI-COR), IRDye 800CW Goat anti-Rabbit IgG (H + L) (LI-COR), IRDye 800CW Donkey anti-Goat IgG (H + L) (LI-COR). For visualization of bands, we used the Odyssey Clx infrared imager system (LI-COR).

***NORAD* conservation analysis.** We tested for conservation of *NORAD* transcription across 11 mammalian species: human, chimpanzee, gorilla, orangutan, rhesus macaque, mouse, rat, ferret, dog, cow and armadillo. Because expression of *NORAD* is highest in human brain[46] we checked for transcription in brain tissue from these 11 species. Raw RNA-sequencing read data were downloaded from previous studies[47–49] and mapped to respective genomes using STAR v.2.5.2a[50], with gene annotations from Ensembl Release 91[51] as a reference guide. For each RNA-sequencing library, a de novo transcriptome was made using Stringtie v.1.3.3b[52], using default parameters. Samples from the same species were then merged using the stringtie–merge option. To find reciprocal best hits, we used nucleotide BLAST with default parameters. Multiple sequence alignment was created using MAFFT[53] with gap penalty reduced to 1.0.

Subcellular fractionation. We prepared nuclear and cytoplasmic extracts from freshly grown HCT116 cells using the PARIS Kit (Thermo Fisher Scientific) by following the manufacturer's instructions. We used ~10 million cells for each fractionation experiment and analysed extracts by western blot or RT–qPCR as described in the corresponding sections.

smRNA FISH. smRNA FISH experiments were performed using the ViewRNA Cell Plus Assay Kit (Thermo Fisher Scientific) and following the manufacturer's instructions. We grew 50,000 cells in black 12-well glass-bottom plates (Cellvis) for 24 h. To induce DNA damage, we supplemented culture medium with doxorubicin (1 μM) or camptothecin (200 nM) for 12 h. We washed cells once with PBS before fixation. Cells were then fixed and permeabilized simultaneously in fixation/permeabilization buffer for 30 min at room temperature on a rotating plate. After three brief washes in PBS, we incubated cells with the appropriate probe set diluted 1:100 in probe set diluent for 2 h at 40 °C, then with preamplifier mix at 40 °C for 70 min, followed by amplifier mix at 40 °C for 70 min, and finally label probe mix at 40 °C for 60 min. For nuclei staining, we incubated the cells for 2 min at room temperature with 1× ViewRNA Cell Plus DAPI in PBS. Cells were then washed three times in PBS and then incubated with Alexa Fluor 647 phalliodin (Cell Signaling Technology) diluted 1:20 in PBS for 15 min at room temperature for staining of actin filaments. After a final set of washes, we covered cells with ProLong Gold Anti-Fade Reagent (Cell Signaling Technology) and stored the plates at 4 °C until imaging. The probe sets and corresponding fluorophores were type 1 – *NORAD* and *MALAT1* (Alexa Fluor 546) and type 4 – *GAPDH* (Alexa Fluor 488). Confocal microscopy was performed using a Nikon Eclipse Ti1 with Andor Yokogawa Spinning Disk Revolution WD system.

Quantification of RNA FISH images. For three-dimensional FISH image analysis, *Z*-stacks were exported such that the top and bottom slices were the beginning and end of DAPI signal in the *z* direction. Quantification of FISH foci was done with FISH-quant[54] in MATLAB (version R2017b) following the software's instructions for mature mRNA quantification. Before spot detection, a dual Gaussian filter was applied to the images in FISH-quant using the default settings. The outline of nuclei and cells were determined automatically with the Cell Segmentation Tool in FISH-quant and a modified version of a Cell Profiler pipeline provided in the FISH-quant repository. In Cell Profiler (v.2.2.0)[55], nuclear boundaries were determined by the Otsu method guided by DAPI staining. Identified nuclei were then used as seeds to identify the boundaries of cells by the watershed method aided with the phalloidin stain. For all probes, the local maximum strategy of spot pre-detection was used. Settings for thresholding pre-detected spots were optimized for each probe separately to account for differences in signal intensity.

In situ proximity ligation assay. In situ proximity ligation assay (PLA) was performed using the Duolink PLA platform (Sigma) and following the manufacturer's instructions. Cells were plated in black, glass-bottom 96-well plates the day before the experiment and allowed to grow overnight at 37 °C. Cells were fixed with 4% paraformaldehyde in PBS for 15 min at room temperature, washed three times in PBS and then permeabilized with 0.5% Triton X-100 in PBS for 15 min at room temperature. Cells were blocked for 1 h at 37 °C in a humidified chamber using the Duolink blocking solution, and subsequently stained with primary antibodies diluted 1:250 in Duolink antibody diluent for 1 h at room temperature. DuoLink PLA probes (Rabbit PLUS and Mouse MINUS) were added for 1 h at 37 °C. The ligation and subsequent amplification steps were performed for 30 min and 100 min, respectively, at 37 °C. Upon completion of the assay, cells were overlaid with Duolink mounting medium with DAPI. Two sets of primary antibody pairs were used: rabbit anti-RBMX (Cell Signaling Technologies #14794) was paired with mouse anti-TOP1 (Thermo Fisher Scientific #435900); mouse anti-Flag (Cell Signaling Technologies #8146S) (targeting Flag–RBMX) was paired with rabbit anti-TOP1 (Bethyl #A302-589A).

Quantification of PLA images. For PLA signal quantification we used Cell Profiler 3.0.0. Separate maximum intensity projections for each channel were exported. Nuclei and PLA-signal segmentation was performed using the minimal cross entropy thresholding method. We applied default settings for nuclei segmentation, whereas the PLA signal detection required more stringent thresholding to distinguish individual spots within clusters. A size filter was applied to exclude overlapping nuclei from the analysis. The total nuclear PLA spot count was normalized to the total nuclear area for each cell.

Immunostaining of cultured cells for anaphase nuclei imaging. We induced knockdown in CRISPRi cells with stably integrated sgRNAs by supplementing cell culture medium with 0.5 μg/ml doxycycline for 48 h. Cells were then trypsinized and plated in multi-well glass-bottom plates (Cellvis), again supplementing culture medium with doxycycline in knockdown cells, and grown for an additional 24 h. We removed culture medium, rinsed each well in PBS and fixed cells in 4% paraformaldehyde (PFA) for 10 min at room temperature. All subsequent manipulation steps were carried out in a humidified chamber. PFA was removed, cells were washed twice in PBS and permeabilized by incubating with PBS + 0.1% Triton X-100 for 10 min at room temperature. Following permeabilization, we blocked cells in PBS containing 4% BSA (Roche), 10% goat serum (Sigma Aldrich) and 0.1% Triton X-100 for 30 min at room temperature. Primary antibodies (anti-α-tubulin—FITC antibody, Sigma Aldrich #F2168 (1:1,000); anti-centromere antibodies, Antibodies Incorporated # 15-234-0001 (1:200)) were diluted in blocking buffer (PBS containing 4% BSA (Roche), 10% goat serum (Sigma Aldrich) and

0.1% Triton X-100) and incubated in the dark for 2 h at room temperature or overnight at 4 °C. Following antibody incubation, cells were washed 3 times in PBS + 0.1% Triton X-100 and incubated 10 min at room temperature in between each washing step. Secondary antibody (Goat anti-Human IgG (H+L) Cross-Adsorbed, Alexa Fluor 568, Thermo Fisher Scientific #A-21090 (1:250)) was diluted in blocking buffer and added to cells for 1 h at room temperature in the dark. Cells were washed 3 times in PBS + 0.1% Triton X-100 and incubated 10 min at room temperature in between each washing step. Following removal of any residual washing buffer, we covered cells with ProLong Gold Antifade reagent containing DAPI (Thermo Fisher Scientific) and allowed them to cure overnight in the dark before imaging. Confocal microscopy was performed using a Nikon Eclipse Ti1 with Andor Yokogawa Spinning Disk Revolution WD system.

Generation of *NORAD trans* rescue constructs and cell lines. We synthesized the full-length *NORAD* cDNA (Genewiz) and cloned it into pDONR221 (Thermo Fisher Scientific). Using Gateway technology, we cloned *NORAD* downstream of a CAG promoter in a destination vector expressing BFP linked by IRES to a hygromycin resistance cassette driven by an EF1α promoter (ClonTech). 5′-truncated *NORAD* was generated by deleting bases 33–898 from *NORAD* in pDONR221 using site-directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, New England Biolabs). Sequence-verified 5′-truncated *NORAD* pDONR221 was cloned into the described destination vector using LR recombination.

Sequence-verified rescue constructs were transfected into CRISPRi cells with stably integrated *NORAD* sgRNAs using FuGene HD (Promega) by following the manufacturer's instructions. We selected cells that stably integrated *NORAD* rescue constructs by selecting with hygromycin B (Sigma Aldrich) at a final concentration of 25 μg/ml. Knockdown of endogenous *NORAD* was achieved by inducing KRAB–dCas9 expression in CRISPRi cell lines stably expressing sgRNAs targeting the endogenous *NORAD* promoter using doxycycline at 0.5 μg/ml. RT–qPCR Primers specific to the 5′ end of *NORAD* (forward primer: CTCTGCTGTGGCTGCCC, reverse primer: GGGTGGGAAAGAGAGGTTCG) or a middle segment of *NORAD* (forward primer: CTCTCCACCACCAACCTGATG, reverse primer: GGAAGTGAGATAACATCAGCTCTAA) were used to verify expression of full-length or 5′-truncated *NORAD* in cells depleted of endogenous *NORAD*.

Cell-cycle analysis. Cell-cycle analysis was carried out by measuring EdU incorporation and total DNA content. CRISPRi cells with stably integrated *NORAD* or *RBMX* sgRNAs and stably integrated rescue cassettes expressing different *NORAD* constructs (full-length *NORAD*, 5′-truncated *NORAD* or empty rescue cassette) were maintained in medium containing hygromycin B (Sigma Aldrich) at a final concentration of 12.5 μg/ml. Induction of KRAB–dCas9 and constitutive expression of rescue cassettes was routinely monitored by fluorescence-activated cell sorting. Medium supplemented with 0.5 μg/ml doxycycline was added to knockdown samples for 48 h. We then trypsinized cells and plated them in 24-well cell culture plates using 100,000 cells per well and incubated them for another 24 h in the presence of doxycycline. We labelled newly replicating DNA by supplementing cell culture medium with 10 μM EdU for 1 h. Cells were washed with PBS, trypsinized and transferred to a 96-well round-bottom plate for improved handling of many samples in parallel. We used the Click-iT Plus EdU Flow Cytometry Assay Kit (Thermo Fisher Scientific) and followed the manufacturer's instructions with the following modifications. For improved multiplexing, we reduced the number of cells per assay by a factor of 10 (1×10^6 cells/ml) and scaled down washing volumes accordingly. The Click-IT reaction was performed using half the recommended reagent volumes per sample. After the last washing step, cells were resuspended in PBS containing FxCycle Far Red Stain (Thermo Fisher Scientific) as well as RNase Cocktail (Thermo Fisher Scientific) and incubated for 30 min at room temperature to stain total DNA. Fluorescence-activated cell sorting was performed on a CytoFLEX S Instrument (Beckman Coulter).

RNAi knockdown experiments. For RNAi knockdown experiments, we plated 50,000 HCT116 cells 24 h before transfection into 24-well tissue culture plates using antibiotic-free medium. We transfected 50 nM short interfering RNAs (siRNAs) into each well using Lipofectamine RNAiMAX (Thermo Fisher Scientific) by following the manufacturer's instructions. Medium was changed the day after transfections and cells were incubated with siRNAs for a total time of 72 h. Cell-cycle analysis and RT–qPCR were performed as described in the above sections (see 'Cell-cycle analysis' and 'RNA extraction and RT–qPCR' above).

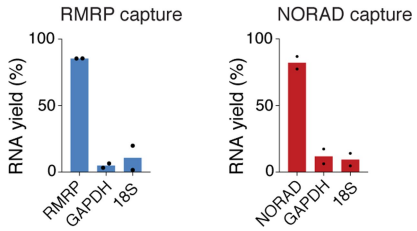
DNA combing. We induced knockdown in CRISPRi cells with stably integrated sgRNAs by supplementing cell culture medium with 0.5 μg/ml doxycycline for 72 h. Knockdown and wild-type cells were labelled with a final concentration of 100 μM CIdU for 70 min. CIdU-containing medium was removed, cells were washed twice with warm medium (which did not contain any thymidine analogues) and then incubated with a final concentration of 100 μM IdU for 70 min. IdU-containing medium was removed, cells were washed twice with warm PBS and trypsinized. We counted cells in triplicates and used 75,000 cells for each experiment. Cells were embedded in agarose plugs using the FibrePrep DNA Extraction Kit (Genomic Vision) by following the manufacturer's instructions. DNA combing, immunodetection, image acquisition and data analysis were performed at specialized service facilities. Only intact replication origins with positive DNA counterstaining were used to measure fibre length and calculate replication fork velocity.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

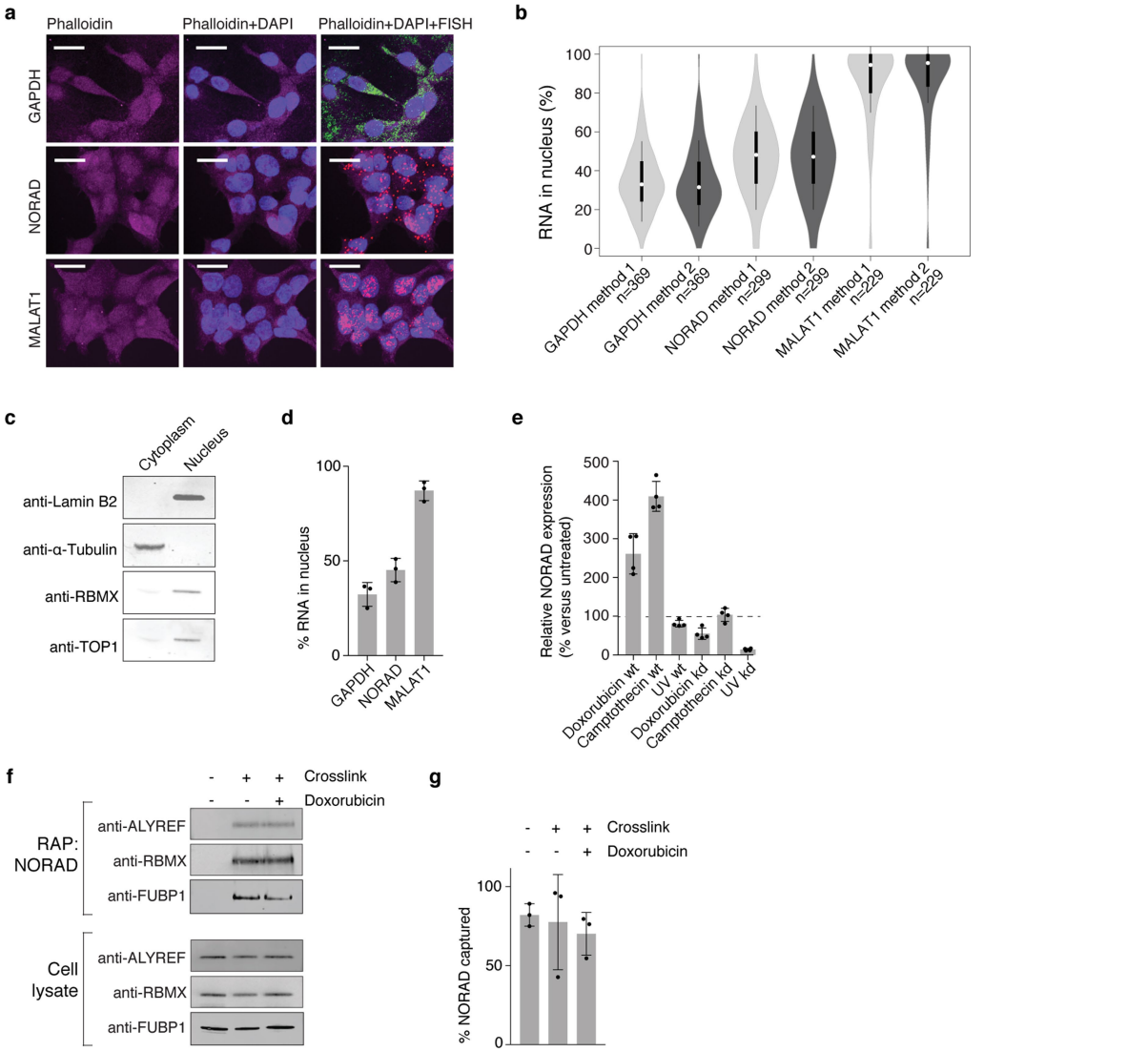
Code availability. Code for the analyses described in this paper is available from the corresponding authors upon request.

Data availability. Sequencing data for this study are available at the Gene Expression Omnibus under the accession number GSE114953. The original mass spectra may be downloaded from MassIVE (http://massive.ucsd.edu) using the identifier: MSV000082561. The data are directly accessible via ftp://massive.ucsd.edu/MSV000082561. All other data are available from the corresponding authors upon reasonable request.

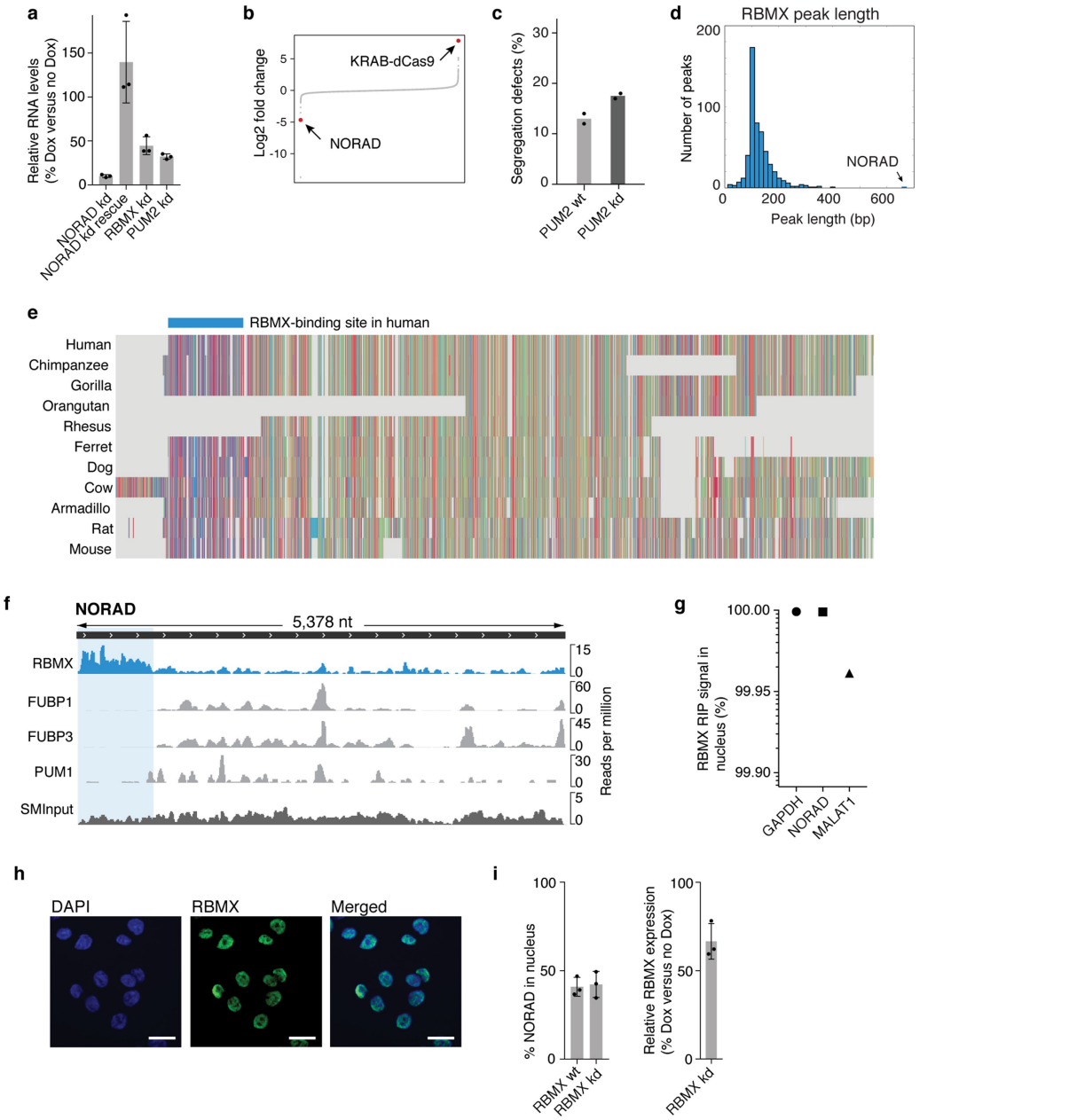
31. Fulco, C. P. et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
32. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
33. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
34. Wiese, S., Reidegeld, K. A., Meyer, H. E. & Warscheid, B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7**, 340–350 (2007).
35. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, article 3 (2004).
36. Engreitz, J. M. et al. RNA–RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* **159**, 188–199 (2014).
37. Zarnegar, B. J. et al. irCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods* **13**, 489–492 (2016).
38. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
39. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
40. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
41. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
42. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
44. Trincado, J. L. et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **19**, 40 (2018).
45. Hu, Y. et al. DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res.* **41**, e39 (2013).
46. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
47. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
48. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
49. Chen, J. et al. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* **17**, 19 (2016).
50. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
51. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2017).
52. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
53. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
54. Mueller, F. et al. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).
55. Kamentsky, L. et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).



Extended Data Fig. 1 | RNA antisense purification of *RMRP* and *NORAD* transcripts. RT–qPCR measurements of RNA yield in *RMRP* and *NORAD* RAP MS experiments. Columns represent the mean of two biological replicate experiments, individual data points are shown.



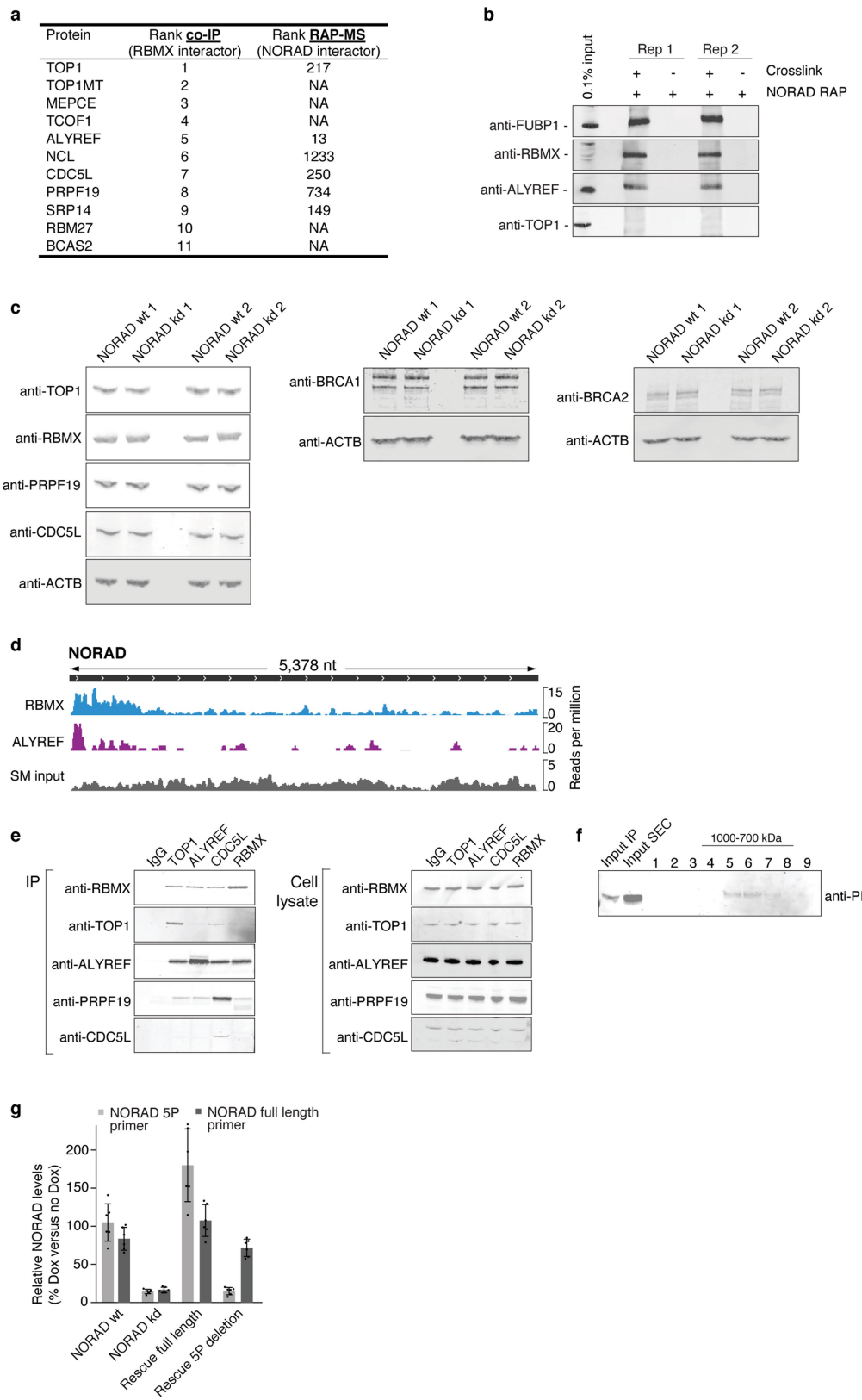
Extended Data Fig. 2 | Subcellular localization of *NORAD* and analysis of *NORAD*–protein interactions with DNA damage. a, smRNA FISH of *GAPDH*, *NORAD* and *MALAT1* in wild-type HCT116 cells. *GAPDH*, cytoplasmic reference; *MALAT1*, nuclear reference. Actin is stained with Alexa Fluor 647-conjungated phalloidin. Scale bar, 20 μm. Images are representative of one experiment; three independent experiments were performed. **b**, Quantification of smRNA FISH experiments. Circles show medians; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range; polygons, extreme values. Method 1: phalloidin-aided cell boundary detection using the watershed method. Method 2: proximity-based cell boundary detection using the distance − N method in Cell Profiler. Sample sizes: *GAPDH* method 1, $n = 369$; *GAPDH* method 2, $n = 369$; *NORAD* method 1, $n = 299$; *NORAD* method 2, $n = 299$; *MALAT1* method 1, $n = 229$; *MALAT1* method 2, $n = 229$. **c**, Subcellular fractionation of HCT116 cells. Lamin B2 and α-tubulin serve as controls for nuclear and cytoplasmic fractions, respectively. Western blots are representative of one experiment; three independent experiments were performed. **d**, RT–qPCR measurements of relative RNA levels in nuclear and cytoplasmic extracts. Quantification relative to *GAPDH*. Percent nuclear extract is calculated relative to the total signal observed in nuclear and cytoplasmic fractions. Values are mean ± standard deviation ($n = 3$). **e**, RT–qPCR measurements of *NORAD* expression upon doxorubicin, camptothecin or ultraviolet treatment in *NORAD* wild-type or knockdown cells. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 4$). **f**, Western blot of *NORAD* RAP experiments with or without DNA damage. Western blots are representative of one experiment; three independent experiments were performed. **g**, RT–qPCR measurements of RNA yield in *NORAD* RAP experiments. Values are mean ± standard deviation ($n = 3$).



Extended Data Fig. 3 | Analysis of *NORAD* knockdown, *NORAD* conservation and *NORAD*–protein interactions. a, RT–qPCR measurements of *NORAD*, *RBMX* and *PUM2* CRISPRi knockdown and *NORAD* rescue experiments. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 3$). **b**, Differentially expressed genes in RNA-sequencing experiments from *NORAD* CRISPRi knockdown cells. **c**, Quantification of chromosome segregation errors in PUM2 wild-type or knockdown cells. One hundred anaphases were scored for each condition. Columns represent the mean of two biological replicate experiments, individual data points are shown. **d**, Histogram of RBMX-binding-site length in CLIP experiments. **e**, Multiple sequence alignment of *NORAD* transcripts, assembled de novo from RNA-sequencing data from 11 mammalian species. Only transcribed sequences are shown. Blue bar indicates RBMX-binding site in human *NORAD*. Alignment colour scheme: A, orange; C, blue; T, green; G, red. **f**, CLIP data plotted across *NORAD* RNA for RBMX, FUBP1, FUBP3 and PUM1. RBMX SM input library is shown. Representative alignments from two biological replicates are shown. **g**, RBMX RIP in nuclear and cytoplasmic fractions. The percentage of nuclear RIP signal is calculated relative to the total signal observed in nuclear and cytoplasmic fractions. **h**, Immunofluorescence imaging of RBMX in HCT116 cells. Scale bar, 20 μm. Representative images from three biological replicates are shown. **i**, Left, RT–qPCR measurements of *NORAD* RNA levels in nuclear and cytoplasmic extracts under RBMX CRISPRi wild-type or knockdown conditions. The percentage of nuclear *NORAD* is calculated relative to the total signal observed in nuclear and cytoplasmic fractions. Right, RT–qPCR measurements of *RBMX* CRISPRi knockdown. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 3$).

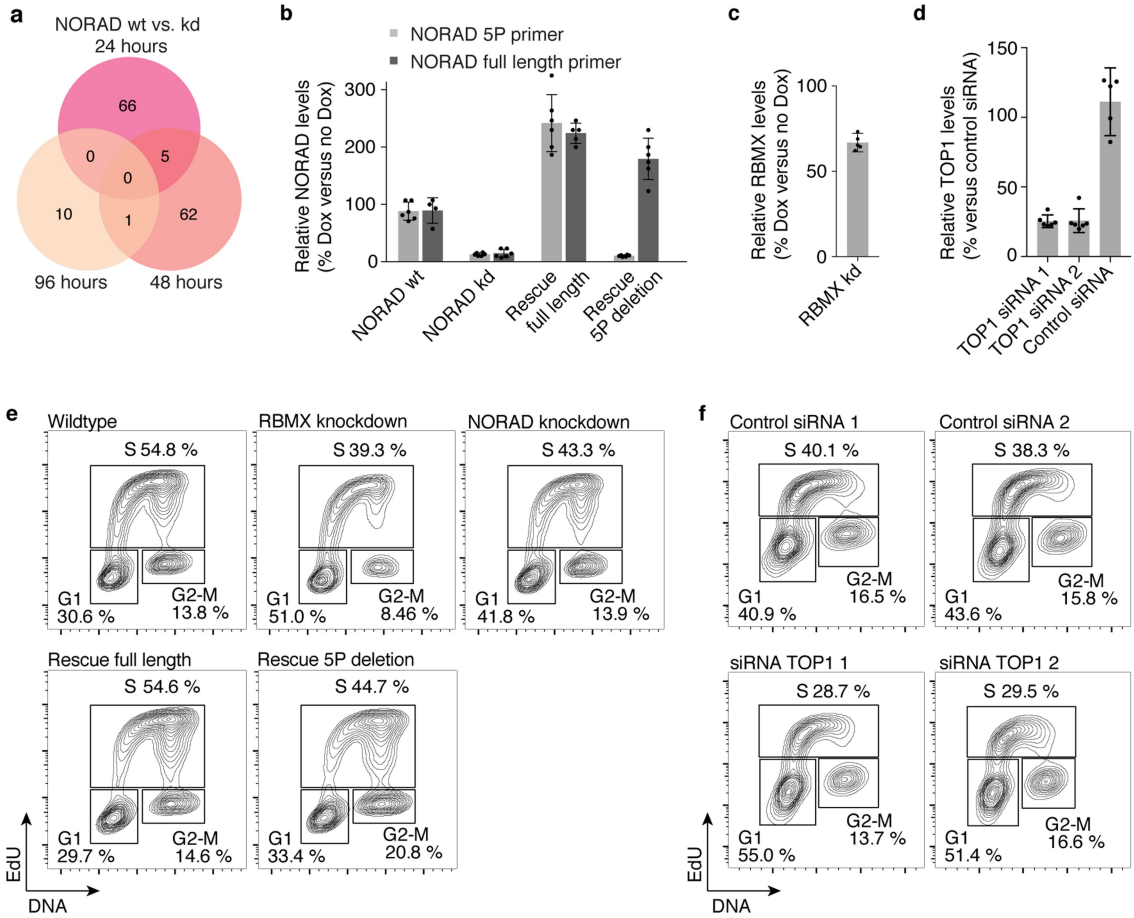
a

Protein	Rank co-IP (RBMX interactor)	Rank RAP-MS (NORAD interactor)
TOP1	1	217
TOP1MT	2	NA
MEPCE	3	NA
TCOF1	4	NA
ALYREF	5	13
NCL	6	1233
CDC5L	7	250
PRPF19	8	734
SRP14	9	149
RBM27	10	NA
BCAS2	11	NA



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Analysis of RBMX protein–protein interactions and their dependency on *NORAD*. **a**, Ranked list of *NORAD*-dependent RBMX-interacting proteins identified by quantitative co-IP MS (Supplementary Table 4) and their respective rank in *NORAD* RAP MS experiments. **b**, Western blot of two independent *NORAD* RAP experiments with or without crosslink. Antibodies were pooled and incubated with the same membrane. Corresponding size regions were cropped for simplicity of presentation. **c**, Western blot of levels of TOP1, RBMX, PRPF19, CDC5L, BRCA1 and BRCA2 proteins in *NORAD* wild-type and knockdown cells from two independent experiments. β-actin serves as loading control. **d**, CLIP data plotted across *NORAD* RNA for RBMX and ALYREF. RBMX SM input library is shown. Representative alignments from two biological replicates are shown. **e**, Co-IP western blot for TOP1, ALYREF, CDC5L, RBMX and IgG control. Inputs are shown on the right. Western blots are representative of one experiment; three independent experiments were performed. **f**, Western blot of Flag–RBMX–V5 co-IP followed by size-exclusion chromatography. Fractions 1–9 are shown. Fractions 10–20 were not probed for PRPF19 owing to overlap with Flag antibody at this size range (Supplementary Note 4). **g**, RT–qPCR measurements of *NORAD* 5′ fragment (light grey) and full-length *NORAD* (dark grey) in rescue experiments using full-length and 5′-truncated *NORAD* rescue constructs. Measurements correspond to cells used for proximity ligation assays. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 6$).



Extended Data Fig. 5 | Analysis of alternative splicing and cell-cycle progression in *NORAD* depleted cells. a, Venn diagram of significant splicing changes (percentage spliced in (PSI) > 20%; FDR < 0.05) in *NORAD* wild-type and knockdown cells at 24, 48 and 96 h (Supplementary Table 6); 89,352, 88,529 and 84,340 events were analysed at 24, 48 and 96 h, respectively. Only six events were consistent between two time points and none were consistent between all three time points **b**, RT–qPCR measurements of *NORAD* 5′ fragment (light grey) and full-length *NORAD* (dark grey) in rescue experiments using full-length and 5′-truncated *NORAD* rescue constructs. Measurements correspond to cells used in cell-cycle analysis. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 5$ or 6). **c**, RT–qPCR measurements of *RBMX* CRISPRi knockdown. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 5$). **d**, RT–qPCR measurements of *TOP1* RNA interference knockdown. Quantification relative to *GAPDH*. Values are mean ± standard deviation (*TOP1* siRNA, $n = 6$; control siRNA, $n = 5$). **e**, Representative fluorescence-activated cell sorting histograms measuring EdU incorporation and DNA content in RBMX and *NORAD* CRISPRi knockdown and *NORAD* rescue cells. Percentage of cells in each cell-cycle phase is indicated. **f**, As in **e**, but for *TOP1* RNA interference knockdown cells.

nature research

Corresponding author(s): Eric S. Lander and Mathias Munschauer

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Spectrum Mill v6.01 pre-release, NIS elements, CytExpert.
Data analysis	Graphpad Prism 7, FlowJo, Microsoft Excel, R studio, MATLAB (version R2017b), FISH-QUANT, Cell Profiler (version 2.2.0 and version 3.0.0), BEDTools, Integrative Genomics Viewer (version 2.3.26), Igvtools 2.3, STAR v2.5.2a, TopHat version 2.0.8, MACS2, Picard's MarkDuplicates, Bowtie 2, DESeq2, SUPPA2, diffSplice, kallisto, Stringtie v1.3.3b, MAFFT.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data associated with Fig. 2, 3, and Extended Data Fig. 3, 4, 5 is available at the Gene Expression Omnibus under the accession number GSE114953.

The original mass spectra associated with Fig. 1, 3, and Extended Data Fig. 4 may be downloaded from MassIVE (http://massive.ucsd.edu) using the identifier: MSV000012345. The data is directly accessible via ftp://massive.ucsd.edu/MSV000012345.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical measures were used to predetermine sample size. We typically performed at least 3 biological replicates for each experiment, unless otherwise noted in the figure legends. Quantitative mass spectrometry experiments using isobaric mass tag labeling include 2 replicate experiments in each 4plex or 6plex mass tag cassette as is common practice.
Data exclusions	Mass spectrometry data were filtered for common laboratory contaminants. We further required a minimal number of unique peptides as defined in the Methods. Otherwise no data were excluded.
Replication	As reported in the figure legends, main text and Methods, the findings were reliably reproduced.
Randomization	This study uses a candidate-based approach to dissect molecular interactions and biochemical mechanisms. No randomization was required because the results of biochemical measurements or sequencing of nucleic acid libraries are not affected by sample randomization.
Blinding	This study uses a candidate-based approach to dissect molecular interactions and biochemical mechanisms. No blinding was required because the results of biochemical measurements or sequencing of nucleic acid libraries are not affected by knowledge of sample identities. DNA combing was performed by scientists at specialized service facilities and sample identities were not disclosed to the person performing the experiment or data analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

RBMX – Cell Signaling (D7C2V) #14794, Lot #1, 2, (IP (6 ug/ml), western blot (1:1000), proximity ligation assay (1:250), immunofluorescence (1:250))
RBMX – Santa Cruz Biotechnology (G17) #sc-14581, Lot #K0614, (IP (6 ug/ml) and western blot (1:1000))
ALYREF – Bethyl #A302-892A, Lot #1, (IP (6 ug/ml))
ALYREF – Santa Cruz Biotechnology (11G5) #sc-32311, Lot # J1016, (western blot (1:1000))
PUM1 – Bethyl #A302-577A, Lot #1, (IP (6 ug/ml))
V5 – Abcam (SV5-Pk1) #ab27671, Lot #GR242588-4, (IP (6ug/ml))
V5 – Abcam #ab9116, Lot #GR256657-12, (western blot (1:1000))
Flag – Cell Signaling (9A3) #8146S, Lot #3, (proximity ligation assay (1:250))
TOP1 – Bethyl #A302-589A, Lot #1, (IP (6 ug/ml), proximity ligation assay (1:250))
TOP1 – Bethyl #A302-590A, Lot #3, (western blot 1:1000)
TOP1 – Santa Cruz Biotechnology (C-21) #sc-32736, Lot #K0515, (western blot (1:1000))
TOP1 – Thermo Fisher Scientific (24.5) #435900, Lot #QC215904, (proximity ligation assay (1:250))
CDC5L – Bethyl #A301-681A, Lot #1, (IP (6 ug/ml))
CDC5L – Santa Cruz Biotechnology (2136C1a) #sc-81220, Lot #C0917, (western blot (1:1000))
PRPF19 – Bethyl #A300-101A, Lot #1, (western blot (1:500))

ACTB – Abcam #ab8226 (western blot (1:1000))
BRCA1 – Bethyl #A301-378A, Lot #2, (western blot (1:1000))
BRCA2 – Bethyl #A300-005A (western blot (1:1000))
Alpha-Tubulin–FITC – Sigma Aldrich #F2168 (immunofluorescence (1:1000))
Anti-Centromere antibodies (ACA) – Antibodies Incorporated #15-234-0001 (immunofluorescence (1:200))
Normal Rabbit IgG - Cell signaling #2729, Lot #8, (IP (6 ug/ml))

Validation

All antibodies were validated by western blot. Commercial antibodies were also validated by the manufactures as indicated on their web sites.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HCT116 cell line was acquired from ATCC.
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	Cell lines tested negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	A detailed description of the sample preparation procedure is given in the Methods sections "Cell cycle analysis".
Instrument	Beckman Coulter Cytoflex S
Software	CytExpert, FlowJo.
Cell population abundance	No post-sort fractions were collected.
Gating strategy	Cell cycle analysis: We applied forward and side scatter parameters (FSC, SSC) to exclude cell debris and doublets as exemplified in Supplementary Figure 1. Typically, 3-4% of events were excluded to remove events that could not be assigned to cells in G1-, S-, and G2-M-phase. The fraction of cells in each phase of the cell cycle was then quantified using indicated gates (contour plot).

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.