SUPPLEMENTAL MATERIAL

UAP56/DDX39B is a major co-transcriptional RNA-DNA helicase that unwinds harmful R loops genome-wide

Carmen Pérez-Calero [1], Aleix Bayona-Feliu [1], Xiaoyu Xue [2,3], Sonia I. Barroso [1], Sergio Muñoz [1], Víctor M. González-Basallote [1], Patrick Sung [2,4] and Andrés Aguilera [1,*]

[1] Centro Andaluz de Biología Molecular y Medicina Regenerativa CABIMER, Universidad de Sevilla-CSIC-Universidad Pablo de Olavide, Sevilla 41092, Spain; [2] Molecular Biophysics and Biochemistry, School of Medicine, University of Yale, New Haven, CT06510, USA; [3] Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas 78666, USA; [4]Department of Biochemistry and Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA.

* *Corresponding and leading author.* E-mail: aguilo@us.es

Inventory of Supplemental Material:

SUPPLEMENTAL MATERIALS AND METHODS

Human cells culture, transfection reagents and plasmid

HeLa (ECACC, 93021013) and HEK293T (ECACC, 12022001) cells were cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO) supplemented with 10% heat-inactivated fetal bovine serum (Sigma Aldrich, Merck KGaA) and 1% antibiotic-antimycotic (BioWEST) at 37ºC (5% CO_2). K562 (ATCC, CCL-243) cells were cultured in Iscove's Modified Dulbecco's medium (IMDM; GIBCO) supplemented with 10% heat-inactivated fetal bovine serum (Sigma Aldrich) and 1% antibiotic-antimycotic (BioWEST) at 37ºC (5% CO_2).

Plasmid transfection was performed using Lipofectamine 2000 or Lipofectamine 3000 (Invitrogen), whereas transient transfection of siRNA was performed using DharmaFECT 1 (Dharmacon) according to the manufacturer's instructions. All assays were performed 48 h after siRNA transfection plus 24 h after plasmid transfection.

Plasmids used for transfections were: the pcDNA3 (Invitrogen) vector and pcDNA3-RNaseH1, containing the full-length RNase H1 cloned into pcDNA3 (ten Asbroek et al. 2002); the pEGFP (Clontech) vector and pEGFP-M27-H1, containing the GFP-fused RNase H1 lacking the first 26 amino acids responsible for its mitochondrial localization cloned into pEGFP for GFP-RNase H1 overexpression (Cerritelli et al. 2003); the pFLAG (Sigma Aldrich) vector and pFLAG-UAP56-WT, pFLAG-UAP56-K95A and pFLAG-UAP56-E197A, containing UAP56-WT, UAP56-K95A and UAP56-E197A cloned into pFLAG, respectively; the pUBC-EGFP vector (Vitor et al. 2019) and pUBC-EGFP-UAP56 containing UAP56-WT cloned into pUBC-EGFP vector. For siRNA depletions, we used the ON-TARGET SMARTpool siRNAs from Dharmacon for UAP56 (L-003805-00), SETX (L-021420-00), DDX23 (L-019861-01), AQR (L-022214-01), FANCD2 (L-016376-00) THOC1 (L-016376-00), Sin3A (L-012990-00) and XPG (L-006626-00). For UAP56 depletion the 4 siRNAs of the SMART pool were checked independently.

Reagents

Cordycepin (C3394; Sigma Aldrich) and TOP2 inhibitor dexrazoxane (D1446; Sigma Aldrich) were used.

Single-cell electrophoresis

Single-cell electrophoresis or comet assay was performed using a commercial kit (Trevigen) following the manufacturer's protocol. Comet slides were stained with SYBRGreen, and images were captured at 10× magnification with a Leica DM6000 microscope equipped with a DFC390 camera (Leica). More than 100 cells were counted in each experiment to calculate the median of the tail moment.

Immunofluorescence

Analysis of DNA damage by γH2AX foci was performed as previously described (Dominguez-Sanchez et al. 2011), using anti-γH2AX (1:500, 613402 Biolegend) and anti-RNase H1 (1:500, 15606-1-AP Proteintech) or anti anti-pFLAG rabbit (1:1000, F7425 Sigma Aldrich) for pFLAG-UAP56 detection. Then, secondary antibodies conjugated with Alexa 488 (1:1000, A21200 Thermo Fisher) and Alexa 594 (1:1000, A21201 Thermo Fisher) were used. For γH2AX nuclear intensity, cells were fixed with PBS + 4% formaldehyde + 0.5% Triton X-100 for 10 minutes at RT, washed with PBS, permeabilized with PBS + 0.2% Triton X-100 for 10 minutes at RT and blocked with PBS + 2% BSA + 0.05% Tween-20 for 30 minutes at RT. Primary anti γH2AX antibody (1:500, ab2893 Abcam) and secondary Alexa Fluor 488 antibody (1:1000) were incubated for 1 hour at RT in PBS + 1% BSA blocking solution. When analyzing γH2AX foci per cell throughout cell cycle, cells were fixed with PBS + 4% formaldehyde + 0.1% Triton X-100 for 10 minutes at RT, permeabilized with PBS + 0.5% Triton X-100 for 5 minutes at RT and blocked with PBS + 3% BSA + 0.1% Tween-20 for 30 minutes at RT. Primary anti-γH2AX antibody (1:1000, ab2893 Abcam) and secondary antibody

were incubated for 1 hour at RT in PBS + 3% BSA + 0.1% Tween-20. In the case of FANCD2 IF, the experiments were performed as reported (Garcia-Rubio et al. 2015). In brief, cells were pre-permeabilized with PBS + 0.25% Triton X-100 for 1 minute on ice and then fixed with PBS + 2% formaldehyde for 20 minutes at RT. Then, cells were incubated with anti-FANCD2 (1:100, sc-20022 Santa Cruz) and with anti-RNase H1 (1:400, 15606-1-AP Proteintech) followed by the secondary antibody conjugated with Alexa 488 and Alexa 546 (1:1000) in PBS + 3% BSA. In pre-permeabilized cells, the overexpressed RNase H1 stained only nucleus and nucleoli since the rest of the protein had been washed out. For XPG staining, cells were fixed with PBS + 4% formaldehyde for 10 minutes at room temperature, washed with PBS, permeabilized with PBS + 0.2% Triton X-100 for 10 minutes at RT and blocked with PBS + 3% BSA + 0.05% Tween-20 for 30 minutes at RT. Primary anti-XPG antibody (1:500, ab99248 Abcam) and secondary Alexa Fluor 488 antibody (1:1000) were used. In the case of EdU staining, cells were previously incubated for 30 minutes and then the click-it reaction was conducted following manufacturer instructions. For H3S10P IF, cells were fixed with PBS + 4% formaldehyde + 0.1% Triton X-100 for 10 minutes at RT, permeabilized with PBS + 0.5% Triton X-100 for 5 minutes at RT and blocked with PBS + 3% BSA + 0.1% Tween-20 for 30 minutes at RT. Primary anti-H3S10P (1:200, 06-750 Millipore) and secondary conjugated-antibody were incubated for 1 hour at RT in PBS + 3% BSA + 0.1% Tween-20.

S9.6 (hybridoma cell line HB-8730) immunofluorescence was performed essentially as described (Garcia-Rubio et al. 2018). Briefly, cells were fixed with 100% ice-cold methanol, blocked with PBS + 2% BSA overnight at 4°C and incubated with S9.6 (1:500) and anti-nucleolin (1:1000, ab50279 Abcam) or anti-pFLAG (1:1000) antibodies overnight at 4°C. Coverslips were washed three times in PBS, and then incubated with Alexa 488 and Alexa 594 conjugated secondary antibodies (1:1000) for 1 hour at RT.

Nuclei in all IFs were counter-stained with DAPI (Sigma Aldrich) and mounted in ProLong Gold AntiFade reagent (Invitrogen). Images of IF were acquired with a Leica DM6000 microscope equipped with a DFC390 camera (Leica) at ×63. The number of cells analyzed in each experiment is mentioned in the correspondent figure legend. Data acquisition was performed with LAS AF (Leica). Metamorph v7.5.1.0 software (Molecular Probes, Thermo Fisher) image analysis software was used to quantify the IF. In the case of the S9.6, only the foci and nuclear S9.6 signal intensity were quantified.

Co-immunoprecipitation and Proximity Ligation Assay

These experiments were performed as described (Salas-Armenteros et al. 2017) in accordance with the manufacturer's instructions. For Proximity Ligation Assay, the following antibodies were used: anti-mSin3A (1:50, sc-5299 Santa Cruz) and anti-UAP56 (1:200, ab47955 Abcam). Likewise, for co-immunoprecipitation assay anti-UAP56 (10 µg, 14798-1-AP Proteintech) was employed. Concordantly, later Western Blot analysis was performed with anti-mSin3A (1:2000, ab3479 Abcam).

DNA combing

DNA combing was performed as previously described (Salas-Armenteros et al. 2017). Cells were transfected with pcDNA3 or pcDNA3 RNase H1 for 48 h. Iododeoxyuridine and chlorodeoxyuridine labels were added for 20 min each. DNA molecules were counterstained with an anti-ssDNA antibody (1:500, DSHB) and an anti-mouse IgG coupled to Alexa 647 (1:50, A21241 Invitrogen). CldU and IdU were detected with BU1/75 (1:20, AbCys) and BD44 (1:20, Becton Dickinson) anti-BrdU antibodies, respectively. Secondary antibodies used were goat anti-mouse IgG Alexa 546 (1:50, A21123 Invitrogen) and chicken anti-rat Alexa 488 (1:50, A21470 Invitrogen). DNA fibers were analyzed on a Leica DM6000 microscope equipped with a DFC390 camera (Leica). Data acquisition was performed with LAS AX (Leica). Representative images

of DNA fibers were assembled from different microscopic fields and were processed as described (Bianco et al. 2012). To measure RF velocity (kb/min), the distance covered by individual forks during the pulse was determined as described (Dominguez-Sanchez et al. 2011). Replication asymmetry was calculated as the ratio "[(longest green tract – shortest green tract) by the longest green tract]" in divergent CldU tracks.

Purification of UAP56 wild-type and mutant proteins

The cDNAs that encode the wild type, K95A and E197A variants of UAP56 were introduced into the pGEX-KG vector to add an N-terminal GST tag to these proteins. The resulting UAP56 expression plasmids were introduced into *E. coli* BL21:DE3 Rosetta cells, which were grown at 37°C to OD600 = 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 (20 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 300 mM KCl and 5 µg/ml each of the protease inhibitors aprotinin, chymostatin, leupeptin and pepstatin, and then subject to sonication (three 1 min pulses). The crude cell lysate was clarified by ultracentrifugation (100,000Xg for 90 min) and then mixed gently with 2 ml of Glutathione Sepharose 4B resin (GE) for 1.5 h. The resin was washed sequentially with 50 ml K buffer containing 1 M KCl, 50 ml K buffer containing 300 mM KCl, 20 ml K buffer containing 300 mM KCl and 1 mM ATP, 20 ml K buffer containing 300 mM KCl and 5 mM $MgCl_2$, and 2 x 50 ml K buffer containing 300 mM KCl. UAP56 was eluted with 12 ml K buffer containing 300 mM KCl and 10 mM reduced glutathione and concentrated to 1 ml (Amicon 10K concentrator, Millipore). The GST tag was cleaved by incubating the concentrated protein pool with 100 µg of thrombin for 12 h. The reaction mixture was diluted with 2 ml of K buffer and applied onto to a 1-ml Mono Q column (GE), which was washed with 5 ml K buffer plus 150 mM KCl and then developed with a 25-ml linear gradient from 150 to 650 mM KCl.

The peak of UAP56, eluting at ~350 mM KCl, was collected, concentrated to 0.5 ml, and fractionated in a Superdex 200 gel filtration column (24 ml, GE) in K buffer containing 300 mM KCl. Fractions containing highly purified UAP56 (1 mg protein) were pooled, concentrated to 1 mg/ml, and stored in small aliquots at -80°C. The UAP56 K95A and E197A mutants were 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 (Shen et al. 2007). RNA-DNA hybrids without and with a 5' or 3' overhang and DNA-DNA duplex with a 5' overhang were prepared by annealing oligonucleotides (with one of the oligonucleotides being labeled with ^{32}P) listed above. In the unwinding reaction, UAP56 (wild type or mutant at the indicated concentration) was incubated with 5 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 100 nM of "trap" RNA or DNA (unlabeled version of the oligonucleotide that was labeled in the substrate) at 37°C (for the RNA-RNA substrates) or 30°C (for the RNA-DNA substrates) for 30 min or the indicated time. Reaction mixtures were deproteinized by treatment with SDS (0.1%) and proteinase K (0.5 mg/ml) for 10 min at 37°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 5' RNA-DNA flap structure that resembles a branch migratable R-loop structure was constructed as described (Schwab et al. 2015). UAP56 (wild type or mutant at the indicated concentration) was incubated with the substrate in reaction buffer (25 mM Hepes, pH 6.5, 1 mM DTT, 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 (40 mM Tris, 20 mM Acetate acid and 1 mM EDTA) at 4°C and analyzed, as above.

Oligonucleotides for unwinding assays

The following oligonucleotides were used in the unwinding assays: R13 (RNA, 13 nt) 5'GCUUUACGGUGCU3' and R13C (RNA, 13 nt) 5'-AGCACCGUAAAGC-3'; R23-5' (RNA, 23 nt) 5'-AAAACAAAAUAGCACCGUAAAGC-3' and R23-3' (RNA, 23 nt) 5'-GCUUUACGGUGCUUAAAACAAAA-3'; D13 (DNA, 13 nt) 5'-GCTTTACGGTGCT-3' and D13C (DNA, 13 nt) 5'-AGCACCGTAAAGC-3' and D23-5' (DNA, 23 nt) 5'-AAAACAAAATAGCACCGTAAAGC-3'; XX1 (DNA, 60 nt) 5'-ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC-3' and XX2 (DNA, 60 nt) 5'-GGGTGAACCTGCAGGTGGGCAAAGATGTCCCAGCAAGGCACTGGTAGAATTCGGCAGCGT-3'; R5'F (RNA, 30 nt) 5'-GGGUGAACCUGCAGGUGGGCAAAGAUGUCC-3'

mRNA quantification

cDNA synthesis and qPCR were performed as previously described (Dominguez-Sanchez et al. 2011). mRNA expression values of the indicated genes were calculated using the $2^{-\Delta Ct}$ method and normalized to the expression of the HPRT housekeeping gene as endogenous control.

RNA primers for real-time qPCR

The following primers were used for real-time quantitative PCR (qPCR):

APOE (forward) 5'- GGGAGCCCTATAATTGGACAAGT -3'

APOE (reverse) 5'- CCCGACTGCGCTTCTCA -3'

RPL13A (forward) 5'-GCTTCCAGCACAGGACAGGTAT-3'

RPL13A (reverse) 5'-CACCCACTACCCGAGTTCAAG-3'

EGR1 (forward) 5'-GCCAAGTCCTCCCTCTCTACTG-3'

EGR1 (reverse) 5'-GGAAGTGGGCAGAAAGGATTG-3'

PDK1 (forward) 5'-CACGGCTTTGCACTCTTCCT-3'

PKD1 (reverse) 5'-GCATGCCATGTAGCCTCTTGA-3'

TSC2 (forward) 5'-TGCATCATAGCCGCTCCAA-3'

TSC2 (reverse) 5'-GCCGGGCAATCCACTTG-3'

NTHL1 (forward) 5'-GCCCTCCTTGAATCCTTTTCC-3'

NTHL1 (reverse) 5'-TGCACCCAACTGCCAAGAA-3'

SNTB2 (forward) 5'-CCAGCCTGCTTGAGGAGTAGA-3'

SNTB2 (reverse) 5'-GCTGCATCCTGTTGTTTTTGG-3'

COG1 (forward) 5'-CCAACCTGCTCATGTACGTGAA-3'

COG1 (reverse) 5'-CATGGCGTCCCGGATTC-3'

RHOT2 (forward) 5'-GTGCCAGGCTGTATTGCTT-3

RHOT2 (reverse) 5'-GGGAAATGCAGACGTGTCAT-3'

DDX23 (forward) 5'-AGCCATTATCCCTGGAGGAG-3'

DDX23 (reverse) 5'-CTTCAGCCTCTCGTTCTGCT-3'

AQR (forward) 5'-TGGGAGAATCTGAACCTAATCC-3'

AQR (reverse) 5'-GCAGGGTAACCAAGTAAACACA-3'

SETX (forward) 5'-CACACTATGGAGAGGGAAGCA-3'

SETX (reverse) 5'-TTAGATCCAAGGCGATCCAG-3'

Western Blot

Western blot experiments were performed following standard procedures. Membrane were incubated with anti-UAP56 (1:1000, ab47955 Abcam), anti-Vinculin (1:5000, V9264 Merck), anti-THOC1 (1:100, ab487 Abcam), anti-PCNA (PC10) (1:500, sc-56 Santa Cruz) anti-FANCD2 (1:100, sc-20022 Santa Cruz) and anti-β-actin (1:1000, ab8227 Abcam).

Flow cytometry assays

For cell cycle analysis, cells were harvested, fixed with PBS + 4% formaldehyde for 10 min at room temperature (RT) and permeabilized with PBS + 0,2% Triton X-100 during 10 min at RT. Finally, DNA was stained with DAPI (1 μg/ml) 4º C overnight in PBS.

For EdU and γH2AX analysis, cells were treated with EdU (10 μM) for 30 min before harvesting and fixed with PBS + 4% formaldehyde for 10 minutes at RT. Cells were permeabilized with PBS + 0.2% Triton X-100 for 10 minutes at RT and blocked in PBS + 1% BSA + 0.05% Tween-20 for 30 minutes at RT. Click-it reaction was conducted following manufacturer instructions. Primary anti- γH2AX antibody (1:100, Millipore clone JBW301, 05-636) and secondary Alexa Fluor 594 antibodies (1:200, Invitrogen) were incubated in blocking buffer for 1 hour at RT in the darkness. DNA was stained with 1 μg/ml of DAPI at 4º overnight in PBS. Between incubations cells were washed with PBS + 0.1% BSA + 0.05% Tween-20.

Cells were analyzed in a BD influx sorter. When needed, plasmid-transfected cells were identified with GFP signal. Data were analyzed in FlowJo 9.3.2 (Tree Star).

Immunofluorescence high-throughput analysis

Large field images were acquired using automated plate on microscope Leica DM6000 microscope equipped with a DFC390 camera to ensure enough cell number in all cell cycle phases. Next, images were submitted to FiJi processing using DNA cell cycle plug-in from MBF collection and cells associated to each phase of the cell cycle according to its DAPI content. Finally, S9.6 mean nuclear intensity quantification and γH2AX foci counting was performed on each cell and results plotted according to the corresponding phase of the cell cycle. *In vivo* validations of this plug-in with EdU staining and H3S10P immunofluorescence were shown in Supplemental Fig. S2A,B.

Statistical information

Statistical parameters including the number of biological replicates (n), standard deviation (SD) or standard error of the mean (SEM) and statistical significance are reported in the figure legends. For γH2AX foci analysis, single-cell electrophoresis, EdU and γH2AX analysis by FACS Student's t-test was used. For γH2AX nuclear

intensity analysis, one-tailed Student's t-test was used. When indicated, paired test was used to reduce experimental variation. For flow cytometry cell cycle analysis One-way ANOVA analysis and Bonferroni post test was used. For γH2AX foci IF throughout the cell cycle, S9.6 IF, DNA combing and certain analysis from DRIPc-seq data, Mann–Whitney U-test two-tailed was performed. For DRIP-qPCR paired Student's t-test was used. For Venn diagrams hypergeometric test was used. For correlations the Pearson correlation coefficient and ANCOVA tests were used. In general, a P-value < 0.05 was considered as statistically significant (****P < 0.0001;***P < 0.001;**P < 0.01; *P < 0.05). Data were analyzed with EXCEL (Microsoft) or GraphPad Prism software.

SUPPLEMENTAL REFERENCES

Bianco JN, Poli J, Saksouk J, Bacal J, Silva MJ, Yoshida K, Lin YL, Tourriere H, Lengronne A, Pasero P. 2012. Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods* **57**: 149-157.

Cerritelli SM, Frolova EG, Feng C, Grinberg A, Love PE, Crouch RJ. 2003. Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol Cell* **11**: 807-815.

Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7**: e1002386.

Garcia-Rubio M, Barroso SI, Aguilera A. 2018. Detection of DNA-RNA Hybrids In Vivo. *Methods Mol Biol* **1672**: 347-361.

Garcia-Rubio ML, Perez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado IV, Aguilera A. 2015. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet* **11**: e1005674.

Salas-Armenteros I, Perez-Calero C, Bayona-Feliu A, Tumini E, Luna R, Aguilera A. 2017. Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J* **36**: 3532-3547.

Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang CC, Cohn MA, Gibbons RJ, Deans AJ, Niedzwiedz W. 2015. The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol Cell* **60**: 351-361.

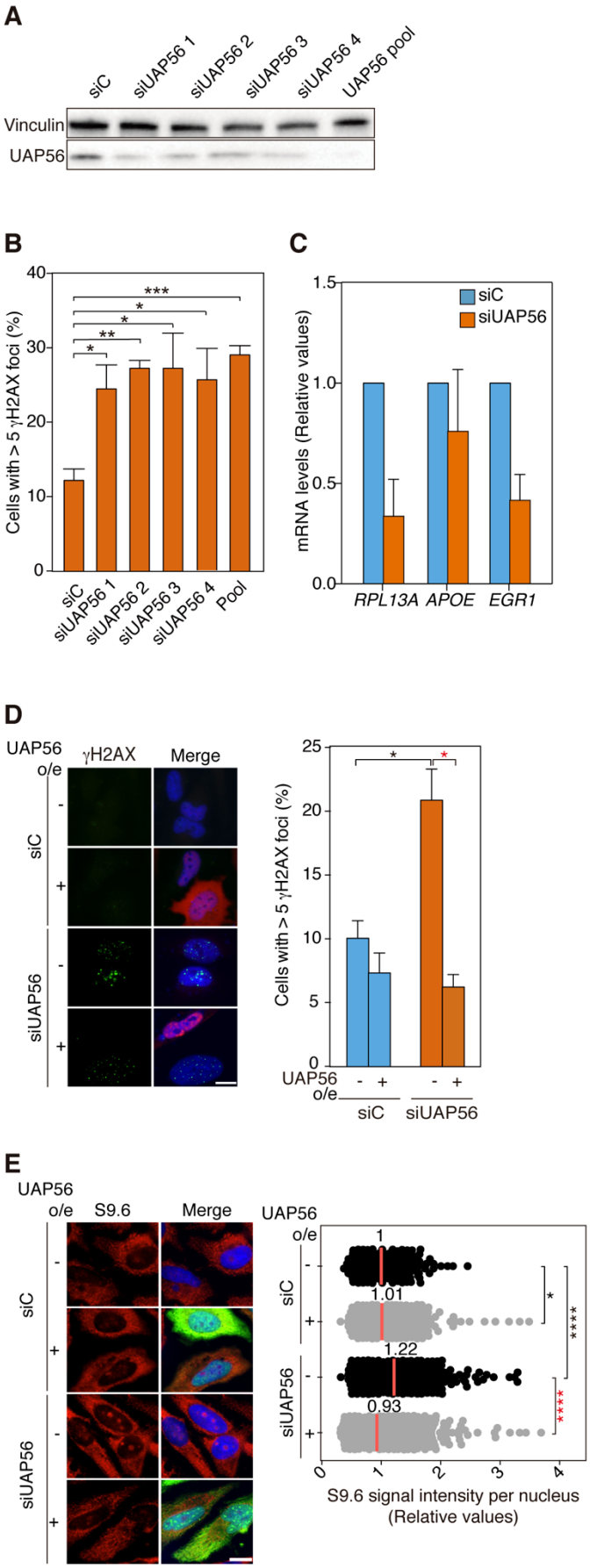
Shen J, Zhang L, Zhao R. 2007. Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. *J Biol Chem* **282**: 22544-22550.

ten Asbroek AL, van Groenigen M, Nooij M, 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.

Vitor AC, Sridhara SC, Sabino JC, Afonso AI, Grosso AR, Martin RM, de Almeida SF. 2019. Single-molecule imaging of transcription at damaged chromatin. *Sci Adv* **5**: eaau1249.

SUPPLEMENTAL FIGURE AND LEGENDS

Supplemental Fig. S1



Supplemental Figure S1. Different validations of UAP56 siRNA efficiency.

(A) Western blot assay showing the knock-down efficiency of each independent siRNA from the UAP56 pool.

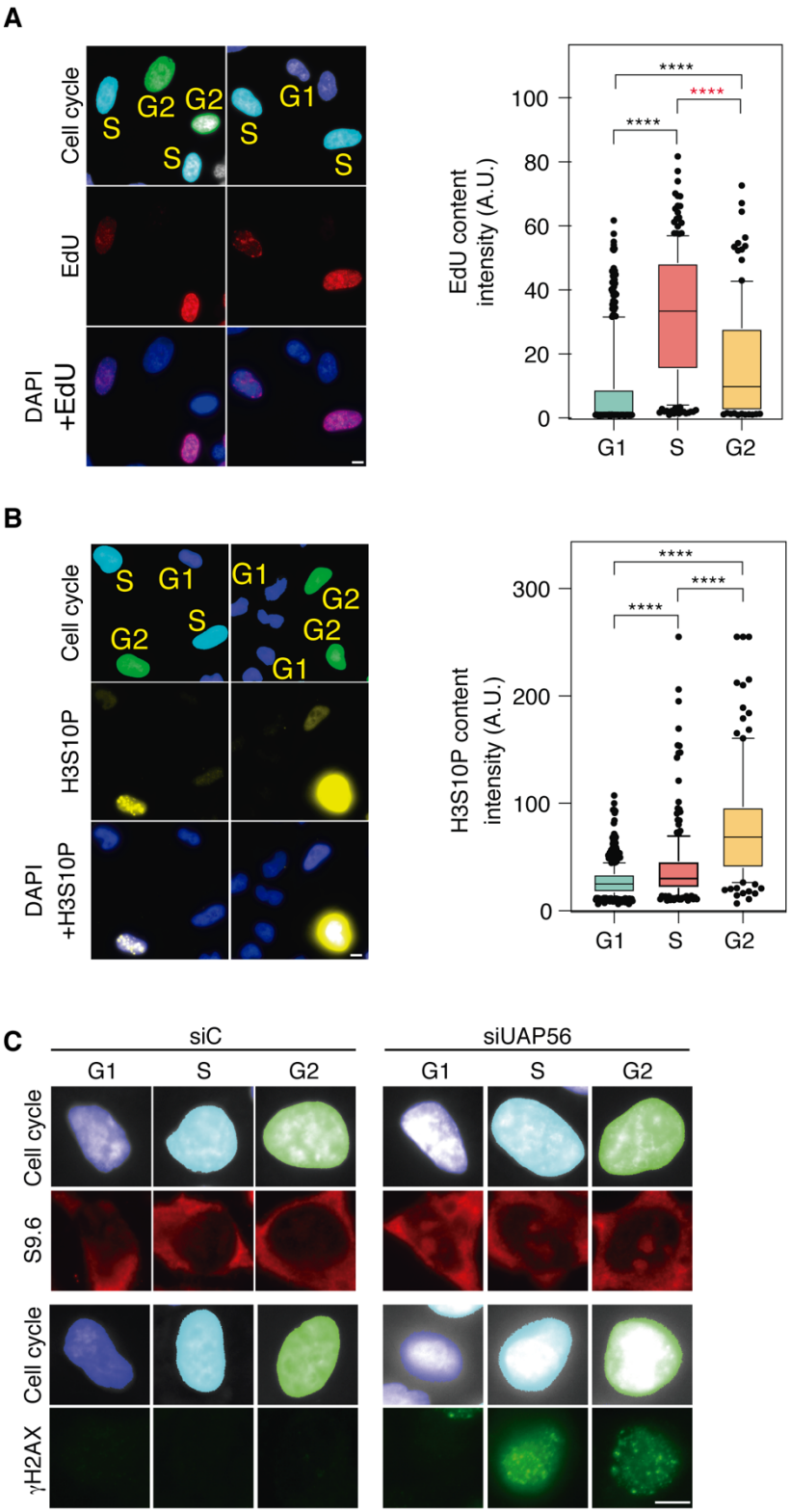
(B) Quantification of γH2AX foci by IF in HeLa cells transfected with siC (control) or different independent siRNA from the UAP56 siRNA pool. The graph shows the quantification of cells containing >5 γH2AX foci. More than 100 cells per condition were counted in each experiment. Data are plotted as mean + SEM (n=3). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (Student's t-test, two tailed).

(C) Relative mRNA quantification of *APOE, RPL13A* and *EGR1* genes in UAP56-depleted HeLa cells. Data are plotted mean + SEM (n=3).

(D) Representative images and quantification of γH2AX foci after ectopic expression of UAP56 in UAP56-depleted cells. Detection of γH2AX foci by IF in siC and siUAP56 HeLa cells transfected with pFLAG (-UAP56) or pFLAG-UAP56 (+UAP56) overexpressing UAP56. Immunostaining with anti-γH2AX antibody (green), anti-FLAG antibody (red) to detect UAP56 overexpression and DAPI (blue) are shown. The graph shows the quantification of cells containing >5 γH2AX foci. More than 100 cells per condition were considered in each experiment. Data are plotted as mean + SEM (n=3). Scale bar, 25 μm. *, $P < 0.05$ (Student's t-test, two-tailed).

(E) Representative images and quantification of nuclear S9.6 signal after ectopic expression of UAP56 in UAP56-depleted cells. Evaluation of S9.6 immunofluorescence signal in siC and siUAP56 HeLa cells transfected with the empty vector pFLAG (-UAP56) or pFLAG-UAP56 (+UAP56) for UAP56 overexpression. Immunostaining using the S9.6 monoclonal antibody (red), anti-FLAG antibody (green) to detect UAP56-FLAG overexpression, and DAPI (blue). The graph shows the median of the relative S9.6 signal intensity per nucleus. Same number of cells per condition was counted in each experiment. More than 600 total cells were considered (n=3). Scale bar, 25 μm. *, $P < 0.05$; ***, $P < 0.001$ (Mann-Whitney U test, two-tailed).

Data information: black stars denote significant increases, whereas red stars denote significant decreases.


A
Cell cycle
EdU
DAPI +EdU
S
G2
G1
EdU content intensity (A.U.)
****
B
H3S10P
DAPI +H3S10P
H3S10P content intensity (A.U.)
C
siC
siUAP56
G1
S
G2
Cell cycle
S9.6
γH2AX

Supplemental Figure S2. IF high-throughput analysis validations.

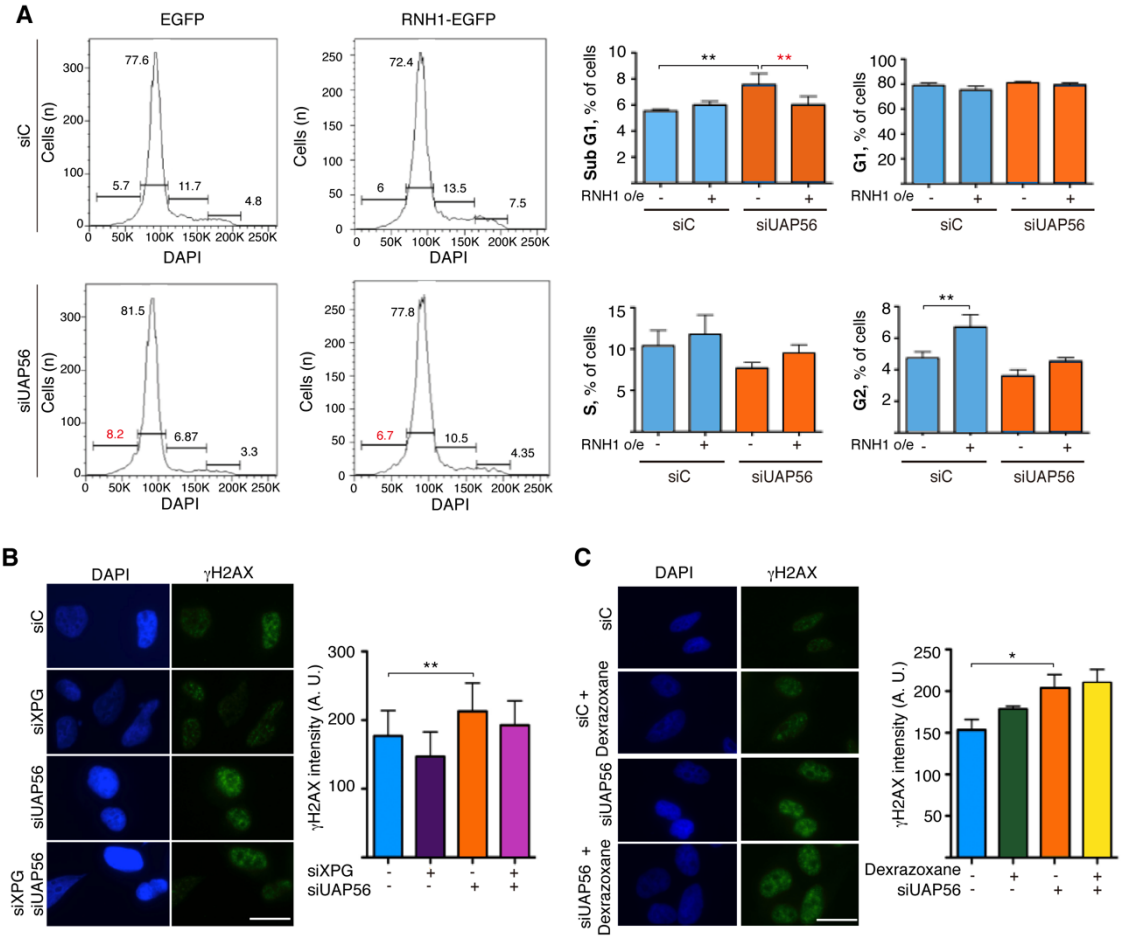
(A) Representative images and quantification of EdU content intensity through all the phases of the cell cycle. DAPI staining was used to classify different stages of the cell cycle, whereas EdU indicates DNA synthesis. The median of each population of the cell cycle is shown. Boxes and whiskers indicate 10-90 percentiles. More than 500 total cells were considered in each experiment (n=3). Scale bar, 5 μm. ****, P < 0.0001 (Mann-Whitney U test, two-tailed).

(B) Representative images and quantification of H3S10P content intensity through all the phases of the cell cycle. DAPI staining were used to classify different stages of the cell cycle, whereas H3S10P was used as a G2/M marker. The median of each population of the cell cycle is shown. Boxes and whiskers indicate 10-90 percentiles. More than 500 total cells were considered in each experiment (n=3). Scale bar, 5 μm. ****, P < 0.0001 (Mann-Whitney U test, two-tailed).

(C) Representative images of nuclear S9.6 signal and γH2AX foci per cell in siC and siUAP56 cells according to different stages of the cell cycle. Scale bar, 5 μm.

Data information: black stars denote significant increases, whereas red stars denote significant decreases.

Supplemental Fig. S3



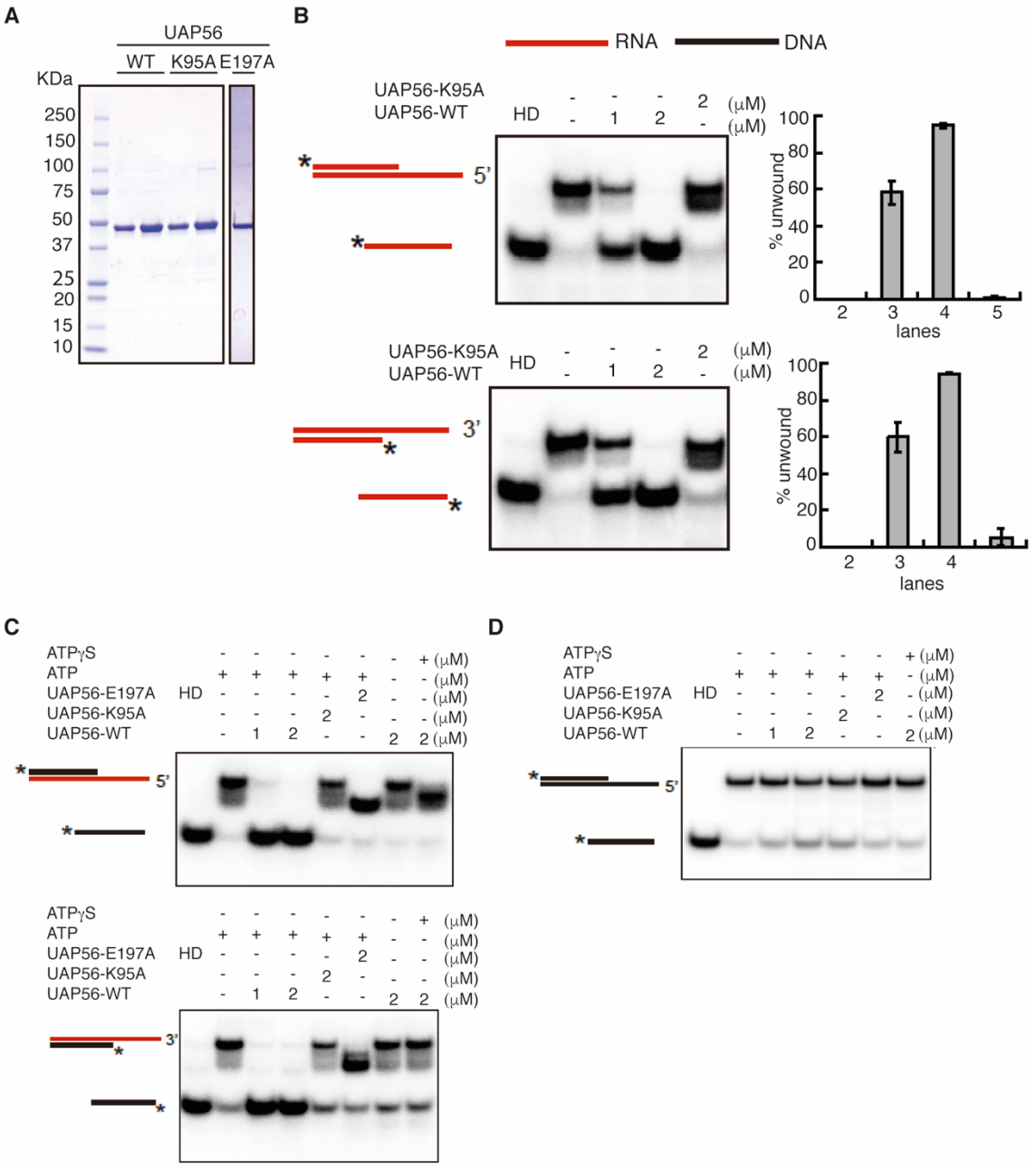
Supplemental Figure S3. Extended cell cycle and DNA damage analysis in UAP56-depleted cells.

(A) FACS analysis in control and UAP56-depleted HeLa cells transfected with empty plasmid (EGFP) or plasmid overexpressing RNase H1 (RNH1-EGFP). Transfected cells were selected by GFP signal and cell cycle was evaluated with DAPI staining. Graphs represent the percentage of cells in each phase of the cell cycle. Data are plotted as mean and SD (n=3). **, P <0.01 (Repeated measures ANNOVA and Bonferroni's post test).

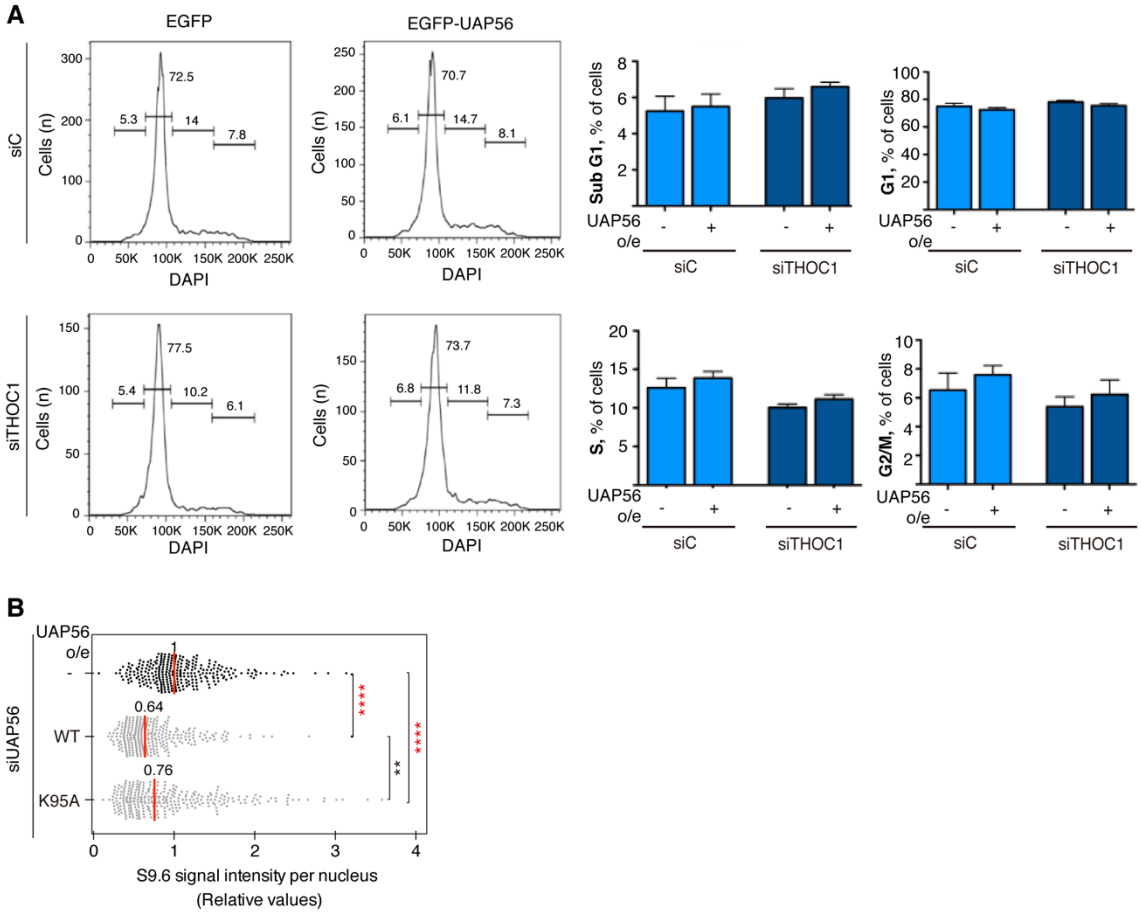
(B) Representative images and quantification of γH2AX nucleus intensity in UAP56- and XPG-depleted HeLa cells. Immunostaining with anti-γH2AX antibody (green) and DAPI (blue) are shown. The graph shows the quantification of relative γH2AX nucleus intensity. More than 230 cells per condition were counted in each experiment. Data are plotted as mean and SD of γH2AX median intensity (n≥3). Scale bar, 25 μm. **, P < 0.01 (paired Student's t-test, one-tailed).

(C) Representative images and quantification of γH2AX nucleus intensity in siUAP56 HeLa cells untreated or treated for 4 hours with 50 μM of the TOPII inhibitor dexrazoxane. Graph shows the quantification of relative γH2AX nucleus intensity under these conditions. More than 350 cells per condition were counted in each experiment. Data are plotted as mean and SD of γH2AX median intensity (n=3). Scale bar, 25 μm. *, P < 0.05 (unpaired Student's t-test, one-tailed).

Data information: black stars denote significant increases, whereas red stars denote significant decreases.


A
UAP56
WT
K95A
E197A
KDa
250
150
100
75
50
37
25
20
15
10
B
RNA
DNA
UAP56-K95A
UAP56-WT
HD
(μM)
5'
3'
% unwound
lanes
C
ATPγS
ATP
UAP56-E197A
UAP56-K95A
UAP56-WT
D

Supplemental Figure S4. Additional *in vitro* analysis of UAP56 helicase activity.
(A) UAP56-WT and UAP56 helicase-dead mutants protein purification. Purified wild type and UAP56 mutants were analyzed on SDS-polyacrilamide gel and stained with Coomassie Blue. First two lines representing UAP56-WT, the next two lines UAP56-K95A mutant and last line UAP56-E197A mutant.
(B) UAP56 and UAP56-K95A helicase-dead mutant RNA-unwinding assay using 5' or 3' overhang dsRNA as a substrate. The positions of duplex substrate and unwound products are indicated at the left, where the stars show the position of the radiolabel. Gels were dried and subject to phosphorimaging analysis. HD, head denatured substrate.
(C) Unwinding assay with UAP56-WT, UAP56-K95A and UAP56-E197A using 5' overhang or 3' overhang RNA-DNA duplex as a substrate. Other details as in (B).
(D) UAP56 DNA-DNA unwinding activity. DNA-unwinding assay with UAP56-WT, UAP56-K95A and UAP56-E197A using 5' overhang DNA duplex as a substrate. Other details as in (B).


A
EGFP
EGFP-UAP56
siC
siTHOC1
Cells (n)
DAPI
72.5
5.3
14
7.8
70.7
6.1
14.7
8.1
77.5
5.4
10.2
6.1
73.7
6.8
11.8
7.3
Sub G1, % of cells
G1, % of cells
S, % of cells
G2/M, % of cells
UAP56 o/e
siC
siTHOC1

B
siUAP56
UAP56 o/e
WT
K95A
0.64
0.76
S9.6 signal intensity per nucleus
(Relative values)

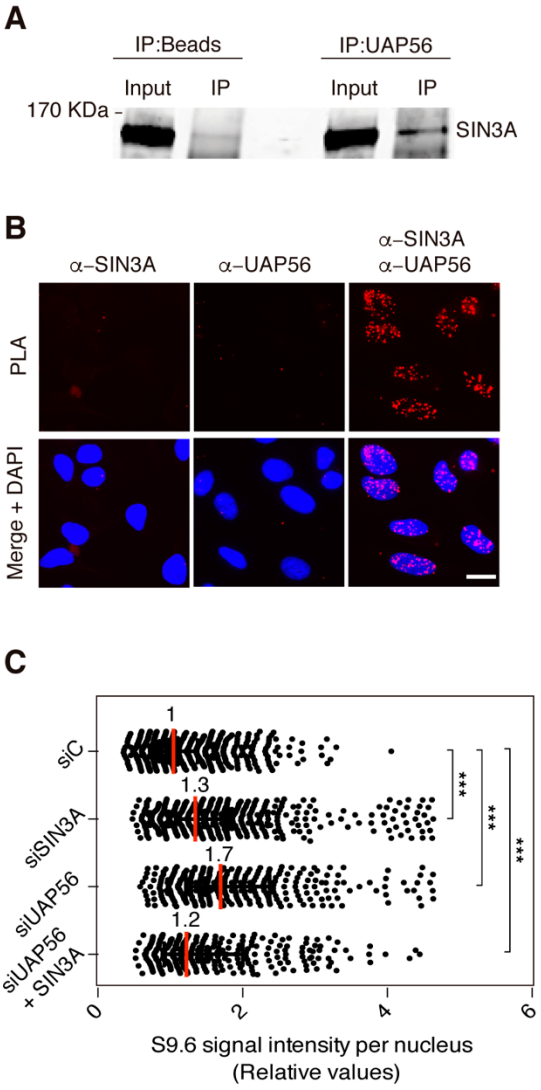
Supplemental Figure S5. UAP56 overexpression effects in cell cycle and RNA-DNA hybrids.

(A) FACS analysis in siC and THOC1-depleted HeLa cells transfected with empty plasmid (EGFP) or plasmid overexpressing UAP56 (EGFP-UAP56). Transfected cells were selected by GFP signal and cell cycle was evaluated with DAPI staining. Graphs represent the percentage of cells in each phase of the cell cycle. Data are plotted as mean and SD (n=3). No statistical significance is reported using One-way ANNOVA and Bonferroni's post test.

(B) Quantification of S9.6 immunofluorescence signal in siUAP56 HeLa cells transfected with pFLAG (-UAP56), pFLAG-UAP56 (+UAP56) for UAP56 overexpression and pFLAG-UAP56-K95A (+UAP56-K95A) for helicase-dead UAP56 overexpression. 100 cells per condition was counted in each experiment. The graph shows the median of the S9.6 signal intensity per nucleus after nucleolar signal removal (n=3). Values were normalized respect to siC median of each experiment. **, $P < 0.01$; ***, $P < 0.001$ (Mann-Whitney U test, two-tailed).

Data information: black stars denote significant increases, whereas red stars denote significant decreases.

Supplemental Fig. S6



Supplemental Figure S6. UAP56 and SIN3 histone deacetylase complex association.

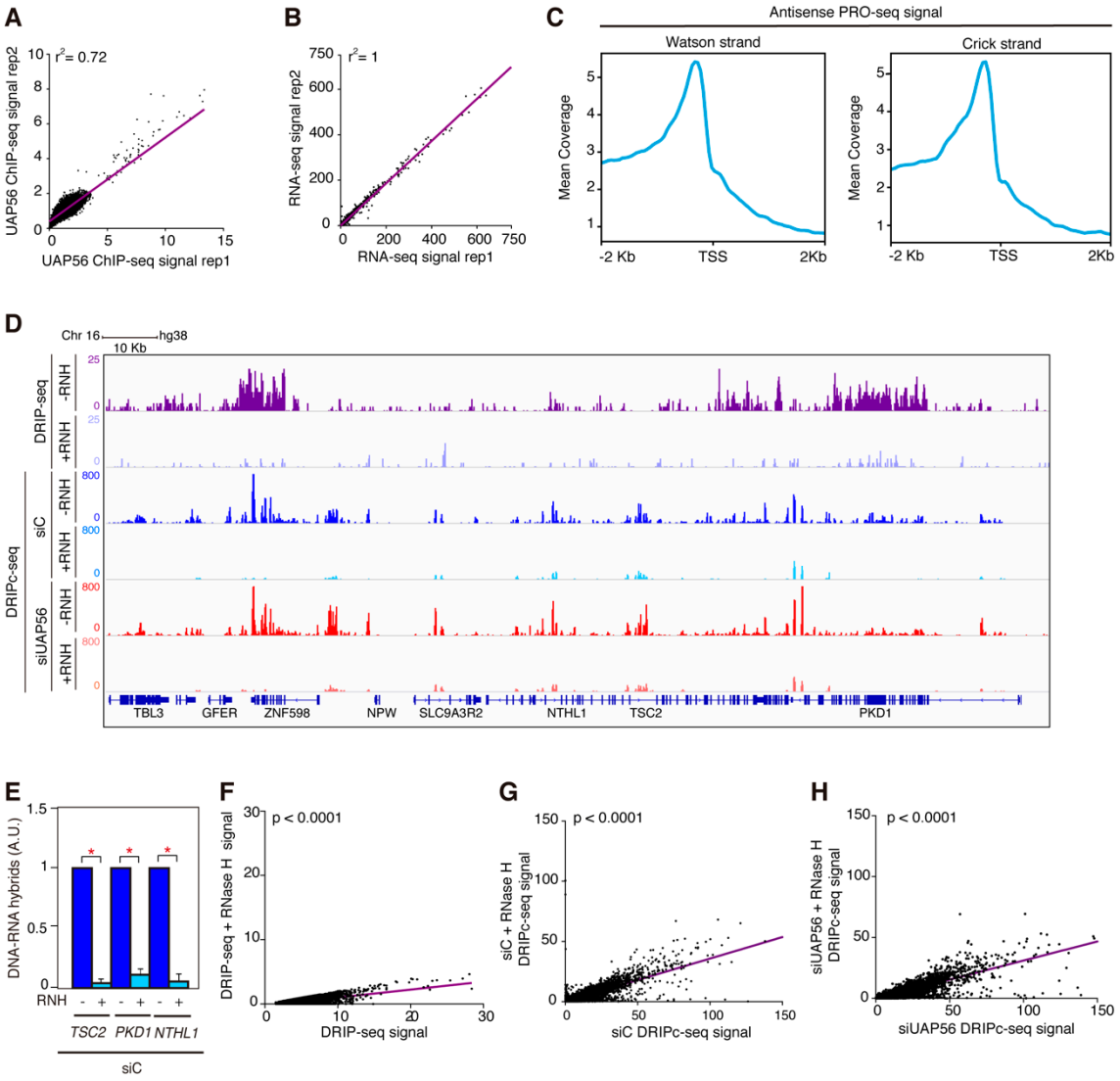
(A) UAP56 association with SIN3 detected by co-IP assays in whole-cell extracts of HEK293T cells with anti-UAP56 antibody. Input extract and total immunoprecipitated (IP) were analyzed by Western blot with anti-Sin3A antibody (n=2).

(B) Representative images of Proximity Ligation Assay (PLA) in HeLa cells showing specific association of UAP56 and SIN3A endogenous protein. PLA signals (red spots) (n=2). Scale bar, 25 μm.

(C) Quantification of S9.6 IF signal in siC, siSIN3A, siUAP56 and siUAP56-siSIN3A HeLa cells. Same number of cells per condition was counted in each experiment. The graph shows the median of the S9.6 signal intensity per nucleus after nucleolar signal removal. More than 90 cells were considered in each experiment. (n=3). Values were normalized respect to siC median of each experiment. ***, $P < 0.001$ (Mann-Whitney U test, two-tailed).

Data information: black stars denote significant increases.

Supplemental Fig. S7



Supplemental Figure S7. Different genome-wide studies validations.

(A) UAP56 ChIP-seq experiments reproducibility in K562 cells. *xy* correlation plot between two ChIP-seq replicates (r^2, Pearson correlation).

(B) Reproducibility of RNA-seq experiments in K562 cells. *xy* correlation plot between two RNA-seq replicates (r^2, Pearson correlation).

(C) PRO-seq metaplot along genes in which UAP56 is localized at promoters corresponding to antisense PRO-seq signal in K562 cells.

(D) Representative screenshot of a genomic region showing the DRIP-seq signal in untreated (purple) and RNase H-treated cells (light purple), and DRIPc-seq signal profiles for untreated siC (blue), siC RNase H-treated cells (light blue) and untreated siUAP56 (red) and RNase H-treated (light red) cells. Track values lower than 3 (RPKM) are shown as background at 0 value in plots.

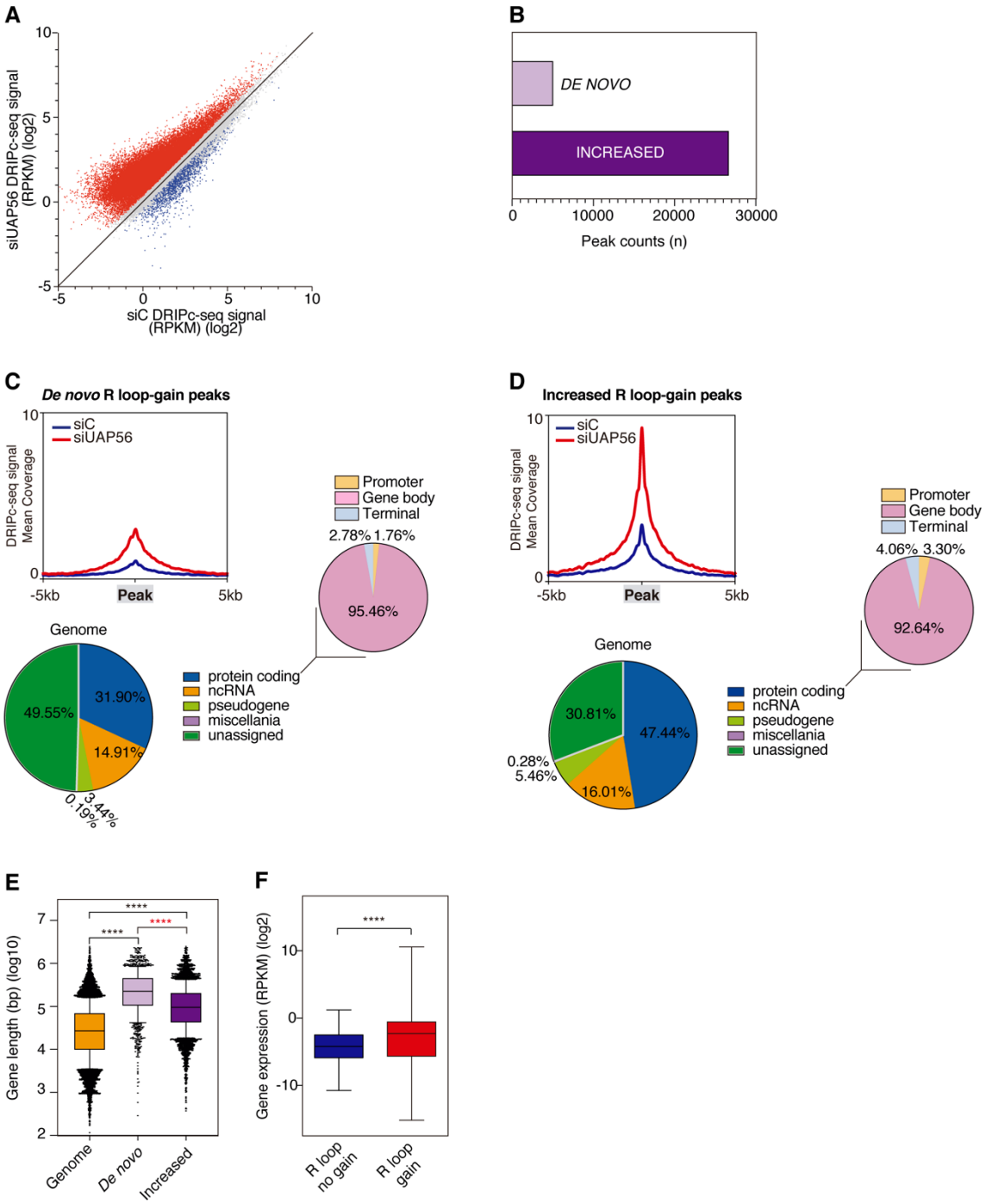
(E) DRIP-qPCR using the anti-RNA-DNA hybrids S9.6 monoclonal antibody in siC K562 cells at *TSC2*, *PKD1* and *NTHL1* genes. Signal values were normalized with respect to the siC control and plotted as mean + SEM (n=3). *, $P < 0.05$ (Paired Student's t-test, two tailed).

(F) RNase H effect on the R loop signal in DRIP-seq experiments. *xy* correlation plot between the DRIP-seq signal obtained from RNase H-treated (*y* axis) and untreated (*x* axis) cells ($P < 0.0001$ compared with a hypothetical perfect line x=y, ANCOVA test).

(G) RNase H effect on the R loop signal in DRIPc-seq experiments of siC cells. *xy* correlation plot between the DRIPc-seq signal obtained from RNase H-treated (*y* axis) and untreated (*x* axis) siC control cells ($P < 0.0001$ compared with a hypothetical perfect line x=y, ANCOVA test).

(H) RNase H effect on R loop signal in DRIPc-seq experiments of UAP56-depleted cells. *xy* correlation plot for UAP56-depleted cells untreated or treated with RNase H ($P < 0.0001$ compared with a hypothetical perfect line x=y, ANCOVA test).

Data information: black stars denote significant increases, whereas red stars denote significant decreases.


A
siUAP56 DRIPc-seq signal (RPKM) (log2)
siC DRIPc-seq signal (RPKM) (log2)
B
DE NOVO
INCREASED
Peak counts (n)
C
De novo R loop-gain peaks
siC
siUAP56
DRIPc-seq signal Mean Coverage
-5kb
Peak
5kb
Promoter
Gene body
Terminal
2.78%
1.76%
95.46%
Genome
31.90%
49.55%
14.91%
3.44%
0.19%
protein coding
ncRNA
pseudogene
miscellania
unassigned
D
Increased R loop-gain peaks
4.06%
3.30%
92.64%
30.81%
47.44%
0.28%
5.46%
16.01%
E
Gene length (bp) (log10)
Genome
De novo
Increased
F
Gene expression (RPKM) (log2)
R loop no gain
R loop gain

Supplemental Figure S8. Extended analysis of R loop-gain peaks after UAP56 depletion.

(A) DRIPc-seq signal from R loop peak correlation analysis between siC and siUAP56 cells. R loop peaks represented present DRIPc signal fold change higher than 1.25X in siUAP56 respect to the siC control cells in two replicates. Diagonal line indicates a perfect correlation. Red dots represent R loop-gain peaks. Grey dots and blue dots correspond to R loop-no gain peaks.

(B) Classification of R loop-gain peaks in *De novo* and increased categories.

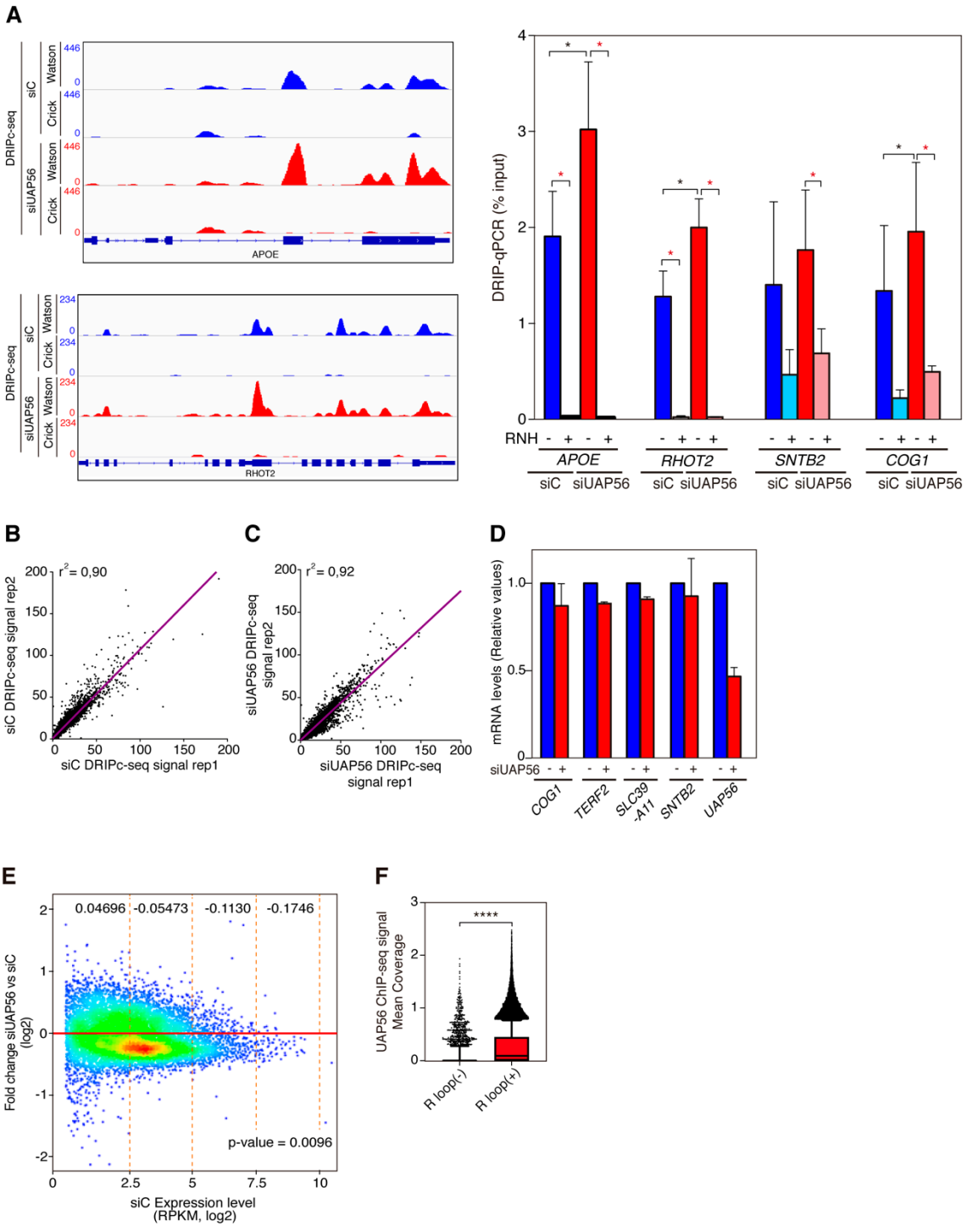
(C) *De novo* R loop-gain peaks. Upper panel: *de novo* R loop-gain peak metaplot for control and UAP56-depleted cells. Bottom panel: *De novo* R loop-gain peaks across several genomic features depicted. Peaks that localize within protein coding genes are also classified in three different categories: promoter, gene body and terminal region.

(D) Increased R loop-gain peaks. Upper panel: increased R loop-gain peak metaplot for control and UAP56-depleted cells. Bottom panel: Increased R loop-gain peaks across several genomic features depicted. Other details as in (C).

(E) Distribution of length values for *de novo* and increased R loop-gain genes. Data are plotted as box and whiskers (10-90 percentile) where median values are indicated. ****, $P < 0.0001$ (Mann-Whitney U-test, two tailed).

(F) Box plot showing gene expression for R loop-gain and R loop-no gain genes where median values are indicated. ****, $P < 0.0001$ (Mann-Whitney U-test, two tailed).

Data information: black stars denote significant increases, whereas red stars denote significant decreases.


A
DRIPc-seq
siC
siUAP56
Watson
Crick
APOE
RHOT2
DRIP-qPCR (% input)
RNH
APOE
RHOT2
SNTB2
COG1
siC
siUAP56
B
C
siC DRIPc-seq signal rep2
siC DRIPc-seq signal rep1
r² = 0,90
siUAP56 DRIPc-seq signal rep2
siUAP56 DRIPc-seq signal rep1
r² = 0,92
D
mRNA levels (Relative values)
siUAP56
COG1
TERF2
SLC39A11
SNTB2
UAP56
E
Fold change siUAP56 vs siC (log2)
siC Expression level (RPKM, log2)
0.04696
-0.05473
-0.1130
-0.1746
p-value = 0.0096
F
UAP56 ChIP-seq signal Mean Coverage
****
R loop(-)
R loop(+)

Supplemental Figure S9. Additional DRIPc-seq validations.

(A) DRIPc-seq validations. Left: strand detection of RNA-DNA hybrids by DRIPc-seq in siC control and UAP56-depleted K562 cells. Representative screenshots of *APOE* and *RHOT2* genes where R loops are mapped at Watson and Crick strand in siC control (blue) and UAP56-depleted cells (red). Right: validation of DRIPc-seq by DRIP-qPCR in siC and siUAP56 cells. Data are plotted as mean + SEM (n=4 for *APOE*, n=3 for *RHOT2*, n=2 for *SNTB2* and *COG1*). *, $P < 0.05$ (Paired Student's t-test)

(B) Reproducibility of DRIPc-seq experiments. *xy* correlation plot between two siC control cells DRIPc-seq replicates.

(C) Reproducibility of DRIPc-seq experiments. *xy* correlation plot between two siUAP56 cells DRIPc-seq replicates (r^2, Pearson correlation).

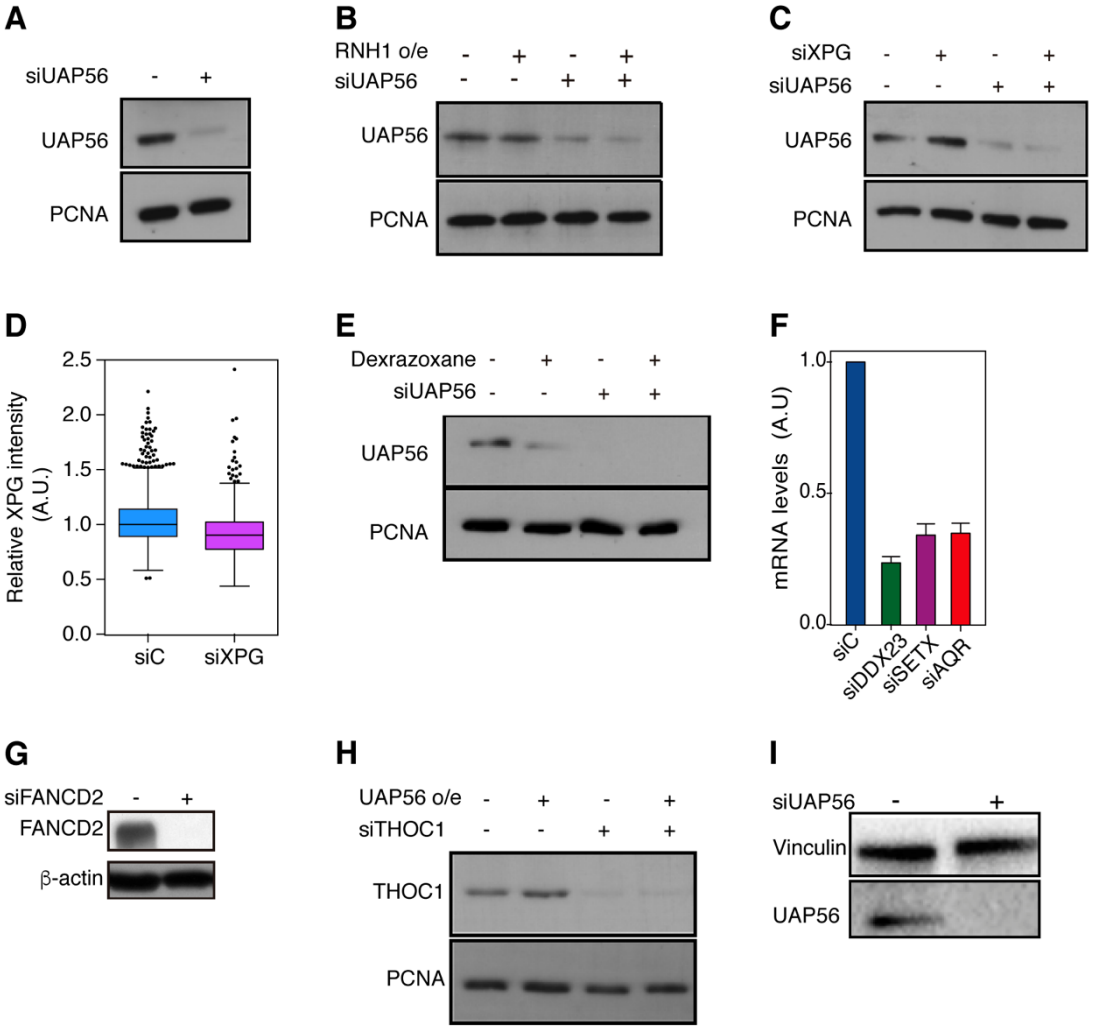
(D) Relative mRNA quantification of *COG1, TERF2, SLC39A11, SNTB2* and *UAP56* genes in control and UAP56-depleted K562 cells. Data are plotted mean + SEM (n=2).

(E) Density scatterplot showing $\log_2$ fold change for siUAP56 against siC as function of $\log_2$ siC expression levels in K562 cells. The red line (y = 0) indicates no change between samples in expression. Orange dot lines separate different expression ranges in which their fold change median is indicated (n=1). Median Pearson correlation associated p-value indicated in the bottom right.

(F) UAP56 ChIP-seq signal (Mean Coverage) over genes according to its R loop content. R loop (-) refers to genes that do not present R loops, while R loop (+) those accumulating them in siC control K562 cells. ****, $P < 0.0001$ (Mann-Whitney U test, two-tailed).

Data information: black stars denote significant increases, whereas red stars denote significant decreases.

Supplemental Fig. S10



Supplemental Figure S10. General analysis of siRNA Knock-down efficiency.

(A) Western blot verifying UAP56 depletion compared to PCNA as a control in UAP56-depleted HeLa cells. (Fig. 2A,B)

(B) Western blot ensuring UAP56 depletion compared to PCNA as control in siC and siUAP56 HeLa cells overexpressing or not RNase H1. (Supplemental Fig. S3A)

(C) Examination of depletion of UAP56 in in UAP56- and XPG-depleted HeLa cells. (Supplemental Fig. S3B)

(D) Quantification of XPG intensity by IF in control and XPG-depleted HeLa cells. Relative XPG intensity is represented (Box and whiskers plot, Tukey's representation). Values were normalized to siC median in each experiment. More than 400 cells per condition were evaluated in each experiment (n=2). (Supplemental Fig. S3B)

(E) Western blot validating UAP56 Knock-down in UAP56-depleted HeLa cells untreated or treated with the TOP2 inhibitor dexrazoxane. (Supplemental Fig. S3C)

(F) Efficiency of siRNA depletion against DDX23, SETX and AQR. Graph showing the mRNA levels by qPCR of siC (control), siDDX23, siSETX and siAQR depleted-cells. mRNA expression values of the indicated genes were normalized with mRNA expression of the HPRT housekeeping gene. Data are plotted as mean + SEM (n=3). (Fig. 4).

(G) Western blot verifying FANCD2 depletion compared to β-actin as a control in FANCD2-depleted HeLa cells. (Fig. 4)

(H) Examination of THOC1 depletion compared to PCNA as a control in THOC1-depleted cells overexpressing or not UAP56. (Supplemental Fig. S5A).

(I) Western blot verifying UAP56 depletion compared to vinculin as a control in UAP56-depleted K562 cells. (Fig. 7)