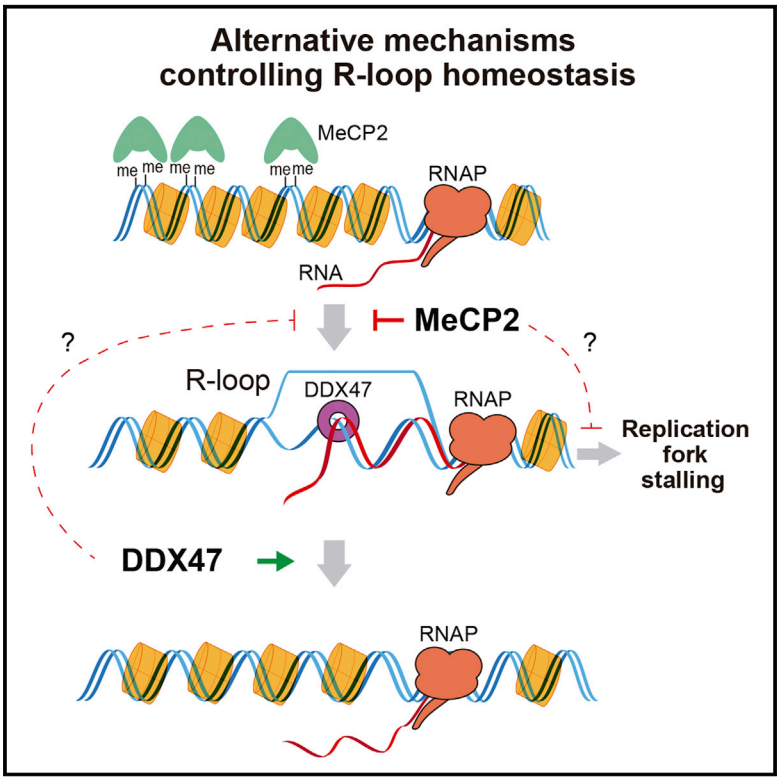
Cell Reports

Article

DDX47, MeCP2, and other functionally heterogeneous factors protect cells from harmful R loops

Graphical abstract



Highlights

- An siRNA screening identifies up to 59 factors affecting R-loop homeostasis
- The methyl-CpG-binding protein MeCP2 prevents R loops and associated DNA damage
- DDX47 unwinds DNA-RNA hybrids and protects cells from harmful R loops
- Dysfunction of different cellular processes increases co-transcriptional R loops

Authors

Esther Marchena-Cruz, Lola P. Camino, Jay Bhandari, ..., Xiaoyu Xue, Rosa Luna, Andrés Aguilera

Correspondence

rlvaro@us.es (R.L.),
aguilo@us.es (A.A.)

In brief

Marchena-Cruz et al. present a functional assay identifying 59 factors that protect against harmful R loops and are known to be involved in different cellular processes, including RNA metabolism and chromatin function. They study two of these factors, DDX47 and MeCP2, and conclude that R loops accumulate upon dysfunction of unrelated processes.

Marchena-Cruz et al., 2023, Cell Reports *42*, 112148
March 28, 2023
https://doi.org/10.1016/j.celrep.2023.112148

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

DDX47, MeCP2, and other functionally heterogeneous factors protect cells from harmful R loops

Esther Marchena-Cruz,[1,4] Lola P. Camino,[1,4] Jay Bhandari,[3] Sónia Silva,[1] José Javier Marqueta-Gracia,[1,2] Shahad A. Amdeen,[3] Cristina Guillén-Mendoza,[1,2] María L. García-Rubio,[1,2] José M. Calderón-Montaño,[1] Xiaoyu Xue,[3] Rosa Luna,[1,2,*] and Andrés Aguilera[1,2,5,*]

[1]Centro Andaluz de Biología Molecular y Medicina Regenerativa-CABIMER, Universidad de Sevilla-CSIC-Universidad Pablo de Olavide, 41092 Seville, Spain

[2]Departmento de Genética, Facultad de Biología, Universidad de Sevilla, 41012 Seville, Spain

[3]Department of Chemistry and Biochemistry, Texas State University, San Marcos, TX 78666, USA

[4]These authors contributed equally

[5]Lead contact

*Correspondence: rlvaro@us.es (R.L.), aguilo@us.es (A.A.)

https://doi.org/10.1016/j.celrep.2023.112148

SUMMARY

Unscheduled R loops can be a source of genome instability, a hallmark of cancer cells. Although targeted proteomic approaches and cellular analysis of specific mutants have uncovered factors potentially involved in R-loop homeostasis, we report a more open screening of factors whose depletion causes R loops based on the ability of activation-induced cytidine deaminase (AID) to target R loops. Immunofluorescence analysis of γH2AX caused by small interfering RNAs (siRNAs) covering 3,205 protein-coding genes identifies 59 potential candidates, from which 13 are analyzed further and show a significant increase of R loops. Such candidates are enriched in factors involved in chromatin, transcription, and RNA biogenesis and other processes. A more focused study shows that the DDX47 helicase is an R-loop resolvase, whereas the MeCP2 methyl-CpG-binding protein uncovers a link between DNA methylation and R loops. Thus, our results suggest that a plethora of gene dysfunctions can alter cell physiology via affecting R-loop homeostasis by different mechanisms.

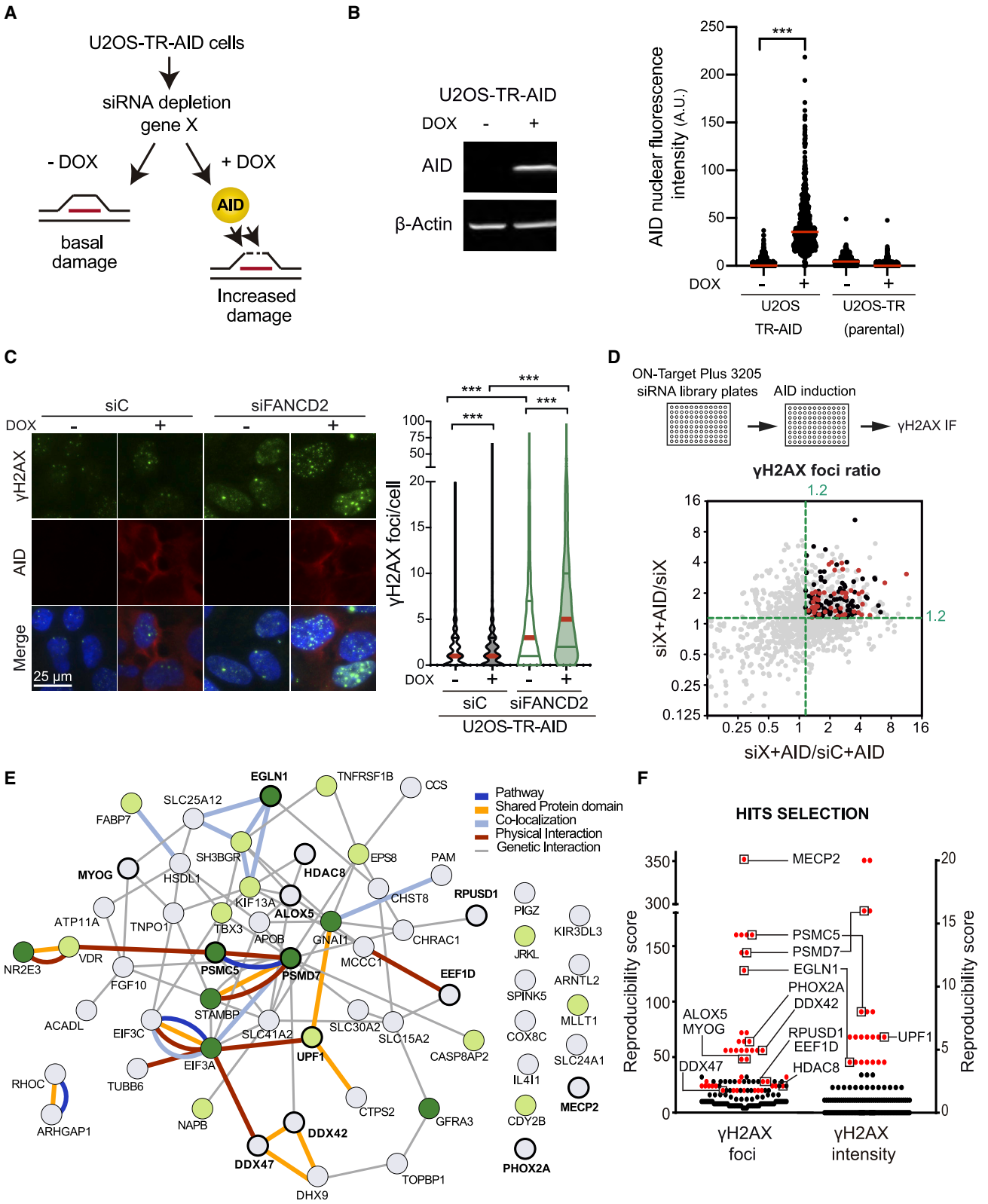
INTRODUCTION

R loops are byproducts resulting from the co-transcriptional hybridization of the nascent RNA with the DNA template. They are dynamic structures that form potentially all over the transcribed genome.[1] R loops participate in different biological processes such as immunoglobulin class-switching recombination, mitochondrial replication, transcription termination, gene regulation, and telomere stability.[2–4] Apart from these specific physiological roles, R loops are an important source of genome instability because of their ability to halt replication or the action of nucleases on the single-stranded DNA (ssDNA) displaced by the DNA-RNA hybrid.[5,6] The recent connection between the accumulation of R loops and human diseases, such as neurodegenerative pathologies, autoimmunity, or cancer, highlights the relevance of the study of R-loop homeostasis.[7–11]

Cells have evolved factors and mechanisms to prevent or resolve harmful R loops. During transcription, R-loop formation is prevented by different RNA-binding and processing factors, including the THO complex,[12] SRSF1,[13] and other factors,[14–16] supporting a model by which those mRNAs suboptimally assembled on ribonucleoprotein complexes (mRNPs) would have a better chance to rehybridize back with the DNA template from which they were copied.[17] However, if formed, R loops can be removed by the conserved ribonuclease RNase H, which specifically degrades the RNA moiety of the DNA-RNA hybrids, or by DNA-RNA helicases. Depletion of SETX, AQR, PIF1, BLM, and members of the DExD/H family of RNA helicases, such as UAP56/DDX39B, DDX1, DDX5, DHX9, DDX19, DDX21, or DDX23, have been shown to lead to R-loop accumulation,[2] even though there is no evidence for direct DNA-RNA hybrid unwinding for all of them in cells.

In addition, DNA replication-associated factors such as BRCA1 and BRCA2 and Fanconi anemia factors prevent R-loop accumulation, R-loop-mediated genome instability, and transcription-replication conflicts.[18–22] Chromatin is also emerging as another key player in R-loop dynamics, as supported by the role of chromatin remodeling or modification factors preventing R-loop-mediated genome instability, such as those involved in histone acetylation/deacetylation,[15,23,24] the histone chaperones or remodelers FACT, ATRX, INO80, and SWI/SNF.[25–29] In addition, a set of histone modifications of varied physiological relevance has been shown to correlate with R-loop-enriched sites.[30–33]

Given the impact of R loops on cell homeostasis and genome integrity, different screenings have been performed to identify genes related to R-loop metabolism. Some are based on the detection of hybrid-induced DNA damage or chromosomal instability, as determined by γH2AX foci, gross chromosomal rearrangements (GCRs), or Rad52 foci in yeast and human cells.[14,15,34] Other studies are based directly either on immunodetection of R loops[16,35] or by pull-down coupled to mass



(legend on next page)

spectrometry with purified DNA-RNA hybrids or with the anti-DNA-RNA hybrid antibody S9.6.[36–38] As a result, different factors related to DNA-RNA hybrids, including different RNA-binding proteins (RBPs), RNA processing, chromatin modification, and DNA damage response (DDR) factors have been identified. However, functional evidence for a physiological role on R-loop homeostasis in cells has not always been provided. Thus, further studies to achieve a deeper knowledge on R-loop metabolism and its impact on genome integrity are needed.

Taking advantage of the ability of the activation-induced cytidine deaminase (AID) to act on the displaced ssDNA of R loops, we designed a strategy to identify factors involved in R-loop homeostasis that might have escaped previous detection. We developed a sensitive immunofluorescence (IF) assay in cells overexpressing the human cytidine deaminase AID that specifically targets the ssDNA of R loops, causing DNA breaks detectable by γH2AX,[39,40] to search for genes that, upon depletion, lead to an R-loop-dependent increase of DNA breaks. A similar strategy was successfully carried out in our laboratory in a screening of *Saccharomyces cerevisiae* to identify histone mutations associated with R loops.[31] Screening of a library of small interfering RNAs (siRNAs) against 3,205 different potentially druggable target factors identified genes involved in transcription, RNA processing, translation, chromatin modification, and others, which supports the notion that the origin and resolution of R loops are heterogeneous. We focused here on DDX47, a DEAD-box RNA helicase, and MeCP2, a DNA-binding protein, as two factors counteracting R-loop accumulation by different mechanisms, with the first potentially acting directly on the hybrid and the second acting via an epigenetic process related to DNA methylation.

RESULTS

Search of human factors counteracting R loops susceptible to AID-mediated DNA breaks

To identify factors involved in R-loop homeostasis, we developed a sensitive system for R-loop detection based on the ability of AID to target ssDNA and lead to DNA damage (Figure 1A). We generated stable U2OS cell lines carrying the human AID gene fused to a chimeric *tet* promoter regulated by doxycycline (DOX) (U2OS-TR-AID-cells). Among all the clones obtained, one that expressed the highest levels of AID, as determined by western blot and IF upon DOX addition, was selected for further studies (Figures 1B and S1A). We validated the system by assaying whether AID increased DNA damage in cells depleted of FANCD2, known to accumulate high R-loop levels.[19,20] As can be seen, siFANCD2 cells, and to a lesser extent siControl cells, showed a significant increase of the number of γH2AX foci/cell upon AID induction (+DOX) compared with non-induced conditions (−DOX) (Figure 1C). This increase in γH2AX foci of siFANCD2 cells was reduced by overexpression of RNase H1 (RNH1), confirming that it was mediated by DNA-RNA hybrids (Figure S1B). The system was further validated in 96-well plates by automated high-throughput microscopy using the ImageX-Express equipment (Molecular Devices) with R-loop-accumulating siFANCD2 and siTHOC1 cells,[39] thus establishing the conditions for the siRNA library screening (Figure S1C).

A collection of 4,796 selected siRNAs targeting 3,205 human genes involved in different processes, including apoptosis, nucleic acid binding, RNA or DNA metabolism, autophagy, and various other nuclear pathways that are potential targets of therapeutic drugs (Dharmacon-ON TARGET Plus-Druggable genome siRNA library) were analyzed for γH2AX foci in induced versus non-induced conditions. The percentage of cells with $\geq$5 or $\geq$10 γH2AX foci were determined by duplicate in 96-well plates with and without AID induction (see STAR Methods for details). As a result, 156 candidate genes whose depletion led to $\geq$1.2-fold higher levels of γH2AX foci upon AID expression compared with AID^{-} cells and siControls were selected (Table S1). Of those, 46 candidate genes were validated in a second round of IF (Figure 1D; Table S2). In parallel, γH2AX intensity was determined as a marker for replicative stress and used as an alternative method to identify candidate genes. 21 genes showed an increase of γH2AX intensity upon AID induction (Table S3), 8 of which had also been selected as positive for

Figure 1. A method for screening for factors controlling R-loop homeostasis

(A) Schematic illustration of the AID-based principle for detecting factors involved in R-loop metabolism. A U2OS-TR-AID stable cell line carrying a tetracycline-regulated AID gene cassette is used as a tool. Upon induction of AID expression, deamination of cytosines on the displaced ssDNA of an R loop leads to an increase in DNA double-strand breaks (DSBs). Comparison of γH2AX foci in siRNA-depleted cells before and after expression of AID allows detection of siRNAs altering R-loop levels.

(B) Quantification of nuclear AID immunofluorescence signal (right panel) and protein levels by western blot (left panel) in the selected U2OS-TR-AID stable clone in the absence or presence of doxycycline for 24 h. n = 3.

(C) Immunofluorescence of γH2AX in siC and siFANCD2-transfected U2OS-TR-AID cells (clone #8) in the absence (−DOX) or presence of AID (+DOX) as indicated. Nuclei were stained with DAPI. AID was overexpressed from a CMV promoter, leading to its accumulation both in the nucleus and cytoplasm, which is more evident in the latter because nuclear import is probably a rate-limiting step. Scale bars, 25 μm. Quantification of the number of γH2AX foci per cell is shown. Red lines in the plots indicate the median. n = 3. The statistical significance of the difference was calculated with Mann-Whitney U-test; *$p < 0.05$; ***$p < 0.001$ (B and C). See also Figure S1.

(D) Candidate high-throughput identification. Analysis of γH2AX immunofluorescence data (first round: 3,205 siRNA in one replicate with a duplicate). The percentages of cells with $\geq$5 or $\geq$10 γH2AX foci was used to calculate the indicated ratios, as represented in both axes. The dotted horizontal lines show the cutoff (1.2) used to designate an siRNA as positive. Two ratios for each siRNA are represented by black and red dots (candidates with $\geq$10 γH2AX foci or $\geq$5 γH2AX foci, respectively) and gray dots (others).

(E) Networks of interacting proteins identified in the screening using GeneMANIA software. γH2AX foci candidates (gray circles), γH2AX intensity candidates (light green circles), and candidates from both screenings (dark green circles). The candidates selected for further study are shown as highlighted circles.

(F) Graphic shows the reproducibility scores of the candidates. Representation of the 156 candidate genes selected for validation. The 46 confirmed hits for the γH2AX foci and γH2AX intensity screenings are marked in red, the rest are plotted in black. Genes selected for further analysis are identified and highlighted in boxes. See also Figure S1.

γH2AX foci increase (Figure S1D; Table S3). Among the candidate genes was DHX9, encoding an RNA helicase that prevents R-loop-associated DNA damage, previously found to be enriched in a DNA-RNA hybrid interactome,[36] and TOPBP1, a protein required for DNA repair and checkpoint control, previously identified in a screening for DDR factors involved in R-loop homeostasis,[35] further validating our screening assay.

To determine the spectrum of biological functions within the selected candidates, we performed an analysis with protein analysis through evolutionary relationships (PANTHER). The analysis revealed a significant enrichment in functions involved in transcription, RNA biogenesis, and cell cycle, consistent with genes controlling R loops and transcription-associated genome instability (Figure S1E). Among the final candidate genes were chromatin-related factors, transcription and translation factors, RNA helicases, and factors involved in protein degradation (Table S4). Finally, the screening identified cytoplasmic factors that were not linked previously to R loops, as those involved in lipid metabolism, transporters, G-protein, and signaling and metabolic pathways (Table S4), whose role in R-loop homeostasis may be mediated by other factors.

A close analysis of the functions of all candidate genes and the network map of their interactions generated with the GeneMANIA plugin (Cytoscape v.3.8.0) (Figure 1E) reveals a close functional relationship for most of the genes, a large fraction of which function in RNA metabolism. We selected a representative sample of 13 genes for further analysis, including candidates with high levels of either γH2AX foci, γH2AX intensity, or both parameters simultaneously, with different reproducibility scores in the screening, and whose function was preferentially nuclear (see STAR Methods; Figure 1F; Table S4).These include genes encoding chromatin-associated factors (MeCP2 and HDAC8); RNA helicases (DDX42, DDX47 and UPF1); proteasome subunits (PSMC5 and PMSD7); and factors involved in transcription (MYOG and PHOX2A), translation (EEF1D and EGLN1), and RNA modification (RPUSD1). In addition, we included a lipid biosynthesis factor (ALOX5) as a candidate with non-related functions in nucleic acid metabolism and gene expression.

We next assayed directly R loops by S9.6 IF in U2OS cells performing the experiments in coverslips with a pool of two siRNAs for each gene. Cells were also immunostained with the anti-nucleolin antibody to exclude the nucleolar signal and specifically measure the nucleoplasm signal, following standard protocols.[41] 8 out of the 13 candidates showed a significant increase in S9.6 nuclear signal respect to control cells (siC) (Figure 2A). Then, we determined R-loop levels by DNA-RNA immunoprecipitation (DRIP) in two human genes (APOE and RPL13A) that have been previously validated for hybrid detection.[25,42] Depletion of 12 out of the 13 candidates increased the DRIP signal, with respect to siC levels, in at least one of the studied genes (Figure 2B). In all cases, the DRIP signal was removed by *in vitro* treatment with RNH, confirming that it was specific for DNA-RNA hybrids. These results validate our screening as an alternative method for the identification of factors involved in R-loop homeostasis. The different functions identified confirm the relevance of cellular R-loop control that explains the vast variety of molecular processes whose alteration may impact on R-loop levels as a potential source of genome instability.

DDX47 and MeCP2, two factors protecting cells from R-loop accumulation

Next, we focused on 2 candidates that function on two different nuclear processes for which a role in R-loop homeostasis is poorly known. These were the methyl-CpG-binding protein 2 MeCP2, a factor with a function related to chromatin and transcription[43] that indeed gave highly reproducible results, and the DEAD-box RNA helicase DDX47, a nucleolar RNA helicase.[44]

Previous analyses performed in U2OS cells were extended to another cell line and with specific siRNAs to confirm our results in a different genetic background. We determined the R-loop levels after siRNA depletion in HeLa cells by S9.6 IF and DRIP analysis (Figures 3A, 3B, S2A, and S2B). A significant enrichment of the nuclear signal by S9.6 IF was observed for both siMECP2 and siDDX47 cells that was suppressed by *in vitro* RNH treatment (Figure 3A). R loops were also confirmed by an RNH-sensitive increase of DRIP signals in the standard human genes (APOE and RPL13A) upon depletion of both factors, and in the case of siDDX47 cells, hybrid accumulation was also observed in the 18S and 28S rDNA loci (Figure 3B).

Consistent with an action of DDX47 and MeCP2 on active genes in which R loops are enriched, we found that both proteins were recruited to these chromatin regions, as determined by chromatin IP (ChIP) analysis (Figure 3C). Importantly, the high DNA-RNA hybrid levels were not due to an increase in transcription provided by the minor and different impact of the mutations on the mRNA levels of the genes tested by qRT-PCR (Figure S2C).

Next, we analyzed DNA damage in HeLa siRNA-transfected cells. A significant increase in γH2AX foci per cell was detected in siMECP2 cells, whereas this increase was mild in siDDX47 cells (Figure S2D). Single-cell gel electrophoresis, however, revealed that depletion of each protein increased DNA breaks, which was partially suppressed by RNH1 overexpression *in vivo* (Figure S2E). Altogether, the data confirm and validate that DDX47 and MECP2 are genes involved in R-loop homeostasis in both U2OS and HeLa cells.

Provided that MeCP2 is a multifunctional modulator of gene expression that plays a role in transcription, RNA processing, and heterochromatin,[43] it is possible that dysfunction of this factor alters R-loop levels by altering more than one process. Therefore, we focused the rest of our work on the DDX47 DEAD-box helicase to get insight into the relevance of this helicase in R-loop metabolism.

DDX47 depletion leads to transcription impairment and transcription-replication conflicts in the nucleolus

DDX47 helicase localizes to the nucleolus, and evidence suggests that it is associated with rRNA processing (Figure S3A).[44] Given the R-loop accumulation and genome instability phenotypes of siDDX47 cells, we decided to study further the possible role of this helicase in the nucleolus. Interestingly, we observed that siDDX47 cells showed a general reduction in total nucleolar area defined by IF nucleolin staining (Figure 4A). Since the primary function of the nucleolus consists of rDNA transcription, we considered the possibility that DDX47 silencing affected RNA polymerase I (RNAPI) function. Analysis of RNAPI occupancy at rDNA upon siDDX47 depletion by ChIP using primers

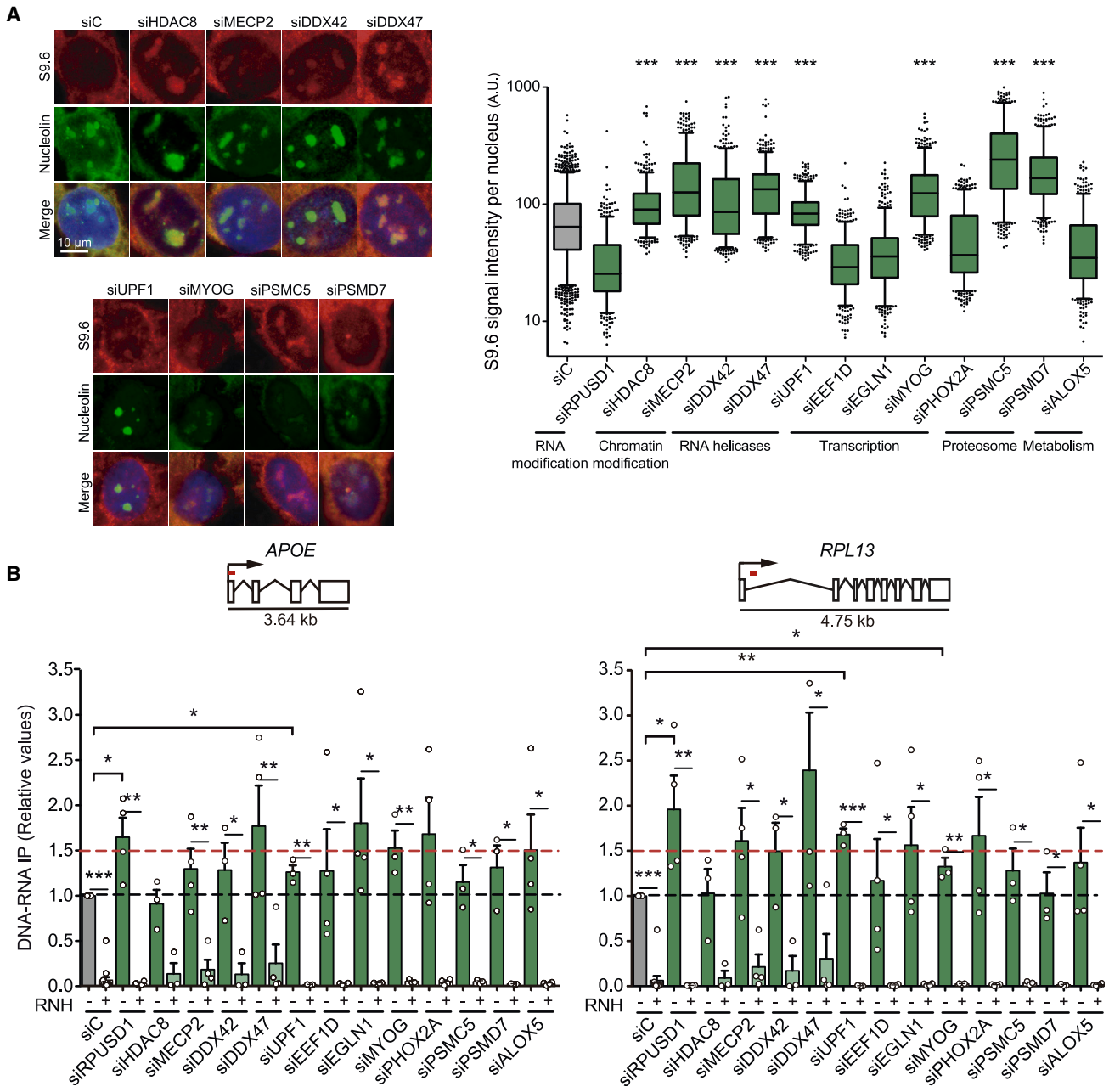


Figure 2. DNA-RNA hybrid analysis upon depletion of selected candidates

(A) Representative images of U2OS cells immunostained with S9.6 (red) and nucleolin (green) antibodies after transfection with at least two siRNAs of each original pool (left panel). The median of S9.6 signal intensity per nucleus after nucleolar signal removal in U2OS cells transfected with the indicated siRNAs is shown (n = 3). Boxes and whiskers indicate 5th–95th percentiles. ***p < 0.001 (Mann-Whitney U-test, two-tailed). A.U., arbitrary units.

(B) Relative DRIP-qPCR signal values at *APOE* and *RPL13A* genes in U2OS cells transfected with the indicated siRNAs and treated *in vitro* with RNase H prior to immunoprecipitation where indicated. Data are plotted as mean ± SEM (n = 3). *p < 0.05; **p < 0.01; ***p < 0.001 (one-tailed paired t test).

spanning the rDNA region revealed a slight, but not statistically significant, reduction of RNAPI from 5′ to 3′ (Figure 4B), so a putative impact on rDNA transcription, if any, would be minor.

To test whether transcription is impaired in siDDX47 cells, we performed a time course 5′ ethynyluridine (EU) incorporation assay by IF. DDX47 depletion significantly reduced EU labeling in the nucleolus, especially at short times of incubation (20 min), and this reduction was observed at different times (Figure 4C). A reduction of EU labeling in the nucleoplasm was also detected but mainly at long incubation times, suggesting a lower impact of DDX47 depletion, if any, on RNAPII transcription, based on the assumption that a late effect suggests an indirect role.

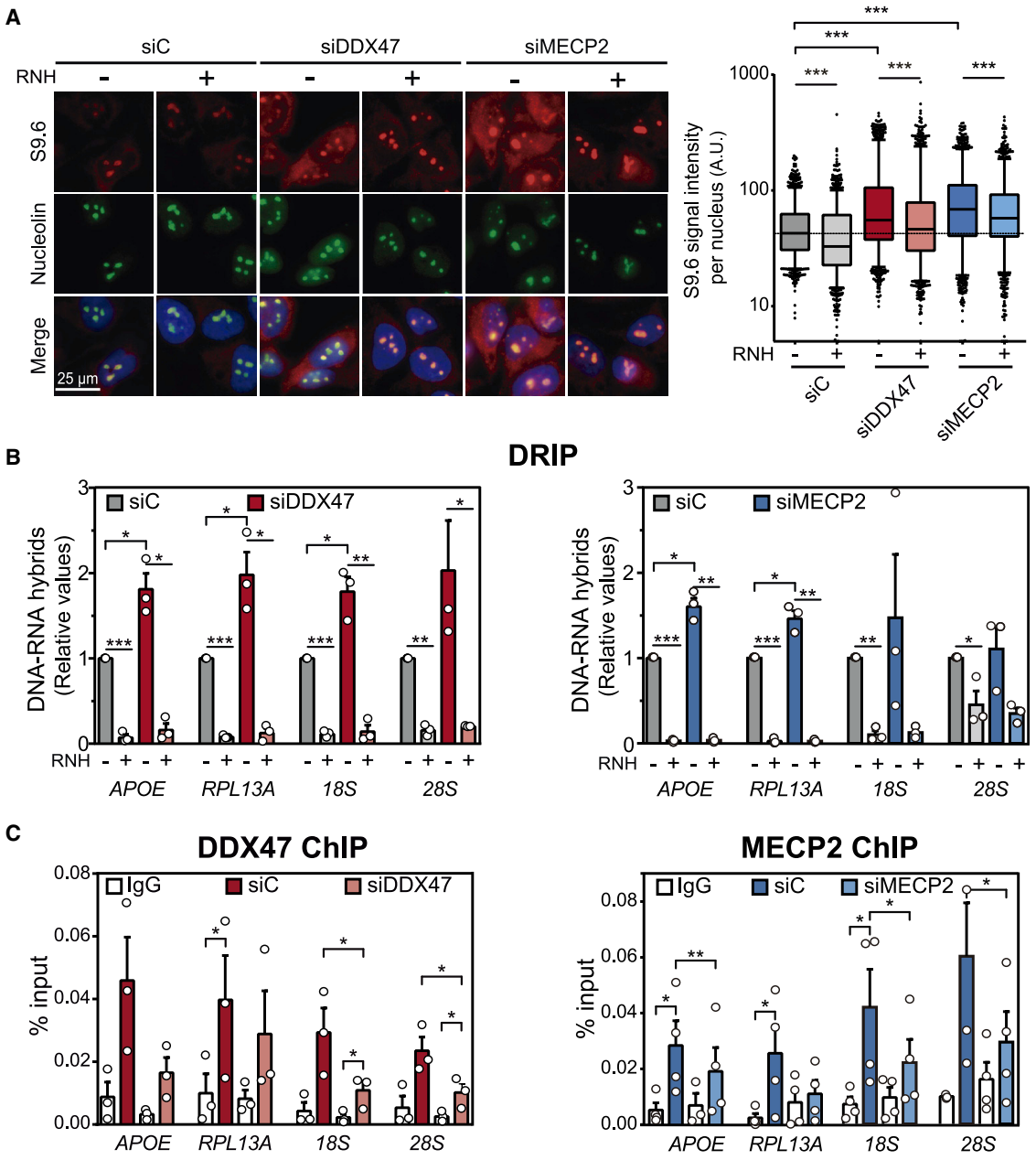


Figure 3. R-loop accumulation in siDDX47 and siMECP2 HeLa cells

(A) Representative images of immunostaining with S9.6 (red) and anti-nucleolin (green) antibodies in HeLa cells upon DDX47 and MECP2 depletion. The median of S9.6 signal intensity per nucleus after nucleolar signal removal in siC, siDDX47, and siMECP2 HeLa cells and treated *in vitro* with RNase H where indicated are shown. Boxes and whiskers indicate 5th–95th percentiles. n = 5. ***p < 0.001 (Mann-Whitney U test, two-tailed). A.U., arbitrary units.

(B) Relative DRIP-qPCR signal values at *APOE*, *RPL13A*, *18S*, and *28S* genes in HeLa cells transfected with the indicated siRNAs and treated *in vitro* with RNase H prior to immunoprecipitation where indicated. Data are plotted as mean ± SEM (n = 3). *p < 0.05; **p < 0.01; ***p < 0.001 (one-tailed paired t test).

(C) ChIP analyses of DDX47 and MeCP2 proteins at APOE, RPL13, 18S, and 28S genes. Values represent the percentage of the precipitated DNA (IP) to input DNA (INPUT) normalized with respect to the siC. Immunoprecipitation in siDDX47 and siMECP2 cells were included as control of specificity. Immunoglobulin G (IgG) was used as negative control. Data are plotted as mean ± SEM (n = 3). *p < 0.05 (one-tailed paired t test). See also Figure S2.

RNAPII transcription of the genes in the rDNA intergenic regions has been shown to generate R loops that impact RNAPI transcription.[45] Consequently, we asked whether the impact of DDX47 depletion on transcription and R-loop accumulation in the nucleolus could be caused by RNAPII-driven transcripts, as is the case in the nucleoplasm. For this, we carried out EU labeling experiments and S9.6IF assays under conditions of RNAPII inhibition. RNA synthesis was reduced in siDDX47 cells

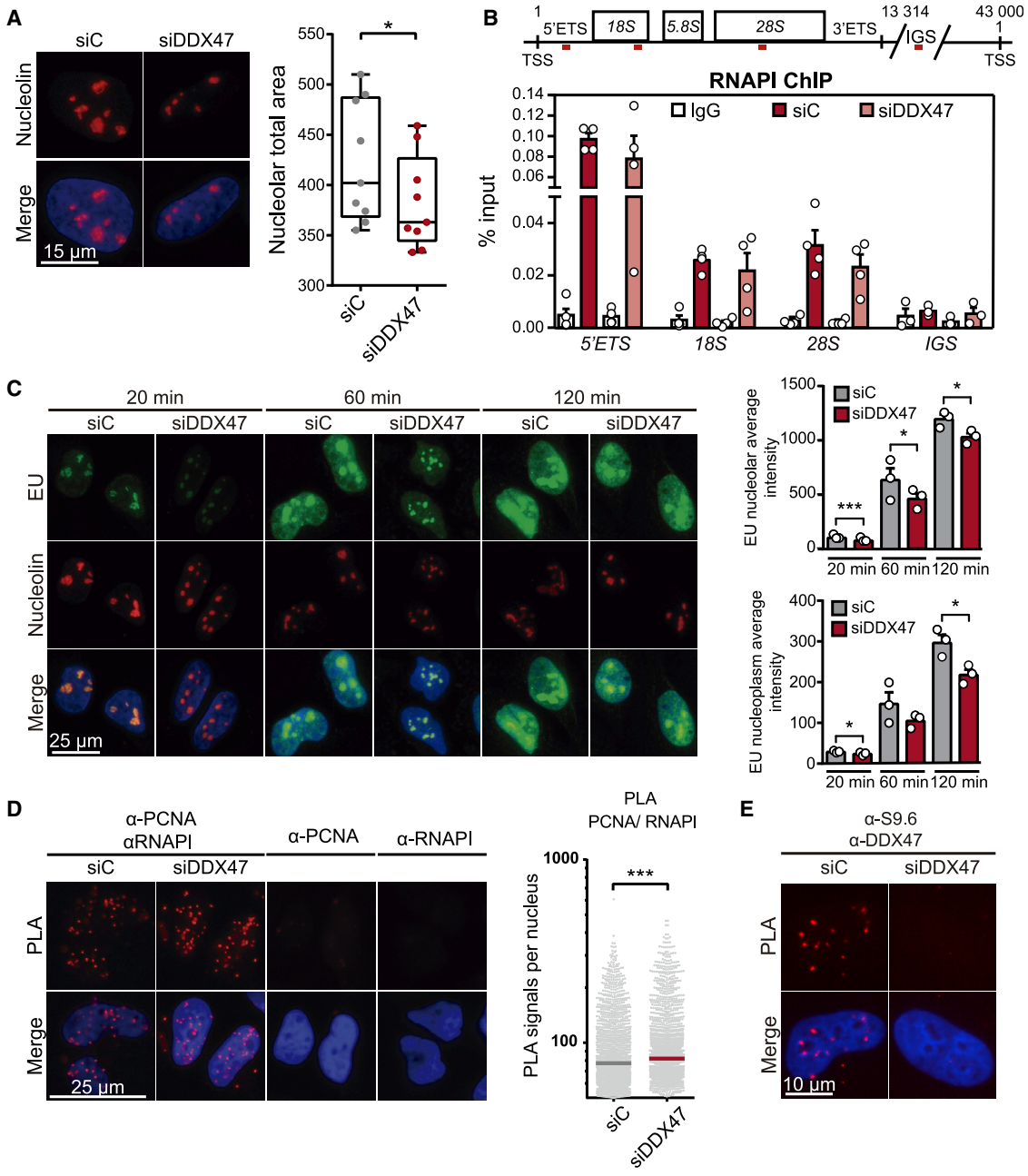


Figure 4. DDX47 depletion causes transcription impairment

(A) Representative images of immunostaining with anti-nucleolin (red) in HeLa cells upon DDX47 depletion. Quantification shows the median of nucleolar total area in HeLa cells transfected with siC and siDDX47 from at least three biological repeats. Boxes and whiskers indicate minimum and maximum. *$p < 0.05$ (two-tailed paired t test).

(B) ChIP analyses of RNAPI protein at 5′ ETS, 18S, 28S, and IGS regions. Values represent the percentage of the precipitated DNA (IP) to input DNA (INPUT). IgG was used as negative control. Data are plotted as mean ± SEM (n = 4) (one-tailed paired t test).

(C) Representative immunofluorescence images of 5-ethynyluridine (EU) incorporation in HeLa cells transfected with siC and siDDX47 after 72 h. Anti-nucleolin was used as a marker of nucleolus. Quantification of EU nucleolar (top panel) and EU nucleoplasm (bottom panel) incorporation signals in siC and siDDX47-depleted cells. Data are plotted as mean ± SEM (n = 3). *$p < 0.05$, ***$p < 0.001$ (one-tailed paired t test).

(D) Proximity ligation assay (PLA) showing interactions of RNAPI and PCNA endogenous proteins. Red spots are indicative of a positive PLA signal. Negative controls with only one of the antibodies are shown. DNA was stained with DAPI. The median of PLA signal intensity per nucleus in siC and siDDX47 is plotted. n = 3. ***$p < 0.001$ (Mann-Whitney U test, two-tailed).

(E) PLAs showing specific interactions of DDX47 and DNA-RNA hybrids (100 cells were scored). Data are representative of two independent experiments. siDDX47 HeLa cells were included as control of detection specificity. See also Figure S3F for negative controls. DNA was stained with DAPI. Other details as in (D). See also Figure S3.

treated with the inhibitor α-amanitin but to a lesser extent than in siC cells, consistent with a role of DDX47 associated with RNAPI transcription (Figure S3B) as expected from previous reports.[44] Interestingly, depletion of DDX47 leads to an increase of R loops in the nucleolus (Figure S3C) that was reduced upon RNAPII inhibition (Figure S3D) supporting the conclusion that DDX47 act on R loops generated by both RNAPI and RNAPII transcription in the nucleolus (Figure S3D).

Since R loops are a source of transcription-replication conflicts (TRCs),[6] we wondered whether TRCs were increased in siDDX47 cells. For this, we carried out proximity ligation assay (PLA) using antibodies against RNAPI and proliferating cell nuclear antigen (PCNA). A significant increase in PLA signal was observed in DDX47-depleted cells as compared with siC (Figure 4D), consistent with the harmfulness of the R loops prevented or resolved by DDX47 and a role of TRCs in the R-loop-mediated genome instability caused by siDDX47. Similar to depletion of other factors preventing harmful R loops,[23,46] this replication problem is not sufficient to cause significant differences in cell-cycle progression of siDDX47 cells upon thymidine synchronization (Figure S3E).

Next, we asked whether DDX47 acted directly at R loops. We first determined if DDX47 associates *in vivo* with DNA-RNA hybrids in the cells by PLA. Using anti-DDX47 and S9.6 antibodies, we observed a strong PLA signal, concluding an association of DDX47 with DNA-RNA hybrids (Figures 4E and S3F). Altogether, these results, together with the increase of R loops in siDDX47 cells, prompted us to explore further the activity of the protein.

DDX47 has DNA-RNA-unwinding activity

We analyzed whether the DEAD-box RNA helicase DDX47 possesses DNA-RNA-unwinding activity. We purified to homogeneity wild-type DDX47 and two mutant forms, DDX47-K74A and DDX47-E175A, that inactivated the highly conserved motifs I (ATP binding) and II (ATP hydrolysis), respectively (Figures 5A and S4A).[47] We observed that DDX47 unwinds RNA duplexes (double-strand RNAs [dsRNAs]) in a protein-concentration-dependent manner, both blunt ended or with either a 5′ or 3′ overhangs (Figures 5B, top panel, and S4B), consistent with its role in RNA metabolism. Notably, we observed that DDX47 unwinds DNA-RNA hybrids, whether blunt ended or with overhangs, in a protein-concentration- and ATP-dependent manner (Figure 5B, bottom panel, and S4C). Importantly, data indicate that DDX47 can unwind DNA-RNA hybrids more efficiently than dsRNA since the percentage of unwound product was up to 4-fold of that obtained for dsRNA under the same conditions (Figure 5B). To confirm that this unwinding activity was intrinsic of DDX47, we carried out the same experiments with DDX47-K74A and DDX47-E175A mutants. As shown in Figure 5C, both mutant proteins were able to unwind dsRNA but with a strongly reduced efficiency compared with wild-type DDX47 (see also Figure S4B). Importantly, these mutants were unable (or impaired) to unwind blunt DNA-RNA hybrids (Figure 5C), even though a remaining unwinding activity was observed in 5′ overhang RNA-DNA hybrids (Figure S4C), which could be due to thermal fluctuation of these substrates or leaky unwinding/binding activity mediated by these DDX47 mutants. Finally, we tested the ability of DDX47 to unwind a half R loop based on a DNA-RNA hybrid with a 5′ ssDNA flap at one end.[20] Wild-type DDX47 clearly unwinds the structure, whereas the activity was reduced in the DDX47-K74A and DDX47-E175A mutants (Figure 5D). Therefore, our results indicate that DDX47 resolves DNA-RNA hybrids and R-loop-mimicking structures *in vitro*.

DDX47 overexpression suppresses the high R-loop accumulation in different mutant cell lines

A number of RNA helicases have been shown to have DNA-RNA unwinding activity *in vitro*,[2] but this does not necessarily mean that this is their main activity in cells. To study whether DDX47 could unwind DNA-RNA hybrids in the cell, we constructed plasmids carrying either the wild-type DDX47 or mutant forms DDX47-K74A and DDX47-E175A tagged with FLAG under the control of the strong promoter CMV. Cells transfected with these plasmids overexpressed DDX47 as confirmed by IF and western blots (Figures S5A and S5B). In this case, we used a different pool of two siRNAs against the 3′ UTR to ensure endogenous DDX47 silencing, but not silencing of plasmid-driven DDX47 expression, and carried out S9.6 IF, confirming that this new siRNA pool leads to similar R-loop accumulation as those previously used (Figures S5C–S5E). Then, we assayed whether overexpression of wild-type DDX47 and helicase-dead mutants could suppress the R-loop accumulation caused by DDX47 depletion. As expected, wild-type DDX47 overexpression suppressed the nuclear DNA-RNA hybrid accumulation of DDX47 depletion, but overexpression of the two mutant forms also reduced hybrids, although to a lesser extent (Figure 6A).

To assay whether DDX47 unwinds any type of DNA-RNA hybrids in the cell, regardless of their origin, we assayed whether overexpression of DDX47 suppressed R-loop accumulation induced by loss of other unrelated proteins, such as depletion of the DDX23 and SETX helicases or the Fanconi anemia repair factor FANCD2.[19,20,48,49] We carried out overexpression experiments in cells depleted of these selected factors as previously reported for the cellular helicase activity of UAP56/DDX39B.[46] S9.6 IF assays show that DDX47 suppressed the accumulation of DNA-RNA hybrids induced by SETX depletion and, to a lesser extent, induced by DDX23 and FANCD2 depletions (Figures 6B and S5F). Then, we extended the analysis to cells depleted of UAP56, a DEAD-box DNA-RNA helicase that localizes to active chromatin and prevents DNA-RNA hybrids over the entire genome.[46] Notably, DDX47 overexpression did not suppress the R loops induced by siUAP56 (Figure 6C). On the contrary, UAP56 overexpression fully rescued the R-loop-accumulation phenotype caused by siDDX47 (Figure 6C), consistent with the reported general activity of UAP56/DDX39B as a DNA-RNA resolvase in human cells. These results support that DDX47 has DNA-RNA-unwinding activity in cells that may not have a wide range of action, in contrast to UAP56/DDX39B, whether due to a more specific role of DDX47 on a subset of R loops or, likely, a functional interaction between DDX47 and UAP56/DDX39B or other factors that precludes its action in the absence of UAP56.

DISCUSSION

We have developed an approach based on the ability of AID to target and cause DNA lesions at R loops to search for factors

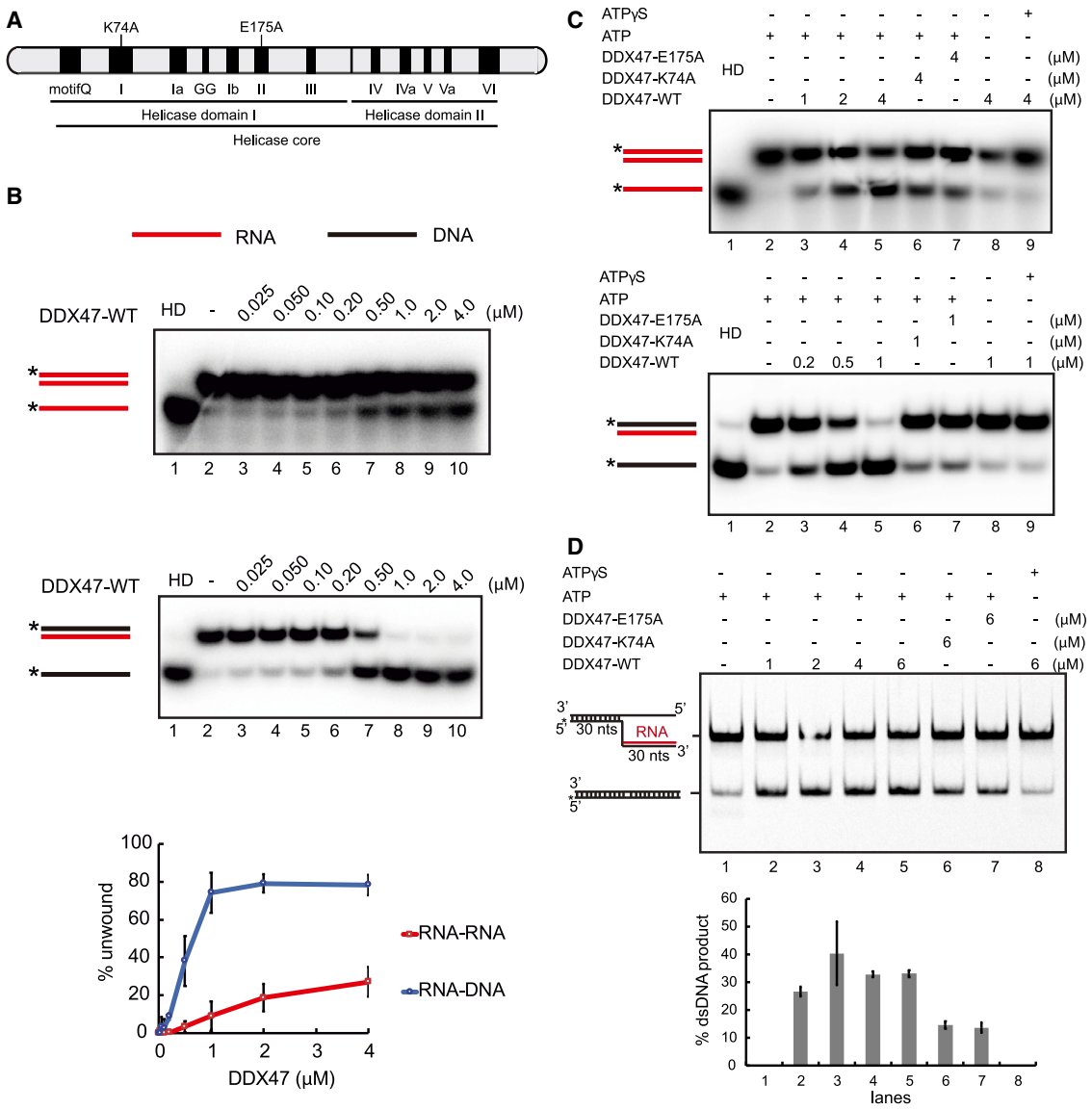


Figure 5. DDX47 is a DNA-RNA helicase and R-loop resolvase

(A) Schematic representation of mutations in the helicase core of DDX47 analyzed in this study.

(B) Comparison between DDX47 dsRNA helicase and RNA-DNA helicase activity using same amount of dsRNA or DNA-RNA duplex with a serial dilution of DDX47-wild-type (WT) protein. The graphic shows the percentage of unwound product respect to the DDX47 in a concentration-dependent manner. The results were quantified and plotted as mean ± SD (n = 3 technical replicates). HD, heat-denatured substrate.

(C) DDX47 RNA-RNA (top panel) and DNA-RNA (lower panel) unwinding activity. Assays were performed using a blunt-ended RNA (dsRNA) or DNA-RNA duplex, respectively, as substrate with different amounts of DDX47-WT, DDX47-K74A, and DDX47-E175A.

(D) DDX47 unwinding activity of DNA-RNA flap structures mimicking R loops. DNA-RNA-unwinding assays with DDX47-WT, DDX47-K74A, and DDX47-E175A using DNA-RNA flap structures. Graph shows the percentage of dsDNA product recovered after the reaction, and the results were quantified and plotted as mean ± SD (n = 2 technical replicates). Note that the spontaneous dissociation of the substrate was normalized to 0 for all reactions. See also Figure S4.

controlling R-loop homeostasis. An siRNA screening of 3,205 genes permitted us to identify 59 candidates that increased AID-mediated DNA damage as determined by γH2AX intensity or foci detected by high-throughput microscopy. Importantly, R-loop accumulation was confirmed in the 13 selected candidates by S9.6 IF or DRIP assays, validating this strategy as an "*in vivo*" approach to identify factors controlling R-loop homeostasis. With the latest estimate of protein-coding genes of a human female genome (19,969),[50] since we have covered only 16.05% of such genes, we could extrapolate that a screening covering all protein-coding genes would lead to a selection of a maximum 369 candidates. Although this number is certainly an overestimation, since the sample used in this study is not random but rather based on a druggable set of genes from

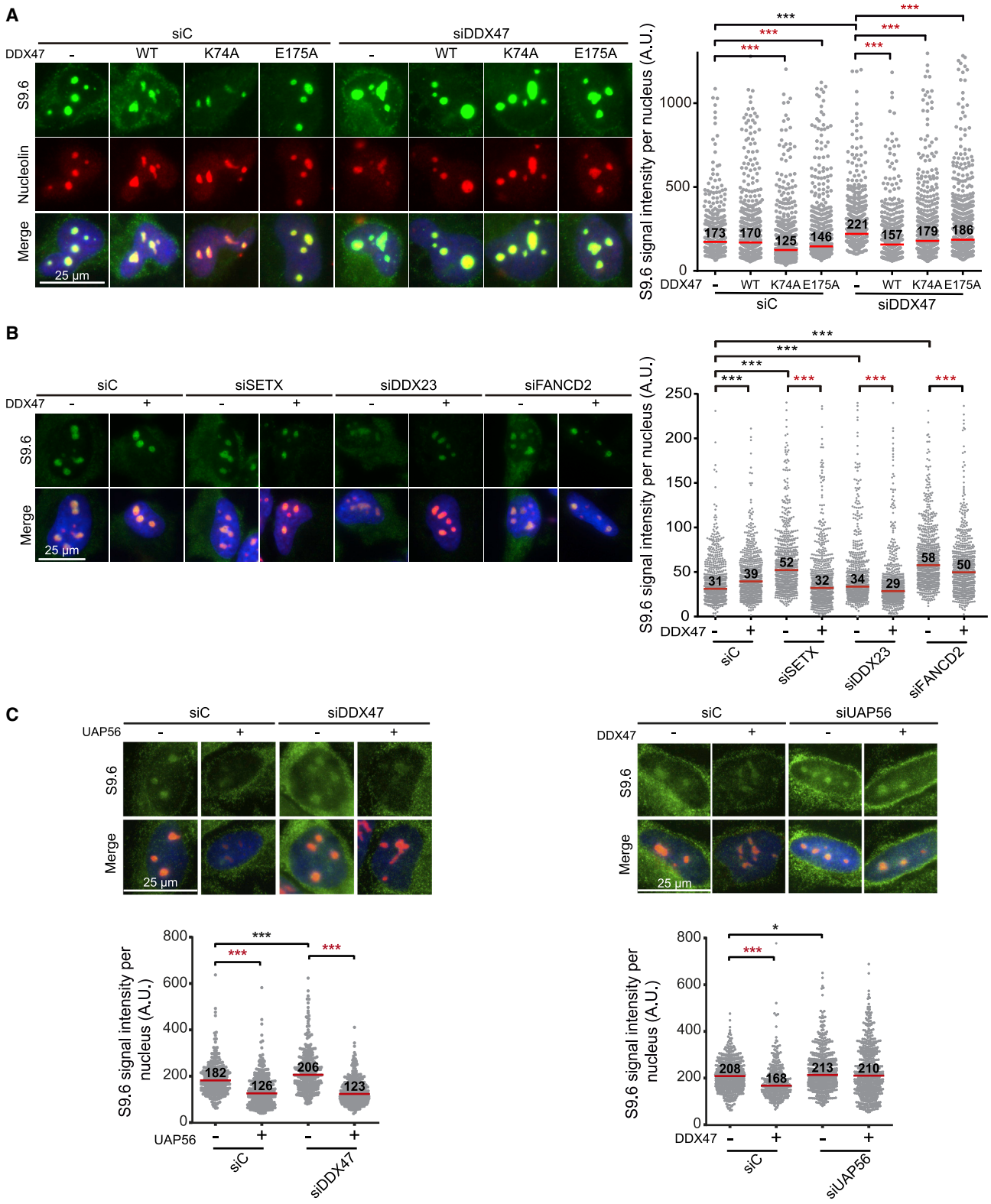


Figure 6. DDX47 overexpression rescues R-loop accumulation

(A) Representative images and quantification of S9.6 immunofluorescence signal in HeLa transfected cells with siDDX47 and the empty vector pFLAG (–), pFLAG-DDX47 (+DDX47) for DDX47-WT overexpression, and pFLAG-DDX47-K74A (+DDX47-K74A) or pFLAG-DDX47-E175A (+DDX47-E175A) for ATP-binding

(legend continued on next page)

which we further focused on nuclear protein-coding genes, our study reveals that harmful R loops can be enhanced by mutations in a large group of genes, consistent with the increasing number of these being reported by different labs in recent years.

Among the selected candidates, we identified RNA-related nuclear proteins involved in transcription (MYOG, PHOX2A), RNA metabolism (DDX42, DDX47, UPF1, RPUSD1), or translation (EEF1D, EGLN1), consistent with the known role of mRNP biogenesis factors in the prevention of R-loop accumulation.[2] The same is the case for epigenetic factors (MeCP2, HDAC8). In addition, proteasome subunits (PSMC5 and PMSD7) were found among the validated candidates, which might be explained by the reported role of the proteasome on transcription elongation and RNA biogenesis.[51] In this sense, it is worth noting that other factors related to protein degradation were also identified in a DNA-RNA hybrid interactome generated by protein pull-down.[36] Altogether, our results confirm the large number of heterogeneous factors and essential cellular processes that, if dysfunctional, lead to a stress condition that would favor accumulation of potentially harmful R loops. Our study reveals the need for an in-depth analysis of factors whose depletion causes DNA-RNA hybrid accumulation to define their mechanism of action on R loops, whether or not direct. In this study, we focused on MeCP2 and DDX47.

In agreement with the emerging role of chromatin in R-loop dynamics, chromatin-related factors such as the DNA-binding protein MeCP2 and the histone deacetylase HDAC8 were identified. High R-loop levels in siHDAC8 cells could be due to a more open hyper-acetylated chromatin, as previously shown for siSIN3 cells,[23] but this needs further analysis. It is worth noting in this sense that the chromatin factor chromatin accessibility complex protein 1 (CHRAC1) (Table S4), which functions in nucleosome positioning during transcription and replication,[52] was also detected among the 59 potential candidates. This would add to the list of histone chaperones and chromatin remodelers previously shown to control R-loop homeostasis,[25,27–29] strengthening the relevance of chromatin in R-loop occurrence and accumulation.

MeCP2, one of the top candidates of the screening, is a protein binding to methylated DNA that is highly expressed in neurons,[53,54] and mutations in the MECP2 gene cause Rett syndrome, a progressive neurologic developmental disorder.[43,55,56] This multifunctional protein modulates gene transcription, microRNA (miRNA) biogenesis, and splicing[57–59] and has recently been shown to be involved in condensation of chromatin.[60] It was proposed that MeCP2 participates in transcriptional repression by its ability to bind to methylated DNA and co-repressor complexes.[43] Our observation that MeCP2 (Figures 3 and S2E) protects cells from R-loop accumulation could be linked to its ability to interact with DNA and transcriptional repressors[57] by promoting a closed chromatin environment less prone to R-loop formation. Indeed, partners of MeCP2 such as the histone deacetylase complex Sin3A and the chromatin remodeler ATRX have also been involved in R-loop prevention.[23,26] Interestingly, it has been recently shown that the RNA-binding activity of ATRX can inhibit R-loop accumulation.[61] So, given the reported interaction between MeCP2 and ATRX,[62] it could be possible that such an interaction could be relevant in R-loop prevention. Nevertheless, since MeCP2 binds RNA *in vitro*,[63] a role mediated by its ability to bind RNA cannot be discarded.

It is worth noticing the recent observation that the oxidation of methylated DNA to 5hmC by ten-eleven translocation (TET) enzymes promotes R-loop accumulation.[64] This, together with our identification of MeCP2 as a factor involved in R-loop homeostasis, supports a functional link between DNA methylation and R loops whose biological meaning would need to be investigated. In summary, our results define MeCP2 as a factor modulating R loops. Further work will be needed to decipher the molecular mechanism and the possible impact of R-loop accumulation in Rett syndrome.

On the other hand, a number of RNA helicases have been previously implicated in R-loop biology, including DHX9, which prevents R-loop-associated DNA damage,[36] and DDX42, which binds G-quadruplex (G4) structures,[65] both found in this study, which validates further our screening assay. In addition, we identified other RNA helicases such as UPF1, involved in non-sense-mediated mRNA decay,[66,67] and the nucleolar DDX47. UPF1 was found as a candidate in the γH2AX intensity screening, suggesting a role in replicative stress in agreement with the requirement of this helicase for S-phase progression and maintenance of genome integrity.[66] Our data suggest a role of this helicase in the control of harmful R loops that could be behind the effects of its depletion on replication. In any case, a recent work suggests that UPF1 promotes the formation of telomeric repeat-containing RNA (TERRA) DNA-RNA hybrids to stimulate DNA resection and repair at telomeric regions.[68] It is thus possible that UPF1 plays a dual role in R-loop metabolism and genome integrity at different genomic regions. A putative dual role in R-loop homeostasis has indeed been reported for DDX1.[69,70]

We show that DDX47 is an RNA helicase with a role resolving DNA-RNA hybrids and thus preventing R-loop accumulation able to trigger genome instability. DDX47 unwinds DNA-RNA hybrids and R-loop-like structures *in vitro* (Figure 5D), and its overexpression in cells suppresses high R-loop levels induced by different conditions (Figures 6A and 6B). Thus, we can conclude that DDX47's DNA-RNA-unwinding activity contributes to maintain low levels of harmful R loops in the cell. However, in contrast to UAP56/DDX39B, a dsRNA and DNA-RNA helicase that works

or ATPase helicase-dead mutant DDX47 overexpression, respectively. Immunostaining shows hybrids by S9.6 monoclonal antibody (green) and nucleolin (red). Nuclei were stained with DAPI. The graphic shows the median of S9.6 signal intensity in the nucleoplasm (n = 3). A.U., arbitrary units. ***p < 0.001 (Mann-Whitney U test, two-tailed). Black stars denote significant increases, whereas red stars denote significant decreases.

(B) DDX47 overexpression rescues R loop in cells. Representative images and quantification of S9.6 immunofluorescence signal in HeLa transfected cells with siSETX, siDDX23, or siFANCD2 and with the empty vector pFLAG (−DDX47) or pFLAG-DDX47 (+DDX47). n = 3. Other details as in (A).

(C) Comparison of UAP56 and DDX47 R-loop resolvase activities *in vivo* as determined by S9.6 immunofluorescence. Overexpression of DDX47 in siUAP56 cells (right panel) and overexpression of UAP56 in siDDX47 cells (left panel) were tested. Representative images and quantification of S9.6 immunofluorescence signals are shown. n = 3. Other details as in (A). *p < 0.05, ***p < 0.001 (Mann-Whitney U test, two-tailed). See also Figure S5.

together with the THO complex during transcription elongation all over the genome,[46] DDX47's ability to resolve R loops in cells is not general. DDX47 is not able to suppress the unresolved R loops accumulated when UAP56 is absent in the cell, whereas, on the contrary, UAP56 overexpression suppresses siDDX47-induced R loops as well as all other R loops accumulated under different conditions. In this sense, it is worth noting that numerous RNA helicases had been reported to be involved in R-loop dynamics,[2] a list that continues increasing.[22,46,71,72] For some of these helicases, it has been shown that they act at R loops, as is the case for SETX and DHX9, shown to resolve R loops accumulated at 3′ end regions and to contribute to transcription termination[33,36]; PIF1, which controls R-loop accumulation at tRNAs[73]; DDX5, which resolves R loops generated in the proximity of DSBs[22,71]; or UAP56/DDX39B, a major co-transcriptional R-loop resolvase acting over the entire genome.[46] Nevertheless, their mechanism of action and direct role in R-loop resolution in cells remain to be proved in most other cases.

DDX47 has DNA-RNA-unwinding activity *in vitro*, but as it may happen with other RNA helicases, this does not imply that it functions as a DNA-RNA-unwinding factor in cells.[2] The observation that overexpression of ATP-binding and ATPase-dead mutant DDX47 variants are still able to reduce hybrids in siDDX47 cells (Figure 6A) supports the view that the *in-vitro*-detected helicase activity is not the main mechanism by which DDX47 limits DNA-RNA hybrid accumulation in cells. We believe that its putative role in RNA metabolism as a dsRNA ATPase resolving either RNA secondary structures or acting as an RNA chaperon, mainly on rRNA, may contribute to prevent R-loop accumulation, as proposed for other RNA-binding factors.[17] Our results confirm that DDX47 is mainly a nucleolar protein. It is involved in nucleolus homeostasis and rDNA transcription (Figure 4). Other nucleolar helicases, like DDX21, a multifunctional enzyme that regulates both RNAPI and RNAPII transcription, is also relevant for rRNA biogenesis and R-loop metabolism.[74,75] DDX47 depletion increases rDNA transcription-replication collision events in siDDX47 cells (Figure 4D). which could be behind the R-loop-dependent genome instability of siDDX47 cells. A recent study has proposed that DDX47 is recruited by FANCD2 and has a role preventing RNAPII TRCs.[76] However, it would be necessary to determine whether such an increase in collisions is suppressed by RNH1 and DDX47 overexpression. Our work would suggest that DDX47 helps prevent R-loop accumulation as a source of such conflicts rather than helping resolve them. In this sense, it is worth noting that DDX47 is overexpressed in different tumor cells (http://firebrowse.org/viewGene.html) and that changes in nucleolar morphology and ribosome biosynthesis have been recognized as hallmark features of many cancer cells.[77]

Our study is consistent with the heterogeneous origin of unscheduled R loops and factors whose depletion causes an accumulation that potentially compromises genome integrity. DDX47 and MeCP2 are two examples functioning differently, one via RNA metabolism and the other via epigenetics yet to be defined. Thus, it seems that harmful R loops may accumulate under different stresses caused by cellular deficiencies that directly or indirectly impair gene expression, chromatin structure, or DNA replication and repair. Further molecular studies are required to establish which factors are directly involved in unscheduled R-loop prevention and resolution.

Limitations of the study

Our model leaves several open questions and limitations that will need to be addressed in the future. Thus, the molecular mechanism by which MeCP2 prevents R-loop accumulation is unclear. The possibility that MeCP2 interaction with histone modifiers and chromatin remodelers is behind its R-loop-protecting role needs to be addressed. On the other hand, although we show that DDX47 efficiently resolves DNA-RNA hybrids *in vitro* and in cells, we cannot exclude that its R-loop-protective role was also mediated by a potential RNA chaperon role linked to its RNA-RNA-unwinding activity, as may also be the case of other RNA helicases. It will ultimately be necessary to generate stable deficient cell lines to test the impact of specific DDX47 mutations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell cultures
- METHOD DETAILS
  - siRNAs, plasmids, and transfections
  - High-throughput siRNA screening methods
  - siRNA library transfection
  - AID induction
  - γH2AX immunofluorescence
  - S9.6 immunofluorescence
  - DNA-RNA hybrid immunoprecipitation (DRIP-qPCR)
  - Chromatin immunoprecipitation (ChIP)
  - Single-cell gel electrophoresis
  - EU labeling
  - Proximity ligation assay (PLA)
  - Microscopy image acquisition and data analysis
  - Expression and purification of wild type and mutant DDX47
  - Nucleic acid unwinding assays
  - Flow cytometry
  - Miscellanea
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gene set enrichment and gene interaction analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112148.

ACKNOWLEDGMENTS

We thank B. Gomez-Gonzalez for critical reading of the manuscript and the Genomic Unit of CABIMER for support in the high-throughput screening.

Research was part of the project I+D+i PID2019-104270GB-I00/BMC, funded by MCIN/AEI/10.13039/501100011033/ to A.A. as the main funding source, plus grants from the European Research Council (ERC2014 AdG669898 TARLOOP); the Spanish Ministry of Economy and Competitiveness (BFU2016-75058-P); the European Union (FEDER) and Foundation "Vencer el Cancer" to A.A., and the National Institutes of Health grants R15 GM139135 and R21 ES028792 to X.X. C.G.-M. and J.J.M.-G. were supported by predoctoral FPU contracts from the Spanish Ministry of Education, Culture and Sports, and S.S. was partially supported by a postdoctoral "Juan de la Cierva" contract from the Spanish Ministry of Science and Innovation. S.A.A. was supported by NIH R25 GM102783 Bridge to Doctorate Program.

AUTHOR CONTRIBUTIONS

E.M.-C., L.P.C., S.S., and J.J.M.-G. carried out the high-throughput screening and candidate analysis. J.M.C.-M., C.G.-M., and M.L.G.-R. contributed with specific constructions and characterizations. E.M.-C. and L.P.C. performed most cellular studies and constructions on DDX47 and MeCP2. J.B., S.A.A., and X.X. performed the *in vitro* assays. A.A. designed and conceptualized the study. A.A. and R.L. wrote the paper. All authors contributed to data analysis and read and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 7, 2022
Revised: December 20, 2022
Accepted: February 6, 2023
Published: February 23, 2023

REFERENCES

1. Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell *46*, 115–124. https://doi.org/10.1016/j.molcel.2012.04.009.
2. García-Muse, T., and Aguilera, A. (2019). R loops: from physiological to pathological roles. Cell *179*, 604–618. https://doi.org/10.1016/j.cell.2019.08.055.
3. Niehrs, C., and Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. Nat. Rev. Mol. Cell Biol. *21*, 167–178. https://doi.org/10.1038/s41580-019-0206-3.
4. Petermann, E., Lan, L., and Zou, L. (2022). Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. Nat. Rev. Mol. Cell Biol. *23*, 521–540. https://doi.org/10.1038/s41580-022-00474-x.
5. Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-loops as cellular regulators and genomic threats. Mol. Cell *73*, 398–411. https://doi.org/10.1016/j.molcel.2019.01.024.
6. Gómez-González, B., and Aguilera, A. (2019). Transcription-mediated replication hindrance: a major driver of genome instability. Genes Dev. *33*, 1008–1026. https://doi.org/10.1101/gad.324517.119.
7. Richard, P., and Manley, J.L. (2017). R loops and links to human disease. J. Mol. Biol. *429*, 3168–3180. https://doi.org/10.1016/j.jmb.2016.08.031.
8. Wells, J.P., White, J., and Stirling, P.C. (2019). R loops and their composite cancer connections. Trends Cancer *5*, 619–631. https://doi.org/10.1016/j.trecan.2019.08.006.
9. Giannini, M., Bayona-Feliu, A., Sproviero, D., Barroso, S.I., Cereda, C., and Aguilera, A. (2020). TDP-43 mutations link Amyotrophic Lateral Sclerosis with R-loop homeostasis and R loop-mediated DNA damage. PLoS Genet. *16*, e1009260. https://doi.org/10.1371/journal.pgen.1009260.
10. Wood, M., Quinet, A., Lin, Y.L., Davis, A.A., Pasero, P., Ayala, Y.M., and Vindigni, A. (2020). TDP-43 dysfunction results in R-loop accumulation and DNA replication defects. J. Cell Sci. *133*, jcs244129. https://doi.org/10.1242/jcs.244129.
11. Cristini, A., Tellier, M., Constantinescu, F., Accalai, C., Albulescu, L.O., Heiringhoff, R., Bery, N., Sordet, O., Murphy, S., and Gromak, N. (2022). RNase H2, mutated in Aicardi-Goutieres syndrome, resolves co-transcriptional R-loops to prevent DNA breaks and inflammation. Nat. Commun. *13*, 2961. https://doi.org/10.1038/s41467-022-30604-0.
12. Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol. Cell *12*, 711–721. https://doi.org/10.1016/j.molcel.2003.08.010.
13. Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365–378. https://doi.org/10.1016/j.cell.2005.06.008.
14. Paulsen, R.D., Soni, D.V., Wollman, R., Hahn, A.T., Yee, M.C., Guan, A., Hesley, J.A., Miller, S.C., Cromwell, E.F., Solow-Cordero, D.E., et al. (2009). A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Mol. Cell *35*, 228–239. https://doi.org/10.1016/j.molcel.2009.06.021.
15. Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol. Cell *44*, 978–988. https://doi.org/10.1016/j.molcel.2011.10.017.
16. Chan, Y.A., Aristizabal, M.J., Lu, P.Y.T., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genet. *10*, e1004288. https://doi.org/10.1371/journal.pgen.1004288.
17. Aguilera, A. (2005). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. Curr. Opin. Cell Biol. *17*, 242–250. https://doi.org/10.1016/j.ceb.2005.03.001.
18. Bhatia, V., Barroso, S.I., García-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature *511*, 362–365. https://doi.org/10.1038/nature13374.
19. García-Rubio, M.L., Pérez-Calero, C., Barroso, S.I., Tumini, E., Herrera-Moyano, E., Rosado, I.V., and Aguilera, A. (2015). The Fanconi anemia pathway protects genome integrity from R-loops. PLoS Genet. *11*, e1005674. https://doi.org/10.1371/journal.pgen.1005674.
20. Schwab, R.A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C.C., Cohn, M.A., Gibbons, R.J., Deans, A.J., and Niedzwiedz, W. (2015). The Fanconi anemia pathway maintains genome stability by coordinating replication and transcription. Mol. Cell *60*, 351–361. https://doi.org/10.1016/j.molcel.2015.09.012.
21. Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K.M., Eaton, M.L., et al. (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Mol. Cell *57*, 636–647. https://doi.org/10.1016/j.molcel.2015.01.011.
22. Sessa, G., Gómez-González, B., Silva, S., Pérez-Calero, C., Beaurepere, R., Barroso, S., Martineau, S., Martin, C., Ehlén, Å., Martínez, J.S., et al. (2021). BRCA2 promotes DNA-RNA hybrid resolution by DDX5 helicase at DNA breaks to facilitate their repairdouble dagger. EMBO J. *40*, e106018. https://doi.org/10.15252/embj.2020106018.
23. Salas-Armenteros, I., Pérez-Calero, C., Bayona-Feliu, A., Tumini, E., Luna, R., and Aguilera, A. (2017). Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. EMBO J. *36*, 3532–3547. https://doi.org/10.15252/embj.201797208.
24. Feldman, J.L., and Peterson, C.L. (2019). Yeast sirtuin family members maintain transcription homeostasis to ensure genome stability. Cell Rep. *27*, 2978–2989.e5. https://doi.org/10.1016/j.celrep.2019.05.009.
25. Herrera-Moyano, E., Mergui, X., García-Rubio, M.L., Barroso, S., and Aguilera, A. (2014). The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. Genes Dev. *28*, 735–748. https://doi.org/10.1101/gad.234070.113.

26. Nguyen, D.T., Voon, H.P.J., Xella, B., Scott, C., Clynes, D., Babbs, C., Ayyub, H., Kerry, J., Sharpe, J.A., Sloane-Stanley, J.A., et al. (2017). The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats. EMBO Rep. *18*, 914–928. https://doi.org/10.15252/embr.201643078.

27. Prendergast, L., McClurg, U.L., Hristova, R., Berlinguer-Palmini, R., Greener, S., Veitch, K., Hernandez, I., Pasero, P., Rico, D., Higgins, J.M.G., et al. (2020). Resolution of R-loops by INO80 promotes DNA replication and maintains cancer cell proliferation and viability. Nat. Commun. *11*, 4534. https://doi.org/10.1038/s41467-020-18306-x.

28. Bayona-Feliu, A., Barroso, S., Muñoz, S., and Aguilera, A. (2021). The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription-replication conflicts. Nat. Genet. *53*, 1050–1063. https://doi.org/10.1038/s41588-021-00867-2.

29. Tsai, S., Fournier, L.A., Chang, E.Y.C., Wells, J.P., Minaker, S.W., Zhu, Y.D., Wang, A.Y.H., Wang, Y., Huntsman, D.G., and Stirling, P.C. (2021). ARID1A regulates R-loop associated DNA replication stress. PLoS Genet. *17*, e1009238. https://doi.org/10.1371/journal.pgen.1009238.

30. Castellano-Pozo, M., Santos-Pereira, J.M., Rondón, A.G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., and Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol. Cell *52*, 583–590. https://doi.org/10.1016/j.molcel.2013.10.006.

31. García-Pichardo, D., Cañas, J.C., García-Rubio, M.L., Gómez-González, B., Rondón, A.G., and Aguilera, A. (2017). Histone mutants separate R loop formation from genome instability induction. Mol. Cell *66*, 597–609.e5. https://doi.org/10.1016/j.molcel.2017.05.014.

32. Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chédin, F. (2016). Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol. Cell *63*, 167–178. https://doi.org/10.1016/j.molcel.2016.05.032.

33. Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439. https://doi.org/10.1038/nature13787.

34. Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev. *26*, 163–175. https://doi.org/10.1101/gad.179721.111.

35. Barroso, S., Herrera-Moyano, E., Muñoz, S., García-Rubio, M., Gómez-González, B., and Aguilera, A. (2019). The DNA damage response acts as a safeguard against harmful DNA-RNA hybrids of different origins. EMBO Rep. *20*, e47250. https://doi.org/10.15252/embr.201847250.

36. Cristini, A., Groh, M., Kristiansen, M.S., and Gromak, N. (2018). RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. Cell Rep. *23*, 1891–1905. https://doi.org/10.1016/j.celrep.2018.04.025.

37. Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., Ó Broin, P., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A., et al. (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenet. Chromatin *8*, 46. https://doi.org/10.1186/s13072-015-0040-6.

38. Wang, I.X., Grunseich, C., Fox, J., Burdick, J., Zhu, Z., Ravazian, N., Hafner, M., and Cheung, V.G. (2018). Human proteins that interact with RNA/DNA hybrids. Genome Res. *28*, 1405–1414. https://doi.org/10.1101/gr.237362.118.

39. Domínguez-Sánchez, M.S., Barroso, S., Gómez-González, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS Genet. *7*, e1002386. https://doi.org/10.1371/journal.pgen.1002386.

40. Gómez-González, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc. Natl. Acad. Sci. USA *104*, 8409–8414. https://doi.org/10.1073/pnas.0702836104.

41. García-Rubio, M., Barroso, S.I., and Aguilera, A. (2018). Detection of DNA-RNA hybrids in vivo. Methods Mol. Biol. *1672*, 347–361. https://doi.org/10.1007/978-1-4939-7306-4_24.

42. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell *45*, 814–825. https://doi.org/10.1016/j.molcel.2012.01.017.

43. Ip, J.P.K., Mellios, N., and Sur, M. (2018). Rett syndrome: insights into genetic, molecular and circuit mechanisms. Nat. Rev. Neurosci. *19*, 368–382. https://doi.org/10.1038/s41583-018-0006-3.

44. Sekiguchi, T., Hayano, T., Yanagida, M., Takahashi, N., and Nishimoto, T. (2006). NOP132 is required for proper nucleolus localization of DEAD-box RNA helicase DDX47. Nucleic Acids Res. *34*, 4593–4608. https://doi.org/10.1093/nar/gkl603.

45. Abraham, K.J., Khosraviani, N., Chan, J.N.Y., Gorthi, A., Samman, A., Zhao, D.Y., Wang, M., Bokros, M., Vidya, E., Ostrowski, L.A., et al. (2020). Nucleolar RNA polymerase II drives ribosome biogenesis. Nature *585*, 298–302. https://doi.org/10.1038/s41586-020-2497-0.

46. Pérez-Calero, C., Bayona-Feliu, A., Xue, X., Barroso, S.I., Muñoz, S., González-Basallote, V.M., Sung, P., and Aguilera, A. (2020). UAP56/DDX39B is a major cotranscriptional RNA-DNA helicase that unwinds harmful R loops genome-wide. Genes Dev. *34*, 898–912. https://doi.org/10.1101/gad.336024.119.

47. Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nat. Rev. Mol. Cell Biol. *12*, 505–516. https://doi.org/10.1038/nrm3154.

48. Sridhara, S.C., Carvalho, S., Grosso, A.R., Gallego-Paez, L.M., Carmo-Fonseca, M., and de Almeida, S.F. (2017). Transcription dynamics prevent RNA-mediated genomic instability through SRPK2-dependent DDX23 phosphorylation. Cell Rep. *18*, 334–343. https://doi.org/10.1016/j.celrep.2016.12.050.

49. Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol. Cell *42*, 794–805. https://doi.org/10.1016/j.molcel.2011.04.026.

50. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44–53. https://doi.org/10.1126/science.abj6987.

51. Geng, F., Wenzel, S., and Tansey, W.P. (2012). Ubiquitin and proteasomes in transcription. Annu. Rev. Biochem. *81*, 177–201. https://doi.org/10.1146/annurev-biochem-052110-120012.

52. Eberharter, A., Ferrari, S., Längst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., and Becker, P.B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. EMBO J. *20*, 3781–3788. https://doi.org/10.1093/emboj/20.14.3781.

53. Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914. https://doi.org/10.1016/0092-8674(92)90610-o.

54. Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol. Cell *37*, 457–468. https://doi.org/10.1016/j.molcel.2010.01.030.

55. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. *23*, 185–188. https://doi.org/10.1038/13810.

56. Good, K.V., Vincent, J.B., and Ausió, J. (2021). MeCP2: the genetic driver of Rett syndrome epigenetics. Front. Genet. *12*, 620859. https://doi.org/10.3389/fgene.2021.620859.

57. Boxer, L.D., Renthal, W., Greben, A.W., Whitwam, T., Silberfeld, A., Stroud, H., Li, E., Yang, M.G., Kinde, B., Griffith, E.C., et al. (2020). MeCP2 represses the rate of transcriptional initiation of highly methylated long genes. Mol. Cell *77*, 294–309.e9. https://doi.org/10.1016/j.molcel.2019.10.032.

58. Cheng, T.L., Wang, Z., Liao, Q., Zhu, Y., Zhou, W.H., Xu, W., and Qiu, Z. (2014). MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. Dev. Cell *28*, 547–560. https://doi.org/10.1016/j.devcel.2014.01.032.

59. Glaich, O., Parikh, S., Bell, R.E., Mekahel, K., Donyo, M., Leader, Y., Shayevitch, R., Sheinboim, D., Yannai, S., Hollander, D., et al. (2019). DNA methylation directs microRNA biogenesis in mammalian cells. Nat. Commun. *10*, 5657. https://doi.org/10.1038/s41467-019-13527-1.

60. Li, C.H., Coffey, E.L., Dall'Agnese, A., Hannett, N.M., Tang, X., Henninger, J.E., Platt, J.M., Oksuz, O., Zamudio, A.V., Afeyan, L.K., et al. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. Nature *586*, 440–444. https://doi.org/10.1038/s41586-020-2574-4.

61. Yan, Q., Wulfridge, P., Doherty, J., Fernandez-Luna, J.L., Real, P.J., Tang, H.Y., and Sarma, K. (2022). Proximity labeling identifies a repertoire of site-specific R-loop modulators. Nat. Commun. *13*, 53. https://doi.org/10.1038/s41467-021-27722-6.

62. Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proc. Natl. Acad. Sci. USA *104*, 2709–2714. https://doi.org/10.1073/pnas.0608056104.

63. Jeffery, L., and Nakielny, S. (2004). Components of the DNA methylation system of chromatin control are RNA-binding proteins. J. Biol. Chem. *279*, 49479–49487. https://doi.org/10.1074/jbc.M409070200.

64. Sabino, J.C., de Almeida, M.R., Abreu, P.L., Ferreira, A.M., Caldas, P., Domingues, M.M., Santos, N.C., Azzalin, C.M., Grosso, A.R., and de Almeida, S.F. (2022). Epigenetic reprogramming by TET enzymes impacts co-transcriptional R-loops. Elife *11*, e69476. https://doi.org/10.7554/eLife.69476.

65. Zyner, K.G., Mulhearn, D.S., Adhikari, S., Martínez Cuesta, S., Di Antonio, M., Erard, N., Hannon, G.J., Tannahill, D., and Balasubramanian, S. (2019). Genetic interactions of G-quadruplexes in humans. Elife *8*, e46793. https://doi.org/10.7554/eLife.46793.

66. Azzalin, C.M., and Lingner, J. (2006). The human RNA surveillance factor UPF1 is required for S phase progression and genome stability. Curr. Biol. *16*, 433–439. https://doi.org/10.1016/j.cub.2006.01.018.

67. Hosoda, N., Kim, Y.K., Lejeune, F., and Maquat, L.E. (2005). CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. Nat. Struct. Mol. Biol. *12*, 893–901. https://doi.org/10.1038/nsmb995.

68. Ngo, G.H.P., Grimstead, J.W., and Baird, D.M. (2021). UPF1 promotes the formation of R loops to stimulate DNA double-strand break repair. Nat. Commun. *12*, 3849. https://doi.org/10.1038/s41467-021-24201-w.

69. Li, L., Germain, D.R., Poon, H.Y., Hildebrandt, M.R., Monckton, E.A., McDonald, D., Hendzel, M.J., and Godbout, R. (2016). DEAD box 1 facilitates removal of RNA and homologous recombination at DNA double-strand breaks. Mol. Cell Biol. *36*, 2794–2810. https://doi.org/10.1128/MCB.00415-16.

70. Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A.E., Sattentau, Q., Meinhart, A., and Proudfoot, N.J. (2018). RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. Mol. Cell *70*, 650–662.e8. https://doi.org/10.1016/j.molcel.2018.04.001.

71. Mersaoui, S.Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F.F., Masson, J.Y., and Richard, S. (2019). Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. EMBO J. *38*, e100986. https://doi.org/10.15252/embj.2018100986.

72. Mosler, T., Conte, F., Longo, G.M.C., Mikicic, I., Kreim, N., Möckel, M.M., Petrosino, G., Flach, J., Barau, J., Luke, B., et al. (2021). R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability. Nat. Commun. *12*, 7314. https://doi.org/10.1038/s41467-021-27530-y.

73. Tran, P.L.T., Pohl, T.J., Chen, C.F., Chan, A., Pott, S., and Zakian, V.A. (2017). PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. Nat. Commun. *8*, 15025. https://doi.org/10.1038/ncomms15025.

74. Song, C., Hotz-Wagenblatt, A., Voit, R., and Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes Dev. *31*, 1370–1381. https://doi.org/10.1101/gad.300624.117.

75. Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249–253. https://doi.org/10.1038/nature13923.

76. Okamoto, Y., Abe, M., Itaya, A., Tomida, J., Ishiai, M., Takaori-Kondo, A., Taoka, M., Isobe, T., and Takata, M. (2019). FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. FEBS J. *286*, 139–150. https://doi.org/10.1111/febs.14700.

77. Lindström, M.S., Jurada, D., Bursac, S., Orsolic, I., Bartek, J., and Volarevic, S. (2018). Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. Oncogene *37*, 2351–2366. https://doi.org/10.1038/s41388-017-0121-z.

78. ten Asbroek, A.L.M.A., van Groenigen, M., Nooij, M., and Baas, F. (2002). The involvement of human ribonucleases H1 and H2 in the variation of response of cells to antisense phosphorothioate oligonucleotides. Eur. J. Biochem. *269*, 583–592. https://doi.org/10.1046/j.0014-2956.2001.02686.x.

79. Subtil-Rodríguez, A., and Reyes, J.C. (2010). BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. EMBO Rep. *11*, 751–757. https://doi.org/10.1038/embor.2010.131.

80. Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., and Landegren, U. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat. Methods *3*, 995–1000. https://doi.org/10.1038/nmeth947.

81. Shen, J., Zhang, L., and Zhao, R. (2007). Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. J. Biol. Chem. *282*, 22544–22550. https://doi.org/10.1074/jbc.M702304200.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-β-Actin	Sigma	Cat# A2066; RRID:AB_476693
Polyclonal antibody Rabbit anti-AID	Abcam	Cat# ab59361; RRID:AB_940171
Polyclonal antibody Rabbit anti-DDX47	Sigma	Cat# HPA014855; RRID:AB_1847567
Monoclonal antibody Mouse anti-DDX47	Santa Cruz	Cat# sc-377333
Monoclonal antibody Mouse anti-FLAG	Sigma	Cat# F3165; RRID: AB_259529
Polyclonal antibody Rabbit anti- FLAG	Sigma	Cat# F7425; RRID: AB_439687
Polyclonal Goat Anti-FLAG	Abcam	Cat# ab1257; RRID: AB_299216
Polyclonal antibody Rabbit anti-Fibrillarin	Abcam	Cat# ab5821; RRID:AB_2105785
Monoclonal antibody anti-Histidine	Sigma	Cat# GE27-4710-01
Polyclonal antibody Rabbit anti-IgG	Sigma	Cat# M7023; RRID: AB_260634
Mouse anti-IgG	Sigma	Cat# I8765; RRID: AB_1163672
Polyclonal antibody Rabbit anti- MECP2	Abcam	Cat# ab2828; RRID: AB_2143853
Polyclonal antibody Rabbit anti-MECP2 ChIP Grade	Abcam	Cat# ab195393; RRID: AB_2848209
Polyclonal antibody Rabbit anti- Nucleolin	Abcam	Cat# ab50279; RRID: AB_881762
Polyclonal antibody Rabbit anti- PCNA	Abcam	Cat# ab18197; RRID: AB_444313
Monoclonal antibody Mouse anti- Phospho Histone H2AX.X (Ser 139)	Biolegend	Cat# 613402; RRID: AB_315795
Monoclonal antibody Mouse anti-RNAPI (RPA194)	Santa Cruz	Cat#sc-48385; RRID: AB_675814
Mouse anti-S9.6 Hybridoma cell	ATCC	Cat # HB-8730; RRID:CVCL_G144
Polyclonal antibody Rabbit anti-UBF	Behtyl	Cat#A301-859A; RRID: AB_1279493
Monoclonal antibody Mouse anti-Vinculin	Sigma	Cat#V9264 Sigma; RRID: AB_10603627
Polyclonal Antibody Horseradish peroxidase anti-Rabbit	Sigma	Cat#A0545; RRID: AB_257896
Polyclonal Antibody Horseradish peroxidase anti-Mouse	Sigma	A4416; RRID: AB_258167
Polyclonal antibody PLA probe anti-Mouse	Sigma	Cat# DUO92004; RRID: AB_2713942
Polyclonal Antibody PLA probe anti-Rabbit	Sigma	Cat# DUO92002; RRID: AB_2810940
Polyclonal Antibody Alexa Fluor 488 Goat anti-Mouse	Molecular Probes	Cat# A11029; RRID: AB_2534088

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Polyclonal Antibody Alexa Fluor 488 Chicken anti-Mouse	Molecular Probes	Cat# A21200; RRID: AB_2535786
Alexa Fluor 594 Chicken anti-Mouse	Molecular Probes	Cat# A21201; RRID: AB_141630
Alexa Fluor 647 Chicken anti-Mouse	Molecular Probes	Cat# A21463; RRID: AB_1500641
Polyclonal Antibody Alexa Fluor 488 Goat anti-Rabbit	Molecular Probes	Cat# A11008; RRID: AB_143165
Polyclonal Antibody Alexa Fluor 555 Donkey anti-Rabbit	Molecular Probes	Cat# A31572; RRID: AB_162543
Alexa Fluor 568 Goat anti-Rabbit	Molecular Probes	Cat# A11011; RRID: AB_143157
Alexa Fluor 647 Donkey anti-Goat	Molecular Probes	Cat# A21447; RRID: AB_141844
Bacterial and Virus Strains		
E. coli BL21:DE3 Rosetta cells	Novagen	Cat# 70954
Chemicals, peptides, and recombinant proteins		
DDX47	This study	N/A
DDX47 K74A	This study	N/A
DDX47 E175A	This study	N/A
Pierce Protease Inhibitor Tablets	Thermo Scientific	Cat# A32963
Phenyl-methylsulfonyl fluoride (PMSF)	Fisher Scientific	Cat# ICN800263
Igepal CA-630	Fisher Scientific	Cat# ICN19859690
Isopropyl-b-D-thiogalactopyranoside (IPTG)	Fisher Scientific	Cat# BP1620-10
Imidazole	Sigma-Aldrich	Cat# 56749-250G
Ni-NTA Agarose resin	Qiagen	Cat# 30230
SP Sepharose Fast Flow resin	Cytiva	Cat# 17-0729-10
Mono S column	Cytiva	Cat# 17516801
Superdex 200 Increase 10/300 GL column	Cytiva	Cat# 28990944
Doxycycline	Sigma	D9891
α-Amanitin	Sigma	A2263
Thymidine	Sigma	T9250
Propidium Iodide	Sigma	P4170
Critical commercial assays		
Duo-link *in situ* Red started Kit (mouse-rabbit). Proximity ligation assay	Sigma	Cat# DUO92101
Comet assay	Trevigen	Cat# 4250-050-K
Mutagenesis Q5 New England	NEB	Cat# E0554S
Click-iT RNA Imaging Kit	Invitrogen	Cat# C10329
Experimental models: Cell lines		
U2OS-TR-AID	This study	N/A
U2OS	ATCC	N/A
HeLa	ECACC	93021013
Oligonucleotides		
ON TARGET Plus-Druggable genome siRNA library	Dharmacon	77G-104655-05
siRNA and DNA oligonucleotides are listed in Table S5	N/A	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pCDNA3	(ten Asbroek et al.,2002)	N/A
pCDNA3-RNaseH1	(ten Asbroek et al.,2002)	N/A
pFLAG-CMV-6A	Sigma	E1900
pCMV-DDX47-FLAG	Origene	RC209448
pCMV-DDX47-K74A-FLAG	This study	N/A
pCMV-DDX47-E175A-FLAG	This study	N/A
pDONR221-DDX47-WT	This study	N/A
pDONR221-DDX47-K74A	This study	N/A
pDONR221-DDX47-E175A	This study	N/A
pET300 NT-His6- DDX47-WT	This study	N/A
pET300 NT-His6-DDX47-K74A	This study	N/A
pET300 NT-His6-DDX47-E175A	This study	N/A
Software and algorithms		
MetaXpress software (version 4.0.0.24 Molecular Devices-granularity application)	Molecular Probes	https://www.moleculardevices.com/products/cellular-imaging-systems/acquisition-and-analysis-software/metaxpress
Prism software	GraphPad	https://www.graphpad.com/scientific-software/prism/
GeneMANIA prediction software (Cytoscape v3.8.0)	N/A	https://cytoscape.org/download.html
TriTek CometScore Professional (version 1.0.1.60)	TriTek Corporation	https://www.bioz.com/result/cometscore%20pro%20software/product/TriTek%20Corp

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Andrés Aguilera (aguilo@us.es).

Materials availability

Plasmids and cell lines generated in this study are available upon request from the lead contact without restrictions.

Data and code availability

- All raw data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell cultures

This study has been performed using U2OS (ATCC) and HeLa (ECACC, 93021013) human cells as model systems.

Cells were cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO, USA) supplemented with 10% heat-inactivated fetal bovine serum (SIGMA Aldrich, Germany), 2 mM L-glutamine and 1% antibiotic-antimycotic (Biowest, France). U2OS-TR-AID stable cell line carrying a tetracycline-regulated AID gene cassette was cultured under the same conditions but supplemented with a special Fetal Bovine Serum (FBS) South America, Tetracycline Free (S181T, Biowest). Cells were grown at 37°C and 5% CO_2.

For RNAPII inhibition, cells were treated with the inhibitor α-amanitin 50 μg/mL for 3 h (Sigma-A2263).

METHOD DETAILS

siRNAs, plasmids, and transfections

Transient transfection of siRNA was performed using DharmaFECT 1 (Dharmacon) or Lipofectamine 3000 (Invitrogene) according to the manufacturer's instructions. Lipofectamine 2000 and Lipofectamine 3000 (Invitrogen) was used for plasmid transfection. Assays were performed 72 h after small interfering RNA (siRNA) transfection and 24 or 48 h after plasmid transfection.

Plasmids used are listed in key resources table. pCMV-DDX47-FLAG (Origene) contains the full-length human DDX47-WT cDNA cloned into the vector pCMV6-Entry with C-terminal Myc-DDK tag. DDX47 mutant forms were generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according manufacturer's recommendations (pCMV-DDX47-K74A-FLAG and pCMV-DDX47-E175A-FLAG). DDX47 WT and DDX47 mutants were cloned into the vector pDONR221 for rapid cloning with Gateway technology and subsequently into bacterial expression vectors with T7 promoter and an N-terminal 6xHis tag for protein purification (pET300 NT-His6- DDX47-WT; pET300 NT-His6- DDX47-K74A; pET300 NT-His6- DDX47- E175A). RNaseH1 expression was driven using the plasmid pcDNA3-RNaseH1 containing the full-length RNase H1 cloned into pcDNA3.[78]

High-throughput siRNA screening methods

ON TARGET Plus-Druggable genome siRNA library (Dharmacon, 77G-104655-05) containing four-siRNA pools targeting human genes considered potential targets for therapeutic drugs was used for the screening. 3205 out of 4796 siRNAs of the library were analyzed (Table S1). siRNA library preparation, transfection and immunofluorescence protocol were carried out using Hamilton Microlab Star.

siRNA library transfection

A reverse transfection protocol was followed by plating U2OS-TR-AID cells onto pre-plated siRNA-transfection mixtures. siRNA pools were reconstituted and diluted to 1 μM in 1x siRNA buffer (60 mM KCl, 6 mM HEPES-pH 7.5, 0.2 mM $MgCl_2$). siRNA-transfection mixtures were prepared mixing 47 μL of each siRNA pool (0.5 μM) in Opti-MEM, and 47 μL of the transfection reagent Lipofectamine 2000 (0.21 μL/well) in Opti-MEM, and then incubated for 20 min at room temperature to allow siRNA-lipofectamine complex formation to occur. For reverse transfection, U2OS-TR-AID cells in optimal growth phase were previously prepared. A final concentration of 33.3 nM siRNA was used to transfect 6000 cells/well. 4 h after incubation, 50 μL/well of high-concentrate (3x) complete medium was added to adjust the normal concentration of the medium. Two transfection-plates (duplicate-plates) were prepared. Positive control FANCD2 siRNA pool and the negative non-targeting control siRNA pool were included in each plate.

AID induction

U2OS-TR-AID is a stable cell line carrying a tetracycline-regulated AID gene cassette. Transcription was activated with doxycycline. For AID induction, 24 h post-transfection U2OS-TR-AID cells were treated or not with doxycycline (6 μg/mL) and 48 h later cells were harvested.

γH2AX immunofluorescence

72 h post-transfection, the reaction was stopped by manually tossing the media and fixing the cells with 4% formaldehyde in PBS for 10 min at RT. Cells were permeabilized with 70% ethanol for 5 min at −20°C, 5 min at 4°C and washed three times in PBS. The immunofluorescence protocol was performed using Hamilton Microlab Star. The plates were blocked by adding 100 μL blocking solution (3% BSA in PBS). After 1 h, the blocking solution was removed and 50 μL of anti-H2A.X Phospho (Ser 139) antibody (1:1000 in blocking solution; Biolegend 613,402) was added and plates were incubated at RT for 1 h. The plates were then washed three times. 50 μL of the secondary antibody goat-anti-mouse antibody Alexa 488 conjugated (1:1000 in blocking solution; Molecular Probes, A11029) was added, and the plates were incubated at RT for 1 h. The plates were then washed three times and 100 μL of Hoechst (AnaSpec Inc. Cat# 83219) was added in each well to stain the nucleus.

Imaging

Data acquisition was performed with ImageXpress Micro Electron (Molecular Devices, 137,239) using a 40x objective. The entire 96-well plate (PerkinElmer) was scanned with a 20 nW 690 nm laser and set up for two channels of acquisition. γH2AX-Alexa-488 fluorescence was acquired using a 472/30 nm excitation filter, and Hoechst fluorescence was acquired using 377/50 nm excitation filter. Randomly fields were acquired and analyzed (≃ 200 cells/well). H2AX foci were quantified by automated scoring using MetaXpress software (version 4.0.0.24 Molecular Devices-granularity application).

γH2AX foci analysis

MetaXpress was used in order to measure Granule Count that characterizes γH2AX foci associated to each siRNA and condition in each well. The percentage of cells with ≥5 γH2AX foci before and after AID induction of each siRNA depletion was calculated. The screening was performed in duplicate, and siRNAs whose depletion leads to an increase (≥1.2) in the percentage of cells with γH2AX foci after AID induction (siRNA + DOX), versus non-induced conditions (siRNA -DOX), that was higher (≥1.2) to that observed in AID-induced-control cells (siRNA + DOX versus siC + DOX) were selected as candidates. To take into account the variability of the data, per-plate and per-date, each condition was always compared versus non-induced conditions and control cells located in the same plate. The same procedure was performed to calculate the percentage of cells with ≥10 γH2AX foci after AID induction. The *Z* score and the reproducibility score were calculated to take into consideration when candidate selection.

Z-score= (%- μ)/σ.

% = percentage of cells with ≥5 γH2AX foci/percentage of cells with ≥10 γH2AX foci.

μ = Mean of each plate.

σ = Standard deviation of each plate.

Reproducibility = $2^y n$

n = Number of times that a candidate meets the criteria described above, taking into account 5 and 10 γH2AX foci analysis.
y = Number of times each gene is positive regardless of replicates taking into account both analysis.

γ-H2AX intensity analysis

Analysis of the γ-H2AX intensity were carried out for the candidates obtained in the first round. We used MetaXpress to measure the Nucleus Average Intensity of each well. A decrease of the fluorescence intensity levels during the image capture along the plate was observed. So, an adjusted intensity value for every well was calculated using a power regression where variable x is the order number of the laser pathway. To select a positive candidate we calculated the intensity for each well and then we used this value for each siRNA following the condition/criteria: siX (+AID) > 0 > siX (-AID). If the previous condition is accomplished the candidate is positive in that plate.

Reproducibility = $2^y n$
n = Number of times that a candidate meets the criteria described above
y = Number of times each gene is positive.

S9.6 immunofluorescence

Cells were cultured on glass coverslips and fixed in methanol (8 min at −20°C) and washed three times in PBS. When required, pre-extraction prior fixation and RNase *in vitro* treatments were used. For RNase treatments cells were incubated in their respective commercial buffers at 1x containing 40 U/mL RNase III (AM2290, Ambion) or 60 U/mL RNase H (M0297S, NEB), for 30 min at 37°C.cells. Next, for S9.6 immunofluorescences, coverslips were blocked in 3% BSA in PBS for 5 h at 4°C and incubated with anti-S9.6 (Hybridoma cell Line HB-8730, 1:500) and anti-nucleolin (ab50279, 1:1000) or anti-FLAG (ab1257, 1:1000) primary antibodies diluted in 3% BSA in PBS o/n at 4°C, washed in PBS three times and incubated with the subsequent secondary antibodies conjugated with Alexa Fluor diluted in 3% BSA in PBS (1:1000) for 1 h at RT in darkness. Washed twice for 5 min, DAPI staining and mounting as described above. At least 100 cells from each experiment were scored.

DNA-RNA hybrid immunoprecipitation (DRIP-qPCR)

DRIP assays were performed by immunoprecipitating DNA–RNA hybrids using the S9.6 antibody from gently extracted and enzymatically digested DNA, treated or not with RNase H (New England Biolabs, USA) *in vitro* as described,[19,25] with minor modifications. After 72 h of siRNA transfection, pellet from a confluent 10-cm plate of cells was collected using accutase, washed in PBS and resuspended in 800 μL of 1X TE. Then, 20.75 μL SDS 20% and 2.5 μL proteinase K (20 mg/mL) were added and pellet was incubated at 37°C overnight. DNA was extracted gently with phenol-chloroform in phase lock tubes (VWR, USA). Precipitated DNA was spooled on a glass rod, washed 2 times with 70% EtOH, resuspended gently in 1X TE and digested overnight with 50 U of HindIII, EcoRI, BsrGI, XbaI and SspI. For the negative control, half of the DNA (16 μg) was treated with 5 μL RNase H overnight. In parallel, S9.6 antibody (3 μg/sample) was incubated overnight at 4°C with Dynabeads Protein A (Invitrogen) (30 μL/sample) in 1X binding buffer (final volume: 60 μL/sample). 5 μg of the digested DNA, untreated or treated with RNase H, were bound to S9.6 antibody-dynabeads complexes during 2 h at 4°C for immunoprecipitation. Next, the beads were washed 3 times with 1X binding buffer. DNA was eluted in 180 μL elution buffer, treated 45 min with 7 μL proteinase K at 55°C and cleaned with the NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, USA). 1 μg of the digested DNA, untreated or treated with RNase H, was used for input DNA of each condition, in which similar proteinase K treatment and purification were performed. Quantitative PCR (qPCR) of immunoprecipitated DNA (IP) fragments and input DNA was performed on a 7500 Fast Real-Time PCR System (Applied Biosystems, Carlsbad,CA). Primers used are listed in Table S5.

10X Binding buffer: 100 mM $NaPO_4$ pH 7.0, 1.4 M NaCl, 0.5% Triton X-100. Elution buffer: 50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS.

DRIP quantification and normalization: Input and immunoprecipitated (IP) were eluted in 150 μL of 1X TE. 2 μL of input and IP were used for qPCR. Changes in the abundance of DNA-RNA hybrids in each region were determined as the percentage of input recovered for each immunoprecipitated sample using the Equation % input = 2[(CtINPUT – log2DF)− CtIP]× 100. Ct = threshold cycle; DF = dilution factor.

Chromatin immunoprecipitation (ChIP)

ChIP assays were performed according to[79] with some modifications. After 72 h of siRNA transfection, HeLa cells were crosslinked, resuspended in cell lysis buffer 1, then centrifuged and nuclei lysis buffer 2 was added. Chromatin was sonicated on the maximum intensity setting, with eight pulses of 30 s on and 30 s off in Bioruptor (Diagenode). For each immunoprecipitation, 30 μg of chromatin were diluted with IP buffer. Chromatin was incubated overnight at 4°C with ChIP-grade antibodies. A negative control with IgG rabbit or mouse antibody was used to calculate the background signal. Chromatin-antibody complexed were immunoprecipitated for 2 h with 30 μL of Dynabeads Protein A (rabbit)/G (mouse) (Invitrogen) at 4°C. Input and immunoprecipitate (IP) were then un-crosslinked and phenol-chloroform purified. Finally, DNA was resuspended in 50 μL of MQ H_2O. Primer pairs used for amplification are listed in Table S5. Samples were run in 7500 Fast Real-time PCR system (Applied Biosystem). Results were analyzed with 7500 System Software V2.0.6. The amount of DNA was determined by Real-Time qPCR and was expressed as percentage of input. It was calculated based on the standard DNA curve and the amplification efficiency. Considering amplification (AF) and dilution (DF) factor ($AF^{(40 - Ct)}$)

DF, we calculate the absolute quantification of each sample. IP/Input ratios in the different regions were calculated and multiplied by 100. The IP signal of the IgG (negative control) was considered as background.

Single-cell gel electrophoresis

Alkaline single-cell gel electrophoresis or comet assays were performed as described.[46] Single-cell alkaline gel electrophoresis was performed with CometAssay Kit (Trevigen) following manufacturer's instructions. Images were acquired with a Leica DM6000 microscope equipped with a DFC390 camera (Leica). Comet tail moments were analyzed using Comet-score software (version 1.5). The median of tail moment from at least independent three biological repeats were calculated and are shown in each case. More than 200 cells from each experiment were scored.

EU labeling

Click-iT RNA Imaging Kit (C10329, Invitrogen) was used to assay the newly transcribed RNA *in vivo* during active synthesis through EU incorporation following manufacturer's instructions. First, HeLa cells were seeded in 6-well plates and transfected with siRNA. After 72 h siRNA transfection, HeLa cells were incubated with the modified uridine analogue EU (5-ethynyluridine) at the final concentration of 1 mM (C10329, Invitrogen). A range of EU incubations from 20 min to 120 min, with an intermediate point at 60 min was tested. In cells treated with α-amanitin EU was added in the last 20 min. Then, samples were fixed, permeabilized and Click-iT reaction performed according to the manufacturer's guidelines. Finally, nuclei were stained with DAPI and mounted in ProLong Gold AntiFade reagent (Thermo). Images were acquired with a Leica DM6000 microscope equipped with a DFC390 camera (Leica) at 63X magnification and LAS AF software software (Leica). Random images were acquired with a 63x objective, and EU intensity was scored using the MetaMorph software.

Proximity ligation assay (PLA)

PLA was performed as in previous reports[23,80] with reagents from Duolink In Situ Red Starter Kit (Sigma) in accordance with the manufacturer's instructions. First, cells were fixed, permeabilized and incubated with primary antibodies diluted in PBS-3%BSA overnight at 4°C (see key resources table). Then, secondary antibody binding, ligation and amplification reactions were performed with Duolink *in situ* PLA probe anti-rabbit PLUS and Duolink *in situ* PLA probe anti-mouse MINUS. Finally, nuclei were stained with DAPI and mounted in ProLong Gold AntiFade reagent (Thermo). For negative controls, everything was performed identically, except that only one of the primary antibodies was added. Images were acquired with a Leica DM6000 microscope equipped with a DFC390 camera (Leica) at 63X magnification and LAS AF software (Leica). PLA foci number per cell was scored using the MetaMorph software.

Microscopy image acquisition and data analysis

A Leica DM6000 microscope equipped with a DFC390 camera (Leica) and the LAS AF software (Leica) were used for fluorescence microscopy images and data acquisition, respectively. Software used for image analysis are listed in key resource table.

Expression and purification of wild type and mutant DDX47

The cDNAs that encode wild type, K74A and E175A variants of DDX47 were introduced into the pET300-NT Destination vector to add an N-terminal His_6 tag to these proteins. The resulting DDX47 expression plasmids (see pET300 NT-His6-DDX47 vectors in key resource table) were introduced into *E. coli* BL21:DE3 Rosetta cells, which were grown at 37°C to OD_{600} = 0.8, and protein expression was induced by the addition of 0.2 mM IPTG and incubation at 16°C for 16 h. Cells were harvested by centrifugation and all the subsequent steps were carried out at 0-4°C. For lysate preparation, a cell pellet (14 g, from 4 L of culture) was suspended in 100 mL K buffer (20 mM KH_2PO_4, pH 7.4, 10% glycerol, 0.5 mM EDTA, 0.01% Igepal, 1 mM DTT) with 500 mM KCl, 1 mM PMSF and 1 tablet of Pierce Protease Inhibitor (Thermo Scientific), and then subject to sonication (six 30 s pulses). The crude cell lysate was clarified by ultracentrifugation (100,000xg for 45 min) and diluted with 2 volumes ($\sim$200 mL) of K buffer. The diluted lysate was loaded onto a chromatography column (Bio-Rad) filled with 10 mL SP Sepharose resin (Cytiva), which was equilibrated with 50 mL of K buffer plus 150 mM KCl. A 140 mL linear gradient from 150 to 850 mM KCl in K buffer was applied to elute DDX47, and the peak fractions ($\sim$400 mM KCl) were pooled and incubated with 2.5 mL of Nickel-NTA resin (Qiagen) for 1 h. The resin was washed once with 50 mL K buffer containing 1000 mM KCl, once with 25 mL K buffer containing 300 mM KCl and 1 mM each of ATP and $MgCl_2$, three times with 50 mL K buffer containing 300 mM KCl and 10 mM imidazole, before being treated with K buffer with 300 mM KCl and 200 mM imidazole to elute DDX47. The eluate (8 mL) was diluted with 2.5 volumes of K buffer and loaded onto a 1 mL Mono S column (Cytiva). The column was washed with 10 mL K buffer containing 150 mM KCl and then developed with a 25-mL gradient from 150 to 800 mM KCl. DDX47 protein eluted at $\sim$400 mM KCl, and the peak fractions were pooled and concentrated to 0.5 mL and further fractionated in a 24 mL Superdex 200 column (Cytiva) in K buffer with 300 mM KCl. The peak fractions were collected and concentrated to 3 mg/mL (Amicon 10K concentrator, Millipore) and stored in small aliquots at -80°C. The yield of highly purified DDX47 was $\sim$1 mg. The DDX47 mutant (K74A or E175A) was purified using the same procedure with a similar yield.

Nucleic acid unwinding assays

RNA-RNA duplexes without or with a 5′ or 3′ overhang were prepared as described.[81] DNA-RNA hybrids without or with a 5′ or 3′ overhang were prepared by annealing oligonucleotides (with one of the oligonucleotides being labeled with ^{32}P) listed in Table S5. In the unwinding reaction, DDX47 (wild type or mutant at the indicated concentration) was incubated with 10 nM substrate in reaction buffer (35 mM Tris-Cl, pH 7.5, 1 mM DTT, 3 mM ATP, 2 mM $MgCl_2$, 60 mM KCl), and 200 nM of "trap" RNA or DNA (the oligonucleotide that was labeled in the substrate) at 30°C for 30 min (for the RNA-RNA substrates) or 28°C for 10 min (for the DNA-RNA substrates). For direct comparison (Figure 5B), DDX47 (0.025-4.0 μM) was incubated with 10 nM substrate (RNA-RNA duplex or DNA-RNA duplex) in the same reaction buffer supplemented with 200 nM of "trap" RNA or DNA both at 28°C for 15 min. Reaction mixtures were then immediately deproteinized by treatment with SDS (0.1%) and proteinase K (0.5 mg/mL) for 6 min at 28°C and then resolved in 15% polyacrylamide gels in TAE buffer (40 mM Tris, 20 mM Acetate acid and 1 mM EDTA) at 4°C. Gels were dried and subject to phosphorimaging analysis.

The half R-loop structure based on a DNA-RNA hybrid with a 5′ ssDNA flap at one end was constructed as described.[20] DDX47 (wild type or mutant at the indicated concentration) was incubated with the substrate in reaction buffer (25 mM Tris-Cl, pH 7.5, 1 mM DTT, 100 μg/mL BSA, 2 mM ATP, 2 mM $MgCl_2$, 60 mM KCl) at 30°C for 20 min. Reaction mixtures were deproteinized before being resolved in 7% polyacrylamide gels in TAE buffer at 4°C and analyzed, as above.

Flow cytometry

For evaluating cell cycle progression, cells were synchronized by treating them with 2mM thymidine for 16 h. Cells were collected, fixed in 70% ethanol (−20°C, >1hour), washed (x3) in PBS and stained (overnight, 4°C) with 50μg/mL propidium iodide (SIGMA) in the presence of 10μg/mL RNase A (Qiagen). Data were acquired in BD FACSCalibur cell analyzer.

Miscellanea

Western-blot were performed using standard procedures. Relative qPCRs were used to determine the relative mRNA levels in human cells. cDNA was obtained from total RNA extracted using RNeasy Mini Kit (Qiagen) (1 μg) by reverse transcription using QuantiTect Reverse transcription (Qiagen). mRNA expression values were normalized to mRNA expression of the Hypoxanthine PhosphoRibosylTransferase (HPRT) or Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) housekeeping genes. Primers used are listed in Table S5.

QUANTIFICATION AND STATISTICAL ANALYSIS

For DRIP and ChIP assay statistical analysis, Student's t-test one tailed was applied for comparisons of two independent groups. In Immunofluorescence, single-cell electrophoresis assays and PLA assays, statistical significant differences were assessed with nonparametric Mann-Whitney U-test two tailed. For nucleolar total area per nucleus and EU labelling experiments, Student's t-test two tailed were used to ensure statistical significant differences. Data were analyzed with EXCEL (Microsoft) or GraphPad Prism software. The statistical test used in each experiment is mentioned in the figure legend. In general, a p-value <0.05 was considered as statistically significant (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$). Specific replicate numbers (n) for each experiment can be found in the corresponding figure legends. In all figures, means are plotted and SEM is represented as error bars.

Gene set enrichment and gene interaction analysis

Gene Ontology (GO) analyses of enriched biological processes for putative R-loop forming factors were performed using PANTHER (p-value calculated using Fischer's exact test).

GeneMANIA plugin for Cytoscape v3.8.0 was used to build a network of known interactions, both physical and genetic, shared protein domains and co-localization.